

Universitat de Lleida

Disponibilidad léxica en la provincia de Lleida: estudio comparado de dos lenguas en contacto

Maribel Serrano Zapata

Dipòsit Legal: L.1698-2014

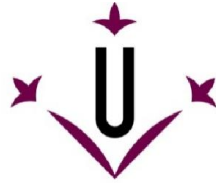
<http://hdl.handle.net/10803/285008>

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

Facultat de Lletres
Departament de Filologia Clàssica, Francesa i Hispànica



Universitat de Lleida

**Disponibilidad léxica en la provincia de
Lleida: estudio comparado de dos lenguas en
contacto**

Tesis doctoral dirigida por M^a Ángeles Calero Fernández

Maribel Serrano Zapata

Lleida, abril de 2014

RESUMEN:

La presente tesis doctoral se enmarca dentro de la disponibilidad léxica, una línea de investigación que cuenta con un notable desarrollo en la actualidad en la lingüística del mundo hispánico. Siguiendo, por tanto, los presupuestos metodológicos del *Proyecto Panhispánico de Estudio del Léxico Disponible*, nuestra aportación concreta supone la incorporación a dicho proyecto de una parcela inexplorada hasta el momento: la provincia de Lleida. Dada la situación de bilingüismo social de esta área, hemos considerado imprescindible extender nuestro análisis a la lengua catalana y, por tanto, el principal objetivo de nuestro trabajo ha sido la elaboración de los diccionarios de disponibilidad en dos de las lenguas oficiales de nuestra comunidad: el español y el catalán.

El segundo de nuestros propósitos ha sido analizar la incidencia que tenían en nuestros datos tanto las variables lingüísticas, esto es los distintos centros de interés, como las cinco variables sociales elegidas (*sexo, tipo de centro, zona geográfica, nivel sociocultural y lengua materna*). Nuestro análisis cuantitativo tiene cierta originalidad por ser el primero que, basándose en una sola muestra —compuesta por 240 informantes y un total de 480 encuestas—, analiza de forma comparada tanto la productividad de las dos lenguas entre sí como la incidencia de las distintas variables lingüísticas y sociales sobre los dos corpus obtenidos. Las variables sociales elegidas se han estudiado de forma independiente en cada corpus y también se han descrito de manera comparada las principales tendencias de cada subgrupo social con relación a la producción de las dos lenguas.

El tercer objetivo de nuestro estudio ha sido analizar la influencia mutua de los dos idiomas, esto es, en qué medida el léxico disponible del castellano, o el del catalán, contiene elementos de la otra lengua y qué peso alcanzan dichas unidades en el conjunto total de los datos. Para ello hemos clasificado todos los fenómenos de contacto lingüístico presentes en los dos corpus y hemos analizado cuantitativamente su valor, en relación con el total de palabras distintas recabadas pero también con la cantidad de respuestas totales aportadas por los informantes en cada lengua.

RESUM:

La present tesi doctoral s'emmarca dins de la disponibilitat lèxica, una línia d'investigació que té un notable desenvolupament actualment en la lingüística hispànica. La nostra aportació concreta segueix els pressupòsits metodològics del *Projecte Panhispànic d'Estudi del Lèxic Disponible* i comporta la incorporació a aquest projecte d'una parcel·la inexplorada fins al moment: la província de Lleida. Atès que la situació d'aquesta àrea és de bilingüisme social, hem considerat imprescindible ampliar la nostra anàlisi a la llengua catalana i, per tant, el principal objectiu del nostre treball ha estat l'elaboració dels diccionaris de disponibilitat en dues de les llengües oficials de la nostra comunitat: el castellà i el català.

El segon dels nostres propòsits ha estat analitzar la incidència que tenien en les nostres dades tant les variables lingüístiques, és a dir els diferents *centres d'interès*, com les cinc variables socials triades (*sexe, tipus de centre, zona geogràfica, nivell sociocultural i llengua materna*). La nostra anàlisi quantitativa té certa originalitat per ser la primera que es basa en una sola mostra —composta per 240 informants i un total de 480 enquestes—, i que analitza, de forma comparada, tant la productivitat de les dues llengües entre elles com la incidència de les diferents variables lingüístiques i socials sobre els dos corpus obtinguts. Les variables socials triades s'han estudiat de forma independent en cada corpus i també s'han descrit de manera comparada les principals tendències de cada subgrup social amb relació a la producció de les dues llengües.

El tercer objectiu del nostre estudi ha estat analitzar la influència mútua dels dos idiomes, és a dir, en quina mesura el lèxic disponible del castellà, o el del català, conté elements de l'altra llengua i quin pes tenen aquestes unitats en el conjunt total de les dades. Per això, hem classificat tots els fenòmens de contacte lingüístic presents en els dos corpus i hem analitzat quantitativament el seu valor, en relació amb el total de paraules diferents recollides però també amb la quantitat de respostes totals aportades pels informants en cada llengua.

ABSTRACT:

This thesis is part of the lexical availability, a line of research which nowadays has a great development within the Hispanic world. Following the methodological requirements of the *Proyecto Panhispánico de Estudio del Léxico Disponible* (*Pan-Hispanic Project of Research of Available Lexicon*), our specific contribution has consisted of including to that project an up to now unexplored geographical area: the Spanish province of Lleida. Taking into account the situation of social bilingualism of this region, it has been considered as essential to extend our analysis to the Catalan language and, therefore, the main objective of our work has been the development of dictionaries of lexical availability in the two official languages of our community: Spanish and Catalan.

The second of our purposes was to analyse the impact that the different linguistic variables (that is, the *areas of interest* and the five selected sociolinguistic variables: *sex, type of school, geographic area, sociocultural level* and *mother tongue*) had on our data. The quantitative analysis is in a certain way original because it is the first one that analyses in a comparative way both the production with both languages and the influence of the different linguistic and social variables on the two linguistic corpora, based a single sample of 240 informants (a total of 480 surveys). The selected sociolinguistic variables have been studied independently in each corpus, and the main trends in each social subgroup, in relation with the production in the two languages, have also been described in a comparative way.

The third objective of our study was to analyse the mutual influence of the two languages, that is, to what extent the available Spanish or Catalan lexicon contains elements of the other language, and what weight do such units have in the overall data. To that effect all the phenomena of language contact present in both corpus have been classified, and their value has qualitatively been analysed, in relation not only with the total different words that have been collected, but also with the number of total responses from informants in each language.

A Joaquín, por hacer de este proyecto también el suyo

A Víctor, *el meu patu(fet)*

A mis padres, por su apoyo incondicional en todo

AGRADECIMIENTOS

Cuando se finaliza un proyecto como el presente son muchas las personas que, de una forma u otra, han colaborado en su consecución. En primer lugar, quiero destacar la labor de dirección de M^a Ángeles Calero Fernández, quien me ha acompañado en las distintas fases de realización de esta tesis y ha resuelto siempre mis dudas, me ha hecho sugerencias inestimables y ha revisado todos mis escritos. Sus consejos, no solo relativos a la elaboración de este trabajo sino también sobre la vida, me han sido siempre de muchísima utilidad.

En segundo lugar, quiero mencionar a todos los miembros del *Proyecto Panhispánico de Estudio de la Disponibilidad*, que han respondido a mis emails, me han mandado datos, me han dado orientaciones valiosísimas y, en definitiva, por su cercanía y amistad. Algunos de sus nombres son María Areta, M^a Luisa Arnal, Manuela Casanova, Luis Escoriza, Begoña Gómez Devís, José Ramón Gómez Molina, Francesc Llopis, Florentino Paredes o la estupenda familia Samper Hernández. Quisiera destacar, entre todos ellos, a Natividad Hernández Muñoz, por una larga conversación telefónica que me sirvió para orientar el análisis y también porque sus hallazgos e investigaciones sobre los factores cognitivos que predicen la disponibilidad me han permitido alcanzar a mí los míos. De modo similar, quiero dejar constancia de que José Antonio Bartol Hernández atendió con paciencia infinita las numerosas dudas que se me plantearon en el volcado de las dos lenguas, sobre todo del catalán, en la página Dispolex.

Este trabajo también se nutre de los aprendizajes y experiencias vividas durante la realización de dos estancias de investigación. La primera de ellas tuvo lugar durante la primavera de 2003 en la Universidad de Cádiz de la que destacamos, desde aquí, la calurosa acogida que nos brindaron todos los miembros del Área de Lingüística General que dirige Miguel Casas, y muy especialmente los profesores Luis Escoriza y Adolfo González. La segunda estancia se realizó en el invierno de 2004 en la Universidad de Turku. Durante cuatro meses inolvidables, Esther Martínez Olmos y Alberto Carcedo fueron mi familia bajo la nieve finlandesa. Todas las experiencias vividas junto a ellos, así como sus largas conversaciones, su amistad, su quehacer científico, su cariño y su recuerdo me acompañarán de por vida. Si algo lamento de haber dilatado tanto en el

tiempo este proyecto, es que ellos ya no estén aquí para leerlo, ya que sé a ciencia cierta que les hubiera gustado hacerlo.

Estoy en deuda, sin duda alguna, con la profesora M^a Àngels Cabasés Piqué, del *Departament d'Economia Aplicada* de la Universitat de Lleida, cuyos amplios conocimientos sobre estadística y, en concreto sobre el programa SPSS, han sido fundamentales para el análisis de los datos. De otra parte, mi amigo Xavier Pascual López, desde Polonia, ha respondido siempre a las numerosas dudas de carácter formal que me asaltaban durante la elaboración de este trabajo. Con respecto al corpus catalán, Josep-Enric Teixidó, en un primer momento, y Francesc Pané, al final, me ayudaron a descifrar algunas palabras relacionadas, principalmente, con el léxico rural, que confieso que nunca en mi vida había oído.

Muchas personas me han animado a continuar en los momentos en que me parecía que la empresa que debía realizar era inabarcable. Quiero mencionar especialmente el apoyo de mis amigas Rosina, Esther y Rocío, y de mi familia, que me ha proporcionado momentos de paz interior indispensables cuando realizas un proyecto de esta envergadura. La familia Rubio-Rivas me ha cedido, durante años y de forma completamente desinteresada, una plaza de parking que me ha permitido ganar un tiempo inestimable en mi condición vital de madre trabajadora. Distintos amigos y familiares me acompañaron por las comarcas leridanas durante los años de aplicación del estudio, ayudándome a vencer así el miedo a la carretera. Agradezco desde aquí también su colaboración a todos los directores de colegios e institutos de secundaria que participaron en el estudio y, sobre todo, a los 240 informantes que me proporcionaron suficiente material como para elaborar tres tesis más y demostraron, una vez más, que el bilingüismo es una realidad en nuestros días. Para acabar, quisiera decir que estoy profundamente agradecida al azar, que me brindó la gran oportunidad de conocer y amar dos lenguas con la misma intensidad.

ÍNDICE

	<i>Páginas</i>
1. INTRODUCCIÓN	29
2. PRESUPUESTOS TEÓRICOS	39
2.1. LA DISPONIBILIDAD LÉXICA	39
2.1.1. Estudios pioneros sobre disponibilidad	47
2.1.2. El Proyecto Panhispánico	54
2.2. LA SOCIOLINGÜÍSTICA	62
2.2.1. La sociolingüística y la disponibilidad léxica	71
2.3. EL ESTUDIO DE LAS LENGUAS EN CONTACTO	76
2.3.1. Fenómenos de contacto lingüístico	88
3. DISPONIBILIDAD LÉXICA Y LENGUAS EN CONTACTO: ESTADO DE LA CUESTIÓN	103
3.1. ESTUDIOS EN TERRITORIOS NO HISPANOHABLANTES	104
3.2. ESTUDIOS SOBRE EL ESPAÑOL	113
3.2.1. El español en contacto con una lengua peninsular	113
3.2.1.1. <i>En contacto con el gallego y el portugués</i>	113
3.2.1.2. <i>En contacto con el euskera</i>	117
3.2.1.3. <i>En contacto con el catalán</i>	122
3.2.1.4. <i>En contacto con otras lenguas</i>	130
3.2.2. El español en contacto con otras lenguas no peninsulares	132
3.2.2.1. <i>En el norte de África</i>	132
3.2.2.2. <i>En EEUU</i>	136
3.2.2.3. <i>En el norte de Europa</i>	143
3.3. ESTUDIOS SOBRE EL CATALÁN (EN CONTACTO CON EL ESPAÑOL)	144
3.4. CONCLUSIONES PARCIALES	149
3.4.1. Perspectivas aplicadas	149
3.4.2. El enfoque comparado de dos lenguas	151
3.4.3. Variables analizadas en comunidades bilingües	153
3.4.4. Transferencias entre sistemas lingüísticos	155

4. MARCO SOCIOLINGÜÍSTICO DE LA INVESTIGACIÓN.....	159
4.1. EL CONTACTO ENTRE EL CASTELLANO Y EL CATALÁN EN CATALUÑA...	159
4.2. RASGOS DEL CASTELLANO EN TERRITORIOS CATALANOHABLANTES....	176
4.3. CARACTERIZACIÓN DE LA PROVINCIA DE LLEIDA.....	190
4.3.1. Contexto geográfico del estudio.....	190
4.3.2. Situación social y lingüística de la provincia	197
5. METODOLOGÍA.....	225
5.1. LA MUESTRA.....	225
5. 2. LAS VARIABLES SOCIALES	234
5. 3. OBTENCIÓN Y PROCESAMIENTO DE LOS DATOS	250
5. 4. CRITERIOS DE EDICIÓN.....	258
5.4.1. Pautas generales.....	259
5.4.2. Pautas relativas a los fenómenos de contacto lingüístico.....	268
5.4.3. Pautas específicas.....	270
5.4.3.1. <i>El cuerpo humano / El cos humà.....</i>	<i>270</i>
5.4.3.2. <i>Ropa y complementos / Roba i complements.....</i>	<i>272</i>
5.4.3.3. <i>Partes de la casa (sin los muebles) / Parts de la casa</i> <i>(sense els mobles)</i>	<i>275</i>
5.4.3.4. <i>Muebles de la casa / Mobles de la casa.....</i>	<i>277</i>
5.4.3.5. <i>Alimentos y bebidas / Aliments i begudes.....</i>	<i>279</i>
5.4.3.6. <i>Objetos colocados en la mesa para la comida/</i> <i>Objectes col·locats a taula pel dinar.....</i>	<i>283</i>
5.4.3.7. <i>La cocina y sus utensilios / La cuina i els seus</i> <i>utensilis.....</i>	<i>286</i>
5.4.3.8. <i>La escuela: muebles y materiales / L'escola: mobles i</i> <i>materials.....</i>	<i>288</i>
5.4.3.9. <i>Iluminación, calefacción y medios para airear un</i> <i>recinto/ Il·luminació, calefacció i mitjans per airejar un</i> <i>recinte.....</i>	<i>290</i>
5.4.3.10. <i>La ciudad/ La ciutat.....</i>	<i>292</i>
5.4.3.11. <i>El campo / El camp.....</i>	<i>294</i>
5.4.3.12. <i>Medios de transporte / Mitjans de transport.....</i>	<i>297</i>

5.4.3.13. <i>Trabajos del campo y del jardín / Treballs del camp i del jardí</i>	299
5.4.3.14. <i>Animales / Animals</i>	301
5.4.3.15. <i>Juegos y distracciones / Jocs i distraccions</i>	304
5.4.3.16. <i>Profesiones y oficios / Professions i oficis</i>	307
5.4.3.17. <i>Colores / Colors</i>	309
6. ANÁLISIS DE LOS DATOS	313
6.1. RESULTADOS GLOBALES DE LAS VARIABLES LINGÜÍSTICAS	316
6.1.1. Total de palabras por centro de interés	321
6.1.2. Promedio de palabras por centro de interés	328
6.1.3. Análisis estadístico de la diferencia en la producción de palabras	335
6.1.4. Total de vocablos por centro de interés	346
6.1.5. Densidad léxica e índice de cohesión de los centros de interés	353
6.1.6. Valoración de los resultados según las variables lingüísticas	361
6.1.7. Comparación con otras sintopías	366
6.2. INCIDENCIA CUANTITATIVA DE LAS VARIABLES SOCIALES	378
6.2.1. Variable social <i>sexo</i>	379
6.2.2. Variable social <i>tipo de centro</i>	390
6.2.3. Variable social <i>zona geográfica</i>	403
6.2.4. Variable social <i>nivel sociocultural</i>	416
6.2.5. Variable social <i>lengua materna</i>	440
6.2.6. Valoración de los resultados según las variables sociales	476
7. ESTUDIO DE LOS FENÓMENOS DE CONTACTO LINGÜÍSTICO	489
7.1. INTRODUCCIÓN	489
7.2. CLASIFICACIÓN DE LOS FENÓMENOS DE CONTACTO LINGÜÍSTICO	499
7.3. ANÁLISIS CUANTITATIVO	538
7.4. ESTUDIO DE LOS PRÉSTAMOS EN LAS DOS LENGUAS	562
8. CONCLUSIONES	571

9. BIBLIOGRAFÍA.....	595
-----------------------------	------------

CONTENIDO DEL CD

10. EL LÉXICO DISPONIBLE EN LENGUA CASTELLANA.....	631
10.1. DISPONIBILIDAD LÉXICA GENERAL.....	631
10.2. DISPONIBILIDAD LÉXICA POR ORDEN ALFABÉTICO.....	782
11. EL LÉXICO DISPONIBLE EN LENGUA CATALANA.....	865
11.1. DISPONIBILIDAD LÉXICA GENERAL.....	865
11.2. DISPONIBILIDAD LÉXICA POR ORDEN ALFABÉTICO.....	1021

APÉNDICES

APÉNDICE 1: LOS INFORMANTES.....	1111
APÉNDICE 2: LA ENCUESTA	1119
APÉNDICE 3: CÁLCULOS REALIZADOS CON EL PROGRAMA SPSS....	1123
A. Resultados generales castellano.....	1123
B. Resultados generales catalán.....	1132
C. Resultados centro a centro castellano.....	1141
D. Resultados centro a centro catalán.....	1168

ÍNDICE DE TABLAS

	<i>Páginas</i>
Tabla 2.1. Configuración de los grados de interferencia, cambio de código y préstamo en función de la integración lingüística y social.....	91
Tabla 5.1. Centros educativos seleccionados con su titularidad y ubicación territorial.....	228
Tabla 5.2. Distribución por <i>sexo</i> según distintos segmentos de población.....	235
Tabla 5.3. Distribución por <i>tipo de centro</i> según distintos parámetros.....	237
Tabla 5.4. Distribución por <i>zona geográfica</i> según distintos parámetros.....	240
Tabla 5.5. Distribución de los informantes según distintos criterios de estratificación social.....	244
Tabla 5.6. Distribución de los informantes según las variables de adscripción lingüística.....	246
Tabla 5.7. Distribución de las variables sociales analizadas en la muestra.....	250
Tabla 6.1. Índices cuantitativos generales en las dos lenguas con respecto al número de palabras.....	318
Tabla 6.2. Índices cuantitativos generales en las dos lenguas con respecto al número de vocablos.....	320
Tabla 6.3. Total de palabras por centro de interés en castellano y en catalán.....	322
Tabla 6.4. Rango de los centros de interés en cada una de las lenguas según la producción de palabras.....	326
Tabla 6.5. Promedio de palabras por informante y centro de interés en castellano y en catalán.....	328
Tabla 6.6. Diferencias entre el promedio y el total de producción de las dos lenguas con respecto al catalán (de más a menos).....	331
Tabla 6.7. Estadísticos de muestras relacionadas.....	336
Tabla 6.8. Correlaciones de muestras relacionadas.....	337
Tabla 6.9. Prueba de muestras relacionadas.....	338
Tabla 6.10. Total de vocablos por centro de interés en castellano y en catalán.....	346
Tabla 6.11. Rango de los centros de interés en cada una de las lenguas según el número de vocablos (de mayor a menor).....	348
Tabla 6.12. Densidad léxica por centro de interés en castellano y en catalán.....	355
Tabla 6.13. Índice de cohesión por centro de interés en castellano y en catalán.....	356
Tabla 6.14. Rango de los centros de interés en cada una de las lenguas según el índice de cohesión (de mayor a menor).....	358

Tabla 6.15. Rangos de los centros de interés en las dos lenguas según distintos parámetros.....	362
Tabla 6.16. Resumen de las diferencias halladas entre las dos lenguas según distintos parámetros.....	364
Tabla 6.17. Promedio de palabras por centro de interés en castellano en las provincias o comunidades autónomas españolas con más de una lengua en contacto.	371
Tabla 6.18. Promedio de palabras en catalán en Valencia y en Lleida.....	376
Tabla 6.19. Promedio de palabras por centro de interés según la variable social <i>sexo</i> .	380
Tabla 6.20. Análisis estadístico de la diferencia según el sexo. Valor de la ‘p’ en las dos lenguas.....	383
Tabla 6.21. Total de vocablos por centro de interés según la variable social <i>sexo</i>	389
Tabla 6.22. Índice de cohesión por centro de interés según la variable social <i>sexo</i>	390
Tabla 6.23. Promedio de palabras por centro de interés según la variable social <i>tipo de centro</i>	391
Tabla 6.24. Análisis estadístico de la diferencia según el centro educativo. Valor de la ‘p’ en las dos lenguas.....	395
Tabla 6.25. Total de vocablos por centro de interés según la variable social <i>tipo de centro</i>	400
Tabla 6.26. Índice de cohesión por centro de interés según la variable social <i>tipo de centro</i>	402
Tabla 6.27. Promedio de palabras por centro de interés según la variable <i>zona geográfica</i>	403
Tabla 6.28. Análisis estadístico de la diferencia según la zona geográfica. Valor de la ‘p’ en las dos lenguas.....	407
Tabla 6.29. Total de vocablos por centro de interés según la variable social <i>zona geográfica</i>	413
Tabla 6.30. Índice de cohesión según la variable social <i>zona geográfica</i>	415
Tabla 6.31. Promedio de palabras por centro de interés según la variable social <i>nivel sociocultural</i>	417
Tabla 6.32. Análisis estadístico de la diferencia según el nivel sociocultural. Valor de la ‘p’ en las dos lenguas.....	425
Tabla 6.33. Total de vocablos por centros de interés según la variable social <i>nivel sociocultural</i>	437
Tabla 6.34. Índice de cohesión por centros de interés según la variable social <i>nivel sociocultural</i>	439
Tabla 6.35. Promedio de palabras por centro de interés según la variable <i>lengua materna</i>	442
Tabla 6.36. Promedio de palabras por centro de interés	443

según la variable <i>lengua materna</i> en lengua castellana.....	
Tabla 6.37. Promedio de palabras por centro de interés según la variable <i>lengua materna</i> en lengua catalana.....	446
Tabla 6.38. Análisis estadístico de la diferencia según la lengua materna. Valor de la ‘p’ en las dos lenguas.....	455
Tabla 6.39. Estadísticos de muestras relacionadas (informantes castellanohablantes).	466
Tabla 6.40. Correlaciones de muestras relacionadas (informantes castellanohablantes).....	466
Tabla 6.41. Prueba de muestras relacionadas (informantes castellanohablantes).....	467
Tabla 6.42. Estadísticos de muestras relacionadas (informantes catalanohablantes)...	469
Tabla 6.43. Correlaciones de muestras relacionadas (informantes catalanohablantes).....	469
Tabla 6.44. Prueba de muestras relacionadas (informantes catalanohablantes).....	469
Tabla 6.45. Total de vocablos por centro de interés según la variable social <i>lengua materna</i>	472
Tabla 6.46. Índice de cohesión por centro de interés según la variable social <i>lengua materna</i>	475
Tabla 6.47. Valor de la ‘p’ en cada corpus según las variables sociales analizadas....	480
Tabla 6.48. Resumen de la significación de las diferencias halladas en cada uno de los dos corpus (número de centros de interés donde $p < 0.05$).....	481
Tabla 6.49. Relación de las variables sociales que inciden en cada uno de los centros de interés en las dos lenguas analizadas.....	482
Tabla 6.50. Comparación de la producción y la riqueza léxicas de los distintos subgrupos en las dos lenguas de la encuesta.....	486
Tabla 7.1. Voces marcadas lexicográficamente como en desuso o antiguas en el corpus castellano.....	495
Tabla 7.2. Voces marcadas lexicográficamente como en desuso o antiguas en el corpus catalán.....	496
Tabla 7.3. Voces con entrada en el DRAE consideradas catalanas en este estudio...	497
Tabla 7.4. El cuerpo humano.....	504
Tabla 7.5. El cos humà.....	505
Tabla 7.6. Ropa y complementos.....	506
Tabla 7.7. Roba i complements.....	507
Tabla 7.8. Partes de la casa (sin los muebles).....	508

Tabla 7.9. Parts de la casa (sense els mobles).....	509
Tabla 7.10. Muebles de la casa.....	510
Tabla 7.11. Mobles de la casa.....	511
Tabla 7.12. Alimentos y bebidas.....	512
Tabla 7.13. Aliments i begudes.....	513
Tabla 7.14. Objetos colocados en la mesa para la comida.....	514
Tabla 7.15. Objectes col·locats a taula pel dinar.....	515
Tabla 7.16. La cocina y sus utensilios.....	516
Tabla 7.17. La cuina i els seus utensilis.....	517
Tabla 7.18. La escuela: muebles y materiales.....	518
Tabla 7.19. L'escola: mobles i materials.....	519
Tabla 7.20. Iluminación, calefacción y medios para airear un recinto.....	520
Tabla 7.21. Il·luminació, calefacció i mitjans per airejar un recinte.....	521
Tabla 7.22. La ciudad.....	522
Tabla 7.23. La ciutat.....	523
Tabla 7.24. El campo.....	524
Tabla 7.25. El camp.....	525
Tabla 7.26. Medios de transporte.....	526
Tabla 7.27. Mitjans de transport.....	527
Tabla 7.28. Trabajos del campo y del jardín.....	528
Tabla 7.29. Treballs del camp i del jardí.....	529
Tabla 7.30. Animales.....	530
Tabla 7.31. Animals.....	531
Tabla 7.32. Juegos y distracciones.....	532
Tabla 7.33. Jocs i distraccions.....	533
Tabla 7.34. Profesiones y oficios.....	534

Tabla 7.35. Professions i oficis.....	535
Tabla 7.36. Colores.....	536
Tabla 7.37. Colors.....	537
Tabla 7.38. Frecuencia absoluta y relativa de los fenómeno de contacto lingüístico en las dos lenguas.....	538
Tabla 7.39. Respuestas con algún fenómenos de contacto lingüístico sobre el total de palabras en ambas lenguas.....	544
Tabla 7.40. Porcentaje de respuestas por centro de interés con algún fenómeno de contacto lingüístico en castellano (frecuencia absoluta y relativa).....	545
Tabla 7.41. Porcentaje de respuestas por centro de interés con algún fenómeno de contacto lingüístico sobre el total de palabras en catalán.....	547
Tabla 7.42. Rango de los centros de interés en cada una de las lenguas según el porcentaje de respuestas con algún fenómeno de contacto lingüístico sobre el total de palabras.....	551
Tabla 7.43. Fenómenos de contacto lingüístico sobre el total de vocablos en las dos lenguas.....	552
Tabla 7.44. Porcentaje de fenómenos de contacto lingüístico sobre el total de vocablos en castellano por centro de interés.....	552
Tabla 7.45. Porcentaje de fenómenos de contacto lingüístico sobre el total de vocablos en catalán por centro de interés.....	554
Tabla 7.46. Rango de los centros de interés en cada una de las lenguas según el porcentaje de fenómenos de contacto lingüístico sobre el total de vocablos (de mayor a menor).....	557
Tabla 7.47. Préstamos en vías de integración en el corpus castellano.....	562
Tabla 7.48. Préstamos integrados en el corpus castellano.....	563
Tabla 7.49. Préstamos en vías de integración en el corpus catalán.....	564
Tabla 7.50. Préstamos integrados en el corpus catalán.....	567

ÍNDICE DE GRÁFICOS

	<i>Páginas</i>
Gráfico 4.1. Conocimiento de lenguas de la población de 15 años y más, en porcentajes.....	162
Gráfico 4.2. Lengua inicial, de identificación y habitual de la población de 15 años y más, en porcentajes.....	163
Gráfico 4.3. Lengua habitual según lugar de nacimiento de la población de 15 años y más, en porcentajes.....	164
Gráfico 4.4. Total de habitantes de las cuatro provincias catalanas en el año 2002 (año de aplicación del estudio) y en el 2011.....	192
Gráfico 4.5. Densidad de población de las cuatro provincias catalanas en el año 2002 (año de aplicación del estudio) y en el 2011.....	192
Gráfico 4.6. Reparto de la población en la provincia de Lleida en el año 2002 (año de aplicación del estudio).....	194
Gráfico 4.7. Reparto de la población en la provincia de Lleida en el año 2002 (año de aplicación del estudio) y en el 2011.....	195
Gráfico 4.8. Evolución de los diferentes sectores económicos en la provincia entre 1960 y 1991.....	197
Gráfico 4.9. Distribución de la población de la provincia de Lleida según sexo en porcentajes. Año 2002.....	199
Gráfico 4.10. Población según sexo y edad en la provincia de Lleida. Año 2002.....	200
Gráfico 4.11. Distribución de la población de la provincia de Lleida según el lugar de nacimiento en porcentajes. Año 2002.....	201
Gráfico 4.12. Distribución de la población del municipio de Lleida según el lugar de nacimiento en porcentajes. Año 2002.....	202
Gráfico 4.13. Distribución de la población de Cataluña según el lugar de nacimiento en porcentajes. Año 2002.....	202
Gráfico 4.14. Población de la provincia de Lleida nacida en Cataluña según la provincia de nacimiento en porcentajes. Año 2002.....	204
Gráfico 4.15. Población de la provincia de Lleida nacida en el resto del Estado según la comunidad autónoma en porcentajes. Año 2002.....	204
Gráfico 4.16. Conocimiento del catalán en la provincia de Lleida en porcentajes. Año 2001.....	205
Gráfico 4.17. Conocimiento del catalán en la provincia de Lleida según la franja de edad en porcentajes. Año 2001.....	205

Gráfico 4.18. Conocimiento del catalán en la provincia de Lleida según la comunidad autónoma de nacimiento en porcentajes. Año 2001.....	206
Gráfico 4.19. Población del <i>Àmbit de Ponent</i> según primera lengua, lengua propia y lengua habitual, en porcentajes. Año 2003.....	208
Gráfico 4.20. Población de <i>l'Alt Pirineu i Aran</i> según primera lengua, lengua propia y lengua habitual, en porcentajes. Año 2003.....	209
Gráfico 4.21. Usos lingüísticos en ámbitos interpersonales en <i>l'Àmbit de Ponent</i> , en porcentajes. Año 2003.....	210
Gráfico 4.22. Usos lingüísticos en ámbitos interpersonales en <i>l'Alt Pirineu i Aran</i> , en porcentajes. Año 2003.....	211
Gráfico 4.23 Usos lingüísticos en ámbitos de consumo y servicio en <i>l'Àmbit de Ponent</i> , en porcentajes. Año 2003.....	212
Gráfico 4.24. Usos lingüísticos en ámbitos de consumo y servicio en <i>l'Alt Pirineu i Aran</i> , en porcentajes. Año 2003.....	212
Gráfico 4.25. Usos lingüísticos en el consumo de los medios de comunicación en <i>l'Àmbit de Ponent</i> . Media de horas. Año 2003.....	213
Gráfico 4.26. Usos lingüísticos en el consumo de los medios de comunicación en <i>l'Alt Pirineu i Aran</i> . Media de horas. Año 2003.....	214
Gráfico 4.27. Conocimientos de aranés de la población de 15 años y más en <i>la Vall d'Aran</i> , en porcentajes. Año 2008.....	222
Gráfico 4.28. Lengua inicial, de identificación y habitual de la población de 15 años y más en <i>la Vall d'Aran</i> , en porcentajes. Año 2008.	223
Gráfico 5.3. Distribución según la variable <i>sexo</i>	236
Gráfico 5.4. Distribución según la variable <i>tipo de centro</i>	237
Gráfico 5.5. Distribución según la variable <i>zona geográfica</i>	240
Gráfico 5.6. Distribución según la variable <i>nivel sociocultural</i>	244
Gráfico 5.7. Distribución según la variable <i>lengua materna</i>	247
Gráfico 5.8. Distribución según la variable <i>lengua habitual</i>	247
Gráfico 5.9. Distribución según el <i>tipo de centro</i> y el <i>nivel sociocultural</i>	248
Gráfico 5.10. Distribución según la <i>zona geográfica</i> y la <i>lengua materna</i>	249
Gráfico 5.11. Distribución según el <i>tipo de centro</i> y la <i>lengua materna</i>	250

Gráfico 6.1. Total de palabras en castellano y en catalán.....	318
Gráfico 6.2. Promedio de palabras por informante y centro de interés en castellano y en catalán.....	318
Gráfico 6.3. Total de vocablos en castellano y en catalán.....	320
Gráfico 6.4. Promedio de vocablos por centro de interés en castellano y en catalán...	320
Gráfico 6.5. Frecuencia relativa del total de palabras por centro de interés en castellano.....	325
Gráfico 6.6. Frecuencia relativa del total de palabras por centro de interés en catalán.	325
Gráfico 6.7. Promedio de palabras por centro de interés en castellano y en catalán....	330
Gráfico 6.8. Total de vocablos por centro de interés en castellano y en catalán.....	350
Gráfico 6.9. Índice de cohesión por centro de interés en castellano y en catalán.....	356
Gráfico 6.10. Promedio general de palabras en castellano en distintas sintopías españolas monolingües.....	368
Gráfico 6.11 Promedio general de palabras en castellano en distintas sintopías españolas donde convive más de una lengua.....	370
Gráfico 6.12. Promedio de palabras por centro de interés en distintas sintopías españolas con más de una lengua en contacto (media) y en Lleida.....	373
Gráfico 6.13. Promedio de palabras por centro de interés en catalán en Lleida y Valencia.....	377
Gráfico 6.14. Promedio de palabras por centro de interés según la variable social <i>sexo</i> en lengua castellana.....	381
Gráfico 6.15. Promedio de palabras por centro de interés según la variable social <i>sexo</i> en lengua catalana.....	382
Gráfico 6.16. Comparación de la producción en castellano y en catalán de los varones.....	386
Gráfico 6.17. Comparación de la producción en castellano y en catalán de las mujeres.....	387
Gráfico 6.18. Promedio de palabras por centro de interés según la variable social <i>tipo de centro</i> en lengua castellana.....	393
Gráfico 6.19. Promedio de palabras por centro de interés según la variable social <i>tipo de centro</i> en lengua catalana.....	394
Gráfico 6.20. Comparación de la producción en castellano y en catalán	398

de los informantes de los centros públicos.....	
Gráfico 6.21. Comparación de la producción en castellano y en catalán de los informantes de los centros privados.....	399
Gráfico 6.22. Promedio de palabras por centro de interés según la variable social <i>zona geográfica</i> en lengua castellana.....	405
Gráfico 6.23. Promedio de palabras por centro de interés según la variable social <i>zona geográfica</i> en lengua catalana.....	406
Gráfico 6.24. Comparación de la producción en castellano y en catalán de los informantes urbanos.....	409
Gráfico 6.25. Comparación de la producción en castellano y en catalán de los informantes rurales.....	411
Gráfico 6.26. Promedio de palabras por centro de interés según la variable social <i>nivel sociocultural</i> en lengua castellana.....	420
Gráfico 6.27. Promedio de palabras por centro de interés según la variable social <i>nivel sociocultural</i> en lengua catalana.....	422
Gráfico 6.28. Comparación de la producción en castellano y en catalán de los informantes del nivel bajo.....	432
Gráfico 6.29. Comparación de la producción en castellano y en catalán de los informantes del nivel medio.....	433
Gráfico 6.30. Comparación de la producción en castellano y en catalán de los informantes del nivel alto.....	435
Gráfico 6.31. Promedio de palabras por centro de interés según la variable <i>lengua materna</i> en lengua castellana.....	444
Gráfico 6.32. Promedio de palabras por centro de interés según la variable <i>lengua materna</i> en lengua catalana.....	447
Gráfico 6.33. Promedios globales de los cuatro grupos lingüísticos en el corpus castellano.....	449
Gráfico 6.34. Promedios globales de los cuatro grupos lingüísticos en el corpus catalán.....	449
Gráfico 6.35. Promedio de palabras por centro de interés según la variable <i>lengua materna</i> en lengua castellana (solo LM: castellano y LM: catalán).....	452
Gráfico 6.36. Promedio de palabras por centro de interés según la variable <i>lengua materna</i> en lengua catalana (solo LM: castellano y LM: catalán).....	454
Gráfico 6.37. Comparación de la producción en castellano y en catalán de los informantes cuya lengua materna es el castellano y el catalán (ambas lenguas).....	462

Gráfico 6.38. Comparación de la producción en castellano y en catalán de los informantes cuya lengua materna es el aranés.....	464
Gráfico 6.39. Comparación de la producción en castellano y en catalán de los informantes cuya lengua materna es el castellano.....	465
Gráfico 6.40. Comparación de la producción de los informantes cuya lengua materna es el catalán.....	468
Gráfico 6.41. Resumen de las diferencias en la producción en lengua castellana según las distintas variables sociales.....	477
Gráfico 6.42. Resumen de las diferencias en la producción en lengua catalana según las distintas variables sociales.....	479
Gráfico 6.43. Resumen de las diferencias en la producción en castellano y en catalán de cada grupo de informantes según las variables sociales.	484
Gráfico 7.1. Distribución de los fenómenos de contacto lingüístico en el corpus castellano.....	539
Gráfico 7.2. Distribución de los fenómenos de contacto lingüístico en el corpus catalán.....	539
Gráfico 7.3. Porcentaje de respuestas con algún fenómeno de contacto lingüístico sobre el total de palabras por centro de interés en las dos lenguas.....	550
Gráfico 7.4. Porcentaje de fenómenos de contacto lingüístico sobre el total de vocablos por centro de interés en las dos lenguas.....	556

ÍNDICE DE MAPAS

	<i>Páginas</i>
Mapa 4.1. Ubicación de la provincia de Lleida y de su capital en Cataluña y España.	191
Mapa 4.2. Ubicación de las comarcas de la provincia de Lleida.....	193
Mapa 4.3. Extensión del catalán <i>nord-occidental</i> y sus distintos subdialectos.....	216
Mapa 4.4. Mapa de <i>la Vall d'Aran</i>	220
Mapa 5.1. Mapa de la ciudad de Lleida dividida en barrios con los centros educativos seleccionados.....	230
Mapa 5.3. Mapa de la provincia de Lleida con los municipios seleccionados.....	233

1. INTRODUCCIÓN

No han sido pocos hasta la fecha los intentos, llevados a cabo desde el ámbito de la lingüística, de establecer el léxico que habitualmente emplea una comunidad de habla determinada. La investigación que aquí presentamos nace también de este deseo y se enmarca dentro de los estudios de léxico-estadística que otorgan, mediante diversos métodos, índices estadísticos a las unidades léxicas que integran una lengua.

En los años 50, un grupo de investigadores franceses intentó determinar, con una finalidad esencialmente pedagógica, cuál era el léxico “fundamental” de la lengua francesa. De este modo daban comienzo los estudios de disponibilidad léxica que cuentan, en la actualidad, con un notable desarrollo en el mundo hispánico. Humberto López Morales continuó la senda abierta por los lingüistas franceses y, con un entusiasmo verdaderamente admirable, asumió el reto de establecer, mediante un ambicioso proyecto coordinado, el léxico disponible general del mundo hispánico. Nuestro trabajo pretende aportar un grano de arena a este proyecto y ocuparse de una parcela inexplorada hasta el momento: el estudio de la disponibilidad léxica en la provincia de Lleida.

Con la finalidad de retratar con fidelidad la realidad lingüística de nuestra comunidad, hemos extendido el análisis del léxico disponible a las dos lenguas usadas y cooficiales en Lleida, el castellano y el catalán, puesto que, cuando se recogieron los datos para elaborar nuestro corpus de trabajo, la lengua catalana no había sido antes estudiada desde este punto de vista y entendíamos que era imprescindible hacerlo. Esta circunstancia, unida a que contamos con una única muestra de informantes que completa la encuesta en las dos lenguas, es una novedad importante en nuestra disciplina, en la que ningún estudio actual analiza de forma comparada, como hacemos en el presente trabajo, el léxico disponible de dos lenguas utilizadas en un mismo territorio. Del mismo modo, el hecho de que sean los mismos informantes quienes completen la encuesta en su lengua materna y en su segunda lengua ofrece posibilidades de análisis interesantísimas, algunas de las cuales exceden los límites estrictos de esta monografía, y que, sin duda, suponen un avance en el terreno del estudio de las lenguas en contacto.

De hecho, la metodología de la disponibilidad, es decir, el control del número de asociaciones con palabras estímulo traducidas en dos idiomas, se ha utilizado como método para conocer los efectos del bilingüismo en la vertiente lingüística y descubrir así cuál de las dos lenguas predomina en el individuo bilingüe. Las asociaciones realizadas en cada lengua pueden diferir en función de los referentes culturales que se asocien a cada una de ellas o según la distribución en ámbitos de las dos lenguas, teniendo en cuenta la situación sociolingüística concreta que se dé. Del mismo modo, los investigadores que nos preceden apuntan a que la disponibilidad léxica de un hablante bilingüe será mayor en una lengua u otra en función del área semántica analizada (Appel y Muysken 1996: 153-154, 170, Baetens 1989: 22). En el contacto entre el castellano y el catalán, ya se ha utilizado la metodología de la disponibilidad, aunque con otro nombre, como instrumento para conocer la competencia lingüística del individuo bilingüe (Bastardas 1986, Boix 1993).

De otra parte, teníamos un gran interés por estudiar la influencia mutua de los dos idiomas, es decir, en qué medida el léxico del español¹, o el del catalán, contiene elementos de la otra lengua y si la presencia de estos varía o no en función del campo léxico analizado. En época temprana, Mackey (1971a) o Humbley (1974: 51-52) ya consideraron que la disponibilidad léxica era una herramienta válida para precisar la extensión de una lengua extranjera en el vocabulario básico de otra lengua o en un campo semántico concreto de la misma. De otra parte, en el caso concreto del contacto entre el castellano y el catalán faltan estudios cuantitativos de frecuencia que permitan valorar si una interferencia está integrada o no, dado que, aunque se conocen cuáles son los elementos más habitualmente transferidos entre las dos lenguas, no hay apenas estudios basados en corpus que analicen la difusión social de estos elementos², hecho que permitiría conocer su grado de integración en el sistema y su distribución social (Hernández García 1998a: 578, Vila i Moreno³ 2004: 43-45).

En este trabajo seguimos un enfoque sociolingüístico, presente desde los mismos inicios de la disponibilidad, a pesar de que los resultados que ofrecen estos estudios,

¹Utilizamos en este trabajo de forma indistinta y como sinónimos exactos, los términos *castellano* y *español*.

² Algunas excepciones son Vila i Moreno (1996) o Galindo (2006).

³ Es costumbre tradicional catalana enlazar los dos apellidos mediante la conjunción copulativa, tal como aquí hacemos, respetando la voluntad del propio autor.

como es sabido, se aplican a numerosas disciplinas, como la psicolingüística, la dialectología, la etnolingüística o la lingüística aplicada. Sin embargo, concretamente desde el enfoque sociolingüístico se han realizado las principales investigaciones españolas de disponibilidad léxica. Hoy en día ha quedado sobradamente demostrado que puede establecerse una estratificación social a partir de los datos que aporta el análisis conjunto de los campos léxicos y de las distintas variables sociales. Por esta razón, partimos de la hipótesis de que las variables lingüísticas —esto es, los distintos *centros de interés*⁴— y las sociales coviarán en menor o mayor medida con los resultados y que, tras realizar el análisis pertinente, podremos llegar a conclusiones significativas en cuanto a la distribución social del léxico.

El corpus de este trabajo lo constituyen las 480 encuestas que completaron 240 informantes preuniversitarios de distintos puntos de la provincia durante el año 2002. Durante cuatro años (de 2000 a 2004) realizamos todo el trabajo de campo y procedimos a la informatización de los datos, gracias a una beca predoctoral concedida por *l'Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR)* de la *Generalitat* de Cataluña. También fruto del trabajo de esos años es nuestro estudio exploratorio realizado con 48 informantes de la ciudad de Lleida (Serrano Zapata 2003b). De las conclusiones alcanzadas en ese estudio, realizado únicamente sobre el castellano, así como de los trabajos ya finalizados en distintos puntos del mundo hispánico (Samper, Bellón y Samper Hernández 2003, Samper Padilla y Samper Hernández 2006) y en otras comunidades bilingües (*vid.* 3.4.) extraemos las siguientes hipótesis:

- 1) Sobre la covariación de los datos con las variables lingüísticas, es decir, los centros de interés, consideramos que:
 - (a) La producción de palabras variará en función de las variables lingüísticas. En este sentido, las dos lenguas mostrarán un patrón de variación semejante en cuanto al número de palabras recabado, y su distribución, en cambio, diferirá más en relación con el número de vocablos (o palabras distintas) obtenido.
 - (b) Tanto la productividad (el número de respuestas hallado bajo cada estímulo) como la riqueza léxica (el número de palabras diferentes) serán mayores en

⁴ Utilizamos las expresiones “centros de interés”, “campos léxicos”, “estímulos verbales”, “categorías semánticas” etc. como sinónimas en esta monografía, aunque reconocemos que no todos estos términos podrían considerarse como equivalentes en algunos casos.

lengua catalana. Este hecho reflejará, de un lado, la gran extensión social que tiene esta lengua en la realidad de la provincia y, de otro, la circunstancia de que toda la muestra ha sido escolarizada en lengua vernácula y, además, la mayor parte de ella también tiene esta lengua como materna.

- 2) En relación con las cinco variables sociales contempladas en esta investigación, creemos que:
 - (a) la estratificación social que se dé en las dos lenguas tenderá a ser similar, esto es, las variables mostrarán significación estadística desde un punto de vista sociolingüístico (o bien en las dos lenguas o bien en ninguna de ellas);
 - (c) el léxico disponible de los dos idiomas coviará de forma significativa con el factor *sexo*, otorgando una productividad mayor a las mujeres;
 - (d) en relación con la variable *tipo de centro*, los promedios serán superiores entre los informantes de los centros privados de la provincia tanto en castellano como en catalán;
 - (e) el factor *zona geográfica* discriminará las respuestas de los informantes a favor del alumnado de las instituciones educativas ubicadas en la capital;
 - (f) la variable *nivel sociocultural* estratificará el léxico disponible en función del estrato social del informante de forma que, a un nivel sociocultural mayor, le corresponderá un promedio superior de respuestas en los dos idiomas y
 - (g) con respecto a la *lengua materna*, es de esperar que en lengua castellana presenten mejores promedios aquellos informantes cuya lengua materna sea el castellano y en lengua catalana suceda lo mismo con el alumnado con el catalán como primera lengua⁵.
- 3) Con respecto a las diferencias que se apreciarán entre los distintos subgrupos sociales en función de la lengua en que completen la encuesta⁶, pensamos que:
 - (a) la mayoría de subgrupos derivados de la muestra (varones, mujeres, informantes urbanos, rurales, etc.) mostrará una preponderancia a favor de la lengua catalana y obtendrá resultados superiores en dicha lengua. Únicamente, la estratificación de la población según la variable *lengua*

⁵ Desde la investigación de Hernández Muñoz (2006), conocemos que las variables cognitivas implicadas en la disponibilidad son la tipicidad, la familiaridad y la edad de adquisición, por tanto, suponemos que estas dos últimas características influirán en que los informantes alcancen mejores índices cuantitativos en su lengua materna respectiva.

⁶ Dado que nuestro estudio es pionero en la comparación de la producción de unos mismos estudiantes en dos lenguas, no contamos con investigaciones anteriores cuyas conclusiones puedan servirnos para extraer las hipótesis.

materna ofrecerá mejores resultados de los informantes castellanohablantes a favor del castellano.

- 4) En cuanto al análisis de los fenómenos de contacto lingüístico en los dos corpus, consideramos que:
- (h) La influencia del léxico de una lengua sobre el de la otra será similar en los dos idiomas. Hay que decir, sin embargo, que los pocos estudios realizados hasta la fecha en los que se analiza, como en el nuestro, las interferencias que se detectan en dos corpus castellano-catalán (*v.gr.* Vila i Moreno 1996 o Galindo 2006), otorgan una mayor influencia del castellano sobre el catalán que en sentido inverso. No obstante, en el territorio concreto que analizamos, la provincia de Lleida, no existe ningún trabajo que estudie mediante dos corpus de forma comparada el peso de los fenómenos de contacto lingüístico y, como es sabido, la situación sociolingüística es muy distinta según el lugar de Cataluña que se considere. En Lleida, el peso de la comunidad cuya lengua materna es la vernácula es de tal envergadura (el 63.95%, *vid.* 5.2), que nos hace creer en una mayor resistencia léxica del catalán.
 - (i) El alcance de los elementos transferidos de un sistema lingüístico al otro en las dos lenguas será muy distinto en función del centro de interés que se analice. Algunos estudios comparados (Bartol 2011) y nuestros propios datos de 2003 ya muestran que algunas áreas son más propicias que otras a la transferencia de elementos.

Por lo tanto, el objetivo principal que nos proponemos al iniciar esta investigación es el de conocer cuál es el léxico disponible en lengua castellana y catalana en diferentes campos semánticos de una muestra representativa de escolares de la provincia de Lleida. Al mismo tiempo, pretendemos conocer la incidencia que tienen ciertos factores sociales en la evocación del vocabulario en una comunidad bilingüe. Para ello seguimos unas pautas metodológicas estrictas ya avaladas por todas las investigaciones que se suman al Proyecto Panhispánico de Estudio de la Disponibilidad y que nos preceden. De modo similar, nos interesa descubrir en qué medida el léxico de cada una de las lenguas en contacto se ve influido por el de la otra lengua cooficial, qué valor alcanza en el total de los datos la presencia de elementos lingüísticos procedentes de la otra lengua oficial en la encuesta —ya sea el castellano o el catalán— y qué tipología de elementos de contacto lingüístico obtenemos tras el análisis.

Con el fin de poder alcanzar estos objetivos, nos hemos ocupado, en el capítulo que sigue, de los presupuestos teóricos en los que se sustenta nuestro trabajo, aunque una característica básica del mismo es precisamente la multidisciplinariedad desde la que se aborda. Por este motivo, junto con una revisión de los fundamentos básicos de la sociolingüística y la léxico-estadística, hemos realizado un panorama general de las investigaciones que se enmarcan en el Proyecto Panhispánico de Estudio de la Disponibilidad, así como un repaso de las obras pioneras que iniciaron esta línea de investigación. El enfoque comparado de las dos lenguas nos ha exigido profundizar en las bases de los estudios sobre lenguas en contacto y, concretamente, en la tipología de fenómenos de contacto habitualmente detectados en este tipo de trabajos.

En el tercer capítulo, hemos elaborado un estado de la cuestión dedicado, con exclusividad, a las obras que analizan la disponibilidad léxica en lugares donde coexisten dos sistemas lingüísticos de forma habitual. De este modo, hemos reseñado tanto los estudios realizados sobre otras lenguas distintas al español y, por ende, llevados a cabo en territorios no hispanohablantes, como los referidos ya a la lengua española en relación con otras lenguas dentro y fuera de la Península Ibérica, y también los que versan sobre la lengua catalana en contacto con el castellano. Son interesantes las conclusiones parciales alcanzadas sobre el análisis de los elementos lingüísticos transferidos, las variables tenidas en cuenta en estas investigaciones y las distintas perspectivas aplicadas en este tipo de trabajos, junto con las condiciones metodológicas que los rodean⁷.

Con el fin de elaborar un panorama global que explicase la situación sociolingüística de nuestra provincia, nos hemos remontado al contacto del español y del catalán a lo largo de la historia haciendo hincapié en las circunstancias actuales que lo explican y ahondando en los rasgos lingüísticos tradicionalmente detectados en los catalanohablantes al hablar en castellano, especialmente los referidos al nivel del léxico.

⁷ Es importante aclarar en este punto que la metodología de la disponibilidad procede de la Psicología empírica y es una prueba de fluencia de categoría semántica, que también contiene elementos de la asociación libre de palabras (Hernández Muñoz 2006: 121-122). Por tanto, hay numerosos trabajos, realizados con bilingües, que utilizan esta técnica, pero nosotros solo nos referimos aquí a los que tienen como objetivo elaborar diccionarios de léxico disponible y se enmarcan, por tanto, dentro de la lingüística.

En el marco de nuestra provincia, además de una breve exposición geográfica de sus características más relevantes, hemos analizado la situación de contacto lingüístico que impera, contando para ello con los datos del censo general de la población del año 2002, con los del censo lingüístico de 2001 y con los de la encuesta de usos lingüísticos del año 2003 y del 2008. Hemos intentado de este modo retratar la situación en los años próximos a nuestro estudio, a la vez que hemos explicado los pequeños cambios que se observan con respecto a la actualidad. La elaboración de este panorama, que constituye el capítulo cuarto de la tesis, ha sido necesaria ante el vacío bibliográfico existente al respecto. Del mismo modo, para poder describir adecuadamente el contacto que se produce en nuestra provincia hemos realizado un resumen de los rasgos dialectales y de las variedades del catalán presentes en Lleida, así como un breve repaso de las obras lexicográficas que se han ocupado de caracterizar esta variedad del catalán⁸.

El capítulo quinto se dedica en exclusividad a la descripción de la metodología llevada a cabo para la aplicación del trabajo de campo y a la edición de los datos que posteriormente conformarán los dos corpus. Como ya hemos indicado, la mayoría de las decisiones metodológicas están condicionadas por nuestra adscripción al Proyecto Panhispánico de Estudio de la Disponibilidad y ya se han aplicado en las numerosas investigaciones que nos preceden. En este capítulo también exponemos con precisión las características de nuestros informantes, comparando su distribución en función de las distintas variables sociales con la del universo del que procede la muestra. Asimismo, detallamos las condiciones de aplicación de la encuesta y dedicamos muchas páginas a la exposición de los criterios de edición, asunto peliagudo donde los haya, en el que nos hemos detenido largas horas por la dificultad que entraña editar dos lenguas a la vez y justificar las decisiones tomadas.

En el capítulo sexto abordamos ya el análisis en sí de los datos, utilizando para ello el programa SPSS (*Statistical Package for the Social Sciences*), con el que determinamos la influencia que tienen las distintas variables lingüísticas —los centros de interés— y las cinco variables sociales elegidas sobre la productividad léxica de los

⁸ La amplitud del capítulo 4, y más concretamente del apartado 4.3.2., se justifica por la ausencia de estudios sobre la situación sociolingüística de la provincia. En general, las investigaciones se han centrado en el área de Barcelona y en el uso de la lengua catalana. El único trabajo sociolingüístico sobre la ciudad de Lleida (Agustí 1998) se limita a la capital y analiza, exclusivamente, el conocimiento del catalán según el barrio y según el perfil de la población.

informantes en lengua castellana y catalana. Partiendo del análisis de las palabras que enumera cada informante bajo cada estímulo verbal, hemos observado si las diferencias entre la producción en cada una de las dos lenguas son significativas, así como si lo son también las halladas entre las medias de cada uno de los subgrupos sociales en todas las áreas semánticas y en cada una de las dos lenguas. Siguiendo la perspectiva comparada en la que nos basamos, hemos cotejado los índices cuantitativos que alcanzan, en distintos parámetros, las dos lenguas estudiadas y hemos comparado estos resultados con los obtenidos en trabajos realizados con anterioridad en otras sintopías. Las variables sociales elegidas se han estudiado de forma independiente en cada corpus (por ejemplo, *sexo: varón/mujer*-corpus castellano, *sexo: varón/mujer*-corpus catalán) y también se han descrito las principales tendencias de cada subgrupo social con relación a la producción de las dos lenguas (por ejemplo, *varón*- castellano/catalán, *mujer*- castellano/catalán). Este último análisis es meramente descriptivo dado que no hemos sometido estos datos a ningún programa estadístico que los avale. Cabe destacar que las características tan peculiares de nuestra investigación —donde los mismos informantes han completado la encuesta en las dos lenguas— ofrecen muchas ventajas en el análisis cuantitativo de los datos.

El análisis cualitativo, que corresponde al capítulo séptimo, se basa en el estudio de todos los elementos de contacto lingüístico hallados en los dos corpus. Hemos clasificado todos los trasvases lingüísticos encontrados, distribuyéndolos en cada centro de interés y analizando el peso que tienen estos elementos en el conjunto de los datos, tanto sobre el total de vocablos como sobre el total de respuestas y, en particular, sobre cada campo semántico. El análisis cualitativo se cierra con un análisis exhaustivo de los préstamos hallados en los dos corpus.

En el capítulo octavo recogemos las conclusiones finales destacando los resultados más relevantes de nuestro estudio. A continuación, incluimos el apartado bibliográfico, que recoge todas las referencias usadas para la elaboración de este trabajo y citadas en el mismo. Los últimos capítulos constituyen los diccionarios de disponibilidad en sí, donde se recoge el elenco de voces que han sido actualizadas en las encuestas, en cada uno de los centros de interés tenidos en cuenta, ordenadas según el índice de disponibilidad y el listado general del léxico disponible por orden alfabético tanto en castellano (capítulo 10) como en catalán (capítulo 11).

Este trabajo ofrece como colofón tres apéndices, en los que presentamos, en primer lugar, las características de cada uno de los informantes que constituyen nuestra muestra; en segundo lugar, el instrumento de investigación utilizado, esto es, la encuesta de disponibilidad, y en tercer lugar, todos los cálculos estadísticos realizados con el programa SPSS y utilizados en el análisis cuantitativo de los datos.

Queremos destacar aquí el gran esfuerzo que ha supuesto asumir el análisis de la disponibilidad léxica de dos lenguas al mismo tiempo. Esa ha sido una de las principales razones por las que este estudio se ha dilatado en el tiempo. Pero creemos que el esfuerzo valía la pena, no solo en aras del rigor científico, sino también para abrir nuevos campos de investigación, nuevas preguntas que nos lleven a seguir trabajando para llegar a conocer más sobre la lengua y su uso.

2. PRESUPUESTOS TEÓRICOS

Una de las características más esenciales del estudio que iniciamos es la multidisciplinariedad desde la que se aborda⁹. Es decir, no es una, sino varias las disciplinas y perspectivas desde las que emprendemos el estudio del léxico disponible, y que aparecen implicadas, al mismo tiempo, en el análisis de la disponibilidad léxica de las lenguas de una comunidad bilingüe. La disponibilidad léxica como tal se enmarca dentro de la léxico-estadística, aunque la covariación efectiva de los datos con las variables sociales correspondientes es una tarea del ámbito sociolingüístico. El hecho de que en la comunidad estudiada convivan dos lenguas, derivará en la presencia de numerosos fenómenos de contacto lingüístico, que también estudiamos y que corresponden a la sociología del lenguaje (*vid.* 2.3.1.).

Por esta razón, en el presente apartado, realizamos, en primer lugar; un panorama histórico que contempla el nacimiento de la disponibilidad léxica y el análisis de las obras pioneras que sentaron las bases de los estudios posteriores. A continuación, detallamos las características del Proyecto Panhispánico de Estudio de la Disponibilidad —al cual nos sumamos con esta investigación— y resumimos de forma concisa, el estado de las investigaciones que forman parte de dicho proyecto. En segundo lugar, trazamos un breve panorama sobre la sociolingüística, haciendo hincapié en la relación que alberga con la disponibilidad léxica. En tercer lugar, repasamos los conceptos básicos que se han manejado en el estudio del bilingüismo y de las lenguas en contacto, y ahondamos en las consecuencias concretas del contacto lingüístico, que se producen en el discurso de los hablantes bilingües.

2.1. LA DISPONIBILIDAD LÉXICA

Los estudios de léxico-estadística¹⁰ cuentan ya con una considerable tradición. Esta disciplina lingüística¹¹ se ocupa de realizar cálculos léxicos mediante procedimientos estadísticos; dicho de otro modo, trata de establecer, por medio de

⁹ El trabajo de Gómez Molina y Gómez Devís (2004), por ejemplo, también se presenta desde una triple perspectiva: léxico-estadística, sociolingüística y psicolingüística.

¹⁰ A pesar de que esta es la denominación más habitual de esta corriente de estudio, también se emplean otros términos sinónimos para designarla como “lexicometría” o “estudios léxico-cuantitativos”.

¹¹ López Morales (1983) considera que la “léxico-estadística” estaría integrada dentro de una disciplina más amplia: la lingüística estadística.

diferentes métodos, índices estadísticos para las distintas unidades léxicas de una lengua. Uno de los principales objetivos de esta corriente de estudio ha sido el de precisar cuál es el vocabulario usual de una comunidad, es decir, qué palabras utiliza habitualmente un hablante a lo largo de toda su vida. En torno a este objetivo, se ha perfilado, desde el comienzo, una finalidad esencialmente pedagógica, por cuanto conocer el léxico usual de una comunidad lingüística permite elaborar una planificación adecuada de cara a la enseñanza de ese idioma, ya sea como lengua materna o como lengua extranjera.

Como señala López Morales (1983: 53-59), suele haber una notable distancia entre las numerosísimas palabras que quedan recogidas en los diccionarios generales y las pocas que utiliza un hablante “corriente” en sus interacciones cotidianas¹². Esto se explica porque estos diccionarios, en su afán por ser lo más exhaustivos posible, pretenden integrar todo el caudal léxico de una lengua y, por consiguiente, incluyen en sus páginas muchos términos —arcaísmos, regionalismos, tecnicismos, cultismos...— que resultan por completo desconocidos para la mayoría de los hablantes de esa lengua.

Con el fin de paliar tales inconvenientes y con el objetivo de descubrir cuál es el léxico más habitual, los estudios de léxico-estadística empiezan su larga trayectoria a principios del siglo XX¹³ confeccionando *listas de frecuencia*. Por aquel entonces, se consideraba que la frecuencia era un buen criterio para elaborar inventarios simplificados del vocabulario compartido por una comunidad determinada. Dicha frecuencia, observada en la mayoría de los casos en un corpus más o menos voluminoso de textos escritos, era establecida por rangos o, lo que es lo mismo, en función del número concreto de apariciones que tenía cada una de las palabras.

Un ejemplo de este tipo de cómputos léxicos es la obra de Rodríguez Bou (1952)¹⁴. El *Recuento de vocabulario español* nació auspiciado por el Consejo Superior de Enseñanza de Puerto Rico y su objetivo primordial era elaborar una lista de palabras,

¹² Parece ser que mientras un hablante “corriente” o “común” no emplea más de 2000 palabras, un hablante culto es capaz de manejar 4000 ó 5000 vocablos (Mateo 1998: 36-37, Carcedo 2000: 11).

¹³ Para ver una panorámica histórica de la disciplina junto a una revisión de las principales obras que, en distintas lenguas, la han ido conformando, consúltense Carcedo (1998a: 5-13) y Mateo (1998: 31-38).

¹⁴ Carcedo (1998a) y Mateo (1998) citan otras obras de léxico frecuente referidas al español como, por ejemplo, *A Graded Spanish Word Book* de Buchanan (1927) o *el Vocabulario usual, fundamental y común* de García de la Hoz (1952).

de acuerdo con la frecuencia de su uso, que pudiera servir para facilitar la enseñanza y el aprendizaje del español como lengua materna¹⁵. Las fuentes principales, orales y escritas, de las que se sirvieron para obtener los más de siete millones de palabras son de tres tipos: vocabulario de expresión (vocabulario oral, asociación libre y controlada y composiciones escritas), vocabulario de reconocimiento (periódicos, programas de radio, material religioso y la lista anterior de Buchanan¹⁶) y vocabulario basado en el juicio de distintos autores (libros de texto y textos de lectura suplementaria). La obra incluye un repaso histórico a los principales recuentos de vocabulario, así como un apartado sobre la importancia y las distintas utilidades de este tipo de estudios, donde se deja claro que la enseñanza del idioma, demasiado abandonada al criterio subjetivo del profesor, debe regirse por pautas científicas (Rodríguez Bou 1952: 3-15). Nos parece especialmente interesante el hecho de que se utilicen pruebas asociativas, muy cercanas al concepto de léxico disponible¹⁷.

Los problemas que acarreaban estos léxicos frecuentes comenzaron a vislumbrarse pronto. A pesar de que sería esperable que las palabras más frecuentes fueran también las más utilizadas y conocidas, los resultados obtenidos no demostraban exactamente eso. La mayor parte de las listas de frecuencia estaban integradas por palabras gramaticales —actualizadores, preposiciones, elementos de relación...— carentes de contenido léxico pleno. Además, parecía existir una relación bastante estrecha entre el más puro azar y las palabras que aparecían en los listados. Si se daba la casualidad de que entre los textos seleccionados había más de uno del mismo tema, la frecuencia de las palabras relativas a ese ámbito temático subía desproporcionadamente¹⁸. No obstante, las listas de frecuencia fueron útiles ya que demostraron que el dominio que tienen los hablantes de su propio vocabulario es

¹⁵ Aunque el autor también señala las utilidades de su obra para la enseñanza del español como lengua extranjera, la mayoría de sus observaciones están dirigidas a las provechosas ventajas que proporcionará su lista de frecuencias a la enseñanza del español en las escuelas de Puerto Rico (Rodríguez Bou 1952: X y 12-13).

¹⁶ Véase la nota 14.

¹⁷ En un primer momento se realizaron pruebas de asociación controlada mediante una serie de “palabras-estímulo”. Con el fin de evitar la limitación expresiva que así se obtenía, se varió el procedimiento por otro de asociación libre, en el que el alumnado debía escribir todas las palabras que venían a su mente durante cinco minutos. Resultaba muy lógico que el primer método no funcionara, sobre todo si tenemos en cuenta que algunas de las “palabras-estímulo” eran “y” o “no” (Rodríguez Bou 1952: 1068 y 1084); los centros de interés, como veremos, darán muchas más posibilidades asociativas.

¹⁸ Esta parece ser la causa por la que en la obra de Rodríguez Bou las palabras “escuela”, “libro” y “maestro” alcanzan frecuencias tan altas. Al parecer, parte del material utilizado para el recuento lo constituían redacciones de escolares que tenían como temas “la escuela” o “nuestra aula” (López Morales 1986: 59).

limitado, toda vez que las palabras que alcanzaban una frecuencia muy alta eran más bien pocas (López Morales 1986: 59).

Más tarde, y con la finalidad de intentar paliar los problemas que presentaba la aplicación única del criterio de frecuencia, surgieron los *léxicos básicos*, que ponderaban el criterio de la frecuencia con el de la dispersión. Para poder relacionar estas dos magnitudes, se dividía el universo léxico en cinco áreas delimitadas bien por el contenido de los textos, bien por las condiciones formales de los mismos. La dispersión indicaba que una palabra aparecía en más de un tipo de texto. De esta manera, al multiplicar la frecuencia por la dispersión (y obtener así el *índice de uso* de las palabras), no solo se tenía en cuenta que la palabra fuera frecuente sino también que apareciera en los diferentes mundos léxicos. Resulta evidente que este nuevo cálculo tenía grandes ventajas puesto que ofrecía un perfil muy aproximado del verdadero vocabulario usual de una comunidad y, por tanto, podía servir como instrumento para una adecuada planificación curricular del léxico (López Morales 1983: 210-212). Sin embargo, aunque los resultados fuesen ponderados, algunas palabras de uso corriente seguían teniendo una frecuencia muy baja y, por tanto, quedaban excluidas de la lengua básica (Mateo 1998: 29).

Otro tipo de análisis que también intenta establecer cuál es el léxico más usual de una comunidad es el de la disponibilidad. La primera obra que utiliza este índice, con la finalidad de identificar los elementos imprescindibles de una lengua para ser aprendida como segundo idioma, en concreto *le français fondamental*, es la de Gougenheim *et al.* (1964)¹⁹. Tal y como sus mismos autores reconocen²⁰, René Michéa tiene el mérito de haber creado la noción de *léxico disponible*, tradicionalmente definido como “el caudal léxico utilizable en una situación comunicativa dada” (López Morales 1986: 62, 1999: 11).

¹⁹ Esta es una versión corregida y ampliada de la edición de 1956, que llevaba por título *L'élaboration du français élémentaire*. Los cambios que se observan con respecto a los objetivos responden, según parece, a motivos políticos (Gougenheim *et al.* 1964: 9-15).

²⁰ “M. R. Michéa, Professeur d'allemand au Lycée de Périgueux, qui s'intéresse, depuis de longues années, aux problèmes des langues de base, a apporté un concours généreux et infatigable à l'élaboration méthodologique de l'ouvrage. C'est à lui que revient le mérite d'avoir mis au point la notion de disponibilité” (Gougenheim *et al.* 1964: 14-15).

Los estudios de léxico disponible surgieron, asimismo, como alternativa a los problemas que ocasionaban las listas de frecuencia²¹. Los profesores franceses, durante el transcurso de diversas investigaciones, se habían percatado de que muchas de las palabras más habituales y usadas comúnmente por cualquier francés no aparecían en las listas de frecuencia. Como ya hemos comentado, las palabras gramaticales dominaban en las listas, casi llegando a constituir el 50%, mientras que el resto lo ocupaban verbos, adjetivos y, por último, sustantivos. La infrecuencia de aparición de una serie de palabras, harto conocidas por un hablante común, como *auto* o *métro* fue explicada por su vinculación a una situación temática determinada. Es decir, mientras que una parte del léxico aparecía fuese cual fuese el asunto tratado, otra parte del vocabulario estaba fuertemente condicionada según cuál fuera el tema que se actualizara en el discurso. Las *palabras atemáticas* se relacionaban con las palabras gramaticales y eran las que, por tanto, ocupaban los primeros puestos en las listas de frecuencia; en cambio, las *palabras temáticas*, que solo se empleaban si la situación así lo requería, guardaban relación con lo que Michéa llamó “palabras disponibles”²².

La noción de disponibilidad fue expuesta con detalle por Michéa en un decisivo artículo de 1953 donde la definió de la siguiente manera:

“Un mot disponible est un mot qui, sans être particulièrement fréquent, est cependant toujours prêt à être employé et se présente immédiatement et naturellement à l’esprit au moment où l’on en a besoin. C’est un mot qui, faisant partie d’associations d’idées usuelles, existe en puissance chez le sujet parlant, dès que ces associations entrent en jeu” (Michéa 1953: 340).

Tras las investigaciones de Michéa, por fin se diferencian entre sí conceptos como *léxico frecuente*, *léxico común* o *léxico usual* que hasta la fecha se utilizaban como equivalentes. Con la disponibilidad empieza a quedar claro que muchos de los términos más usuales no tienen por qué ser frecuentes y, de hecho, no lo eran, tal y como se había observado en las listas de frecuencia. La disponibilidad recoge un tipo de léxico que está presente en la mente de los hablantes y, por tanto, resulta familiar y

²¹ Los estudiosos franceses no parecen conocer los léxicos básicos. Esto resulta lógico puesto que la primera edición de su obra es de 1956 y la primera obra lexicográfica que utiliza índices de uso es la de Juilland y Chang Rodríguez, *Frequency Dictionary of Spanish Words*, que se publica en 1964 (Carcedo 1998a: 11, Mateo 1998: 34).

²² “Les mots «athématiques» sont pour M. Michéa les mots que l’on peut trouver à peu près régulièrement dans n’importe quel texte [...]. Au contraire les mots thématiques sont liés à un thème ou à un genre de thèmes donnés. Ils désignent les êtres et les objets et ce sont, pour la plupart, des mots concrets. Leur présence dans les listes de fréquence dépend du choix des textes dépouillés” (Gougenheim *et al.* 1964: 144).

habitual para ellos, pero que solo entra en juego si la situación comunicativa lo requiere²³.

La primera diferencia que se observa, siguiendo esta nueva metodología, es la referente al tipo de categoría gramatical que se obtiene. Como ya hemos dicho, en las listas de frecuencia abundaban las llamadas *palabras gramaticales*, por el contrario, en las de disponibilidad, el predominio de los sustantivos, palabras de contenido semántico específico, es casi absoluto. Esto nos lleva a concluir que ambos métodos no son para nada excluyentes, sino complementarios. Además, los léxicos básicos, como hemos visto, recogen la norma culta, ya que se basan en producciones escritas pertenecientes a distintos mundos temáticos y, por tanto, conllevan una cierta formalidad. Ahora bien, el léxico disponible indaga en una serie de campos que resultan muy cotidianos para los hablantes y, a pesar de que, normalmente, se emplean pruebas escritas para su obtención, se intenta acceder a una parcela del léxico antes oral que escrita o literaria.

La segunda diferencia esencial es que los léxicos básicos recogen el léxico estable de una comunidad, es decir, la proporción real del uso de las diferentes clases de palabras; sin embargo, la disponibilidad señala el vocabulario concreto que el hablante podría utilizar en un contexto comunicativo dado. De la suma de ambos repertorios obtendremos el léxico fundamental de una lengua (López Morales 1986: 63).

Sin duda alguna, para recabar el léxico disponible había que idear otro método distinto al utilizado con anterioridad por la léxico-estadística. Tras varios experimentos realizados con escolares²⁴, Michéa pergeñó unas pruebas asociativas, denominadas *centros de interés*, en torno a las cuales los encuestados producirían unidades léxicas. Como bien señalan los investigadores franceses, esta noción ya era conocida en los estudios pedagógicos, aunque a Michéa debemos, una vez más, el hecho de haberla relacionado con los estudios de léxico disponible.

²³ Como comenta acertadamente López Morales (1986: 63), es muy poco probable que palabras como *carta* y *sello* salgan en nuestra expresión a no ser que estemos hablando de asuntos de correo postal, aunque no de correo electrónico.

²⁴ En el primero de estos experimentos, después de leer un fragmento de una obra literaria a unos escolares, el profesor les solicitó que escribiesen las palabras que más les habían llamado la atención. En otra prueba posterior, con niños mayores, Michéa pidió que escribieran las 20 primeras palabras que se les ocurrieran al pensar en un viaje en tren. En ambos experimentos, se recababan principalmente sustantivos de valores frecuenciales bajos (Gougenheim *et al.* 1964: 146-149).

Los centros de interés que se utilizaron en la investigación francesa fueron los siguientes (Gougenheim *et al.* 1964: 152-153):

- | | |
|---|--|
| 1) <i>les parties du corps;</i> | 9) <i>le chauffage et l'éclairage;</i> |
| 2) <i>les vêtements (peu importe que ce soient des vêtements d'homme ou de femme);</i> | 10) <i>la ville;</i> |
| 3) <i>la maison (mais pas les meubles);</i> | 11) <i>le village ou le bourg;</i> |
| 4) <i>les meubles de la maison;</i> | 12) <i>les moyens de transport;</i> |
| 5) <i>les aliments et boissons des repas (à tous les repas de la journée)</i> | 13) <i>les travaux des champs et du jardinage;</i> |
| 6) <i>les objets placés sur la table et dont on se sert à tous les repas de la journée;</i> | 14) <i>les animaux;</i> |
| 7) <i>la cuisine, ses meubles et les ustensiles qui s'y trouvent;</i> | 15) <i>les jeux et distractions;</i> |
| 8) <i>l'école, ses meubles et son matériel scolaire;</i> | 16) <i>les métiers (les différents métiers et non pas les noms qui se rapportent à un seul métier)</i> |

Estando así las cosas, había que conseguir un artificio matemático capaz de medir el *grado de disponibilidad*, que los mismos autores franceses definían como “la présence plus ou moins immédiate de ces mots dans notre mémoire” (Gougenheim *et al.* 1964: 152). De este modo, pasaron a confeccionarse listas de léxico disponible en las que las palabras se ordenaban según rangos, es decir, dependiendo de la frecuencia con que habían sido nombradas por los distintos informantes. Más tarde, se idearon diversas fórmulas matemáticas que ponderaron el criterio de frecuencia con el lugar de aparición de las unidades léxicas en los listados, para que quedara constancia de que las palabras que aparecían más veces en los primeros lugares eran más disponibles²⁵.

Los estudios de disponibilidad léxica, desde su nacimiento con los pioneros franceses, han tenido un notable desarrollo, sobre todo en el mundo hispanico. En su

²⁵ Los pioneros franceses solo tuvieron en cuenta la frecuencia. A pesar de que Charles Müller fue el primero en llamar la atención sobre la importancia del orden de las palabras en las listas, tenemos que esperar a 1983 para que Lorán y López Morales elaboraran una fórmula matemática que ponderara ambos parámetros. Tras observar que este cálculo presentaba algunos problemas al descender en los rangos de los análisis de las listas, López Chávez y Strassburguer establecieron otra fórmula en 1987, en la que, actualmente, se basan los programas informáticos de disponibilidad (López Morales 1999a: 15-19).

concepción teórica, la tesis de Natividad Hernández Muñoz (2006) supone un hito importante al responder a cuestiones intrínsecas relacionadas con el concepto mismo de *disponibilidad* que no habían sido resueltas hasta entonces. Por ejemplo, la autora opina que la definición tradicional de *léxico disponible* como “caudal léxico utilizable en una situación comunicativa dada” tendría que ser matizada, toda vez que no podemos comprobar que un hablante que incluye una palabra en la lista de léxico disponible, sea capaz de utilizarla en un determinado contexto real o, dicho de otra forma, ser capaz de nombrar una palabra no equivale a ser capaz de usarla (Hernández Muñoz 2006: 29-32, 46-47). Además, la investigadora —que sitúa la disponibilidad léxica dentro de las Ciencias Cognitivas²⁶— comprueba cómo las variables cognitivas que afectan significativamente al léxico disponible son, por este orden, la *tipicidad*, la *familiaridad* y la *edad de adquisición*. Es decir, “las palabras más típicas de una categoría, las más familiares y las aprendidas más temprano son más disponibles que las palabras menos típicas, menos familiares y aprendidas más tarde” (Hernández Muñoz 2006: 188).

Por otro lado, lo que más puede sorprendernos, en un primer momento, es la considerable inmovilidad de la disciplina, sobre todo en la metodología. Para la obtención del léxico disponible continúan utilizándose idénticas pruebas asociativas —con algunas ligerísimas variaciones en los enunciados— a las utilizadas en los estudios pioneros de los años 50. Parece ser que la razón de esta continuidad se halla en que los estudios de léxico disponible se concibieron, desde el trabajo de Mackey (1971a), como un instrumento muy adecuado para realizar comparaciones interlingüísticas, a pesar de que cabría preguntarse si la comparación interdialectal justifica por sí sola la inmovilidad metodológica. Por ejemplo, los centros de interés que se utilizan actualmente para recabar el léxico disponible son, *grosso modo*, los mismos que utilizaron los pioneros franceses²⁷. Estos creyeron que los centros representaban sectores semánticos universales²⁸; no obstante, hoy en día sabemos que tal cosa no es así y que los centros denotan un cierto relativismo²⁹. Ya en su día, Mackey (1971a: 28-

²⁶ Según la autora, en la teoría general del léxico disponible también están implicadas otras disciplinas como la lingüística, la filosofía del lenguaje, la psicolingüística y la neurolingüística (Hernández Muñoz 2006: 40).

²⁷ Para conocer con detalle cuáles han sido los centros de interés utilizados a lo largo de toda la historia de los estudios de disponibilidad, véase López Morales (1999a: 32-37).

²⁸ Esto es lo que se afirma al respecto en no pocas ocasiones. No obstante, en ningún lugar de su obra Gougenheim *et al.* (al menos en la edición de 1964) así lo expresan.

²⁹ Gallison (1979: 9), un autor de la tradición francesa de estudios lexicológicos aplicados a la enseñanza, opina: “Le centre d’intérêt est donc un moyen d’investigation incomplet (sélectionnant des rubriques sans

31) se preguntaba por qué los centros debían ser exactamente esos y no otros. Sin embargo, a pesar de estas reticencias, como ya hemos dicho, en la gran mayoría de los estudios se ha optado por mantener los dieciséis centros clásicos en aras de una posible comparación de los resultados³⁰.

2.1.1. Estudios pioneros sobre disponibilidad

Bajo este epígrafe pretendemos describir con cierto detenimiento los primeros trabajos que se dedicaron al estudio de la disponibilidad léxica, para ver de qué forma tanto el enfoque sociolingüístico, como las posibilidades de aplicar este instrumento a comunidades bilingües, se llevaron a cabo desde sus comienzos. En primer lugar, nos referiremos al trabajo colectivo de Gougenheim *et al.* (1964), al que siguió en América, Mackey (1971a).

El primer trabajo sobre disponibilidad léxica, como ya señalamos en el apartado anterior, es la obra conjunta *L'élaboration du français fondamental*. Este estudio no solo es el primero que se realiza sobre este tema, sino que además tiene el mérito de haber sentado las bases conceptuales y metodológicas sobre las que se desarrollarán la mayoría de los estudios posteriores. Tal y como explican sus autores en la introducción, la obra nace por una iniciativa de la UNESCO y es apoyada en todo momento por el gobierno francés a través de diversas instituciones. El principal objetivo que se persigue, con la reedición de 1964, es el de lograr una gramática y un léxico simplificado de la lengua francesa con vistas a su utilización en la enseñanza del francés como lengua extranjera. Los autores, tras criticar el escaso interés que ha tenido la pedagogía tradicional en establecer una adecuada planificación de la enseñanza del léxico, apuestan por que este *français fondamental* permitirá una rápida difusión de la lengua francesa y, de este modo, se dará respuesta al numeroso público interesado, por aquel entonces, en aprender francés, ya sea en sus respectivos países o porque residan en Francia —por motivos académicos o laborales— o en una antigua colonia francesa. Las innovaciones más importantes de esta obra con respecto a otros estudios de lenguas

éviter le fatras lexical) et subjectif (procédant de la seule intuition de l'auteur), et un moyen de présentation dangereux "tordant" le langage qu'il propose comme objet d'étude".

³⁰ Otra solución adoptada por muchos es la de mantener los dieciséis clásicos pudiendo añadir los que el investigador desee según los objetivos de su estudio. Con respecto a estos problemas, López Morales (1999a: 33) afirma: "los problemas siguen planteados a la espera de soluciones teóricas; entre tanto, las investigaciones empíricas continúan su curso, aun reconociendo que el andamiaje teórico presenta algunas fracturas notables".

de base reside en el hecho de combinar tres parámetros: la frecuencia, la disponibilidad y el empirismo racional (Gougenheim *et al.* 1964: 7-15).

La primera parte de esta obra (Gougenheim *et al.* 1964: 19-56) constituye una revisión exhaustiva sobre la simplificación de las lenguas y de sus vocabularios. Los autores examinan de forma crítica el origen de este tipo de estudios y las principales obras publicadas en diversas lenguas sobre este asunto. Tanto los estudios que aplican un método lógico —como el *Basic English*—, como las obras que se basan en la estadística de frecuencias, presentan inconvenientes importantes. Entre los más graves, los profesores franceses constatan las repetidas ausencias de términos muy habituales en las listas de frecuencia, el relativo subjetivismo a la hora de escoger los textos en los que se basarán los recuentos, y el hecho de que, al proceder muchos de dichos textos del ámbito de la literatura, los diccionarios que los utilizan como fuente recogen un léxico muy literario e incluso anticuado, para nada acorde con la vida cotidiana del momento³¹.

La segunda parte (Gougenheim *et al.* 1964: 61-134) está dedicada a las listas de frecuencia que los autores han obtenido. Tras repasar los escasos materiales existentes en aquel entonces sobre el francés hablado, se detallan las pautas metodológicas que se siguieron a la hora de llevar a cabo sus grabaciones directas de lengua oral, las circunstancias que rodearon las entrevistas realizadas y algunas de las características de los informantes. Después se describe cómo se han efectuado las transcripciones y qué método siguieron para obtener la lista de frecuencias. A continuación, los autores presentan la larga lista de frecuencias integrada por un total de 1063 palabras³². Al analizar las mismas, los autores constatan el predominio, en primer lugar, de las denominadas palabras gramaticales, seguidas por los verbos, los adjetivos y, en un último lugar, los sustantivos. Esta circunstancia se explica por la mayor estabilidad de las palabras gramaticales y de los verbos, que aparecen en un gran número de circunstancias diferentes. Esta segunda parte finaliza con la comparación de las palabras obtenidas en las listas de frecuencia anteriores a esta obra sobre el francés fundamental,

³¹ Merecen una atención especial los comentarios de los pioneros franceses al *Recuento de Vocabulario Español* (cfr. con lo dicho en el apartado anterior). A pesar de reconocer el mérito de haber utilizado pruebas de asociaciones de ideas para elaborar el *Recuento*, lamentan “que les auteurs aient amalgamé ce vocabulaire au vocabulaire fourni par la fréquence, au lieu de mettre en regard du vocabulaire de la fréquence.” (Gougenheim *et al.* 1964: 58).

³² Solo la integran las palabras de frecuencia superior a 20, pues este es el criterio que los autores consideran conveniente para establecer un vocabulario fundamental o de base (Gougenheim *et al.* 1964: 69).

basadas en la lengua escrita, y las de los propios autores. Una vez más los profesores franceses señalan la baja frecuencia obtenida, esta vez en sus investigaciones, por algunas de las palabras más usuales y modernas.

Al empezar la tercera parte (Gougenheim *et al.* 1964: 137-194), los autores se interrogan sobre la utilidad del criterio de frecuencia³³ y, después de haber constatado que los términos *usual* y *frecuente* no son en absoluto sinónimos, resuelven aceptar la coexistencia de dos vocabularios distintos: el frecuente y el disponible³⁴. Estos dos conceptos están relacionados con los de *palabras atemáticas* y *temáticas*. Como ya dijimos, mientras que las *palabras atemáticas*, principalmente palabras gramaticales, son aquellas presentes en cualquier texto con independencia de su contenido; las *temáticas*, palabras concretas, están muy relacionadas con un tema determinado. A continuación, se detallan los experimentos realizados por Michéa en algunos institutos, donde este investigador pudo corroborar que mediante pruebas asociativas se recababan mayoritariamente sustantivos. En las páginas siguientes se ofrecen las características de los cuestionarios que cumplimentó el alumnado de distintas escuelas³⁵ así como dos encuestas más realizadas con adultos, una de ellas entre los lectores de una revista especializada a los que se les solicitaba un total de 20 verbos de acciones corrientes. Los autores realizan, en otro capítulo, un análisis de las diferencias sociológicas y geográficas a las que llegaron después de vaciar las encuestas³⁶. Al respecto, advierten diferencias mínimas en función de la región geográfica donde se haya realizado la encuesta, a no ser por algunos vocablos determinados que deben su aparición a características de esa región en particular. En cuanto a las diferencias sociológicas, estos investigadores constatan el contraste más significativo entre los escolares del medio urbano y del rural, atribuible también a distintas “realidades locales”. Por otro lado, las diferencias entre niños y niñas fueron de menor importancia, principalmente de

³³ También aprovechan para desmontar algunas teorías, como la de Verlée, quien argumenta que solo las palabras frecuentes, y no las más usuales, son útiles a un extranjero (Gougenheim *et al.* 1964: 141).

³⁴ Según los autores “une combinaison du vocabulaire de fréquence et du vocabulaire disponible nous donne le vocabulaire nécessaire” (Gougenheim *et al.* 1964: 152).

³⁵ Los profesores pidieron explícitamente que el alumnado escribiera los nombres comunes que, a su criterio, resultaba más útil saber en función del centro de interés que se daba. El tiempo máximo de respuesta era de un cuarto de hora por centro; el sistema empleado era el de listas cerradas de 20 palabras. La información que se solicitaba a cada informante era la siguiente: nombre, apellido, escuela, sexo, fecha de nacimiento, curso, fecha, profesión de sus padres o de las personas encargadas de su educación y el nivel del alumno, que completaba el maestro. Los encuestados tenían entre 9 y 12 años.

³⁶ Con una intuición “sociolingüística” verdaderamente admirable si se tiene en cuenta la fecha de publicación.

preferencia de algunos campos léxicos en los centros referentes a la ropa o a la casa³⁷. Finalmente, los investigadores incluyen las listas con la dispersión³⁸ y el grado de disponibilidad de las palabras obtenidas en los dos primeros centros, a saber, *les parties du corps* y *les vêtements*.

En la cuarta y última parte del libro, los autores explican los criterios que finalmente han seguido para editar la lista de vocabulario francés básico³⁹. Dejan constancia de las palabras que han eliminado y de las que han añadido a las listas obtenidas y de los motivos que les han llevado a actuar de cada forma⁴⁰. La mayor parte de estas decisiones se apoya en el hecho de que este *français fondamental* ha de posibilitar la expresión en pocas palabras de cualquier noción esencial en lengua francesa. Finalmente, el vocabulario del francés fundamental queda constituido por una lista de 1475 palabras diferentes, entre las que destacan, por este orden, los nombres —procedentes muchos de ellos de las listas de disponibilidad—, los verbos, las palabras gramaticales, los adjetivos y, por último, los adverbios. En el segundo capítulo, estos investigadores exponen los resultados de la encuesta estadística gramatical llevada a cabo de manera análoga a la del vocabulario y detallan qué nociones sobre gramática francesa deben estudiarse en un primer nivel de francés. Para finalizar la obra, se comprueba si el *français fundamental* que ellos proponen capacita realmente a un extranjero para poder comprender un texto en francés. Los autores comprueban la eficacia de su método contrastándolo con diversas fuentes: entrevistas orales, artículos de periódico, textos novelísticos y científicos. La obra se cierra con un apéndice con diversos extractos de manuales de francés basados en el *français fondamental* y con una bibliografía crítica.

³⁷ Según los estudiosos, los vestidos de niñas fueron más citados por estas y los de niños por estos. En cuanto a la casa, las niñas mostraron con sus respuestas “un gusto por el confort e incluso el lujo” y, los varones; en cambio, “por los aparatos que guardan alimentos” y “por los pequeños accesorios” (Gougenheim *et al.* 1964: 176-178).

³⁸ Este concepto se correspondería con el actual *vocablo*, es decir, “le nombre de mots différents cités pour le centre d'intérêt” (Gougenheim *et al.* 1964: 189).

³⁹ Las decisiones más importantes sobre este tema son la de no considerar las palabras de frecuencia igual o superior a 29 ni las palabras que no figuren al menos en cinco textos (Gougenheim *et al.* 1964: 197-198).

⁴⁰ Concretamente, eliminan interjecciones, conectores, vulgarismos y todas aquellas palabras que consideran que no es necesario incluir en un primer nivel de francés fundamental. Por otro lado, estos investigadores, con un subjetivismo que contrasta con la objetividad de la metodología expuesta a lo largo de todo el libro, añaden todos los términos que creen imprescindibles en un francés básico y que las listas no proporcionan. Algunas de las decisiones tomadas al respecto no coinciden con las de la edición anterior (Gougenheim *et al.* 1964: 197-210), lo cual demuestra la volubilidad de los criterios utilizados en este caso.

La segunda obra que supone un hito en los estudios de disponibilidad léxica versa asimismo sobre la lengua francesa y también vale la pena detenerse en ella. Mackey, lingüista canadiense, publicó en 1971 un trabajo en dos volúmenes que lleva por título *Le vocabulaire disponible du français*. Tomando como ejemplo la obra de Gougenheim *et al.* (1964), este autor pretende utilizar la disponibilidad como medida de diferencias culturales entre dos regiones distintas de habla francesa: Francia y Canadá. Aunque se inscribe en una investigación de mayor alcance sobre el vocabulario común de toda la América francófona, este trabajo presenta solo los resultados de una región canadiense, Acadia⁴¹. Cabe señalar en este punto que esta obra de Mackey no es un estudio cuantitativo o cualitativo como tal sino que, salvo algunos breves comentarios sobre diferencias culturales, el autor se limita a ofrecer todos los datos necesarios en espera de que otros realicen análisis de orden psicológico, lingüístico o sociológico.

En el primer capítulo del libro (1971a: 17-22), Mackey revisa los principales índices utilizados al establecer léxicos —la *frecuencia*, la *dispersión* y la *valencia*⁴²— para terminar constatando una vez más la poca utilidad de estos a la hora de determinar qué sustantivos concretos son los que se usan habitualmente. Por esta razón, el segundo capítulo está dedicado a la noción de *disponibilidad*, tal y como la concibió René Michéa, y a las técnicas e instrumentos creados para su obtención. A pesar de los inconvenientes que Mackey observa en el estudio de Gougenheim, Michéa, Rivenc y Sauvegot⁴³, decide mantener las principales opciones metodológicas adoptadas en este con el fin de poder comparar posteriormente los resultados.

En el tercer capítulo, se explican las decisiones que, con respecto a este estudio, fueron tomadas por un comité interuniversitario creado con la finalidad de homogeneizar las investigaciones que se estaban llevando a cabo en distintos puntos de

⁴¹ En su trabajo de 1976 el autor lleva a cabo esta comparación a partir de los porcentajes de mención de cada palabra en Francia y en Acadia; también compara los índices de cada concepto general (dentro de cada centro de interés) para dar cuenta de las diferencias culturales que pueda tener una misma lengua al desarrollarse en dos lugares distintos (Mackey 1976: 338-368).

⁴² El autor entiende por palabras *polivalentes* (de ahí el término *valencia*) aquellos vocablos cuyas posibilidades de combinación en el discurso son muy altas, que incluyen otros conceptos o que son capaces de funcionar como sinónimos de un gran número de términos.

⁴³ Uno de estos inconvenientes es, como ya dijimos, el relativismo que atañe a los dieciséis centros de interés. Tal y como señala Mackey, “[l]’importance relative des centres d’intérêt change avec l’âge, le métier et le degré d’instruction des intéressés.”(Mackey 1971a: 29). Véanse las preguntas que Mackey se formula sobre la metodología en la p. 13 de la misma obra.

Canadá, relativas a la forma del cuestionario, a la edad de los informantes y a los centros de interés⁴⁴ siguiendo fielmente las decisiones anteriores tomadas por Gougenheim *et al.* (1964). A continuación, se describen las características de la encuesta llevada a cabo en Acadia, la única región cuyos datos estaban recogidos en el momento de publicar ese trabajo; de dicha encuesta, solo se vacía la información correspondiente a los escolares de entre 9 y 13 años y a los dieciséis centros clásicos⁴⁵. Tras analizarla y compararla con la obtenida en Francia, Mackey constata algunas diferencias culturales, como, por ejemplo, las palabras *cinéma*, *théâtre* o *buffet*, que alcanzan posiciones altísimas en las listas francesas y, en cambio, posiciones muy bajas en las de la región canadiense⁴⁶.

En el capítulo cuarto, se describen las labores relacionadas con el vaciado de la encuesta; es decir, el proceso de corrección llevado a cabo y los problemas relacionados con los criterios de edición⁴⁷. Al finalizar este capítulo, se presentan los documentos que componen la parte fundamental del libro, o sea, los resultados de la investigación. En el quinto capítulo de la obra de Mackey se explica con todo lujo de detalles los procedimientos informáticos relativos al tratamiento de los datos. Como el mismo autor señala en reiteradas ocasiones, este trabajo pretende proporcionar un *modus operandi* a las investigaciones que, sobre disponibilidad léxica, se desarrollen con posterioridad, lo

⁴⁴ A los dieciséis centros utilizados en Francia se añadieron seis más: *L'église et la vie paroissiale*, *Les sports et l'équipement sportif*, *Les outils*, *L'automobile et ses parties*, *Les actions courantes* y *Les affaires*. En cuanto a la edad, utilizaron, como sus predecesores franceses, informantes de 9 a 12 años, aunque consideraban posible extender la encuesta hasta los primeros años de estudios secundarios. Las condiciones de la encuesta fueron las mismas que las francesas: veinte palabras en quince minutos. Tanto Gougenheim como Mackey, solicitaban las palabras precedidas de un artículo; sin embargo, al menos en el caso de Mackey, este se suprime automáticamente en una etapa posterior (1971a: 67 y 486).

⁴⁵ Faltan, por tanto y como el mismo autor reconoce, vaciar y publicar los datos de las encuestas de los informantes de entre 13 y 18 años así como los cinco centros de interés (Mackey parece olvidar “les affaires”) añadidos a los dieciséis clásicos (*cf.* nota anterior), que fueron completados por informantes bilingües en francés y en inglés.

⁴⁶ A pesar de estas diferencias, el autor también comprueba que se dan coincidencias más que notables (Mackey 1971a: 41-42).

⁴⁷ Algunas de las decisiones que adoptó son introducir con exactitud los datos en el ordenador y corregirlos posteriormente. Con respecto a los criterios de edición, se opta por conservar todas las categorías gramaticales que aparezcan, no añadir más de 19 caracteres por elemento, incluir las expresiones formadas por un verbo y un nombre común en la categoría del nombre, conservar tanto las palabras que no corresponden al enunciado del centro como aquellas que resultan ininteligibles, dejarse guiar por los criterios de los diccionarios en cuanto a la inclusión de los sintagmas y a la grafía de algunas palabras, priorizar el singular si se hallan las dos posibilidades de número en la encuesta y conservar los anglicismos y los dialectalismos del francés canadiense (Mackey 1971a: 44-49). Sobre las divergencias con respecto a los criterios de edición adoptados por Gougenheim *et al.*, véase Mackey (1971a: 51-52).

que justifica las numerosas páginas que se dedican a los aspectos metodológicos⁴⁸. En este primer volumen se incluye un documento con los índices de disponibilidad en Francia y Acadia⁴⁹, dos documentos más con las palabras que no aparecen en cada uno de los dos lugares y un diccionario alfabético del léxico disponible general en lengua francesa.

El segundo volumen, como ya hemos dicho, está dedicado exclusivamente a los problemas y particularidades de la encuesta en Acadia. Tras una breve descripción del origen de la población que allí reside y de los pormenores de la encuesta, se retrata someramente la situación lingüística de la región, con especial atención a los fenómenos de bilingüismo. En este segundo volumen se incorpora un diccionario con el vocabulario francés disponible en Acadia, en el que se indican los porcentajes de disponibilidad alcanzados según la edad, así como otro documento donde se rinde cuenta de las diversas grafías halladas en las listas para los términos ingleses, los dialectalismos del francés en Canadá y todas las palabras inexistentes o no identificables. Al igual que en el primer volumen, el autor se limita a presentar los datos y dar algunas indicaciones, y deja para otros el estudio minucioso del vocabulario disponible según el factor generacional o en función del bilingüismo imperante (Mackey 1971a, vol. II: 6)⁵⁰.

Además de estas dos obras pioneras que acabamos de comentar, existe un autor alemán, Pfeffer (1964 *apud* Knock 1979: 17), que también realiza un estudio donde pretende establecer el léxico fundamental alemán para la enseñanza y aprendizaje de esta lengua. Para seleccionar las palabras que deben formar parte de este léxico, Pfeffer utiliza varios criterios, entre ellos el de la disponibilidad. Con una finalidad didáctica, el autor realiza encuestas en 82 escuelas de 48 ciudades de la zona de habla germana de Europa central (Alemania, Austria y Suiza). El investigador solicita a los alumnos 20 sustantivos, 12 verbos y 8 adjetivos en cada centro de interés. Los estudiantes que completan la encuesta, unos 2000 de entre 15 y 16 años, cuentan con diez minutos para completar 21 campos semánticos. Finalmente, el autor no incluye todas las palabras

⁴⁸ Además recordemos que el trabajo de Mackey es el primero que establece la disponibilidad con la ayuda de un ordenador (1971a: 11).

⁴⁹ Por aquel entonces la disponibilidad se calculaba simplemente según el número de veces que se repetía la palabra en la lista (Mackey 1971a: 52).

⁵⁰ Volveremos sobre esta obra más adelante (*vid.* 3.1.)

recabadas en los listados sino únicamente aquellas cuyo índice de disponibilidad es superior a 100.

2.1.2. El Proyecto Panhispánico

Con respecto a los estudios sobre la lengua española, corresponde a Humberto López Morales el mérito de haber iniciado e impulsado los estudios de léxico disponible en todo el mundo hispánico. Este lingüista llevó a cabo en fecha muy temprana un estudio sobre disponibilidad léxica en Puerto Rico para el que se sirvió de 63 individuos que procedían de tres escuelas representativas de diferentes niveles socioeconómicos. El material, que giraba en torno a diez centros de interés⁵¹, fue recogido de forma oral y escrita en tres grados escolares distintos. Algunos de los resultados de esta investigación se publicaron en diversos estudios. En López Morales (1978) se subraya una vez más la importancia de la disponibilidad léxica como instrumento de planificación curricular. En 1979 este autor relacionó los resultados de esta investigación con la *teoría del déficit* de Bernstein. Tras analizar los resultados, López Morales concluyó que existía una asociación fuerte entre disponibilidad y nivel socioeconómico con una ventaja cuantitativa evidente con respecto a la producción léxica para los niveles altos⁵².

El trabajo de López Morales abrió el camino para que muchos otros estudios vieran la luz en el ámbito hispanoamericano. De este modo, empezaron a sucederse distintas investigaciones como la de Justo Hernández (1986), Mena Osorio (1986), Cañizal Arévalo (1987)⁵³, Ruiz Basto (1987), etc., que son, sin lugar a dudas, interesantes ejemplos de aplicaciones de la disponibilidad léxica. Sin embargo, todos ellos cuentan, en conjunto, con el inconveniente de que presentan diferencias metodológicas sustanciales. Esto implica que sus resultados no son comparables puesto que, si las condiciones de reacción, los enunciados de los centros de interés o la edad de

⁵¹ *Alimentos, Animales, La casa, La cocina, El cuerpo humano, Materiales, Muebles, Naturaleza, La ropa y Transporte* (López Morales 1979: 177).

⁵² En palabras de López Morales (1979: 181): “hay base empírica para adelantar, como hipótesis de trabajo al menos, la existencia de desniveles de disponibilidad léxica, estrictamente ligada a factores de índole socioeconómico”. El investigador puertorriqueño no iba nada desencaminado con esta afirmación puesto que esta hipótesis se ha visto corroborada en todas las investigaciones posteriores sobre disponibilidad léxica, como más adelante veremos.

⁵³ Por tratarse de memorias de licenciatura inéditas de muy difícil acceso, citamos estas tres obras según las referencias halladas en los diversos estados de la cuestión existentes, donde el lector interesado podrá encontrar más información sobre las mismas (Carcedo 1998a, López Morales 1999a y Samper, Bellón y Samper Hernández 2003).

los informantes varían, es totalmente previsible que los resultados también cambien, incluso de manera notable⁵⁴.

Por esta razón, desde principios de los años 90⁵⁵ existe un Proyecto Panhispánico de Estudio de la Disponibilidad cuyo principal objetivo es elaborar diccionarios de disponibilidad léxica para las diversas zonas del mundo hispánico que, además, sean comparables. Hasta la fecha, se han realizado cuatro encuentros entre los investigadores dedicados a este campo con la finalidad de discutir aspectos del Proyecto y ver cómo se desarrolla en las distintas comunidades de habla⁵⁶. El coordinador del Proyecto y director de dichos encuentros es el sociolingüista académico Humberto López Morales, quien —como ya hemos señalado— es el impulsor de los estudios hispánicos de disponibilidad. Las fases de elaboración de cada diccionario son las siguientes: selección de la muestra, recogida de materiales mediante la encuesta, transcripción de los materiales y homogeneización de los mismos, cálculo de la disponibilidad léxica global y por campos y grupos de informantes y, finalmente, publicación del diccionario. La uniformidad de criterios que tienen los estudios resultantes permite establecer comparaciones de todo tipo (ya sea sociolingüísticas, como etnográficas, dialectológicas o de otro tipo). Toda la información sobre las pautas del Proyecto, así como un panorama de los estudios realizados hasta el momento, puede consultarse en www.dispalex.com, la página oficial del Proyecto, auspiciada por la Universidad de Salamanca.

Los principales acuerdos metodológicos a los que se acogen todas las investigaciones que se inscriben en el Proyecto son los siguientes (Bartol 2006: 380-381):

⁵⁴ Por ejemplo, si los centros de interés son distintos, es completamente imposible comparar los resultados obtenidos; si la edad cambia las respuestas podrán ser muy divergentes por las experiencias vitales propias de cada grupo generacional; y aún en el caso de que los centros y las edades fueran las mismas en dos investigaciones paralelas, si el tiempo concedido para responder a cada estímulo no es exactamente el mismo, tampoco serán equiparables las respuestas. Mackey (1971a: 34) tuvo ya muy en cuenta este hecho puesto que, incluso reconociendo las limitaciones de la investigación de Gougenheim *et al.* (1964), toma decisiones en sintonía con ella para poder comparar los resultados.

⁵⁵ En López Morales y García Marcos (1995) se detallan las características del Proyecto circunscritas a Andalucía, la mayoría de las cuales son extrapolables al Proyecto Panhispánico general. En este mismo artículo se dice que la fecha exacta en que da comienzo el Proyecto es 1989, a pesar de que no contemos con ninguna publicación o estudio sobre el tema correspondiente a esa fecha. De otra parte, en Fernández Smith *et al.* (2012) puede verse una perspectiva general de los estudios de disponibilidad y de las características del Proyecto Panhispánico.

⁵⁶ Estos se han realizado en la Universidad del País Vasco (1999), en San Millán de la Cogolla (2003 y 2005) y en Salamanca (2011).

- los informantes son estudiantes del curso que precede al ingreso en la universidad⁵⁷,
- los centros de interés son los dieciséis que ya se utilizaron en las primeras investigaciones francesas⁵⁸,
- se conceden dos minutos por centro para escribir cuantas palabras se quiera,
- se incorporan a la encuesta las variables “sexo”, “nivel sociocultural”, “zona geográfica” y “tipo de centro”⁵⁹,
- se adoptan unas pautas de edición similares⁶⁰
- y se calcula la disponibilidad léxica mediante el mismo programa informático.

Por tanto, como acabamos de ver, el enfoque sociolingüístico está presente ya en la misma concepción metodológica de los estudios sobre léxico disponible en este Proyecto. A pesar de que el objetivo del mismo es solo la elaboración del diccionario, muchas veces se acompaña este con el análisis sociolingüístico de los datos obtenidos. Los resultados de los estudios de disponibilidad tienen diversas aplicaciones en ámbitos muy diversos, lo que provoca algún que otro inconveniente, ya que —algunas veces— los objetivos del Proyecto Panhispánico tienen unas implicaciones metodológicas que pueden no estar en consonancia con las requeridas en las investigaciones particulares de cada lingüista⁶¹.

Por otro lado, en la actualidad, la mayoría de las investigaciones que se desarrollan se hallan integradas en estudios globales de alcance regional o nacional. Como bien afirman Samper Padilla y Samper Hernández (2006), nos encontramos en

⁵⁷ En realidad, López Morales pretende con estos estudios conocer la norma léxica adulta común a todo el territorio hispanohablante.

⁵⁸ *El cuerpo humano, La ropa, Partes de la casa, Los muebles de la casa, Comidas y bebidas, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, La escuela (muebles y materiales), Calefacción, iluminación y medios de airear un recinto, La ciudad, El campo, Medios de transporte, Trabajos del campo y del jardín, Los animales, Juegos y distracciones y Profesiones y oficios.* Algunos investigadores modifican ligeramente algún enunciado o añaden otras áreas temáticas en función de los intereses particulares de sus investigaciones.

⁵⁹ Sobra decir que todos estos acuerdos son abiertos y que nada impide que un investigador, siguiendo su propio criterio, amplíe la nómina de centros de interés o las variables por estudiar, siempre y cuando mantenga las del Proyecto.

⁶⁰ Se siguen, generalmente, las indicaciones de Samper Padilla (1998), aunque las soluciones adoptadas divergen en algunos puntos (*cf.* 5.4)

⁶¹ Esto es, por ejemplo, lo que ocurre con respecto a los criterios de edición. El Proyecto toma al respecto algunas determinaciones que hace que pueda perderse información relevante desde algún punto de vista (*cf.* con lo que opinan Galloso 2000 y Hernández Muñoz 2006: 266-273).

una segunda etapa dentro del estudio de la disponibilidad, superada ya una primera fase anterior donde predominaron las investigaciones individuales y aisladas. En este sentido, el proyecto más importante es el de la creación del “Léxico disponible de España”, dirigido desde la Universidad de Salamanca y que cuenta ya con unos criterios definidos de elaboración y estratificación de la muestra⁶². Además, otros proyectos regionales están en marcha como el “Léxico disponible de Andalucía” o el “Léxico disponible de Canarias” o el “Léxico disponible de Castilla León”.

Los trabajos en las distintas zonas de España e Hispanoamérica se hallan en fases bastante diversas. Algunos de ellos exceden los límites del Proyecto, como ya hemos comentado, y se dedican a aspectos de diversa índole. A continuación, resumimos, de manera panorámica y sin ánimo de exhaustividad, la situación en la que se encuentra el Proyecto en cada una de las distintas provincias de España y países hispanoamericanos⁶³. En el próximo capítulo, ahondamos en los estudios de disponibilidad realizados en territorios bilingües. Seguimos un orden cronológico, a pesar de que, en muchas ocasiones, las fechas de las publicaciones no coinciden exactamente con el momento en que los estudios empezaron a llevarse a cabo. Además, en muchas comunidades las investigaciones se han sucedido de forma simultánea.

En primer lugar, los estudios de disponibilidad cuentan ya con cierta antigüedad en algunas regiones, como Madrid, Las Islas Canarias o Andalucía. En la primera de estas comunidades, los trabajos se iniciaron ya en 1992 (Benítez 1992a y b) y —aunque no contamos aún con los datos publicados— se han realizado distintos estudios interesantes con estudiantes de Alcalá de Henares, abordando cuestiones como la ortografía o los extranjerismos (Paredes 1999 y 2001b).

De otra parte, las investigaciones sobre disponibilidad también han tenido un gran desarrollo en Gran Canaria, donde —desde fechas muy tempranas—, un grupo de investigadores se encargó de la obtención y el estudio del léxico disponible de esta comunidad. Los resultados se han ido publicando en distintos trabajos (García Domínguez *et al.* 1994, Samper Padilla y Hernández Cabrera 1997) y, a la par, también

⁶² Para más información sobre este proyecto se puede consultar el artículo de Bartol y Hernández Muñoz (2006) o el de Bartol (2006).

⁶³ Toda la información de manera exhaustiva y perfectamente detallada se encuentra en el brillante artículo de Samper Padilla y Samper Hernández (2006).

se han ido realizando distintos estudios dedicados a otros aspectos como los criterios de edición, el cotejo dialectal, etc. (Samper Padilla 1998, 1999). Son destacables, asimismo, los artículos en los que se ha recopilado y estudiado tanto los avances de la disciplina como la incidencia de las distintas variables sociales en cada uno de los trabajos publicados sobre disponibilidad (Samper, Bellón y Samper Hernández 2003, Samper Padilla y Samper Hernández 2006 o Samper Padilla 2006)⁶⁴. En Tenerife, también se ha iniciado el estudio del léxico disponible (Torres y García 2009).

El estudio de la disponibilidad está ya completado en la comunidad andaluza. La primera provincia cuyos resultados vieron la luz fue Almería (Mateo 1998), donde la autora analizó con gran exhaustividad el corpus obtenido. Cádiz es la siguiente provincia en la que se finalizó el estudio de la disponibilidad según las directrices del Proyecto Panhispánico (González Martínez 2002). Desde la propia Universidad de Cádiz, también se ha llevado a cabo el estudio de la disponibilidad de Ceuta (Ayora 2006) y, más recientemente, de Melilla (Fernández Smith *et al.* 2008). Paulatinamente, se han ido publicando los resultados de las distintas provincias: Córdoba (Bellón 2011), Huelva (Prado y Galloso 2005), Málaga (Ávila 2006), Jaén (Ahumada 2006), Sevilla (Trigo 2007), siendo la incorporación más reciente Granada (Pastor y Sánchez 2008).

Los autores de estos estudios han analizado distintos aspectos en los datos recogidos como la incidencia de algunas variables (González Martínez y Orellana Ramírez 1999b, Prado y Galloso 2008), los dialectalismos (González Martínez 1999a) o los anglicismos (González Martínez y Orellana Ramírez 2006). También son interesantes los estudios aplicados a otras franjas de edad (García Megía 2003, Prado y Galloso 2006a), o a otros colectivos (Sánchez-Saus 2011)⁶⁵, sobrepasando los confines propios del Proyecto.

En el norte de la Península, se desarrolló pronto un estudio en el País Vasco (Etxebarria 1996b, 1997 y 1999), centrado en el área metropolitana de Bilbao, que

⁶⁴ Mención especial merece la investigación de Samper Hernández, quien —fuera de los límites del Proyecto— ha realizado estudios muy interesantes ya sea sobre la disponibilidad léxica de estudiantes extranjeros de español (Samper Hernández 2001, 2001-2002, 2002) o sobre la progresión de la disponibilidad en distintos niveles educativos (Samper Hernández 2009).

⁶⁵ En esta tesis doctoral, la autora analiza, de forma teórica, la concepción semántica de los centros de interés, y examina los resultados obtenidos tras aplicar la prueba de léxico disponible a una muestra integrada por 322 estudiantes de español como lengua extranjera.

analizaba el peso de distintas variables, aunque los diccionarios nunca fueron publicados. Al parecer, la misma autora ha completado las encuestas en las tres provincias con la finalidad de adscribirse al Proyecto (Samper Padilla y Samper Hernández 2006: 16). Tras el trabajo de la investigadora vasca, Carcedo (2001) publicó el léxico disponible de Asturias donde analizó la incidencia de las distintas variables y realizó un estudio sobre la compatibilidad de cada uno de los inventarios léxicos según las distintas variables. Además, el mismo autor comparó sus resultados con los de otras comunidades (2003, 2004) y recopiló de forma admirable los pilares bibliográficos de la disciplina (1998a)⁶⁶.

En la comunidad valenciana también está prácticamente finalizado el estudio del léxico disponible. Con respecto a la provincia de Castellón, contamos con algunos trabajos sobre el resultado de algunas variables (Blas y Casanova 2001-2003, 2003a, 2003b) y sobre los anglicismos y valencianismos (Casanova 2008, 2013). Gran desarrollo ha tenido la disponibilidad léxica en la provincia de Valencia, donde ya todos los datos han sido clasificados y analizados estadísticamente (Gómez Molina y Gómez Devís 2004). Además, se han publicado diversos estudios (Gómez Devís 2005, Gómez Molina 2006) y las investigaciones se han extendido a la lengua catalana (Gómez Devís 1995, Llopis 2009 y Gómez Devís y Llopis 2010). Por otro lado, Martínez Olmos finalizó en 2008 el estudio y análisis del léxico disponible en las comunidades de habla alicantinas desde una perspectiva claramente sociolingüística (Martínez Olmos 2008).

De forma casi paralela al estudio de la disponibilidad valenciana, un equipo de la Universidad de Zaragoza culminó con éxito en el 2004 el establecimiento del léxico de la comunidad aragonesa (Arnal *et al.* 2004), al que se acompañó más tarde de distintos estudios a partir del corpus obtenido sobre los extranjerismos, la ortografía o el cotejo dialectal con otras comunidades (Arnal ed. 2008). Especialmente provechosa ha sido la aportación de Arnal (2008a, 2008b y 2009) sobre los aragonesismos hallados en las encuestas.

⁶⁶ Por otro lado, Carcedo cuenta con una amplia trayectoria en el campo de la disponibilidad léxica aplicada a alumnado extranjero, concretamente a estudiantes finlandeses de español con los que realizó distintos estudios (Carcedo 1998b, 1999a, 1999b, 1999c, 2000a, 2000b, 2000c).

Desde la Universidad de Salamanca, un grupo muy prolífico de investigadores ha realizado el análisis del léxico disponible tanto en Castilla León (Cruz Alonso 2012) como en Castilla-La Mancha (Hernández Muñoz 2006). Con respecto a Castilla León, con anterioridad, Galloso (2003a) se ocupa de las provincias de Ávila, Salamanca y Zamora; Bartol (2004) estudia Soria; Fernández Juncal (2008), Burgos; y Santiago Guervós (2008), Segovia. También se han publicado estudios interesantes que plantean desde las diferencias según la franja de edad analizada (Borrego 2002, Borrego y Fernández Juncal 2003) hasta posibles variaciones metodológicas (Galloso 2002, Galloso 2003b⁶⁷, Hernández Muñoz y Borrego 2004) pasando por aspectos como la comparación dialectal (Bartol 2008, Borrego y Fernández Juncal 2002) o los anglicismos (Bartol 1998). Cabe destacar, por su importancia, el enfoque psicolingüístico de la tesis doctoral de Hernández Muñoz (2006), que constituye toda una novedad en la trayectoria de la disciplina. Además, como ya hemos indicado anteriormente, desde la Universidad de Salamanca se gestiona la página web del Proyecto, herramienta imprescindible para el volcado y análisis de los datos, y se coordina el proyecto de “El léxico disponible de España”.

Una incorporación relativamente reciente al Proyecto ha sido el léxico disponible de Galicia, publicado en 2008 (López Meirama), que amplía el panorama de regiones bilingües de la Península que cuentan con este tipo de estudios. A este volumen le sigue otro con el análisis de distintos aspectos del corpus obtenido como los galleguismos, las interferencias, la ortografía o las creaciones léxicas, cuya edición corre a cargo de la misma autora (López Meirama ed. 2011).

Una de las últimas provincias españolas en culminar su estudio del léxico disponible ha sido Navarra. El análisis de todos los materiales recogidos constituye la tesis doctoral de Areta (2009); asimismo se han detallado en varios artículos las pautas metodológicas e informáticas que se han utilizado (Saralegui y Taberner 2008, Mangado y Areta 2008) y, Taberner (2008a) ha realizado un estudio de los regionalismos en el léxico disponible navarro. Con respecto a otra provincia norteña, como Cantabria, su estudio acaba de finalizarse (Fernández Juncal 2013). De otra parte,

⁶⁷ Ya en trabajos anteriores (Galloso 1999b) la misma autora mostraba sus reticencias y dudas sobre todo con respecto a los criterios de edición del Proyecto, ya que estos necesariamente varían según el propósito que se persiga.

en La Rioja, lugar de encuentro tradicional entre los investigadores de la disponibilidad, los trabajos se han iniciado.

Dentro de la comunidad autónoma de Cataluña, la primera provincia donde se iniciaron las investigaciones fue Lleida. En Serrano Zapata (2003a) se expusieron las bases metodológicas de la investigación y, más tarde, (Serrano Zapata 2003b y 2004) se analizó la incidencia de las variables sobre una parte del corpus, concretamente la referida a los datos recabados en la capital de la provincia. También nos hemos dedicado al estudio y clasificación de las transferencias léxica halladas en algunos centros de interés (Serrano Zapata 2006) y el resultado final es esta tesis doctoral que aquí se presenta para su defensa, de la que sin duda podremos continuar bebiendo para trabajos posteriores dada la enorme cantidad de información que queda recogida en los diccionarios (*vid.* capítulos 10 y 11) y en los archivos originales sin las modificaciones exigidas por los criterios comunes de edición. Un grupo de investigadores de la Universidad de Barcelona ha iniciado el estudio de la disponibilidad de esta provincia así como el de las restantes catalanas y el de las Islas Baleares, aunque los trabajos se hallan aún en fase temprana (Pérez Edo 2007). Para finalizar con el panorama peninsular, solo nos falta añadir que desde la Universidad de Huelva se dirige, en la actualidad, la investigación sobre Extremadura, también relativamente reciente.

En Hispanoamérica también contamos con notables contribuciones al Proyecto Panhispánico. Orlando Alba emprendió hace tiempo la tarea de recopilar el léxico disponible de la República Dominicana (Alba 1995a), y también se dedicó a estudiar en él la incidencia de los anglicismos y los factores que pueden favorecer su aparición (Alba 1995b, 1996, 1999). Por otra parte, el director del Proyecto Panhispánico, Humberto López Morales, estableció el léxico disponible de Puerto Rico (López Morales 1999a), después de haber realizado estudios relacionando la disponibilidad con la sociolingüística (1978, 1979) —que hemos señalado más arriba—, además de dedicarse a buscar un modo de calcular la disponibilidad (1983, 1986, 1999a) y a revisar —como otros autores— el peso de los anglicismos (1999b).

Finalmente, debemos comentar que, aunque no contamos aún con los datos, los trabajos sobre el léxico disponible de muchos otros países hispanoamericanos como Colombia, Cuba, Honduras, Panamá, Paraguay y Uruguay ya han sido iniciados. Por

otro lado, aunque queden fuera de los límites del Proyecto, son dignos de mención los trabajos que se han realizado en México acerca del cálculo de la disponibilidad (López Chavez y Strassburguer 1991) o estableciendo comparaciones dialectales (López Chávez 1992); en Chile, sobre el léxico disponible de estudiantes chilenos de educación secundaria (Echeverría *et al.* 1987, Valencia y Echeverría 1999b) o sobre el desarrollo de un programa computacional destinado al análisis psicolingüístico de la disponibilidad (Echeverría *et al.* 2008); y en Costa Rica, dedicados al estudio de la disponibilidad léxica de alumnado preescolar (Murillo Rojas 1993 y 1994).

2.2. LA SOCIOLINGÜÍSTICA

Tal y como señala la mayoría de los autores, la sociolingüística nació alrededor de los años 60 en Estados Unidos (López Morales 1989: 8, Moreno Fernández 1998: 295). Al parecer, fueron diversas las circunstancias que propiciaron la aparición de una serie de trabajos cuyo denominador común era el estudio de las relaciones entre la sociedad y la lengua⁶⁸. Por aquel entonces, la sociolingüística como disciplina lingüística carecía de unos fundamentos teóricos y metodológicos claros y, en consecuencia, muchos de los trabajos que se englobaron bajo esta etiqueta pertenecerían hoy a la esfera de otras disciplinas de contenidos muy afines. Sin embargo, en la actualidad, la situación ha cambiado ostensiblemente y son ya innumerables los trabajos e investigaciones que se enmarcan dentro del ámbito estricto de la sociolingüística. Concretamente, en el mundo hispánico, desde finales de los 80, se han venido publicando diferentes manuales en los que se perfilan con rigurosidad los conceptos, los métodos y los objetivos primordiales de esta disciplina⁶⁹.

Moreno Fernández (1998: 301), partiendo de una concepción amplia de la sociolingüística, considera que pueden delimitarse tres escuelas o *niveles sociolingüísticos*. El primero de ellos sería el *nivel sociológico*, donde se incluirían cuestiones referentes al estudio sociológico de las lenguas y de los fenómenos lingüísticos en las comunidades de habla, es decir, todo aquello que constituye objeto de

⁶⁸ Para conocer con más exactitud cuáles fueron estas circunstancias, así como para tener más detalles sobre el origen y la delimitación de la disciplina, consúltese Fernández Rodríguez (1997).

⁶⁹ *V.gr.* solo hasta finales de los 90, Rotaetxe (1988), López Morales (1989), Silva-Corvalán (1989), García Marcos (1993 y 1999), Moreno Fernández (1998) o Almeida (1999).

investigación de la *sociología del lenguaje*⁷⁰. El segundo nivel sería el *etnográfico*, que abordaría el estudio de las normas de conducta comunicativa propias de una comunidad de habla específica; como es lógico, la *etnografía de la comunicación* sería la disciplina (o tendencia sociolingüística) encargada de emprender tales estudios. Por último, el autor distingue un tercer *nivel lingüístico*, del que se ocuparía la *sociolingüística estricta* en el que tendrían cabida todas las cuestiones referentes a la variación de la lengua en su contexto social⁷¹.

La sociolingüística comparte su interés por el estudio del habla en su uso real con muchas otras disciplinas desarrolladas también durante el siglo XX, como la etnolingüística, la etnografía de la comunicación, la pragmática o el análisis del discurso, o incluso anteriores, como la dialectología. En este sentido, estas disciplinas suponen una superación de las limitaciones de las corrientes teóricas de la lingüística desarrolladas en ese mismo siglo. Recordemos que tanto Saussure (1945) como Chomsky (1957) habían centrado su objeto de estudio en la *langue* o la *competence*, respectivamente, desechando todo lo que tuviera relación con la *parole* o la *performance*. De este modo, las principales escuelas lingüísticas del siglo pasado, la estructuralista y la generativista, se concentraron en detallar las características de la lengua como sistema colectivo y homogéneo localizado en el cerebro de un hablante-oyente ideal⁷².

La sociolingüística, en cambio, propone un paradigma teórico nuevo al dar una importancia básica y esencial a las producciones lingüísticas reales de un conjunto de individuos que, con características muy diferentes, conforman comunidades de habla heterogéneas. De este modo, se focaliza el interés en el estudio del habla o del comportamiento social de una comunidad y se dirige la mirada hacia todas aquellas

⁷⁰ No todos los autores consideran que la sociología del lenguaje pueda englobarse dentro de la sociolingüística. Muchos consideran que se trata de dos disciplinas distintas aunque, por razones obvias, estén estrechamente relacionadas. Tal y como señala Hudson (1981: 15), la diferencia principal entre ambas parece ser, simplemente, que se enfatice más el estudio del lenguaje o el de la sociedad. Véase un resumen de las distintas opiniones vertidas en López Morales (1989: 26, nota 5).

⁷¹ Para elaborar esta clasificación, Moreno Fernández (1998: 301) sigue, en parte, las propuestas de Fishman.

⁷² A pesar de que esta fue la postura mayoritaria entre los lingüistas, es posible encontrar algunos antecedentes de la nueva concepción sociológica del lenguaje como André Meillet (Moreno Fernández 1998: 297). En el estructuralismo, ese hablante-oyente ideal es el hablante culto, el propio lingüista que se toma a sí mismo como objeto de análisis o como modelo del que tomar ejemplos, si no lo hace de las autoridades de la lengua, como hacía la gramática tradicional.

cuestiones que se englobaban bajo la etiqueta de la *lingüística externa*, a las que el estructuralismo y el generativismo habían hecho caso omiso⁷³. En concreto, la escuela *variacionista* o la denominada *sociolingüística estricta*⁷⁴ se ocupa, por vez primera en la historia de la lingüística, del estudio sistemático de la variación. Hasta entonces ninguna escuela lingüística anterior había negado la existencia de la variación lingüística⁷⁵ (López Morales 1989: 182), pero se la relacionaba con el habla y, por consiguiente, con todo lo considerado fortuito y externo al sistema mismo. Por el contrario, la sociolingüística demuestra que es posible estudiar la variación desde una nueva óptica y relacionarla con distintos factores sociales y lingüísticos. Además de determinar las pautas que rigen la variación en las comunidades de habla, el variacionismo tiene el mérito de haber creado un modelo gramatical nuevo que incorpora la variación como fenómeno regular y sistemático. El ejemplo más característico de este nuevo modelo gramatical es, sin duda alguna, el análisis de la regla variable⁷⁶. Consiste en la aplicación de una prueba estadística que mide cómo un conjunto de factores lingüísticos y extralingüísticos determina la aparición de las distintas variantes de una variable lingüística. Para el cálculo de las distintas probabilidades de aparición de cada una de las variantes se parte de la frecuencia observada en un determinado corpus representativo de una comunidad de habla determinada. La regla variable indica la probabilidad que tiene cada una de las variantes de covariar con los factores lingüísticos y sociales condicionantes. Este análisis, como muchos autores ya han señalado, más que explicar, lo que hace es describir la competencia sociolingüística de una comunidad de habla (Moreno Fernández 1998: 307-318, Almeida 1999: 52-62).

⁷³ Esta es la razón por la que Labov (1983: 23 y 235) rechaza el término de *sociolingüística* o lo considera redundante, puesto que implicaría aceptar que puede existir una teoría o práctica lingüística que no fuera social, y el lenguaje es, por definición, una forma de comportamiento social. Las teorías lingüísticas anteriores se centraban solo en analizar una parte de la lingüística, como ya hemos dicho, la referente al que se considera el sistema interno de la lengua.

⁷⁴ Esta escuela nació en los años sesenta en EEUU alrededor de la figura de William Labov, aunque se desarrolló posteriormente también en otros lugares como Canadá o Inglaterra. Esta corriente ha tenido desde sus comienzos dos preocupaciones esenciales: el estudio de la lengua en su contexto social y el cambio lingüístico.

⁷⁵ Entre otras cosas porque la variación es imprescindible para poder explicar cómo se produce el cambio lingüístico (Coseriu 1978).

⁷⁶ Este análisis forma parte de una teoría gramatical más amplia conocida con el nombre de gramática de la variación. Esta gramática utiliza tres reglas para explicar los hechos lingüísticos: la categórica u obligatoria, la semicategórica o semiobligatoria y la variable. Esta última viene a sustituir la regla opcional del generativismo, donde se englobaba la llamada “variación libre”. La sociolingüística cree que tal variación no existe y demuestra que toda variante lingüística esta condicionada por unos determinados factores sociales y lingüísticos en mayor o menor medida.

Las nociones de *variable* y *variante*, como acabamos de ver, son básicas para poder llevar a cabo un análisis sociolingüístico. El variacionismo considera que una *variable* es un elemento, rasgo o unidad que puede manifestarse de formas distintas. Cada manifestación o expresión de una variable recibe el nombre de *variante*. Las variables y, por consiguiente, las variantes pueden ser de dos tipos: lingüísticas⁷⁷ y extralingüísticas. El objeto principal de un estudio sociolingüístico es la *variable dependiente* y es siempre un fenómeno de lengua; mientras que los elementos lingüísticos y sociales que se tienen en cuenta para explicar ese fenómeno son *las variables independientes o explicativas* del estudio en cuestión⁷⁸. Las investigaciones sociolingüísticas realizadas hasta la fecha vienen mostrando de forma unánime que la variación siempre resulta condicionada, en mayor o menor medida, por las distintas variables lingüísticas y sociales. Cuando, tras haber realizado un análisis sociolingüístico, podemos afirmar que existe una relación entre la variable dependiente estudiada y las distintas variables independientes, eso significa que tal variable *covaría* significativamente con alguna o algunas de las variables lingüísticas y sociales examinadas. La finalidad de cualquier estudio sociolingüístico de corte variacionista es la de demostrar que tal relación realmente existe.

La variación puede hallarse en cualquier nivel lingüístico. Sin embargo, el fonético fue el primero en estudiarse y el que, tradicionalmente, ha contado con un mayor número de trabajos. A esto ha contribuido, sin duda, el hecho de que, en el nivel fonético, resulta muy fácil demostrar que diversos elementos lingüísticos son realmente variantes de una variable, es decir, que las distintas variantes resultan totalmente equivalentes puesto que no conllevan un cambio de significado. Otras características existentes en la variación fonética posibilitan que las unidades de este nivel sean adecuadas para la realización de estudios variacionistas. Entre estas características destacan: la frecuencia con que se dan en el discurso, la integración en sistemas cerrados y la posible correlación con factores no solo sociales sino también estilísticos (Moreno Fernández 1998: 21). No obstante, no todo son ventajas en este campo. La

⁷⁷ Una de las definiciones de *variable lingüística* más conocida es la de Henrietta Cedergren. Esta autora considera que una variable lingüística “define un conjunto de equivalencia de realizaciones o expresiones patentes de un mismo elemento o principio subyacente” (Cedergren 1983: 150).

⁷⁸ Un ejemplo sería el análisis que realiza Samper Padilla (1990: 134 y 137) sobre la aspiración de la -s/ en Las Palmas de Gran Canaria o el de Calero (1993) sobre el mismo fenómeno en la ciudad de Toledo. Esta variable lingüística *depende* de otras variables independientes, ya sean lingüísticas (la posición interior o final de la -s, el sonido colindante o la categoría gramatical de la palabra) o sociales (el sexo, la edad o el nivel sociocultural).

variación fonética comporta el problema del establecimiento de los distintos tipos de sonidos que van a considerarse como variantes puesto que ello supone tomar decisiones que implican un cierto grado de arbitrariedad.

La variación gramatical, que abarca el nivel morfológico y sintáctico, trae consigo más problemas teóricos a la hora de realizar análisis sociolingüísticos. Si bien la morfología denominada “gramatical” presenta características similares a las de la variación en el nivel fonético (la frecuencia de aparición, los inventarios cerrados...), tanto la morfología “léxica” como la variación en el nivel sintáctico se enfrentan a la dificultad de establecer si las distintas variantes son realmente equivalentes entre sí. Es decir, al contrario de lo que ocurre con la variación fonética, donde parece imposible dudar de que las distintas variantes equivalen, no puede afirmarse tajantemente que las diversas variantes sintácticas no supongan algún cambio de significado⁷⁹. Con el fin de salvar estos inconvenientes, se ha distinguido entre dos tipos de equivalencia: la lógica (de verdad o referencial) y la pragmática. En general, se considera que, para que pueda realizarse un estudio sociolingüístico de variación sintáctica, es suficiente con demostrar la equivalencia de las variantes en el nivel referencial (Moreno Fernández 1998: 121-122)⁸⁰.

La variación léxica presenta problemas muy similares a los de la variación gramatical. La mayoría de los sociolingüistas defiende la existencia de la sinonimia, al menos en el nivel referencial, aunque es evidente que resulta muy difícil demostrar que dos variantes tienen idéntica función pragmática. Además, supone una ardua tarea determinar cuáles son los límites entre las distintas variables léxicas y, por tanto, qué unidades deben ser consideradas como variantes de una misma variable⁸¹. Otro problema con el que se enfrenta el estudio de la variación léxica es de índole metodológica. Resulta muy complicado recoger datos suficientes a partir del discurso

⁷⁹ Lo que subyace en este dilema es la antigua polémica sobre si existe o no la sinonimia perfecta. Para ampliar información sobre las diferencias que plantea la variación sintáctica en comparación con la fonética, véase Silva-Corvalán (1989: 98).

⁸⁰ Sin embargo, no todos los autores están de acuerdo en este punto. Las opiniones van desde posturas conciliadoras, como la de David Sankoff o Carmen Silva-Corvalán, hasta tajantes, como la de Beatriz Lavandera, para quien es imposible extender el estudio de la variación sociolingüística más allá de los límites de la fonología (López Morales 1989: 91-105, Moreno Fernández 1998: 122-124, Almeida 1999: 40-47).

⁸¹ La variación léxica puede abordarse desde el plano del contenido o desde el de la expresión. Para conocer con exactitud cuáles serían todas las posibilidades teóricas a la hora de establecer distintas variantes léxicas, puede verse Escoriza (2009, 2012).

natural hablado. Por esta razón, exceptuando algunos estudios de corte etnográfico llevados a cabo mediante la convivencia continuada, el resto de trabajos sobre la variación en este nivel lingüístico se han llevado a cabo por medio de encuestas y cuestionarios⁸². Todos estos inconvenientes han provocado que los estudios sociolingüísticos de variación léxica sean aún muy escasos⁸³. No obstante, un buen ejemplo de ellos podría ser el conjunto de investigaciones que se desarrollan en el mundo hispánico sobre frecuencia léxica o disponibilidad, eje fundamental de este trabajo (López Morales 1978).

Los estudios variacionistas realizados hasta el momento han demostrado que unas variables independientes tienen más incidencia que otras en función del nivel lingüístico que se estudie. De esta forma, mientras que la variación fonético-fonológica y la de tipo morfológico se ven determinadas principalmente por factores lingüísticos y sociales, la variación sintáctica se ve influida mayormente por factores lingüísticos, y la variación léxica por factores extralingüísticos (Moreno Fernández 1998)⁸⁴.

Con respecto a las variables sociales, lo primero que debiéramos señalar es que resulta imposible saber *a priori* con total exactitud qué tipo de variables van a actuar sobre el elemento lingüístico objeto de estudio. Esto se debe a que los factores sociales se configuran de distinta manera en cada comunidad de habla y, además, actúan sobre la lengua de forma irregular. Por esta razón, antes de iniciar un estudio sociolingüístico es imprescindible poseer información sobre la comunidad en cuestión. Esta información nos la puede proporcionar un estudio sociológico existente previo o la propia experiencia del investigador en caso de que sea miembro de la comunidad estudiada⁸⁵.

Una de las variables sociales que se ha tenido tradicionalmente en cuenta ha sido la del factor *sexo*, que hoy algunos autores llaman *género*. Según se ha comprobado, en

⁸² A pesar de que con ello se contradiga el principio de que se debe estudiar el comportamiento lingüístico cuando el hablante no se siente sistemáticamente observado. Es lo que William Labov denominó “paradoja del observador” (Labov 1983: 266).

⁸³ Blas Arroyo (2005: 79) comenta al respecto: “el estudio sociolingüístico del léxico es, sin lugar a dudas, la gran asignatura pendiente del variacionismo hispánico y en general, de todos los dominios lingüísticos”.

⁸⁴ Sin embargo, no hay que olvidar que los factores sociales están siempre supeditados a los lingüísticos, puesto que actúan solo donde el sistema lo permite (López Morales 1989: 141).

⁸⁵ El primer método es objetivo y el segundo, en cambio, subjetivo. Los dos enfoques presentan inconvenientes.

las comunidades más flexibles y de mayor interrelación entre varones y mujeres, como los grandes núcleos occidentales, las diferencias suelen ser principalmente de preferencias de uso. Es una opinión generalmente extendida que las mujeres acostumbran a utilizar con mayor frecuencia que los varones ciertas unidades lingüísticas como los eufemismos, los diminutivos o ciertos prefijos, aunque obviamente las diferencias entre los sexos pueden hallarse en cualquier nivel lingüístico⁸⁶.

La dialectología fue la primera disciplina que se interesó por el estudio de las diferencias lingüísticas en relación con el sexo, y en concreto, por conocer los rasgos del habla de las mujeres⁸⁷. A pesar de este hecho, han sido mayoritariamente las investigaciones sociolingüísticas las que han llegado a conclusiones más relevantes sobre las diferencias lingüísticas entre ambos sexos. Desde esta perspectiva se ha demostrado que el comportamiento de las mujeres es siempre más ajustado a las normas prestigiosas de la comunidad mientras que los varones, por el contrario, prefieren las formas más alejadas de la norma⁸⁸. Parece ser que estas diferencias son debidas a los distintos roles asignados a ambos colectivos en la sociedad. Desde esta perspectiva, llamada *teoría del dominio*, se explica el apego de las mujeres a las variantes de prestigio porque, a través de ellas, adquieren o se les asigna un estatus dentro de su comunidad que no pueden conseguir por otros medios⁸⁹.

Otro de los factores que se tiene en cuenta en cualquier estudio variacionista es la *edad*. Este factor, como ya había hecho notar también la dialectología, es uno de los factores sociales más importantes. La sociolingüística ha señalado que se producen diferencias notables en función de la edad. Concretamente, el habla de los más jóvenes ha sido caracterizada como innovadora en comparación con las manifestaciones

⁸⁶ Véase, por ejemplo, Calero (2007, capítulos 3 y 4), donde la autora realiza un repaso teórico de los fundamentos de la covariación entre el sexo y los rasgos lingüísticos, así como de los diferentes trabajos que se han ocupado de las diferencias lingüísticas y comunicativas entre varones y mujeres.

⁸⁷ El volumen de la revista *Orbis* del año 1952 se dedicó a dilucidar si era conveniente utilizar a mujeres en los estudios dialectológicos y si su habla era innovadora o más bien de carácter conservador.

⁸⁸ Esta aparente paradoja se ha resuelto mediante la siguiente explicación: mientras que las mujeres se apegan al prestigio abierto o manifiesto (*overt prestige*), aquel que se relaciona con las formas estándares, los hombres prefieren el prestigio latente o encubierto (*covert prestige*), vinculado a las formas subestándares (Trudgill 1974, López Morales 1989: 118-128, Silva-Corvalán 1989: 69-71, Moreno Fernández 1998: 34-40, Almeida 1999: 69-79).

⁸⁹ Para un resumen de la incidencia del *sexo* y de otros factores sociales en las comunidades de habla hispánicas, puede consultarse Blas Arroyo (2005: 157-249).

lingüísticas de los ancianos. López Morales (1989: 117-118) ha resumido cuáles son los patrones que se dan en función del grupo generacional: una estratificación clara de elementos léxicos en competencia, un mayor conservadurismo en el empleo de palabras tabuizadas en las generaciones mayores y un uso abusivo de los términos indefinidos, de las metaforizaciones festivas y de los nuevos términos en la generación joven.

Otra perspectiva desde la que se presta atención a la variación en función del factor edad es la del cambio lingüístico. Los datos sobre las diferencias intergeneracionales pueden ayudarnos a predecir qué tipo de cambios se van a dar en el futuro. Si un determinado fenómeno lingüístico conforma un patrón regular de estratificación continua al relacionarlo con el factor edad, es casi seguro que nos hallamos ante un cambio en marcha que acabará por generalizarse en esa comunidad⁹⁰. Los estudios sociolingüísticos, generalmente, no estiman conveniente recoger datos de informantes menores de 14 ó 15 años salvo en análisis sobre adquisición de lenguas o sobre grado de dominio lingüístico. A partir de esta edad, se distribuyen los informantes en tres o cuatro grupos generacionales. Los límites de los distintos grupos pueden establecerse en función de que las categorías sean equivalentes numéricamente o de que los informantes se hallen en circunstancias sociales parecidas.

Por otro lado, todas las sociedades se estratifican según criterios culturales y económicos conformando así diversos grupos o *niveles socioculturales*. Las diferencias lingüísticas que pueden darse entre los distintos estratos de una misma comunidad están relacionadas con el mayor o menor grado de distanciamiento social que exista entre los diversos grupos. En las sociedades occidentales es habitual que los sujetos de los distintos grupos interactúen entre sí a diario; en este caso, la estratificación es laxa y fluida y, por lo tanto, los sociolectos se diferenciarán poco entre sí. Sin embargo, eso no significa que este factor no tenga un gran peso a la hora de establecer estratificaciones de fenómenos lingüísticos en estas y otras comunidades, como han demostrado los estudios sociolingüísticos⁹¹. Los trabajos variacionistas también han constatado que los

⁹⁰ Esta óptica de estudio del cambio lingüístico se denomina “estudio en tiempo aparente”. Cuando además coincide con la existencia de un patrón curvilíneo con respecto al nivel sociocultural, es prácticamente imposible que ese rasgo lingüístico no acabe por imponerse, puesto que el estrato medio es el principal impulsor de los cambios lingüísticos (Labov 1983).

⁹¹ López Morales (1989: 133) afirma tajantemente que “no existen sociedades complejas en las que el factor sociocultural no incida, a veces fuertemente, en numerosos casos de variación a través de todos los niveles de la lengua”.

miembros de una comunidad comparten los mismos criterios acerca de qué variantes son mejor consideradas y cuáles peor. Además, cada una de las variantes de una variable puede cargarse de connotaciones positivas y, por tanto, convertirse en prestigiosa o, por el contrario, teñirse de valores negativos y estigmatizarse en función del estatus social que ostenten los grupos que las reproduzcan con mayor frecuencia. Labov (1983: 154-166) demostró que todos los grupos sociales tienden a usar las formas más reputadas conforme se pasa de los estilos más informales a los más formales. También llegó a la conclusión de que el deseo de ascensión social de las clases intermedias, especialmente del estrato inmediatamente inferior al más alto, les lleva en muchas ocasiones a usar las formas de prestigio incluso por encima de los valores numéricos que alcanza la clase alta⁹².

A pesar de la importancia que, como acabamos de ver, tiene esta variable social, son también muchos los problemas metodológicos que plantea. Para empezar, no está del todo claro qué parámetros deben integrar el nivel sociocultural, ni tan siquiera si esos deben ser los mismos o deben variar según la comunidad de habla que estudiemos. En general, las dimensiones que han integrado este factor han sido las referentes al nivel de instrucción, la cantidad de ingresos o la ocupación laboral del informante. Otra cuestión por resolver es si cada parámetro debe tener la misma importancia o no a la hora de establecer el nivel socioeconómico⁹³. Además, la utilización de diferentes dimensiones podría ocultar la importancia primordial de alguna de ellas.

Todos estos inconvenientes han provocado que algunos autores hayan buscado alternativas a esta variable social. En este sentido, las principales soluciones se han encontrado en las nociones de *mercado lingüístico* o de *red social* (Sankoff 1978, Milroy 1978 *apud* Moreno Fernández 1998: 50). El primero de estos conceptos consiste en la asignación de una puntuación a cada informante en función de su profesión, habida cuenta de que en algunas profesiones se tiende a usar una variedad más estándar de la lengua que en otras. La segunda alternativa, las redes sociales, consiste en un estudio detallado de la densidad y complejidad de estas, conformadas por individuos que se relacionan lingüísticamente entre sí.

⁹² Tal comportamiento lingüístico, descubierto por Labov en su estudio sobre el inglés de Nueva York, recibe el nombre de *patrón de hipercorrección* (Labov 1983: 167-188).

⁹³ Esto suele resolverse estableciendo un coeficiente distinto para cada uno de ellos que permite ponderar los datos.

Otras variables sociales que se han tenido en cuenta en algunos estudios sociolingüísticos son la *procedencia* y la *raza* o la *etnia*. Con respecto a la variable *procedencia*, se ha demostrado que su inclusión es adecuada cuando se trabaja con comunidades de habla cuya composición demográfica cuenta con un número importante de inmigrantes. En tal caso es posible estudiar en qué grado influye el origen geográfico o dialectal del informante en relación con la variación de un fenómeno lingüístico concreto. Por otro lado, la raza del informante como variable social nos informa de qué grupos étnicos promueven las distintas variantes de una variable lingüística⁹⁴.

2.2.1. La sociolingüística y la disponibilidad léxica

Ya desde los inicios de la disponibilidad, los investigadores franceses se percataron de que estos estudios podían ser útiles a otras disciplinas⁹⁵, aunque recordemos que su objetivo inicial, la enseñanza del francés como lengua extranjera, ya se relaciona con la lingüística aplicada. En el mundo hispánico, López Morales ha recalado en muchas ocasiones (1978, 1986, 1995-96 y 1999a)⁹⁶ las innumerables utilidades que pueden tener los resultados de los estudios de disponibilidad, no solo con respecto a la lingüística aplicada —ya sea a la enseñanza de la lengua materna o a la de lenguas extranjeras— sino también con relación a otras disciplinas. Por ejemplo, desde la psicolingüística, se pretende —a través de estas pruebas— saber más sobre la estructura del lexicón mental y sobre las relaciones asociativas que se establecen en el seno del mismo. Algunos de los ejemplos más recientes llevados a cabo desde esta óptica son los trabajos de Gómez Molina (2009) y Gómez Devís (2010), que utilizan el programa *Dispógrafo*, creado por Echeverría (Echeverría *et al.* 2008) para ejemplificar las redes asociativas existentes entre los vocablos más disponibles⁹⁷. Sin embargo, el

⁹⁴ Hay que decir en este punto, aunque sea obvio, que la raza (o la etnia), por sí misma, no condiciona el uso de una variedad lingüística determinada. Cuando este factor resulta relevante en un estudio sociolingüístico es porque se acompaña de otras condiciones sociales como una determinada posición social y cultural del grupo étnico dentro de la comunidad (López Morales 1989: 64, Moreno Fernández 1998: 64).

⁹⁵ “Elle [cette notion de disponibilité] offre donc un intérêt à la fois linguistique, psychologique et pédagogique” (Gougenheim *et al.* 1964: 152). El paréntesis es nuestro.

⁹⁶ Samper, Bellón y Samper Hernández (2003: 105-119) y Martínez Olmos (2008: 29-50) glosan las diferentes aplicaciones de la disponibilidad hasta la fecha. Una perspectiva más reciente tanto de los desarrollos teóricos como de los metodológicos puede verse en Paredes (2012).

⁹⁷ Además, el mismo autor creó el *índice de cohesión* (Echeverría *et al.* 1987), que evalúa el grado de coincidencia entre las respuestas de los informantes y que se utiliza en todos los estudios actuales.

intento más completo hasta la fecha de abordar el campo de la disponibilidad léxica desde la psicolingüística es la tesis doctoral de Hernández Muñoz (2006)⁹⁸.

Por otro lado, los frutos de la disponibilidad se aplicaron pronto a las comparaciones interdialectales con la intención de comprobar qué léxico era compartido por todos los hispanohablantes. Lejos quedan ya los bajos porcentajes de compatibilidad arrojados por las primeras incursiones en la materia, realizadas por López Chávez (1992), que se basaban en la comparación del total de vocablos. Más tarde, otros estudios demostraron que, al establecer medidas de corte en el cotejo⁹⁹, los resultados con respecto al léxico compartido eran muy superiores (Alba 1998, Samper Padilla 1999, González Martínez y Orellana Ramírez 2000 y 2000-2001, Carcedo 2003 y 2004, Samper Hernández 2008, etc.). Menos fructífera ha sido la vertiente etnolingüística de los estudios de disponibilidad. Únicamente, Carcedo (2000c) ha analizado —sirviéndose para ello de los listados de léxico disponible— las diferencias culturales existentes entre diversas comunidades hispanas, españolas o hispanoamericanas, por un lado, y, por el otro, un grupo de aprendices finlandeses de español.

Con respecto al campo de la enseñanza de la lengua, López Morales (1978) en fecha temprana ya propone la utilidad de la disponibilidad, ya que los diccionarios de léxico disponible pueden servir para planificar el vocabulario que debe ser aprendido en las escuelas. Con la intención de comprobar las distintas fases en la adquisición del vocabulario, se han realizado investigaciones que trabajan con distintos niveles educativos, como por ejemplo, la de García Megía (2003) o el premiado trabajo de Samper Hernández (2009)¹⁰⁰. Muy fecunda también ha sido la aplicación de la disponibilidad a la enseñanza del español como lengua extranjera. Entre las aportaciones destacan los trabajos pioneros de Carcedo, centrados en analizar la disponibilidad léxica de los estudiantes finlandeses de español (1998b, 2000a), pero que

⁹⁸ Esta autora revisa los trabajos que, desde el marco de la disponibilidad, aluden al paradigma de la psicología del lenguaje, y los divide en estudios que tienen como objetivo la elaboración de normas de referencia para otras investigaciones, estudios de carácter teórico sobre la organización léxica y estudios sobre la psicología evolutiva y el bilingüismo (Hernández Muñoz 2006: 41-45).

⁹⁹ Normalmente estas medidas de corte se basan en la *posición* de las palabras (las 50 primeras), en el *índice de disponibilidad* (en 0.02 ó 0.05) o el porcentaje de *frecuencia acumulada* (75%). El debate sobre cuál debe ser el criterio adoptado en las comparaciones intramuestrales o intermuestrales está presente en varios de los estudios de disponibilidad (v.gr. Bellón 2003, Bartol 2001 o Fernández Smith *et al.* 2008).

¹⁰⁰ En la provincia de Huelva también se ha recogido en varias ocasiones el léxico disponible de otros niveles educativos (v.gr. Prado y Galloso 2006a).

también abordan otros aspectos, como la evolución del aprendizaje (1999a), el análisis de los errores más frecuentes (1999c) o el cotejo de los resultados con los de los hablantes nativos (1999b). También Samper Hernández (2002) y, más recientemente, Sánchez-Saus (2011), se han dedicado al análisis de la disponibilidad léxica de los estudiantes de E/LE. En Jiménez Catalán (2014) pueden hallarse distintos estudios realizados desde esta perspectiva, tanto en español como en inglés como L1, así como una revisión general de las investigaciones dedicadas a esta rama de la disponibilidad (Samper Hernández y Jiménez Catalán 2014)¹⁰¹.

A pesar de todas las aplicaciones que acabamos de enunciar, la sociolingüística es la principal perspectiva desde la que se aborda el estudio de la disponibilidad léxica. De hecho, ambas disciplinas están relacionadas desde el inicio de la disponibilidad, como hemos comentado más arriba. A pesar de que la sociolingüística se desarrolla, como ya dijimos, en la década de los años sesenta y que, por tanto, los creadores de la disponibilidad no pudieron conocerla como tal, sus bases ya están presentes en los primeros estudios de léxico disponible. De hecho, los pioneros francófonos dieron cuenta de la importancia que algunos factores sociales como el sexo o el medio geográfico cobraban al analizar las palabras obtenidas en las listas de disponibilidad (Gougenheim *et al.* 1964: 163-182).

López Morales (1995-96: 252), con respecto a las relaciones entre la sociolingüística y la disponibilidad, afirma que:

“la sociolingüística ve en la disponibilidad un instrumento que le permite establecer estratificaciones de comunidades de habla. Somete a análisis de covariación a la disponibilidad con los factores sociales que le interesa manejar y determina la caracterización léxica de los sociolectos que integran la comunidad”.

Es decir, se considera, en la actualidad, que los estudios de disponibilidad léxica se enmarcan dentro de la variación léxica y constituyen un modo especial de variación en el que las respuestas de los informantes son las distintas variantes. De esta manera, se estudia la relación que hay entre estas y unos factores lingüísticos determinados (*los centros de interés*) y unos factores sociales (*sexo, nivel sociocultural, tipo de centro*, etc.). De este modo, al igual que en la regla variable, se intenta predecir cómo será el

¹⁰¹ Hace relativamente poco que se ha iniciado una nueva línea de estudio dedicada a la disponibilidad léxica de colectivos inmigrantes. En este sentido, hasta la fecha, se ha trabajado con adultos (Fernández Merino 2011, 2013) y con escolares no hispanohablantes (Jiménez Berrio 2013).

léxico que evoque un sociolecto determinado, a la vez que se describen las características que diferencian la producción léxica de cada grupo social, tanto cuantitativa como cualitativamente¹⁰². Para comprobar que las diferencias entre los subgrupos no son fruto del azar, casi todos los trabajos que se realizan en la actualidad (Hernández Muñoz 2006, Fernández Smith *et al.* 2008 o Areta 2009) analizan los datos con programas estadísticos que prueban que las diferencias halladas son realmente significativas¹⁰³. En cuanto al análisis cualitativo, predominan los cotejos entre los distintos subgrupos de cada variable para determinar cuál es el léxico compartido y cuál el divergente¹⁰⁴ y, más allá, se observa qué palabras concretas difieren de un sociolecto a otro y se buscan explicaciones que justifiquen este hecho.

Hasta la fecha, los factores más comúnmente analizados han sugerido ya la existencia de tendencias comunes en los estudios hispánicos¹⁰⁵. Por ejemplo, con respecto a la variable *sexo*, por lo general, se constata que —aunque no con grandes diferencias— las adolescentes alcanzan mejores promedios que los varones, sobre todo en aquellas áreas temáticas relacionadas con el rol tradicional de la mujer como *La cocinay sus utensilios* o *La ropa*. También llama la atención las diferencias cualitativas que se observan en algunos centros, fruto —sin duda— de los condicionamientos culturales del entorno¹⁰⁶. En relación con la variable *nivel sociocultural*, la gran mayoría de los trabajos publicados coincide en señalar su destacada importancia frente a otras variables. En general, la mejor producción léxica se relaciona con los estratos más altos de la sociedad, definidos habitualmente en función del nivel de instrucción y de la profesión de los progenitores¹⁰⁷. López Morales (1979), como hemos mencionado más arriba, fue el primero en llamar la atención sobre las diferencias léxicas que se

¹⁰² Aunque se realizan principalmente análisis cuantitativos, eludiendo así los inconvenientes conceptuales que ocasionaría el estudio de la auténtica variación léxica (*cf.* con lo dicho en el punto 2.2 y con López Morales 1999a: 25).

¹⁰³ Normalmente se usa para este fin el programa estadístico SPSS.

¹⁰⁴ Estos casi siempre se realizan acotando el total de *vocablos* de algún modo, para comparar realmente el léxico de disponibilidad alta y no todo el corpus. Un criterio habitual es el de contrastar las cincuenta primeras palabras por centro como hacen Gómez Molina y Gómez Devís (2004), Martínez Olmos (2008) o Llopis (2009), aunque también se han utilizado otros.

¹⁰⁵ Seguimos, a partir de ahora, la síntesis de Samper, Bellón y Samper Hernández (2003: 70-87) y Samper Padilla y Samper Hernández (2006: 45-79).

¹⁰⁶ Lagüens (2008: 119), por ejemplo, detalla cómo, en *Juegos y distracciones*, los varones tienen preferencia por términos como “fútbolín” o “videojuego”, y en cambio, las mujeres optan por “comba” o “escondite”. Otro centro donde también se suelen hallar diferencias es *Profesiones y oficios*.

¹⁰⁷ En un principio se tenía también en cuenta el nivel de ingresos de la familia, aunque se ha acabado descartando en muchas comunidades por la dificultad que implica obtener esta información y confirmar la veracidad de la misma. La mayor parte de estudios actuales sigue la propuesta de Samper Padilla y Hernández Cabrera (1997).

producían, tanto cuantitativas como cualitativas, entre los distintos niveles socioculturales y relacionó —aunque con cautela— sus datos con la *teoría del déficit* de Bernstein. Todos los estudios de disponibilidad realizados corroboran la importancia de esta variable, siendo, a veces, incluso la que arroja diferencias más significativas entre grupos sociales¹⁰⁸. Los trabajos en los que esta variable no cobra relevancia constituyen realmente excepciones dentro de la tendencia general (v.gr. Llopis 2009 o Prado *et al.* 2009, con respecto a la muestra en castellano)¹⁰⁹.

Otra de las variables estudiadas en el Proyecto es la *zona geográfica*, clasificada normalmente en la dicotomía urbana/rural según la ubicación del centro educativo. En función de la comunidad este factor se divide en diversas regiones o, simplemente, diferencia la capital del resto de la provincia. También según el tipo de comunidad se considera necesario diferenciar la ubicación del centro educativo de la residencia de los progenitores. En general, los estudiantes de las zonas urbanas muestran una productividad léxica mayor, aunque la importancia de esta variable difiere según la comunidad de la que se trate y tampoco faltan ejemplos donde se corrobore lo contrario (*cf.* Gómez Molina y Gómez Devís 2004). Por otro lado, también se han estudiado, aunque en menor medida que las cuantitativas, las diferencias cualitativas en relación con esta variable. Un buen ejemplo de ello puede hallarse en Enguita (2008).

Para finalizar, la última variable que incorporan todos los estudios adscritos al Proyecto es el *tipo de centro educativo*. Hasta ahora predominan las investigaciones donde, al contrastar los datos, los centros privados o concertados alcanzan mejores resultados cuantitativos¹¹⁰. Se ha discutido mucho sobre la independencia de este factor, muy a menudo relacionado con el nivel sociocultural de la familia, aunque en la actualidad se opta por incluirlo de forma separada, pues no siempre hay una equivalencia total entre el alumnado de los estratos más altos y el de los centros privados. Desde el punto de vista cualitativo también se observan algunas diferencias,

¹⁰⁸ Samper Padilla (2006: 117) dice al respecto: “es raro encontrar una comunidad donde no sea perceptible su incidencia; por otro lado, en muchas ocasiones constituye el factor que establece la discriminación social más acusada”.

¹⁰⁹ En nuestro trabajo (Serrano Zapata 2004) sobre la ciudad de Lleida es el estrato medio el que ofrece los mejores promedios, aunque, como ya indicamos en aquella ocasión, el tamaño de la muestra puede estar desvirtuando los resultados.

¹¹⁰ Uno de los lugares donde esta variable adquiere mayor relevancia, siempre a favor de los centros privados, es en Valencia, tanto en castellano (Gómez Molina y Gómez Devís 2004) como en catalán (Llopis 2009).

por ejemplo, Martínez Olmos (2008: 261) comprueba cómo en el campo *Profesiones y oficios* los alumnos de centros privados incluyen un número mayor de profesiones que precisan de estudios universitarios (como “periodista”, “farmacéutico” o “químico”).

Como hemos podido ver, la sociolingüística y la disponibilidad léxica han dado frutos muy interesantes hasta el momento. Nuestro trabajo pretende ir más allá y analizar sociolingüísticamente el léxico disponible en las dos lenguas de la comunidad estudiada, esto es, la provincia de Lleida.

2.3. EL ESTUDIO DE LAS LENGUAS EN CONTACTO

Existen en el mundo, en la actualidad, más de 5000 lenguas, lo que significa que la mayoría de estados, aunque no lo reconozcan oficialmente, son bilingües o plurilingües. Estas lenguas, generalmente habladas por las mismas personas, son las que entran en contacto. Por tanto, a pesar de que el interés por esta materia es relativamente reciente, la situación que se describe es más habitual de lo que podría imaginarse en un principio.

El análisis de los fenómenos derivados del contacto de lenguas, y, por ende, del bilingüismo o plurilingüismo, no constituye en sí mismo una disciplina científica, sino que es un campo de estudio que puede abordarse desde distintas perspectivas. Por ejemplo, la *psicolingüística* se interesa por los mecanismos que subyacen en el cerebro para poder almacenar dos lenguas, en los efectos cognitivos que provoca este hecho o en el proceso de adquisición mismo de una segunda lengua. Desde una perspectiva *sociológica*, se pone acento en los factores sociales que determinan la elección y las funciones sociales de una determinada lengua, en las actitudes lingüísticas de los hablantes, o en otros aspectos más generales como la planificación de las lenguas o la política lingüística. Por otro lado, los efectos producidos en los distintos niveles lingüísticos por la influencia de otras lenguas son estudiados desde un punto de vista estructural o *lingüístico*. La *sociolingüística*, además de examinar las consecuencias sociales de la interacción entre hablantes bilingües, propone que todos los fenómenos de contacto lingüístico pueden describirse y pronosticarse, una vez establecidas las condiciones externas e internas que rigen la convivencia entre las lenguas. Vemos, pues,

que esta área de estudio es muy fructífera y permite muchas y distintas aproximaciones en su análisis (Appel y Muysken 1996: 18-19, Almeida 1999: 183-184)¹¹¹.

Aunque hemos afirmado que el interés por este campo de investigación es relativamente reciente, sus orígenes se remontan al siglo XIX de la mano de Schuchardt (1880) y Whitney (1881)¹¹², en el marco de la lingüística comparativa. Sin embargo, el estudio del contacto lingüístico como tal, nace en los años 50 en EEUU, con las obras de Haugen (1953) y, sobre todo, con la tesis de Weinreich (1953), que abre el enfoque sociolingüístico de la materia y sienta sus principales bases teóricas¹¹³.

Dos lenguas, o más, entran en contacto, cuando existen hablantes bilingües que las usan alternativamente¹¹⁴. Desde esta perspectiva individual, uno de los mayores problemas ha sido, precisamente, definir qué se entiende por *bilingüismo*¹¹⁵. A lo largo de la bibliografía hay muchas definiciones al respecto que van desde “el dominio nativo de dos lenguas” hasta “el uso de expresiones completas y con significado en otras lenguas”¹¹⁶. Por tanto, uno de los problemas subyacentes consiste en determinar el grado de dominio de las dos lenguas que implica el bilingüismo, puesto que sabemos que el caso al que se refiere la primera definición, el *ambilingüismo*, es muy difícil de lograr y casi inexistente en el mundo¹¹⁷. Por otro lado, tampoco parece fácil medir el nivel de competencia del hablante bilingüe, que, además, siempre tendrá alguna habilidad lingüística más desarrollada en uno de los dos idiomas.

Baetens (1989: 22) afirma, en relación con los objetivos de nuestro estudio sobre disponibilidad léxica en hablantes bilingües catalán-castellano, que:

¹¹¹ En el presente apartado nos centramos en los conceptos que son de interés para esta tesis doctoral, fundamentalmente el *bilingüismo individual y social*, puesto que sería muy amplio, como acabamos de ver, tratar todos los campos de estudio de esta materia.

¹¹² Schuchardt (1880) funda los estudios sobre las lenguas criollas y Whitney (1881) analiza el papel de los préstamos (*apud* Appel y Muysken 1996: 16-17).

¹¹³ En los años 60 la disciplina sigue desarrollándose, principalmente, en Norteamérica y Gran Bretaña.

¹¹⁴ Aunque en el presente apartado hablamos de *bilingüismo*, nos referimos también a aquellas situaciones en las que están implicadas más lenguas, las llamadas situaciones de *plurilingüismo* o *multilingüismo social*.

¹¹⁵ Baetens (1989: 15) comenta de forma acertada la paradoja que se da alrededor de este concepto, toda vez que todo el mundo sabe lo que es, pero nadie lo define de forma satisfactoria.

¹¹⁶ La primera definición corresponde a Bloomfield (1933) y la segunda a Haugen (1953) (*apud* Moreno Fernández 1998: 211).

¹¹⁷ El individuo bilingüe tenderá a usar las dos lenguas en situaciones distintas, con lo que las habilidades lingüísticas diferirán en algún grado. El concepto contrario es el *semilingüismo* o la incapacidad de un individuo de usar con corrección alguna de las dos lenguas que habla. Aunque este concepto no está exento de polémica e incluso algunos autores han llegado a negar que exista (Baetens 1989: 30-31).

“de hecho, un hablante que tenga un alto nivel de competencia en dos lenguas, como por ejemplo un intérprete profesional, seguramente pasará como hablante nativo de las dos sin mostrar ningún rasgo de interferencia, pero la disponibilidad léxica en áreas semánticas determinadas será mayor en una lengua que en la otra”¹¹⁸.

Algunas definiciones que intentan evitar la polémica del grado de dominio por parte de los bilingües de las dos lenguas son las de Béziers y Van Oberke —“el bilingüismo es un medio doble de comunicación eficaz entre dos o más mundos, escogido libremente o bien impuesto, basado en el uso de dos sistemas lingüísticos” (*apud* Baetens 1989: 65)— o la de Siguán y Mackey (1986: 17) —el bilingüe es la “persona que, además de su primera lengua, tiene una competencia parecida en otra lengua y que es capaz de usar una u otra en cualquier circunstancia con parecida eficacia”—. De esta última definición se deduce la capacidad de separar las dos lenguas, de pasar de una a otra sin esfuerzo y de expresar los mismos significados en ellas (Moreno Fernández 1998: 212, Etxebarria 2002: 30-31).

Se puede establecer una amplia tipología de bilingüismo según el parámetro que sea tenido en cuenta¹¹⁹. En primer lugar, en función del contexto donde se aprenda la segunda lengua hablaremos de un *bilingüismo natural*, cuando este se produce en el entorno debido a las circunstancias que rodean al individuo, o un *bilingüismo secundario*, al que se llega mediante un aprendizaje formal en un aula. Según el dominio que tenga el hablante de las dos lenguas, se distingue el *bilingüe equilibrado*, cuyos conocimientos son aproximadamente equivalentes en ambas lenguas, del *bilingüe dominante*, cuya competencia en una de las dos lenguas es mayor. Asimismo, las destrezas lingüísticas sirven para diferenciar a los *bilingües receptivos* o *pasivos* de los *bilingües productivos* o *activos*. Mientras los primeros son únicamente capaces de entender o leer en una segunda lengua, los segundos pueden también escribir o hablar en ella. Hay que decir que el *bilingüismo receptivo* se adquiere más rápidamente y sus efectos son más perdurables que el *bilingüismo productivo*, que puede adquirirse en una determinada época de la vida y perderse en otra.

¹¹⁸ Citamos la edición de 1989, traducción al catalán de M. Argay. La traducción del catalán de esta y de las siguientes citas del mismo autor a lo largo de este capítulo es nuestra.

¹¹⁹ Seguimos a Baetens (1989: 13-73) en esta clasificación. Véanse también los ejemplos ilustrativos de Hoffmann (1991: 16-26).

Por otra parte, según el desarrollo que tenga el bilingüismo en la persona, podemos hablar de un *bilingüismo incipiente*, en el primer estadio del aprendizaje cuando se empiezan a descodificar las estructuras del otro idioma, de un *bilingüismo ascendente*, cuando sus habilidades en la segunda lengua se desarrollan y, finalmente, de un *bilingüismo recesivo*, que implica pérdida de competencia en uno de los dos idiomas. La edad es otro factor determinante, siendo los *bilingües precoces* los que aprenden una segunda lengua antes de la adolescencia y los *bilingües tardíos*, los que la aprenden pasada esa etapa¹²⁰. La última de las clasificaciones que vamos a abordar, es la clásica entre *bilingüismo aditivo* y *sustractivo*, que se relaciona con las consecuencias que tiene para la primera lengua el hecho de aprender la segunda. En el *bilingüismo aditivo* la segunda lengua aporta al hablante un conjunto de aptitudes cognitivas y sociales que no afectan a las que ya ha adquirido en la lengua materna; en cambio, en el *bilingüismo sustractivo*, la segunda lengua se adquiere a costa de perder aptitudes en la primera. Salta a la vista que el *bilingüismo aditivo* —en el que la lengua materna del hablante normalmente es de prestigio— tendrá siempre efectos positivos, mientras que el *sustractivo* —en el que la primera lengua suele ser minoritaria—, no (Appel y Muysken 1996: 151-171).

Mucho se ha discutido sobre cuáles son los efectos del bilingüismo tanto en el nivel lingüístico, como en el cognitivo o sobre la personalidad. La mayoría de los trabajos se ha dedicado a comparar las habilidades de los bilingües y de los monolingües en ambas lenguas y a establecer qué lengua es la dominante en los bilingües. Con respecto a este segundo objetivo, son varias las técnicas empleadas para ello, como pruebas de traducción en las dos lenguas, presentación de afirmaciones y petición de respuestas en los dos idiomas, medición de la velocidad de respuesta a instrucciones recibidas en las dos lenguas, etc. También se han utilizado, con esta finalidad, las pruebas asociativas de palabras, con las que se pretende comparar el número de asociaciones aportado en cada lengua por los hablantes ante palabras estímulo equivalentes (Appel y Muysken 1996: 153-154). Esto último es lo que precisamente hacemos en la tesis doctoral que presentamos aquí y que supone una aportación al estudio del bilingüismo catalán-castellano en Lleida.

¹²⁰ Otros autores prefieren distinguir entre el bilingüismo de infancia, de adolescencia y de edad adulta. Dentro del *bilingüismo precoz* se distingue entre el *simultáneo*, cuando el niño aprende dos lenguas maternas a la vez, y el *consecutivo*, cuando aprende una después de la otra (Etxebarria 2002: 34).

Sobre los resultados de los estudios que analizan los efectos del bilingüismo, es curioso comprobar cómo, en una primera época —anterior a los años 60—, los investigadores coincidían en señalar los efectos nocivos que el hecho de ser bilingüe acarrea desde la infancia. Un vocabulario restringido, unas estructuras gramaticales limitadas, errores en la morfología, formas “desviadas” en el habla, etc., eran algunas de las conclusiones recabadas en estos trabajos. Sin embargo, las investigaciones realizadas tras esta etapa, ofrecieron resultados muy distintos y descubrieron en los bilingües una mayor conciencia de las diferencias lingüísticas, una mayor capacidad para aprender otras lenguas e incluso se insinuó que podían tener una capacidad intelectual superior. Después del éxito de estos resultados, otros investigadores fueron más cautos y señalaron, además de errores en la metodología de muchos estudios llevados a cabo con anterioridad, que la mayor parte de ejemplos que corroboraban resultados negativos se asociaba a situaciones de *bilingüismo sustractivo*, así como, a la inversa, el *bilingüismo aditivo* estaba presente en los trabajos que destacaban los efectos positivos. Por tanto, no era el bilingüismo en sí lo que conllevaba unos efectos beneficiosos o perjudiciales sino, más bien, el prestigio o desprestigio de las lenguas implicadas, el nivel socioeconómico del hablante o el programa educativo seguido.

Hoy en día, se asume que sí pueden señalarse algunas consecuencias favorables al desarrollo en los bilingües, principalmente una mejor flexibilidad cognitiva, lo que supone un mejor uso del pensamiento divergente, y una mayor capacidad de abstracción. No obstante, estos beneficios se supeditan a que el hablante haya alcanzado un cierto nivel de competencia bilingüe en ambas lenguas, la denominada *hipótesis del umbral* formulada por Cummins (1979, *apud* Appel y Muysken 1996: 156-161). Esto supone que es necesario, tal y como aconsejó la Unesco en el año 1951¹²¹, que la escolarización se haga en la lengua materna del hablante, especialmente cuando esta es una lengua minoritaria en la comunidad. Se ha demostrado que la enseñanza bilingüe en escolares de minorías lingüísticas tiene resultados positivos en todas las áreas. Además, cuando no solo la lengua, sino también la cultura se incluyen en el currículo, el

¹²¹ El estudio de la UNESCO, además, recomienda que la educación en lengua materna se extienda hasta la etapa educativa más avanzada posible (Fasold 1996: 435). Fasold (1996: 440-454) plantea varias objeciones a estas recomendaciones como, por ejemplo, la imposibilidad de llevarlas a cabo en la mayor parte de lugares del mundo o la ausencia de investigaciones adecuadas que las avalen.

alumnado suele mostrar menos problemas sociales y emocionales¹²². Por otro lado, en un programa bilingüe, los niños de lenguas mayoritarias también se benefician del bilingüismo, siempre y cuando la segunda lengua se introduzca en una fase temprana¹²³ (Grosjean 1982: 220-227, Appel y Muysken 1996: 151-174, Moreno Fernández 1998: 218-222, Baetens 1989: 13-73, Hoffman 1991: 118-156).

Otro aspecto muy interesante es la distribución psicolingüística de las dos lenguas en el cerebro del hablante. La clasificación de Weinreich (1996 [1953]: 44-49) inicia este campo de estudio proponiendo la existencia de un *bilingüismo coordinado*, *compuesto* o *subordinado*. En el primer caso, el del *bilingüismo coordinado*, dos palabras equivalentes en dos lenguas se asocian a significados diferentes; en el *bilingüismo compuesto*, en cambio, las dos palabras comparten un mismo significado; y en el *bilingüismo subordinado*, una de las dos lenguas es la dominante y los significados se interpretan siempre a partir de dicha lengua. Aunque, en un primer momento, surgieron multitud de investigaciones encaminadas a probar esta distinción, hoy en día ha caído en desuso y se considera que, en realidad, todos los bilingües son en, mayor o menor medida, *compuestos* o *coordinados*, entendiéndose como dos extremos de un mismo *continuum*¹²⁴. A partir de los años 80, los estudios se centran en las operaciones psicolingüísticas que permiten llevar a cabo el *cambio de código* y en las diferencias que presentan los bilingües respecto a la percepción, el reconocimiento de palabras o los procesos sintácticos y semánticos (Appel y Muysken 1996: 109-121, Grosjean 1982: 240-268).

¹²² Una de las consecuencias negativas en el desarrollo de la personalidad de los bilingües es la *anomia*, una sensación de desorientación personal o aislamiento social, provocada por la incapacidad de encontrar una solución a las exigencias incompatibles de dos lenguas y dos culturas. Este trastorno, que es potencialmente perjudicial para el desarrollo psicológico, aparece cuando las condiciones sociales no son favorables ni al bilingüismo ni a los individuos bilingües, y es más frecuente en adultos y adolescentes y en los estratos sociales bajos de la sociedad (Moreno Fernández 1998: 219, Baetens 1989: 233).

¹²³ Hay muchos tipos de programas educativos bilingües, uno de los más conocidos es el de *inmersión lingüística*, que consiste en realizar toda la instrucción en una de las dos lenguas y se usa cuando la otra lengua, aquella en la que no se instruye, es de prestigio o mayoritaria. Muy lejos de este concepto está la *enseñanza por sumersión*, expresión con la que se denomina el hecho de que los niños de una lengua minoritaria se escolaricen en la lengua mayoritaria, sin tener en cuenta su lengua materna, lo que sucede habitualmente con los colectivos de inmigrantes.

¹²⁴ El tercer tipo, el *bilingüismo subordinado*, se desecha en muchas ocasiones por considerar que los individuos que responden a este patrón no son realmente bilingües. Normalmente el *bilingüismo coordinado* se asocia a la existencia de distintos contextos de aprendizaje y el *compuesto* a un solo contexto. Por otro lado, algunos autores defienden que los tres tipos de bilingüismo descritos por Weinreich (1953) pueden hallarse en los estadios iniciales de adquisición de una segunda lengua a partir de la instrucción en la primera.

Otro interés clásico de los psicolingüistas ha sido cuestionarse la cantidad de lexicones mentales que existen en el cerebro de un hablante bilingüe. Aún no hay una respuesta clara para este interrogante, a pesar de que algunos trabajos han aportado pruebas tanto de un lexicón común como de más de uno; otros autores, en cambio, apuntan a que hay una única memoria semántica en el individuo bilingüe conectada a dos almacenes léxicos¹²⁵. Además, también se ha demostrado que los bilingües usan más el hemisferio derecho del cerebro que los monolingües, cuya capacidad lingüística se focaliza en el hemisferio izquierdo, aunque esta afirmación es cierta, sobre todo, cuando el bilingüismo es tardío, se da en un contexto informal, en los primeros estadios del aprendizaje y cuando se usa principalmente un método deductivo (Baetens 1989: 222-227, Grosjean 1982: 266-268).

Silva-Corvalán (2001: 272) propone una hipótesis general, basada en la psicología cognitiva, según la cual los hablantes bilingües desarrollarían diferentes estrategias para poder sobrellevar mejor el peso cognitivo que acarrea recordar y usar dos o más sistemas lingüísticos. Estas estrategias serían: (a) simplificar categorías gramaticales, oposiciones léxicas y fonológicas y diferencias fonéticas, (b) desarrollar construcciones perifrásticas, (c) transferir ítems léxicos, rasgos fonéticos, estructuras sintácticas o funciones pragmáticas, (d) hipergeneralizar formas lingüísticas; y (e) intercambiar de código. Profundizaremos más adelante en algunas de estas estrategias, como la *transferencia* o el *cambio de código* (vid. 2.3.1.).

Cuando dos lenguas están en contacto, la mayor parte de los miembros de la comunidad son bilingües y pueden expresarse más o menos adecuadamente en cualquiera de las dos lenguas. Sin embargo, se activa una serie de procesos, denominados conjuntamente *elección de lengua*, por los cuales los hablantes eligen, de manera inconsciente, utilizar una lengua u otra en cada una de las situaciones comunicativas en las que se encuentran. Estos procesos, que abordamos aquí desde la perspectiva del individuo, han sido estudiados desde distintas ópticas. En primer lugar, la sociología o la sociología del lenguaje han establecido *ámbitos* o *dominios* para

¹²⁵ La existencia de uno o dos lexicones dio lugar a las hipótesis del *almacenaje único* o *doble*. Otro interrogante era cómo conseguía el bilingüe mantener separadas las dos lenguas en su cerebro. Para responder a esta pregunta se postuló la existencia de una especie de interruptor cerebral, cuyo funcionamiento en segundos llegó a estipularse. Esta teoría ha sido superada en la actualidad ante la evidencia del *cambio de código*, que el hablante bilingüe realiza de forma automática.

intentar determinar el comportamiento del hablante, es decir, qué lengua empleará en cada ocasión. El concepto de *ámbito*, más amplio que el de *situación*, engloba el lugar, el momento, los participantes, los temas y las condiciones pragmáticas de la interacción (Blas Arroyo 2005: 422, Moreno Fernández 1998: 218). Mackey (1976: 375-384), además, indica que el grado de competencia que se adquiriera en cada lengua dependerá de las funciones o usos que haga el hablante de ambas lenguas. Entre estas funciones externas estarán los contactos que tenga el bilingüe con la familia, la comunidad, el medio escolar, los medios de comunicación de masas y la correspondencia¹²⁶.

Una forma de representación gráfica de los diferentes dominios es el *árbol de decisión*, donde factores como el grado de formalidad, el entorno o la etnia determinan la actuación del hablante¹²⁷. Esta teoría ha sido crítica toda vez que reduce de forma simple las opciones y porque fracasa en muchas ocasiones, lo que prueba que no es tan fácil predecir *a priori* cuál va a ser la elección lingüística del hablante.

En segundo lugar, desde la psicología social, se han enunciado otras teorías para tratar de explicar el proceso de elección de lenguas como la *negociación interaccional* o la *teoría de la acomodación*. La primera de estas teorías tiene en cuenta las necesidades del hablante, las características de su grupo social (situación subyacente o ausente) y los rasgos psicológicos y sociales de los interlocutores (situación inmediata). La segunda teoría se basa en la mejor evaluación que obtiene el individuo si su actuación lingüística se asemeja a la de su interlocutor, lo que acaba determinando muchas de sus elecciones lingüísticas. El hablante usará la convergencia para aproximarse a su interlocutor y la divergencia, para enfatizar las diferencias que haya entre ellos (Fasold 1996: 275-320, Blas Arroyo 2005: 421-437)¹²⁸.

¹²⁶ Entre los factores internos, Mackey (1976: 390-395) cita la edad, la aptitud —que engloba la inteligencia, la memoria, la actitud y la motivación— y el uso interior de la lengua. La función de la “correspondencia” podría sustituirse hoy por “correo electrónico”.

¹²⁷ Grosjean (1982: 135-145) revisa cada uno de los factores (participantes, situación, contenido del discurso y función de la interacción) y señala que la importancia de los mismos variará en función de la comunidad de habla.

¹²⁸ La *teoría de la negociación interaccional*, o de las situaciones que se superponen, es de Herman (1968) y la de la *acomodación* fue enunciada por Giles (1977) (*apud* Fasold 1996: 275-320). Fasold (1996: 316-317) pone el acento en las teorías antropológicas que explican la elección de lenguas mediante metodologías distintas a la sociología o a la psicología social. En su opinión, estas teorías, basadas en la observación participativa son las más útiles y las que analizan con mayor profundidad la elección entre dos o más lenguas. Esta perspectiva antropológica analiza este fenómeno como ejemplo de la visión que el hablante tiene de sí mismo como miembro de una única comunidad o como miembro de dos

Hasta ahora nos hemos referido al bilingüismo individual, como proceso que atañe a una persona. El *bilingüismo social* —entendido como el uso de dos o más lenguas por parte de una misma sociedad— ha sido frecuentemente estudiado desde la sociología del lenguaje¹²⁹. Appel y Muysken (1996: 10-11) distinguen tres tipos de sociedades bilingües: en el primer caso, la sociedad está constituida por dos grupos diferentes monolingües; en el segundo, las dos lenguas son habladas por todos los hablantes; y, en el tercero, un grupo es monolingüe y el otro, bilingüe¹³⁰. Las causas —generalmente factores históricos— que originan estas sociedades bilingües o plurilingües son diversas. Fasold (1996: 37-42) señala la migración, el imperialismo, el federalismo y las zonas fronterizas. Siguán y Mackey (1986), por su lado, nombran la expansión, la unificación, las situaciones poscoloniales y el cosmopolitismo¹³¹.

Cuando dos lenguas o más conviven en una misma sociedad, se pueden originar distintas situaciones según la valoración social que reciba cada una de esas lenguas por parte de los miembros de la comunidad. Puede ser que las dos (o más) lenguas mantengan un estatus social similar, o que tengan diferentes estatus pero sean valoradas positivamente, o que una de las dos lenguas tenga un estatus superior a la otra. Esta última situación conduce a que la lengua dominante vaya ocupando esferas que antes estaban reservadas para la lengua dominada (Almeida 1999: 184-186). Mackey (1976: 27-35) cita como causas que pueden provocar que una lengua se sitúe por encima de otra, la ocupación y la colonización, el comercio, la superioridad demográfica, el poder y el prestigio, la expansión y la ascendencia, la educación, la influencia económica, la religión y los medios de comunicación.

El contacto entre dos lenguas, además, puede tener como consecuencia el nacimiento de otras lenguas (los *sabires* o *criollos*) o, por el contrario, la sustitución y

comunidades al mismo tiempo. Blas Arroyo (2005: 423-437) glosa las investigaciones que se han realizado desde estos presupuestos teóricos en algunas comunidades hispanicas.

¹²⁹ Sin embargo, como afirma Gómez Molina (2000: 291), no se puede separar la noción de *bilingüismo individual* de la de *bilingüismo social*, puesto que es imposible explicar la interacción lingüística de un individuo sin considerar el contexto social en que esta se desarrolla.

¹³⁰ No obstante, como los mismos autores reconocen, es difícil hallar esta tipología en forma pura; lo más común es que se mezclen las distintas situaciones (Appel y Muysken 1996: 11).

¹³¹ Hoffman (1991: 158-163) distingue entre factores históricos y contemporáneos. Grosjean (1982: 30-36) cita como factores los movimientos humanos, el nacionalismo y la política del federalismo, la educación y la cultura, y otras razones como la industrialización o la religión.

pérdida de una de ellas. No obstante, hay factores que pueden contribuir a la vitalidad de una lengua dominada como son el estatus económico, el peso demográfico, el apoyo institucional y la disimilaridad cultural respecto a la lengua dominante (Appel y Muysken 1996: 52-59). Con respecto a la relación entre las dos lenguas, Baetens (1989: 19) distingue entre varios tipos de bilingüismo: (a) el *horizontal*, donde las dos lenguas distintas tienen el mismo rango lingüístico; (b) el *diagonal*, cuando los hablantes utilizan un dialecto o una lengua no estandarizada y una lengua estándar que no guarda relación genética con el dialecto; y (c) el *vertical*, que se da cuando en un mismo hablante coexisten un dialecto y una lengua estándar. Este último tipo ha recibido tradicionalmente el nombre de *diglosia*.

La *diglosia* es un concepto que se usó, en un principio, para definir la situación lingüística de Grecia y de muchos países del mundo árabe, donde conviven dos variedades muy diferentes de una misma lengua. Fue Ferguson (1959), quien en un artículo fundamental creó la noción de *diglosia*¹³², basándose en cuatro casos de comunidades diglósicas (suizo-alemana, griega, árabe y haitiana). Este autor define la diglosia del siguiente modo:

“una situación lingüística relativamente estable en la cual, además de los dialectos primarios de la lengua (que puede incluir una lengua estándar o estándares regionales), hay una variedad superpuesta, muy divergente, altamente codificada (a menudo gramaticalmente más compleja), vehículo de una considerable parte de la literatura escrita ya sea de un periodo anterior o perteneciente a otra comunidad lingüística, que se aprende en su mayor parte a través de una enseñanza formal y se usa en su forma oral o escrita para muchos fines formales, pero que no es empleada por ningún sector de la sociedad para la conversación ordinaria” (Ferguson 1959, *apud* Moreno Fernández 1998: 228).

Las principales características de las dos variedades implicadas (la A: alta y la B: baja) se hallan ya implícitas en la misma definición: las dos variedades tienen distintas funciones en la sociedad, la variedad A goza de un mayor prestigio y es la lengua literaria, los hablantes adquieren la B como lengua materna y aprenden la A en la escuela, la lengua A es una lengua estándar más compleja normativa y gramaticalmente, que posee más tecnicismos y cultismos, los sistemas de sonidos de A y B constituyen una estructura fonológica simple, y, por último, la diglosia tiene estabilidad en la comunidad lingüística.

¹³² Aunque el término, como tal, lo había acuñado un helenista francés (Blas Arroyo 2005: 397).

Posteriormente, Fishman (1979: 119-133) amplía el concepto haciendo posible su uso para referirse a dos variedades lingüísticas ya sean únicamente registros de la misma lengua o incluso dos lenguas completamente distintas¹³³. También relaciona este concepto con el del bilingüismo y establece cuatro hipotéticas situaciones:

- (a) bilingüismo y diglosia: en estas comunidades los hablantes dominan dos lenguas que realizan diferentes funciones,
- (b) bilingüismo sin diglosia: se produce en comunidades donde las dos lenguas son habladas indistintamente por los hablantes,
- (c) diglosia sin bilingüismo: aparece cuando cada lengua es hablada por un grupo social diferente, y
- (d) ni diglosia ni bilingüismo: se da en comunidades pequeñas, primitivas y monolingües.

Tras esbozar esta clasificación, podemos argüir que el caso (a) se corresponde con la segunda situación a la que aludían Appel y Muysken (1996: 10-11, *supra* p. 84), mientras que la situación (c) se corresponde con la primera situación descrita por los mismos autores. Por otro lado, la situación (b) es transitoria, puesto que una de las dos lenguas acabará por imponerse¹³⁴ y la glosada en (d) es irreal e infrecuente como el mismo autor reconoce¹³⁵.

Fasold (1996: 84-102) muestra sus dudas ante algunos puntos de las definiciones de Fishman y Ferguson, redefine el concepto de *diglosia* y lo extiende al hecho de que sean más de dos las variedades que se distribuyan funcionalmente en una comunidad. A este hecho lo denomina *poliglosia* y, según este autor, puede ocurrir en tres situaciones: la *diglosia con doble superposición de lenguas* —realmente hay dos diglosias para tres

¹³³ Recordemos que, para Ferguson, la diglosia no podía aplicarse en el caso de una lengua y sus respectivos dialectos pero, al mismo tiempo, A y B debían ser dos variedades de una misma lengua. Algunos autores han criticado que en los mismos ejemplos que aportaba Ferguson en su artículo (1959) había casos en que esta circunstancia no se daba (*vid.* Blas Arroyo 2005: 400). Fishman (1979: 120) señala que es obra de Gumperz la ampliación del concepto a dialectos y registros. Por otro lado, Fasold (1996: 71-108) compara y extrae sus propias conclusiones ante las divergencias entre el concepto de diglosia de Fishman y el de Ferguson.

¹³⁴ Fishman (1979: 132) afirma: “Sin normas y valores apartados aunque complementarios para establecer y mantener la separación funcional de las variedades lingüísticas, aquella lengua o variedad que tiene la fortuna de ser asociada con la corriente predominante de las fuerzas sociales tiende a desplazar a las otras”.

¹³⁵ Véanse en Pueyo (1991: 76-78) algunas reticencias a la clasificación de Fishman, que el sociolingüista leridano tilda de “poco operativa y dudosamente eficaz” (Pueyo 1991: 78).

lenguas que se superponen—, la *diglosia de esquema doble* —existe dentro de A dos subvariedades y dentro de B, otras dos— y la *poliglosia lineal o continua* —muchas variedades lingüísticas ordenadas según el grado de formalidad—.

Por otro lado, López Morales (1989: 64-83) considera la diglosia como un tipo de estratificación social extrema y rechaza la existencia de auténtica diglosia en el mundo hispánico, toda vez que, en su opinión, algunos de los ejemplos propuestos en la bibliografía son solo casos de bilingüismo donde una lengua se ha impuesto por razones culturales, sociales o políticas a otra.

Respecto al trasvase de elementos lingüísticos entre las dos lenguas implicadas, asunto que debatiremos en el próximo apartado (*vid.* 2.3.1.), Silva-Corvalán (1989: 179) opina que la transferencia de rasgos lingüísticos de una lengua a otra será más frecuente en el caso del bilingüismo no diglósico, que en el caso de una sociedad con diglosia, donde las lenguas se especialicen claramente en función de la situación y el tema.

Como ya hemos visto, la diglosia se asocia a una situación estable que, incluso, puede durar siglos. Por esta razón, el término se ha sustituido a menudo por otro, el de *conflicto lingüístico*, con el que se alude a situaciones más dinámicas. El concepto fue acuñado por Aracil en 1965 para denominar la situación conflictiva que genera el contacto de lenguas cuando los dos o más sistemas lingüísticos compiten entre sí (Aracil 1986). Por otro lado, Ninyoles (1969) afirma que el conflicto puede darse de forma inconsciente o de forma explícita. En este último caso, los hablantes suelen tomar medidas para repararlo. Por otra parte, el *conflicto lingüístico* siempre se relaciona con tensiones sociales más profundas, que evidencian problemas socioeconómicos, políticos, religiosos o históricos entre los individuos o los grupos sociales. El conflicto no tiene por qué tener connotaciones negativas, muchas veces la consciencia del mismo favorece las lenguas minoritas, que son así impulsadas por los ciudadanos e, incluso, por los organismos oficiales (Nelde 1997, *apud* Blas Arroyo 2005: 413-416).

2.3.1. Fenómenos de contacto lingüístico

Los fenómenos derivados del contacto de lenguas pueden ser de diversa índole según tengan su origen en el contacto de dos sistemas, en el uso de varias lenguas o bien estemos ante variedades resultantes de la convivencia de lenguas (Moreno Fernández 1998: 258). En el presente apartado nos ceñiremos a los fenómenos que surgen del contacto entre sistemas, es decir, la *interferencia*, la *convergencia*, el *préstamo* y el *calco* y algunos de los fenómenos originados por el uso de más de una lengua, como el *cambio* o la *mezcla de códigos*. Relegamos pues, ya que queda fuera de los límites de este trabajo, el análisis tanto del proceso de elección o sustitución de una lengua como de las variedades originadas por el contacto lingüístico, a saber, los *pidgin* y los *criollos*.

De entrada, hay que advertir que el panorama terminológico existente referido a los fenómenos de contacto entre sistemas no deja de ser un tanto desolador para el investigador que se enfrenta a él. Abundan en la bibliografía diferentes términos que, con frecuencia, se refieren a conceptos semejantes, mientras que, por otro lado, los mismos términos no tienen para todos los autores idénticos límites semánticos, con lo que se hace necesario redefinir en cada trabajo los conceptos para conocer el alcance de los mismos, tal como también hacemos nosotros (*vid.* capítulo 7)¹³⁶.

Dicho esto, comenzamos abordando la noción de *interferencia*, que, desde la definición clásica de Weinreich (1996 [1953]: 29), se cubre de cierto tinte peyorativo¹³⁷. Este autor la define como “desviación respecto de las normas de cualquiera de las dos lenguas, como resultado del contacto lingüístico” y, en su obra, que supone un verdadero hito en los estudios de contacto lingüístico, sienta las bases para su estudio en los diferentes niveles del lenguaje —fónico, gramatical y léxico—¹³⁸ mediante su descripción cuantitativa y cualitativa. Otros autores prefieren sustituir este término por el de *transferencia* (Clyne 1967, Silva-Corvalán 1989), revistiéndolo así de

¹³⁶ Gómez Molina (2000: 292) afirma al respecto: “En la actualidad no existe unanimidad ni en la delimitación conceptual de interferencia y préstamo ni en las denominaciones terminológicas”, o Romaine (1995: 180): “problems of terminology continue to plague the study of language contact phenomena”.

¹³⁷ El término, procedente del campo de la física, fue introducido por Sandfeld (1938), aunque fue Weinreich (1953) quien lo popularizó (*apud* Blas Arroyo 1993 y Gimeno y Gimeno 2003). Citamos la edición de 1996, traducción de F. Martínez, de la obra clásica de Weinreich. La traducción del catalán es nuestra.

¹³⁸ Con anterioridad a Weinreich, el estudio de los fenómenos de contacto lingüístico solía ceñirse al nivel fonético-fonológico, e incluso algunos autores negaban su posible existencia en otros niveles del lenguaje.

connotaciones más neutras¹³⁹; y no son pocas, por otro lado, las veces en las que se usa este último término, *transferencia*, para referirse de forma general a todos los procesos que suponen una modificación de la estructura de una lengua por influencia de otra (Almeida 1999: 204)¹⁴⁰, tal como hacemos en este trabajo (*vid.* capítulo 7). De este modo, se diferencia la *transferencia* de la *interferencia*, por referirse esta última únicamente a los casos en que los resultados evidencian estructuras agramaticales en la lengua receptora¹⁴¹.

Por otro lado, Silva-Corvalán (2001: 269) distingue estos dos conceptos según la estabilidad que tengan dentro del sistema lingüístico al que se incorporan, y afirma que, mientras la *interferencia* es algo pasajero, ocasional e inestable, la *transferencia* supone que los elementos transferidos cobran cierta estabilidad en la lengua recipiente. La autora la define como “la incorporación de rasgos de un idioma a otro, con una consecuente reestructuración de los subsistemas involucrados” y considera que puede darse una *transferencia* siempre que se den las siguientes circunstancias: (a) el reemplazo de una forma en la lengua S con una forma de la lengua F o la incorporación de una forma antes ausente en S, (b) la incorporación del significado de una forma R de la lengua F a otra forma, parecida estructuralmente a R, en el sistema S, (c) el uso más frecuente de una forma en la lengua S y (d) la pérdida de una forma en la lengua S (Silva-Corvalán 2001: 273 y 274).

Cabe recordar que el mismo Weinreich (1996: 49-51) ya diferencia las *interferencias* en el habla de las que se dan en la lengua, en función de su grado de integración y estabilidad en el sistema lingüístico. También Mackey (1976), desde una perspectiva similar, relaciona la *interferencia* con el dominio del habla y reserva el nombre de *préstamo* para los elementos procedentes de otra lengua que se emplean de forma colectiva y sistemática. Según este autor, la diferencia entre los dos conceptos radica en que el *préstamo* se usa por desconocimiento del vocablo equivalente en la

¹³⁹ El término procede de la psicología y hace referencia a la capacidad del individuo de proyectar su conocimiento de una lengua a otra (Boix y Vila 1998: 234). Otro de los términos utilizados para referirse globalmente a los fenómenos de contacto lingüístico es el de *marca transcódiga* (Lüdi 1987).

¹⁴⁰ En cambio, Payrató (1984) propone para este mismo uso el término de *interferencia*, tras repasar la terminología generalmente empleada en los contactos interlingüísticos y exponer sus reticencias a algunos de estos términos, como el de *préstamo*.

¹⁴¹ No obstante, para otros autores como López Morales (1989: 165) los dos términos son sinónimos en el sentido de que los dos hablan de la influencia de una lengua sobre otra, que produce estructuras agramaticales en esta última.

lengua que se está hablando y sin conciencia por parte del hablante de que está utilizando una palabra de otra lengua¹⁴².

En general, se considera que una de las diferencias fundamentales entre el *préstamo* y la *interferencia* es que, mientras esta última es propia del habla de los bilingües, el *préstamo* puede extenderse en el sistema, adaptándose a la gramática de la lengua receptora, y siendo usado incluso por los hablantes monolingües. El préstamo léxico suele concebirse, actualmente, como un proceso según su grado de integración en el sistema y se puede diferenciar, de este modo, entre *préstamos en vías de integración* o *préstamos consolidados*, si situamos estos conceptos en una línea imaginaria en uno de cuyos extremos hallaríamos la noción ocasional de *interferencia* (Gómez Molina 1999-2000: 317)¹⁴³.

En cuanto a los factores que pueden predecir el futuro de los préstamos y de las transferencias, Silva-Corvalán (1989: 174) cita tres: (1) la prominencia del rasgo transferido, (2) las actitudes subjetivas hacia la otra lengua y sus hablantes, y (3) factores socioculturales de gran complejidad que determinan que las transferencias se evalúen como prestigiosas o como rasgos estigmatizados. En general, se acepta hoy en día que el prestigio de las lenguas es un factor fundamental que sirve para caracterizar tanto la permeabilidad de estas hacia las transferencias como la dirección que cobrarán los trasvases lingüísticos. El hecho más habitual es que estos se produzcan de la lengua más prestigiosa a la menos, aunque también se ha constatado el ejemplo contrario¹⁴⁴; y algunos autores han demostrado empíricamente que, en un bilingüismo estable y continuo, las interferencias pueden ir en ambas direcciones (Baetens 1989: 95-96, Blas Arroyo 1993a: 39-40).

Gómez Molina (2000: 293) glosa las principales características que, comúnmente, se le otorgan al *préstamo*, esto es: (1) su relación con los intercambios económicos y culturales, (2) la adopción por parte de la lengua receptora, (3) la

¹⁴² Mackey (1976) distingue dos fases en el proceso de influencia interlingüística: la *interferencia* y la *integración*.

¹⁴³ Ya Poplack (1993: 256) distingue entre *préstamos establecidos* (*established loanwords*) y *préstamos ocasionales* (*nonce borrowings*). Mientras que los primeros están plenamente integrados, los segundos están en proceso de expansión.

¹⁴⁴ Baetens (1989: 96) cita estudios concretos que corroboran este hecho, e incluso, alguno que comprueba cómo, en situaciones de contacto, las expresiones obscenas y vulgares son fácilmente transferibles de una lengua a otra y, concretamente, de la menos valorada a la más prestigiosa.

transferencia de una lengua a otra con una probable alteración en la forma y, finalmente, (4) la concepción del fenómeno como un proceso que admite distintos grados de integración lingüística y social¹⁴⁵. El mismo autor señala que el grado de integración lingüística se constata evaluando los distintos niveles de la lengua en los que se produce la integración de la palabra así como midiendo las posibilidades de combinación con formantes de la lengua receptora, mientras que, para poder evaluar la integración social, tenemos que comprobar la frecuencia de uso, el nivel de homogeneidad y la aceptación del término. Estos dos parámetros sirven para determinar si nos hallamos ante una *interferencia* o ante un *préstamo*.

Partiendo de los presupuestos del multilingüismo social y asumiendo que todas las lenguas son producto, en mayor o menor medida, del contacto interlingüístico; Gimeno y Gimeno (2003: 79, figura 2.4.) elaboran un cuadro clarificador con la comparación entre los conceptos de *interferencia*, *préstamo no integrado* e *integrado*, y *cambio de código*, —concepto que veremos a continuación—, basándose también en el grado de integración lingüística y social del vocablo.

Tabla 2.1. Configuración de los grados de interferencia, cambio de código y préstamo en función de la integración lingüística y social.

	INTEGRACIÓN LINGÜÍSTICA	INTEGRACIÓN SOCIAL
Préstamo integrado	+	+
Préstamo no integrado	+ -	- +
Cambio de código	-	+ -
Interferencia	-	-

Fuente: Gimeno y Gimeno (2003: 79, figura 2.4).

De este modo, la *interferencia*, por ejemplo, se halla en el extremo inferior de la clasificación por su grado nulo tanto de integración lingüística en el sistema como de integración social en la comunidad de habla. En el polo opuesto se sitúa el *préstamo integrado*, dotado tanto de integración lingüística como de adopción social. Y, a medio camino, se hallan tanto el *préstamo no integrado*, que supone una adaptación parcial a

¹⁴⁵ Esta es la visión actual del fenómeno, que contrasta con la que se tuvo en un primer momento, cuando se analizaba el *préstamo* únicamente considerándolo como resultado, y no como un proceso.

la lengua receptora pero está lejos de ser utilizado por la mayoría de los hablantes de la comunidad, como el *cambio de código*, que, careciendo de integración lingüística, presenta un grado parcial de integración social.

Poplack (1987: 55-56) glosa los criterios fundamentales utilizados generalmente para caracterizar los *préstamos*, a saber, (1) la frecuencia de uso, (2) el desplazamiento de sinónimos de la lengua prestataria, (3) la integración morfofonológica y sintáctica, y (4) la aceptabilidad.

Desde el inicio de los estudios de los fenómenos resultantes del contacto lingüístico, los *préstamos léxicos* han sido la parcela más analizada puesto que, además de ser innumerables los ejemplos que se dan en este nivel de la lengua, cuando hay un contacto estable entre sistemas, también resultan en este nivel más fácilmente observables los trasvases lingüísticos. Un aspecto que ha interesado a los estudiosos ha sido cómo se adapta lingüísticamente el préstamo en la lengua receptora, ya que la lexía en cuestión puede integrarse únicamente en algunos niveles de la lengua —como el fonético, el morfológico o el sintáctico— o en varios simultáneamente, y adoptar, por ejemplo, la morfología y la fonética propias de la lengua receptora (Moreno Fernández 1998: 266)¹⁴⁶.

Ahora bien, en cuanto a qué motiva que una lengua tome palabras de otra, se han argüido distintas causas. La principal es aquella que engloba lo que se ha denominado *préstamos culturales*, es decir, aquellos que sirven para denominar cualquier elemento de otra cultura hasta entonces desconocido en la cultura de la lengua receptora y que, por tanto, carece de denominación¹⁴⁷. Es lo que Baetens (1989: 97) llama *préstamo de necesidad* y, según este autor, dichos préstamos pueden darse tanto en el sentido lengua de mayor prestigio a lengua de menor prestigio, como en sentido contrario¹⁴⁸.

¹⁴⁶ Para ampliar este punto, véase Gómez Capuz (1998: 223-251).

¹⁴⁷ El término proviene de Bloomfield (1933), quien lo diferencia del *préstamo íntimo*, que se produce entre dos lenguas habladas en un mismo territorio, o del *préstamo dialectal*, que se observa entre las variedades de una misma lengua.

¹⁴⁸ Cuando los préstamos sirven para enriquecer y reforzar el acervo léxico de una lengua, este autor los denomina *transferencias positivas*, en cambio, las *transferencias negativas* se dan cuando los préstamos afectan negativamente a las normas de la lengua receptora (Baetens 1989: 98).

Sin embargo, hay muchísimos casos en los que los préstamos no se producen por esa razón, sino que ocupan parcelas reservadas a elementos léxicos que ya existen en la lengua de origen para nombrar determinadas realidades. En estos casos, las causas pueden ser otras. Silva-Corvalan (2001: 290) cita, por ejemplo, factores de tipo psicolingüístico como la dificultad de recordar palabras de uso poco frecuente en una de las lenguas o la evaluación de algunos ítems como más expresivos en una lengua que en otra. Además el hablante puede considerar que una palabra de otra lengua transmite de forma más precisa lo que quiere decir, o, como ya hemos indicado anteriormente, el mayor prestigio conferido a una de las dos lenguas puede motivar el uso de determinados vocablos, como símbolo de su estatus social¹⁴⁹.

Baetens (1989: 98-100) enumera los criterios que condicionan la aparición de préstamos y considera que, en primer lugar, el tipo de actividad puede ser un condicionante, puesto que, si el bilingüe suele hablar de un tema determinado en una lengua, la disponibilidad léxica que tenga al utilizar la segunda lengua para el mismo ámbito temático será menor. En segundo lugar, el autor alude a la frecuencia de los contactos entre las dos lenguas, es decir, al uso más frecuente de una de las dos lenguas en una determinada actividad, a pesar de que la disponibilidad léxica del hablante para los términos equivalentes sea similar. Finalmente, considera que la frecuencia con que el hablante bilingüe usa cada lengua puede también influir en las transferencias, ya que la lengua de uso menos frecuente es más susceptible a las interferencias o los préstamos (incluso en el caso de que sea la lengua materna del hablante). Estos tres factores son además, según el autor, responsables de que la transferencia adopte la forma de *préstamo puro*, *parcial* o *calco*.

Una de las clasificaciones clásicas y formalistas del *préstamo* es la de Haugen (1969 [1953]: 383-411)¹⁵⁰. El autor aborda desde una perspectiva amplia la noción de *préstamo* como “the attempt by a speaker to reproduce in one language patterns which he has learned in another” (Haugen 1969: 363), y distingue entre *préstamo* (*loanword*) y

¹⁴⁹ Weinreich (1996: 138-145) aduce algunas de estas razones para justificar la existencia de préstamos, como la asociación simbólica de la lengua fuente con unos determinados valores sociales. A su vez, arguye factores lingüísticos internos como la baja frecuencia de algunas palabras, la homonimia perjudicial y la necesidad de sinónimos. También llama la atención sobre el hecho de que algunas palabras concretas nunca se transfieren, sin poderse saber por qué. Sobre las causas que originan el préstamo lingüístico, puede consultarse Gómez Capuz (1998: 255-279).

¹⁵⁰ Las clasificaciones formalistas, que se centran en el grado de alteración del modelo prestado, surgieron en la primera mitad del siglo XX a partir de los estudios de Werner Betz (Gómez Capuz 2004: 35).

calco (*loanshift*), según se dé más un proceso de importación de nuevos lexemas o de sustitución de lexemas nativos, respectivamente¹⁵¹. Dentro de los *préstamos*, distingue entre los *puros* —donde la importación es total— y los *híbridos* —donde además de haber importación, se da también una sustitución parcial—. También subdivide el *préstamo puro* según el grado de adaptación (y de este modo, hay *préstamos no asimilados*, *parcialmente asimilados* y *asimilados*); y los *préstamos híbridos*, según la parte de la palabra que se importe, dando lugar a *préstamos de raíz híbrida*, a *derivados híbridos* o a *compuestos híbridos*. Bajo esta última denominación se agrupan lexías complejas que combinan la transferencia de algunos de sus elementos y la reproducción de otros: si la parte importada es el núcleo, son *compuestos híbridos nucleares*, y si no, *compuestos híbridos marginales*¹⁵².

Con respecto a los *calcos*, Haugen distingue entre *extensiones*, que suponen la ampliación del significado de una unidad léxica ya existente en la lengua receptora, y *creaciones*, que son traslaciones nuevas en la lengua realizadas a partir de un modelo extranjero. Las *extensiones*, a su vez, se dividen en *homófonas*, *homólogas* y *sinónimas*, según la similitud entre el vocablo nativo y el foráneo; y las *creaciones*, en *literales* y *aproximadas*, en función de si siguen o no el orden exacto del modelo importado.

Es decir, Haugen (1969) agrupa bajo una misma etiqueta dos fenómenos distintos a los que, comúnmente, se les denomina *calcos*¹⁵³. Ambos fenómenos comparten la idea de “adopción” de significados o estructuras extranjeras, y se diferencian del *préstamo*, en que este último transfiere un elemento extranjero completo, mientras que en los *calcos* se utilizan lexías propias de la lengua receptora, y, simplemente, se asume un significado foráneo para una forma ya existente en la lengua o se crea una lexía compuesta nueva a imitación de un modelo de la lengua fuente. Al primer caso se le denomina *calco* o *préstamo semántico*, y, al segundo, *calco léxico* o *sintáctico*.

¹⁵¹ En realidad, Haugen habla de “morfemas”, aunque esta noción equivale al “monema” europeo (Gómez Capuz 2004: 38).

¹⁵² También Weinreich (1996: 127) se refiere a este préstamo, que él engloba en el tercer tipo de interferencia léxica referida a las palabras compuestas y frases. El investigador lituano cita el famoso ejemplo de “pelota de fly” (*fly ball*) para ilustrar esta subcategoría.

¹⁵³ En esta clasificación está presente la influencia de Betz (1949, *apud* Gómez Capuz 2004: 35-40), quien distingue entre *préstamos externos*, o integrales, y *préstamos internos*, donde incluye todos los tipos de calcos y préstamos semánticos.

El *calco* o *préstamo semántico* corresponde al segundo caso que cita Silva-Corvalán en su clasificación (2001: 271, *supra*), esto es, la incorporación del significado de una forma R de la lengua F, que puede ser parte del significado de una forma P en la lengua S, a otra forma, estructuralmente similar a R, en el sistema S. Una duda que ha asaltado a los estudiosos, en este punto, es cómo comprobar si las extensiones semánticas que se producen en este tipo de calcos son debidas a la influencia de otra lengua o bien suponen un desarrollo semántico propio. Generalmente, se acepta que, a mayor distancia semántica entre los dos significados, el tradicional y el añadido, es más probable la existencia de un *préstamo semántico* (Gómez Capuz 1998: 70-72).

En general, se distinguen, siguiendo la clasificación formalista de Haugen (1969), tres casos de *préstamos semánticos*, según haya similitud formal o semántica entre el término importado y el patrimonial¹⁵⁴. En primer lugar, hallamos los *análogos*, que aparecen cuando hay una semejanza tanto formal como semántica entre los dos términos. Son lo que se conoce como *cognados*, y que se dan, sobre todo, entre lenguas románicas, cuando dos palabras comparten el mismo étimo latino. En segundo lugar, nos encontramos con los *homólogos*, casos en los que solo existe similitud conceptual entre los dos términos pero no formal ni etimológica. Este hecho suele darse entre lenguas que no están emparentadas y algunos autores lo denominan *calco semántico* y lo diferencian del caso anterior, al que se refieren con el nombre de *préstamo semántico*¹⁵⁵. En último lugar, hallamos los *homófonos*, palabras que únicamente muestran un parecido formal, pero sus significados no guardan ninguna relación entre sí (Gómez Capuz 1998: 70-76 y 2004: 46-52).

Los *calcos* o *préstamos semánticos* son habituales en el habla de los bilingües, porque estos tienden a identificar los usos contextuales de los términos equivalentes. Baetens (1989: 100) reconoce que el análisis de las consecuencias semánticas de las interferencias es más difícil que la simple constatación de interferencias léxicas. A menudo, no hay ningún rasgo formal que las identifique, y solo con el tiempo el

¹⁵⁴ Cabe recordar que Haugen (1969) denomina a los préstamos semánticos, *extensiones*. La clasificación que mostramos reproduce fielmente la de Humbley (1974: 58-62). A pesar de ser heredera de la de Haugen (1969), hay diferencias importantes con respecto a la terminología.

¹⁵⁵ Otros autores creen que es una confusión ubicar el *préstamo semántico* dentro de la categoría de *calco* (Gómez Capuz 1998: 76). Por otro lado, con respecto a los *homólogos*, Humbley (1974: 59) comenta que es raro que este tipo de préstamos se dé entre lenguas estrechamente emparentadas, puesto que también aparecería una analogía en la forma.

interlocutor de un hablante bilingüe puede percatarse de que este utiliza la semántica de los términos de forma distinta al monolingüe, en especial si la interferencia es de tipo connotativo. Baetens (1989: 95) también afirma que “este tipo tan sutil de interferencia será mayor cuanto mayor sea el grado de similitud entre las dos lenguas”.

Por otro lado, el *calco estructural* o *léxico* se define como la “traducción literal de una lengua a otra de un lexema compuesto, un modismo u otras expresiones” (Silva-Corvalán 1989: 175) o “la sustitución de una palabra polimorfemática o compuesto nominal de la lengua modelo [...] por medio de morfemas o lexías simples, ya existentes por sí solos en la lengua receptora” (Gómez Capuz 2004: 53). Las clasificaciones formalistas de este tipo de *calco* se basan en el grado de fidelidad al modelo extranjero que se toma y son bastante semejantes entre sí. Ya Weinreich (1996: 125) distingue entre *calcos propiamente dichos*, en los que el modelo se reproduce exactamente, e *interpretaciones*, en las que el patrón de la otra lengua solo proporciona una indicación general. La misma distinción está presente en Haugen (1969), que denomina a este fenómeno *creaciones*, y distingue entre *literales* y *aproximadas*. También Humbley (1974), en la misma línea, distingue —según el grado de divergencia entre el modelo y el resultado del préstamo— entre (a) calcos que siguen el orden de los elementos de la lengua fuente, (b) calcos que invierten dicho orden y (c) traducciones dinámicas. Con respecto al *calco aproximado* o *imperfecto*, Gómez Capuz (1998: 67-68) considera que este puede producirse por *calco asimétrico*, es decir, porque se traduce literalmente solo una parte del compuesto; por *calco contraído*, cuando un compuesto puede reproducirse mediante una palabra simple, o por *calco expandido*, cuando una palabra simple puede expandirse en un compuesto¹⁵⁶. Como variante más libre del *calco estructural*, hallamos las *creaciones nativas inducidas* (Haugen 1969: 403), en las que se acuña un nuevo término que se corresponde con una palabra de otra lengua, pero la lexía creada no tiene ningún paralelismo formal con la palabra extranjera.

Otro de los fenómenos resultantes del contacto lingüístico, constatado en numerosas ocasiones por los investigadores, es la denominada *convergencia*. Esta se define, comúnmente, como “el proceso que conduce a la adquisición por parte de los

¹⁵⁶ El autor sigue, en parte, a Lewandowski (1990, *apud* Gómez Capuz 1998: 67-68) en esta clasificación.

hablantes de una lengua de estructuras que resultan gramaticales en las dos lenguas” (Almeida 1999: 204) o “el logro de similitud estructural en un aspecto dado de la gramática de dos o más lenguas” (Silva-Corvalán 2001: 274). Es decir, la convergencia consiste en igualar estructuralmente dos lenguas que están en contacto, y puede afectar solo a una de las dos lenguas, o a las dos o más lenguas que se hallen implicadas en el contacto (Boix y Vila 1998: 257).

El fenómeno de la *convergencia* se circunscribe al nivel morfosintáctico del lenguaje, aunque también puede afectar a otros niveles como el léxico o el sistema fonológico (Silva-Corvalán 2001: 190). Tratándose de la convergencia sintáctica, López Morales (1989: 167) señala que, el caso más frecuente y documentado, es el de la *convergencia indirecta*, donde se sustituyen las estructuras de una de las lenguas por otras que resultan comunes a ambas lenguas. La *convergencia* puede llegar a ser muy profunda, hasta el punto de que dos variedades de familias distintas acaben diferenciándose únicamente por el vocabulario (Boix y Vila 1998: 258).

En la *convergencia* subyace, por un lado, la tendencia del hablante a sacar partido de las semejanzas entre las dos lenguas que domina, y, por otro, el propio desarrollo del sistema lingüístico hacia una dirección u otra. Por esta razón, en algunas ocasiones, puede ser difícil dilucidar si un cambio en la estructura de una lengua ha sido causado por la influencia de otro sistema lingüístico, o si, por el contrario, responde a una tendencia interna de la propia lengua. Puede ocurrir también que el contacto interlingüístico simplemente haya acelerado un cambio interno de un idioma (Payrató 1984: 56, n. 13, Blas Arroyo 1993a: 30, Silva-Corvalán 2001: 274). Por otro lado, la *convergencia* se diferencia del *préstamo* en que no adapta unos rasgos de otra lengua sino que produce una generalización de esquemas que ya existen en la lengua receptora (Moreno Fernández 1998: 264)¹⁵⁷.

En último lugar, abordamos el fenómeno del *cambio de código* —traducción del inglés *code-switching*— que también se ha denominado en la sociolingüística hispánica

¹⁵⁷ Hernández García (1998a: 48-62) cita como ejemplos de convergencia lingüística en el español de Cataluña, entre otros, los siguientes: el queísmo, el uso del verbo *haber* más artículo determinado, la ausencia de preposición ante el objeto directo, o la confusión entre los usos de *deber* y *deber de*.

conmutación, alternancia o intercambio de códigos (Blas Arroyo 1993b: 225-226)¹⁵⁸. Se alude con este nombre al “uso alternativo, dentro de una misma intervención o en los cambios de turno, de oraciones o fragmentos de oraciones en lenguas (o dialectos) diferentes siempre que el resultado sea gramatical según L1 y L2” (Gómez Molina 1999-2000: 324)¹⁵⁹. Este último matiz resulta importante porque distingue este concepto de la simple amalgama o *mezcla de códigos* —*code-mixing*—, que se da cuando un hablante incorpora de forma sistemática elementos de otra lengua en su discurso, sin respetar las restricciones gramaticales de ambas lenguas¹⁶⁰.

Aunque no siempre se ha pensado así¹⁶¹, hoy en día se acepta, generalmente, que solo los bilingües con plena competencia pueden llevar a cabo eficazmente un *cambio de código*, puesto que este fenómeno exige el conocimiento profundo de las dos lenguas, toda vez que los fragmentos lingüísticos que alternan deben cumplir ciertas restricciones en las dos gramáticas implicadas y el intercambio solo es posible en los puntos en los que las estructuras superficiales de las lenguas coinciden. Poplack (1980) analiza estas limitaciones gramaticales con la finalidad de poder predecir en qué punto del discurso va a producirse un cambio de código, y establece que son de dos tipos: (1) de morfema dependiente y (2) de equivalencia estructural. La primera restricción determina que no es posible alternar unidades lingüísticas dependientes, y la segunda, que la alternancia solo tiene lugar en aquellos puntos del discurso en los que hay una situación de equivalencia, de manera que no se quebranten las reglas gramaticales de ninguna de las dos lenguas¹⁶².

La misma autora (Poplack 1981, 1987) clasifica los cambios de código según el nivel lingüístico en el que se produce la alternancia y establece que hay tres tipos: (a) *los cambios tipo “etiqueta”*, interjecciones o elementos expresivos que pueden aparecer

¹⁵⁸ En el ámbito catalán se ha traducido como *alternança de llengües* o *canvi de codi* (Gómez Molina 1999-2000: 324).

¹⁵⁹ La definición más clásica del fenómeno es la de Gumperz (1982: 59): “conversational code switching can be defined as the juxtaposition within the same speech exchange of passages of speech belonging to two different grammatical systems or subsystems”.

¹⁶⁰ Sin embargo, este concepto resulta confuso, ya que para algunos autores es sinónimo de *cambio de código* y, para otros, denomina un *cambio de código* específico, el intraoracional (Gimeno y Gimeno 2003: 124). Payrató (1996: 162), en cambio, considera que este término es útil porque evita tener que especificar cuál es la lengua base desde la cual el hablante produce el cambio.

¹⁶¹ Weinreich (1996: 170) afirma que el bilingüe ideal no cambia de lengua si no varía la situación de habla ni tampoco a lo largo de una misma oración.

¹⁶² No obstante, algunos autores han propuesto numerosos contraejemplos a las restricciones de Poplack (vid. Gimeno y Gimeno 2003: 125, Almeida 1999: 217).

en cualquier lugar, (b) *los cambios oracionales* o *interoracionales*, cuando se alternan oraciones completas en una y otra lengua, y (c) *los cambios intraoracionales*, cuando el cambio se produce en el seno de una misma oración gramatical. En este último caso, es imprescindible un dominio eficaz de las estructuras de ambas lenguas para no realizar cambios de código agramaticales, esto es, en puntos del discurso no permitidos¹⁶³. Del mismo modo, la autora clasifica los cambios según el modo en que se producen y distingue entre (a) *cambios fluidos* —cambios intraoracionales que respetan las fronteras sintácticas de ambas lenguas—, (b) *cambios señalizados* —marcados por pausas, dudas o repeticiones— y (c) *inserción de constituyentes* —elementos de una lengua que se insertan en otra, cuya estructura interna está determinada por la gramática de la lengua fuente, pero su lugar en la frase se rige por la gramática de la lengua receptora— (Poplack 1993: 255-256).

Como acabamos de ver, Poplack se centra en el análisis lingüístico y distribucional del *cambio de código*, pero no faltan investigadores, como Romaine (1989) o Baetens (1989), que abordan el fenómeno desde otra perspectiva y analizan las funciones pragmáticas y discursivas del mismo¹⁶⁴. Desde esta óptica hallamos los estudios de Gumperz (1982)¹⁶⁵ quien, tras definir el *cambio de código* conversacional, establece seis funciones principales de dicho cambio en el discurso: las citas, la especificación del destinatario, la interjección, la reiteración, la cualificación del mensaje y la personalización frente a la objetivización¹⁶⁶. Appel y Muysken (1996: 177-180), por su lado, explican las posibles causas de la alternancia de lenguas relacionando este fenómeno con las funciones del lenguaje, esto es, *referencial, directiva, expresiva, fática, metalingüística y poética*¹⁶⁷.

¹⁶³ Sin embargo, cuando el contacto se da entre lenguas con gran afinidad genética, como el castellano y el catalán, no es necesario este conocimiento profundo de la gramática de las dos lenguas (Gómez Molina 2000: 304).

¹⁶⁴ Para las fases en el estudio del cambio de código puede consultarse Appel y Muysken (1996: 183-192) y Gimeno y Gimeno (2003: 111-112).

¹⁶⁵ En un trabajo anterior (Blom y Gumperz 1972, *apud* Gumperz 1982: 60 y ss.) los autores distinguen entre cambios situacionales, determinados por factores como el tema o los interlocutores, y cambios metafóricos, según lo que pretenda conseguir el hablante.

¹⁶⁶ Véase en Gómez Molina (1999-2000) la aplicación de esta clasificación y de la anterior de Poplack a un corpus basado en ejemplos de alternancia entre catalán y castellano en Valencia. También Boix (1993) o Turell (1995) estudian el intercambio de códigos entre estas dos lenguas; en el primer caso, en el área metropolitana de Barcelona, y, en el segundo, en un barrio de la periferia de Tarragona.

¹⁶⁷ Otros autores argumentan que es imposible determinar todas las funciones que puede llegar a realizar el *cambio de código*, ya que como estrategia discursiva, depende de la creatividad del hablante (Boix y Vila 1998: 241).

Silva-Corvalán (1995: 9-10) considera que el *cambio de código* es una de las estrategias de las que se sirve el hablante bilingüe para reducir la carga cognitiva que le supone recordar y usar dos lenguas. Paralelamente a estos conceptos, acuña otro, el de *sustitución de código*, para hablar de aquellas situaciones en las que no hay un intercambio fluido de distintas lenguas sino que el hablante bilingüe, con una lengua claramente dominante, cambia con frecuencia a esta por conocimiento insuficiente de la segunda lengua. Este fenómeno sucede por diversas razones como, por ejemplo, para reparar un olvido pasajero o compensar el desconocimiento de material lingüístico en la lengua en que se habla. También puede darse con la finalidad de clarificar un mensaje o evaluarlo (Silva-Corvalán 1989: 182-185).

Otro de los puntos a los que se ha prestado atención, es la necesaria distinción de los conceptos de *cambio de código* y *préstamo*. Una diferencia fundamental entre estos dos fenómenos es que, mientras el *préstamo* puede darse en el discurso de los hablantes monolingües, el *cambio de código* —como el *calco*— requiere siempre algún grado de competencia bilingüe (Gimeno y Gimeno 2003: 109), como ya hemos mencionado. Por otro lado, el *préstamo* supone algún tipo de integración, ya sea fonológica, morfológica o sintáctica, en la lengua receptora; en cambio, el *cambio de código* carece de integración lingüística de cualquier tipo en la lengua receptora (Grosjean 1982: 308-309). Boix y Vila (1998: 251), a partir de un ejemplo entre el castellano y el catalán, concluyen que puede establecerse una oposición clara entre los *préstamos*, unidades claramente establecidas, bien asimiladas a las pautas de la lengua receptora y difundidas en toda la comunidad lingüística, y las *alternancias*, unidades actualizadas sin adaptación a la lengua base y producidas de manera ocasional por un hablante concreto en un contexto determinado y por unos motivos definidos¹⁶⁸. Por otro lado, Gimeno y Gimeno (2003: 119) advierten que las lenguas de tipología semejante favorecen los cambios de código basados en el principio de equivalencia, mientras que las de tipología diferente tienden hacia los préstamos ocasionales o inserciones de constituyentes. A pesar de la claridad de los planteamientos anteriores, no siempre resulta fácil diferenciar los dos conceptos, sobre todo cuando se trata de una única

¹⁶⁸ Según Gómez Molina (1999-2000: 316) esta problemática surge porque no siempre coinciden los factores lingüísticos y extralingüísticos entre las comunidades bilingües que se analizan.

unidad léxica que se transfiere a otra lengua (Poplack 1987: 55-56, Gómez Molina 1999-2000: 325, Gimeno y Gimeno 2003: 111 y 118)¹⁶⁹.

Finalizamos, aquí, la revisión epistemológica de las principales consecuencias del contacto lingüístico. En el capítulo 7, dedicado al estudio de los elementos que se transfieren de una lengua a otra en nuestros corpus, veremos cómo es posible relacionar estos mismos fenómenos con los que nos encontramos en la disponibilidad léxica. Una vez analizados los pilares teóricos fundamentales de las distintas disciplinas implicadas en la presente monografía, pasamos ahora a abordar el estado de la cuestión, donde resumimos el conjunto de investigaciones realizadas hasta la fecha desde la perspectiva de la disponibilidad léxica y referidas al contacto entre dos o más lenguas.

¹⁶⁹ Por otro lado, tampoco faltan autores que consideran que es innecesario establecer una diferencia radical entre estos dos conceptos (Boix y Vila 1998: 252). En Almeida (1999: 214) puede verse un resumen de las principales diferencias argüidas para delimitar los dos fenómenos. Hay más sobre la misma cuestión en Poplack (1993: 279-282) o en Romaine (1995: 142 y ss).

3. DISPONIBILIDAD LÉXICA Y LENGUAS EN CONTACTO: ESTADO DE LA CUESTIÓN

En el estado actual de la disciplina sería ilógico realizar un *status quaestionis* que abarcara todas las aportaciones realizadas hasta la fecha con respecto a la disponibilidad léxica. Para empezar, esta extensa tarea ya ha sido brillantemente realizada en diversos capítulos de libros y artículos que han ido recogiendo los avances de los estudios de léxico disponible y a los que nos remitimos en este capítulo¹⁷⁰. Tal y como otros autores ya han señalado, parece innecesario volver a repetir la retahíla de datos ya incluidos en los trabajos anteriormente citados¹⁷¹. Sin embargo, no existe, que sepamos, ningún trabajo que se dedique a revisar concretamente todas las obras que, desde los inicios de la disciplina, se han realizado en territorios bilingües o con hablantes bilingües, vacío que intentamos subsanar con el presente capítulo.

A pesar del gran desarrollo que, como acabamos de ver (*vid.* 2.1 y 2.2.1), ha tenido la disponibilidad en tantos lugares y en tantas distintas ramas del saber; es aún relativamente escasa su aportación al estudio de las comunidades bilingües, aunque esta situación parece estar cambiando en la actualidad. En el presente apartado, como ya hemos indicado, pretendemos describir todas las obras que han estudiado el léxico disponible en lugares donde coexisten dos sistemas lingüísticos de forma habitual. En primer lugar, analizamos —de forma cronológica— los estudios realizados sobre otras lenguas distintas al español y, por ende, llevados a cabo en territorios no hispanohablantes; en segundo lugar, detallamos los estudios referidos ya a la lengua española —en relación con otras lenguas dentro y fuera de la Península Ibérica—. Para finalizar, abordamos las investigaciones realizadas sobre la lengua catalana, en contacto con el castellano¹⁷².

¹⁷⁰ Nos referimos a Carcedo (1998a), López Morales (1999a) y a los exhaustivos resúmenes de Samper, Bellón y Samper Hernández (2003) y Samper Padilla y Samper Hernández (2006).

¹⁷¹ Martínez Olmos (2008: 16) opina al respecto: “es tarea penosa para el estudioso de la disponibilidad encontrarse, investigación tras investigación, con un *status quaestionis* que repite, casi puntualmente, datos ya glosados anteriormente”.

¹⁷² Por tanto, no nos referimos al estudio de la disponibilidad desde la perspectiva de alguien que domina dos lenguas como consecuencia de su estudio en una edad adulta —enfoque que ha dado aportaciones muy interesantes, *vid.* 2.2.1— ni tampoco a la influencia más general de una lengua sobre otra, como pudiera resultar el estudio de los anglicismos en el léxico de los estudiantes de Castellón, por poner solo un ejemplo (Casanova 2008). De otra parte, la metodología de la disponibilidad léxica, con distintos nombres, ha sido frecuentemente empleada en estudios de todo tipo (*vid.* Hernández Muñoz 2006: 85-108) —de hecho, es una técnica procedente de la Psicología empírica—; tampoco es nuestro objetivo

3.1. ESTUDIOS EN TERRITORIOS NO HISPANOHABLANTES

El primer estudio al que nos vamos a referir es una obra prácticamente olvidada en la mayoría de los trabajos recopilatorios sobre la bibliografía de la disponibilidad. Se trata de la investigación de O’Huallachain (1966, *apud* Knock 1978: 17-20) sobre el léxico disponible en inglés y en irlandés. Este autor pretendió establecer qué elementos lingüísticos debían tenerse en cuenta en la enseñanza del irlandés, sobre todo en los primeros niveles. Para ello, dividió su obra en tres partes: el estudio de las estructuras sintácticas, el del vocabulario y la morfología, y el de las palabras disponibles (denominadas por él mismo *utility words*). Para las dos primeras partes, el autor se basó en la grabación de conversaciones en una situación comunicativa habitual en las diferentes zonas de *Gaeltacht*¹⁷³.

Con la finalidad de establecer el léxico disponible, el autor manejó dos muestras de 260 informantes irlandeses e ingleses de Irlanda, respectivamente. Estos eran desde estudiantes —ya fuera de nivel primario, secundario o de escuelas comerciales— hasta informantes de edad adulta. Fueron elegidos en función de los siguientes factores: *edad, sexo, residencia* (urbana o rural), *ocupación y distribución geográfica*. El autor solicitó veinte palabras por cada centro de interés (1. *The body and its parts*, 2. *Physical exercises, athletics, sports*, 3. *Recreation hobbies, indoor pastimes*, 4. *Illnesses (Kinds of sicknesses), injuries, cures*, 5. *Religion, the church*, 6. *Clothes of all Kinds (men, women, boys, girls)*, 7. *Food and drink*, 8. *Meals, utensils for each meal and for preparing the meals*, 9. *The home or house: furniture, fittings, heat, light*, 10. *Business: buying, selling*, 11. *Life in town and city*, 12. *Life and work in the country and the landscape*, 13. *Animals, birds, fish, insects*, 14. *Trees, plants, flowers, weeds, crops*, 15. *Trades and Kinds of occupation or work*, 16. *Travel, means of transport and things connected with them*, 17. *The family, relatives and people’s feelings about themselves and others*, 18. *School life, activities, furniture*, 19. *The ruling of the country* y 20. *The weather, the time of day, seasons*)¹⁷⁴. Tras realizar el análisis informatizado de los datos,

reseñar aquí estos trabajos, sino que únicamente nos ceñimos a las obras cuyo objetivo principal es establecer la nómina de vocablos disponibles y confeccionar, así, un diccionario.

¹⁷³Se denomina así a ciertos núcleos de la República de Irlanda, como los condados de *Donegal, Mayo, Galway* o *Kerry*, donde el gaélico irlandés es, o era tradicionalmente, la lengua de uso mayoritaria de la población.

¹⁷⁴ Para elegir los centros de interés, el autor se inspira en los trabajos de Gougenheim *et al.* (1964) y Pfeffer (1964). En la estratificación de la muestra, O’Huallachain también tiene en cuenta la religión y utiliza informantes católicos, protestantes y judíos. Tanto los informantes que hablan irlandés como los ingleses son unilingües (O’Huallachain 1966, *apud* Knock 1978: 19 y 23).

O’Huallachain estableció las listas de las palabras disponibles en irlandés y en inglés, ordenadas alfabéticamente y según su frecuencia y distribución en los centros¹⁷⁵.

Como ya comentamos (*vid.* 2.1.1.), el trabajo de Mackey (1971a) sobre el léxico disponible francés de Canadá es una de las obras pioneras. En realidad, como veremos a continuación, muchas de las posibilidades que ofrece la disponibilidad en los territorios bilingües están ya implícitas en los planteamientos de este autor. Para comprobarlo, analizamos, a continuación, el segundo volumen de su obra donde se presentan los resultados de las encuestas realizadas en Acadia, una región marítima en la que el 40% de población era de procedencia francesa. Uno de los objetivos del estudio de Mackey era analizar el efecto del bilingüismo sobre el vocabulario disponible (1971a, vol. II: 6) y por ello insistió en la idoneidad de la disponibilidad como instrumento adecuado para detectar los préstamos, dado que aborda un sector del vocabulario donde estos suelen manifestarse¹⁷⁶:

“Puisque la disponibilité est une manifestation de la langue et non seulement de la parole, c’est l’instrument tout à fait indiqué pour déceler les emprunts, non seulement parce que le sujet bilingüe fait l’inventaire de ses connaissances d’une langue à la fois, mais aussi parce que la disponibilité touche surtout le secteur du vocabulaire où les emprunts se manifestent d’abord, à savoir, le vocabulaire concret” (1971a, vol. II: 16).

El estudioso comprobó la esperable presencia en las listas tanto de palabras inglesas, incluidas bajo múltiples grafías, como de calcos, “anglicismos” para Mackey —como por ejemplo, *station à feu* (que proviene del inglés *fire station*)—, aunque estos últimos aparecieron en un número mucho menor. En conjunto, se aportaron 1397 palabras en inglés, el 13.3% del total. En la distribución por áreas léxicas se observó que los centros que más palabras inglesas incorporaron fueron 15. *Les jeux et distractions*, 11. *Le village* y 7. *La cuisine et les utensiles que s’y trouvent*; por el contrario, la presencia de estos términos descendió radicalmente en 1. *Les parties du corps*, 14. *Les animaux* y 9. *Le chauffage et l’éclairage*¹⁷⁷.

¹⁷⁵ Citamos toda la información relativa a esta obra a partir del resumen hallado sobre ella en el trabajo de Knock (1978: 17-20). Sería deseable conocer más detalles de la misma, pero su localización nos ha resultado imposible.

¹⁷⁶ El autor entiende por *préstamo* el elemento lingüístico foráneo usado por un bilingüe debido a que desconoce el equivalente en la lengua que está utilizando en ese momento (Mackey 1971a, vol. II: 15).

¹⁷⁷ El contraste mayor se dio entre *Les parties du corps* y *Les jeux et distractions (sauf les sports)*. En el primero los informantes aportaron 11 palabras en inglés y, en cambio, en el segundo, 189.

A continuación, Mackey pretendió comprobar hasta qué punto las palabras inglesas habían penetrado en el francés de los canadienses y por ello comparó los índices de disponibilidad de las palabras francesas y sus correspondientes sinónimos ingleses hallados en las encuestas, método que también se utiliza en los estudios actuales de disponibilidad en territorios bilingües (*vid.* 3.2 y 3.3) y que también aplicamos en este trabajo (*vid.* 7.4.). El autor observó que había muy pocas palabras inglesas que superaran en proporción las francesas (algunas de ellas eran *flashlight*, *map*, *office...*) y que la mayor parte de ellas contaba con su correspondiente francés en las encuestas, que era el que —mayoritariamente— citaban los informantes. Finalmente, llegó a la conclusión de que el porcentaje de sustitución (de un término francés por otro inglés) variaba según el individuo y el centro de interés, que la mayor parte de sustituciones léxicas eran producidas por pocos informantes y que, en ningún caso, se sustituía completamente el vocabulario de base francés de un centro de interés por el inglés (1971a: 18-27).

Mackey también manifestó un gran interés por conocer cuál era la muestra numéricamente idónea para determinar el léxico disponible de una comunidad. Por esta razón, dedicó una obra (Mackey 1971b) a realizar una serie de operaciones matemáticas e informatizadas para decidir cuántos informantes y cuántas palabras por centro eran necesarias para calcular, de forma apropiada, eficaz y fiable, las palabras más disponibles¹⁷⁸. El autor se planteó esta cuestión ante la evidencia de que el número de informantes y datos utilizados en sus estudios anteriores debía disminuir para poder extender la encuesta a otros lugares francófonos y concluyó, tras una serie de probaturas, que una muestra de 200 informantes que aporten 20 palabras por centro resulta la más adecuada.

En otra obra posterior (1976), el autor volvió al tema de la disponibilidad como medida no solo de la distancia cultural entre dos regiones francófonas —que era su objetivo inicial¹⁷⁹—, sino también como instrumento para medir el grado de integración

¹⁷⁸ Concretamente, el autor usó isogramas de correlación y una simulación de muestreo aleatorio que consistía en ir reduciendo progresivamente la extensión de la muestra y comprobando si este hecho hacía variar los resultados de forma significativa (1971b: 1-3).

¹⁷⁹ Con respecto a este objetivo, el investigador también comparó los porcentajes de préstamos hallados en Acadia con los hallados en Francia por Gougenheim *et al.* (1964) para acabar afirmando, como era de esperar, que había muchos más préstamos en el francés de los canadienses, y que estos eran prácticamente inexistentes en el francés de Francia (1976: 357).

del vocabulario extranjero en la lengua que se estudia. El investigador consideró que el test de léxico disponible, junto con otras pruebas como el test de aceptabilidad y el de traducibilidad, podían determinar de forma clara la integración real en el código de los préstamos. Con este propósito, Mackey creó una fórmula matemática para calcular con exactitud la probabilidad de integración de una palabra inglesa en el francés de los canadienses¹⁸⁰. No obstante, el autor concluyó, tras los análisis pertinentes, que la probabilidad de integración de las palabras más importantes (más disponibles) no era tan elevada como para hacer desaparecer los equivalentes franceses (1976: 334). Es decir, reiteró las conclusiones alcanzadas ya en 1971.

A pesar del interés que tienen, de por sí, los estudios de Mackey sobre los préstamos, el propio autor intuyó otros posibles análisis que, sin embargo, no llevó a cabo. Para empezar, en la encuesta aplicada al alumnado (Mackey 1971a), aparte de datos relativos al sexo, la edad, la profesión del padre o el lugar donde se vive, también se recogieron otros como los usos lingüísticos. En concreto, se solicitó expresamente la lengua que utilizaban el padre, la madre y el alumno o alumna en casa; la lengua en la que los progenitores leían periódicos, revistas o libros y la lengua en la que el alumno o alumna veía la televisión, escuchaba la radio y jugaba (Mackey 1971a, vol. I: 487-488). Hubiera sido muy interesante, tal como se hace en la actualidad, estudiar el peso de tales variables en los resultados para comprobar si realmente, como sería esperable, había más préstamos en las respuestas dadas por los informantes que tenían el inglés como lengua materna y/o habitual que en las dadas por los francófonos.

Por otro lado, el profesor canadiense previó la posibilidad de que algunos informantes bilingües completaran cinco centros de interés en inglés y en francés. En su obra de 1971 se da a entender que estos cuestionarios se habían completado, pero el mismo autor reconoce que no se habían podido incorporar aún los resultados (Mackey 1971a, vol. I: 37, vol. II: 5). Sin embargo, en su trabajo de 1976 (Mackey 1976: 335-337) obtenemos información adicional sobre este estudio anterior. En primer lugar, el autor ya menciona qué áreas léxicas han sido completadas por los mismos informantes en las dos lenguas (*actions, instruments, transport, religion y sport*), así como las

¹⁸⁰ La fórmula consistía en la división del porcentaje de aparición de la palabra inglesa entre la suma de dicho porcentaje y el de la palabra correspondiente en francés. Una palabra que obtuviera una probabilidad $\leq 1,00$ indicaba que el préstamo había sido integrado casi por completo en el código local (Mackey 1976: 332).

condiciones de reacción (quince minutos por centro de interés y algunos meses transcurridos entre las dos pruebas). Tras examinar los resultados, Mackey ultimó que la proporción de conceptos aportados en las dos lenguas dependía del campo semántico, puesto que; por ejemplo, la religión se asociaba más a la lengua francesa (y presentaba, por tanto, un índice superior en palabras y vocablos en francés y en préstamos al realizar la encuesta en inglés), mientras que los conceptos relativos a los instrumentos, el transporte o los deportes tenían mayor peso en inglés (en todos los índices citados anteriormente). Para realizar este análisis, el autor comparó en cada uno de los cinco centros de interés tanto el total de palabras y vocablos en cada una de las dos lenguas como los préstamos detectados en cada lengua y el porcentaje que suponían los mismos con respecto al total de vocablos. El estudio llegó a la conclusión de que el análisis de la interpenetración de las lenguas podía hacerse calculando el porcentaje de palabras de una lengua que una comunidad bilingüe utiliza en la otra lengua (a partir de las respuestas dadas en las encuestas de disponibilidad, claro está).

Siguiendo el camino iniciado por Mackey y también con el apoyo del CIRB (*Centre International de Recherches sur le Bilinguisme*) de Québec, apareció —algunos años más tarde— la obra de P.E. Knock (1978) centrada en el estudio comparado de la disponibilidad en francés y basaa de hablantes bilingües en una región de Camerún. Los objetivos del estudio de Knock eran varios. Para empezar, dado que los escolares francófonos africanos manejaban, al menos, dos lenguas, el autor pretendió establecer, mediante las pruebas de disponibilidad léxica, cuál era el universo familiar del niño camerunés, estableciendo para ello la lista de palabras que este conocía tanto en su lengua materna (basaa) como en la lengua vehicular de la escuela (francés). Al mismo tiempo que el autor comparó la riqueza léxica de estos dos repertorios, también cotejó estos datos con los proporcionados por Gougeinheim *et al.* (1964) sobre el francés de Francia. Cabe decir que, para llevar a cabo su investigación, realizó un estado de la cuestión sobre los estudios de disponibilidad léxica publicados hasta ese momento y sobre otros trabajos dedicados al vocabulario que se habían realizado en el entorno africano¹⁸¹. El autor también dejó constancia de las carencias y problemas que tiene en

¹⁸¹ Es muy interesante el trabajo que cita de M. Fall (1976) sobre el wolof y el francés en Senegal, que, al parecer, sigue de cerca las pautas metodológicas de Knock. La encuesta abarcaba diez centros de interés (*Les parties du corps, La nourriture et les boissons, Les vêtements, Les jeux et les distractions, Les animaux, Les métiers, La maison et son mobilier, Les moyens de transport, La ville y La cuisine*) y fue aplicada a alumnado de nivel primario de escuelas de Dakar y alrededores. Lamentablemente, Knock no

sí la enseñanza en África, destacando que, al realizarse en lenguas de países europeos, se olvida que estas son incapaces de describir de forma completa el medio local africano, como resulta obvio.

En cuanto a la metodología empleada, Knock manejó una muestra de 220 alumnos bilingües y escolarizados en un nivel primario¹⁸². La muestra fue elegida de forma estratificada para que fuera demográficamente representativa¹⁸³. Los mismos alumnos completaron la encuesta dos veces, primero en basaa y luego en francés, con un intervalo de dos semanas aproximadamente entre ambas encuestas. En cada sesión no se completaron más de tres centros y los informantes contaron con quince minutos para cada una de las áreas léxicas. Los centros de interés utilizados fueron diez, prácticamente los mismos que había en los trabajos pioneros franceses, con alguna excepción que pretendía reflejar la idiosincrasia del pueblo africano: *Les parties du corp*, *La nourriture et les boissons*, *Les vêtements*, *La maison: matériaux de construction et meubles*, *Les travaux de champs*, *La cuisine*, *La salle de classe*, *Les moyens de locomotion*, *La musique et la danse-genre* y *Les jeux et les divertissements*. En la encuesta se solicitó expresamente el nombre, la edad y el sexo del informante; y también se incluyó una serie de preguntas destinadas a conocer cuál era su entorno lingüístico (dónde había vivido los últimos tres años, las lenguas que hablaba y las aprendidas antes de entrar en la escuela)¹⁸⁴.

Knock analizó sus datos mediante un proceso manual de fichas y otro informatizado. De este modo, llegó a establecer los dos listados de palabras disponibles en las dos lenguas y también realizó una serie de cálculos destinados a conocer el porcentaje de población que había citado un mismo concepto en cada lengua. De

da más detalles sobre este estudio que constituye la tesis de maestría de la autora, puesto que se acabó mientras se imprimía el suyo (Knock 1978: 26-27).

¹⁸² Knock (1978: 78-79) consideró de forma independiente el nivel educativo y la edad, puesto que los estudiantes del primer nivel de primaria tenían desde 2 hasta 9 años, e incluso había clases donde el alumnado podía llegar a tener 17 años. Esto sucedía tanto por el alto abandono escolar como por las frecuentes repeticiones de nivel.

¹⁸³ Para ello el autor tuvo en cuenta el porcentaje de población escolar en los dos departamentos en los que aplicó la encuesta y la distribución del alumnado según el sexo y según el tipo de enseñanza (oficial, católica, protestante o laica) (Knock 1978: 89-90).

¹⁸⁴ El relato de la cronología del trabajo de campo de Knock es realmente una odisea por las condiciones tan duras a las que se enfrentó el investigador: mal estado de las vías de acceso, problemas con el carburante, la lluvia que imposibilitaba las carreteras...; a ello se suma que, a causa de ciertas creencias locales, muchas veces tenía que ser primero el pastor o cura del poblado quien convenciera a los padres para que sus hijos completaran la encuesta (1978: 67 y 282-284).

manera exhaustiva, el estudioso agrupó los vocablos de cada centro de interés en distintas categorías conceptuales¹⁸⁵ y cotejó, tras esta clasificación, la disponibilidad de los conceptos en basaa y en francés, por un lado, y estos resultados con los del francés recabados en Francia con la finalidad de constatar en qué lengua se podía describir mejor el entorno familiar africano¹⁸⁶.

El autor concluyó, como era esperable, que el francés de Francia no podía describir la realidad sociocultural africana; de otra parte, el basaa —y las lenguas africanas, en general— eran incapaces de describir el África moderna, porque necesitarían crear nuevas palabras para ello. Por último, el investigador llegó a la conclusión de que el bilingüismo africano-europeo era un factor de enriquecimiento, puesto que un bilingüe contaba con más palabras y, por ende, con más recursos para describir su realidad. A pesar de esta visión positiva, el autor reconoció que este bilingüismo se sostiene en detrimento de las lenguas africanas nacionales, que van perdiendo sus términos científicos al realizarse toda la escolarización en lengua francesa. En cuanto a la metodología, el autor destacó —con mucha intuición— la importancia del rango asociativo de las palabras, es decir, que en el cálculo de la disponibilidad debía tenerse en cuenta, no solo las menciones, sino también la posición que ocupaban las palabras en los listados, tal y como se hace en la actualidad¹⁸⁷.

Años más tarde y en otro lugar muy alejado —la antigua Yugoslavia, antes de la guerra de los Balcanes—, otro investigador (Dimitrijevic 1981) emprendió un análisis comparado entre estudiantes monolingües en serbocroata y estudiantes bilingües en húngaro-serbocroata. Los resultados de este estudio se exponen en un resumidísimo artículo, donde se comparan estos datos con otros recabados en Edimburgo en un

¹⁸⁵ Por ejemplo, en el primer centro estableció: “el cuerpo”, “la cabeza”, “el cuello”, “el tronco”, “los miembros”, etc. y, dentro de cada categoría, determinó otras; por ejemplo, dentro de “cabeza”, “cabeza, cabellera o cara” (Knock 1978: 152-153).

¹⁸⁶ Uno de los problemas con los que se encontró Knock fue el gran número de palabras desconocidas en lengua basaa, y que, como los informantes solo las conocían de forma oral, resultaba difícil decidir a qué palabra equivalía lo que había escrito el estudiante. Se ha de tener en cuenta que el basaa tiene una ortografía normalizada solo desde 1970. Por otro lado, como era de esperar, el autor constató préstamos en ambas listas, es decir, palabras francesas en las listas en basaa y a la inversa, aunque no analizó ninguna.

¹⁸⁷ Es sorprendente ver cómo algunos aspectos que comenta el autor son muy actuales; la decisión de mantener los datos en bruto por su utilidad en el estudio de la ortografía, el hecho de deducir los significados de los términos polisémicos según las palabras que los rodean en los listados o, incluso, la vigencia de las conclusiones cuantitativas, a saber, que hay centros que estimulan más la producción que otros pero que este hecho no es correlativo con la cantidad de palabras distintas —es decir, de *vocablos*— (Knock 1978: 113,130 y 135).

trabajo anterior del propio Dimitrijevic (1969) sobre disponibilidad léxica a partir de una muestra de 185 alumnos de 14 años. Cabe señalar que una de las pautas metodológicas de este trabajo fue fundamental para los estudios posteriores: la decisión de no solicitar un número determinado de palabras sino que las listas fueran abiertas con la única limitación del tiempo, en su caso cinco minutos (Dimitrijevic 1969: 14-15).

Los centros de interés que se utilizaron en esta primera investigación fueron once, concretamente: *Animals, Countryside, Town, Entertainment, Jobs and Professions, Science, Means of transports, Politics, Parts of the house, Food and drink* y *Clothes* (Dimitrijevic 1969: 15). El enfoque dado al análisis, más similar a los estudios actuales, reveló diferencias cuantitativas y cualitativas según el *sexo* o el *tipo de escuela* de los informantes. Los resultados mostraron una mayor producción léxica por parte de los varones, aunque las diferencias fueron aún más importantes en relación con el *tipo de escuela*. Los estudiantes de las *Senior Secondary Schools* obtuvieron mejores resultados en prácticamente todos los centros de interés. Es importante señalar que la inteligencia del alumnado era un factor determinante para que los alumnos acudieran a estas escuelas o a las *Junior Secondary School*.

Como decíamos, en el estudio realizado con posterioridad en Novi Sad, Dimitrijevic (1981) comparó los datos obtenidos de los estudiantes escoceses con los recogidos en esta ciudad serbia. Para ello manejó una muestra de 228 alumnos, divididos según la franja de *edad* (11 y 14 años)¹⁸⁸ y según su *lengua materna* (monolingües en serbocroata y bilingües en húngaro-serbocroata). Además de estas, el autor tomó en consideración otras variables como el *sexo*, la *inteligencia* y las *calificaciones*. Las condiciones de la encuesta fueron las mismas que el autor había aplicado en Edimburgo, a saber, cinco minutos por centro de interés. A los centros usados en la prueba escocesa, se añadieron únicamente dos más: *Parts of the body* y *War and Peace*¹⁸⁹.

¹⁸⁸ El autor elige estas edades, ya que a los 11 años empieza la enseñanza más formal de la lengua materna y a los 14 es la edad en la que terminaba la enseñanza elemental en la Yugoslavia de entonces (Dimitrijevic 1981: 113).

¹⁸⁹ Como el mismo autor reconoce, la selección de las áreas léxicas está parcialmente influida por Gougenheim *et al.* (1964). Algún centro de interés, como *Politics*, se incluyó en las dos investigaciones con la intención de comparar dos países con sistemas políticos y sociales muy distintos (Dimitrijevic 1969: 18, 1981: 112).

Como conclusiones relevantes y significativas, aparecieron que el alumnado de más edad (14 años), tanto monolingüe como bilingüe, tenía un nivel mayor de producción léxica en comparación con la otra franja de edad (11 años), tal como era esperable. Por otro lado, la muestra de alumnos de más edad y monolingües manejaba más palabras que los bilingües y también más que los estudiantes escoceses de Edimburgo¹⁹⁰. En cuanto al *sexo*, las informantes tenían más palabras que los informantes, aunque los chicos las aventajaban en relación con el número de palabras diferentes. Del mismo modo, la variable *inteligencia*, determinada partir del grado de coeficiente intelectual del alumno, pareció ser decisiva en el estudio cuantitativo de la disponibilidad léxica¹⁹¹. Con respecto a los resultados cualitativos, el autor no constató que el vocabulario de chicas y chicos fuera muy distinto, con alguna excepción en algún centro de interés (*Professions, Food and drinks* y *Clothes*). En relación con la *edad*, el investigador observó menos palabras que tuvieran un alto grado de disponibilidad en las listas proporcionadas por los estudiantes más jóvenes. Por otro lado, y tal como el autor imaginaba, se observaron diferencias en el vocabulario respecto a los sujetos que procedían de dos sistemas culturales y sociales muy distintos —Inglaterra y Yugoslavia—, en particular en las áreas temáticas *Politics, Food and drinks* y *Professions*.

Con respecto a la *adscripción lingüística*, el investigador llegó a la conclusión de que, en la antigua Yugoslavia, los monolingües y los bilingües no diferían demasiado en las palabras que alcanzaban un mayor grado de disponibilidad. Al comparar los resultados —en las dos franjas de edad— de las listas en lengua materna y en la segunda lengua, el autor solo observó algunas diferencias poco relevantes¹⁹². Además, los bilingües proporcionaron un gran número de palabras dialectales al completar la encuesta en su segunda lengua, es decir, en serbocroata¹⁹³. El autor finaliza su artículo

¹⁹⁰ El investigador solo compara los estudiantes mayores con los escoceses, puesto que son los únicos que coinciden en la edad.

¹⁹¹ El autor tuvo en cuenta otras variables, que no mostraron ninguna relevancia en los datos finales obtenidos, como las notas de clase, el estatus del alumno, el lugar de nacimiento que ocupaba dentro de la familia o la profesión y educación de los progenitores (Dimitrijevic 1981: 115). Llama especialmente la atención que este último parámetro no influya, puesto que ha demostrado tener una gran importancia en la mayor parte de las investigaciones hispánicas sobre disponibilidad (Samper, Bellón y Samper Hernández 2003: 75-80).

¹⁹² Se entiende que aquí el estudioso se refiere únicamente a los datos aportados por los bilingües.

¹⁹³ Aunque Dimitrijevic habla de “palabras dialectales” se sobrentiende que son palabras de su propia lengua materna. A la vista de esta afirmación y de los resultados que el autor incluye en el artículo (solo las primeras cinco palabras en cada submuestra) se deduce que los bilingües completaron la encuesta dos

indicando las investigaciones que podrían realizarse en el futuro, como, por ejemplo, aplicar la encuesta en otros países anglófonos, trabajar con informantes bilingües pero de lenguas más cercanas, e incluso, utilizar los resultados de la disponibilidad en estudios neurolingüísticos¹⁹⁴.

3.2. ESTUDIOS SOBRE EL ESPAÑOL

En el presente apartado nos referimos con exclusividad a los estudios llevados a cabo en aquellos lugares en los que el español convive con otra lengua, ya hayan sido realizados desde una perspectiva comparada entre las dos lenguas o recojan simplemente la influencia de la segunda lengua en el léxico disponible en castellano.

3.2.1. El español en contacto con una lengua peninsular

3.2.1.1. En contacto con el gallego y el portugués

El estudio de la disponibilidad léxica en la comunidad autónoma de Galicia se ha acometido en los últimos años (López Meirama 2008). En dicha comunidad el gallego es la lengua habitual, aunque se detecta en la actualidad una mayor preferencia, por parte de la gente joven, a usar el castellano. Además, también se prefiere el gallego en algunas comarcas —las orientales de la provincia de A Coruña, las del sur de Lugo y las de la provincia de Ourense— y, sobre todo, en los núcleos de población más pequeños. Por otro lado, cabe recordar que el castellano era la única lengua de la escuela en Galicia hasta hace relativamente poco, los años 70 (López Meirama 2008: 10-15).

Para la investigación sobre el léxico disponible se utiliza una muestra de 800 alumnos de 2º de bachillerato que se estratifican según las siguientes variables sociales: *sexo*, *tipo de población*¹⁹⁵ (diferenciando entre la del centro educativo y la de la residencia familiar), *tipo de centro*, *nivel sociocultural*¹⁹⁶ y *lengua habitual*¹⁹⁷. En

veces, una en cada lengua, aunque se desconoce si fueron exactamente los mismos alumnos o el tiempo transcurrido entre las dos pruebas (Dimitrijevic 1981: 115).

¹⁹⁴ Es realmente una pena que, debido a la falta de espacio, como el mismo autor reconoce en algún momento, no pueda detenerse más en algunos aspectos que hubiera sido interesante conocer (Dimitrijevic 1981: 112).

¹⁹⁵ Dadas las características de la distribución geográfica de la población se establecen cuatro tipos: la población urbana (de más de 50000 habitantes), la *periurbana* (las conurbaciones que rodean las ciudades), las villas (de más de 15000 habitantes) y las poblaciones rurales (de menos de 15000 habitantes).

¹⁹⁶ Para determinarlo únicamente se tiene en cuenta el nivel educativo de los progenitores.

¹⁹⁷ Se establecen, en principio, cinco categorías según el uso preferente o exclusivo de una lengua u otra, aunque finalmente se reducen a tres en el análisis cuantitativo (gallego/castellano/ambas).

cuanto a los estímulos verbales, estos son los dieciséis habituales de cualquier trabajo en el marco del Proyecto Panhispánico, además de *Los colores* y *El mar*, este último indispensable en un entorno geográfico como el de Galicia.

En la edición de los materiales, la autora da cuenta de los múltiples errores ortográficos que son debidos, como es lógico, a la influencia de la lengua gallega. Asimismo, decide marcar con el símbolo “[g]” los numerosos galleguismos hallados en su corpus, aunque es importante aclarar que bajo esta etiqueta se señalan tanto las palabras escritas en gallego propiamente como los dialectalismos del español que están marcados en los diccionarios como propios de Galicia o de zonas afines. La presencia de estas voces en los listados varía según el centro de interés del que se trate. En algunos centros, como el *El campo* o *Trabajos de campo y del jardín*, —precisamente donde más galleguismos se dan— aparecen incluso variantes dialectales, que se unifican según el diccionario normativo de la *Real Academia Galega*¹⁹⁸.

Al realizar el análisis cuantitativo según las variables sociales¹⁹⁹, se observa que las mujeres y los informantes de los centros educativos urbanos alcanzan índices superiores, a pesar de que estas variables —así como el *lugar de residencia familiar*— no parecen tener una incidencia importante en la productividad léxica. Lo mismo sucede con la variable *lengua habitual*, aunque consigue mejores promedios el grupo castellanohablante. Los resultados más significativos, a pesar de que no lo son estadísticamente, son los que se dan sobre el *tipo de centro* y el *nivel sociocultural*, otorgando unos mejores resultados a los centros privados y al alumnado del nivel social más alto. Como se demuestra al final del capítulo, hay una relación evidente entre estas dos últimas variables, y es el *nivel sociocultural* el factor que realmente incide en los elementos léxicos aportados. La autora también analiza la compatibilidad entre distintos grupos de informantes, es decir el grado de coincidencia en las respuestas²⁰⁰.

El estudio de López Meirama (2008) se complementa, años más tarde, con un volumen conjunto —editado por la misma autora (López Meirama ed. 2011)— donde

¹⁹⁸ También se observan voces gallegas castellanizadas, como “mugidora” > *muxidoira* o lexías complejas en las que se mezclan los dos idiomas (*enrestrar* cebollas).

¹⁹⁹ Para ello la autora utiliza los promedios y la técnica estadística de ANOVA.

²⁰⁰ Al respecto, halla la mayor coincidencia entre hombres y mujeres y las mayores diferencias entre el léxico de los centros públicos y privados.

varios especialistas analizan otros aspectos del léxico disponible del español de Galicia. De este modo, por ejemplo, se compara, cuantitativa y cualitativamente, el léxico de los estudiantes según procedan de centros de enseñanza rurales o urbanos (López Meirama 2011); se analizan los errores ortográficos más frecuentes (Blanco 2011) o se describen los procedimientos mediante los cuales los informantes crean neologismos (García Gondar 2011).

El análisis más interesante, desde nuestra perspectiva, es el que efectúa Álvarez de la Granja (Álvarez de la Granja 2011) sobre los fenómenos de contacto lingüístico recogidos en el corpus del léxico disponible del español de Galicia²⁰¹. La autora basa su estudio en las palabras marcadas como gallegas²⁰², salvo contadas excepciones (Álvarez de la Granja 2011: 21), y las analiza desde una doble perspectiva: formal y funcional. Desde la primera óptica, diferencia entre formas gallegas adaptadas al castellano y formas sin adaptar. Estas últimas, a su vez, pueden estar integradas lingüísticamente o no según si poseen alguna característica formal que pone en evidencia su procedencia del gallego —lo que ocurre en las que no están integradas—, o si carecen de ella —como sucediendo cuando se ha producido la integración—. En general, las palabras adaptadas al español son escasas en el corpus y conviven con la correspondiente forma sin adaptación. Por otro lado, la autora también atestigua la presencia de calcos semánticos²⁰³ y morfológicos, junto con expresiones de carácter híbrido y posibles influencias del gallego en distintos fenómenos lingüísticos.

En segundo lugar, la autora tiene en cuenta la distribución funcional de los fenómenos de contacto lingüístico y clasifica, de nuevo, los datos del corpus en: *préstamos integrados*, *interferencias léxicas*, *lapsus bilingües* y *cambios de código*. En el primer grupo, los *préstamos integrados*, sitúa las unidades léxicas con un cierto número de realizaciones, concretamente un mínimo de diez, hecho que puede indicar la integración social o idiolectal del vocablo. En el análisis cuantitativo de estos casos, se toma en consideración la frecuencia total del concepto, es decir, la suma de las actualizaciones del vocablo en gallego y del término correspondiente en castellano.

²⁰¹ Sería injusto no mencionar aquí el gran interés que cobra, para nuestro ámbito de estudio, el análisis de Bartol (2011) en este mismo volumen del que hablaremos más adelante (*vid.* 3.4)

²⁰² Se trata de las palabras que aparecen etiquetadas con una “[g]” (*vid. supra*).

²⁰³ En estos casos las entradas no vienen señaladas con la etiqueta [g], toda vez que se refieren a significantes castellanos que adquieren nuevos significados por influencia de una forma gallega con la que tienen proximidad semántica y/o formal (Álvarez de la Granja 2011: 33-35).

Además, identifica las posibles causas que determinan la aparición de préstamos en las encuestas, como, por ejemplo, la intención de mencionar elementos relacionados con la idiosincrasia cultural gallega que no existen en el resto de la Península, el hecho de que los términos equivalentes en castellano tengan diferencias expresivas con respecto a las formas gallegas o no sean usuales en Galicia, etc.²⁰⁴ Las *interferencias léxicas* son, para la investigadora, unidades que suponen una aproximación al castellano, desde el gallego, creadas en el momento de aplicación de la encuesta (Álvarez de la Granja 2011: 29). Buena parte de las formas adaptadas y de los calcos, según la clasificación formal, puede formar parte de este grupo.

En último lugar, hallamos los *lapsus bilingües* y los *cambios de código*, que corresponden a palabras o expresiones gallegas cuyo equivalente castellano es conocido sobradamente por los informantes. La diferencia radica en que, mientras los *lapsus bilingües* se insertan involuntariamente en los listados, los *cambios de código* se realizan de forma totalmente consciente con una motivación pragmática. En el análisis cuantitativo, que se realiza conjuntamente de los dos fenómenos²⁰⁵, se comprueba cómo el alumnado que declara hablar solo gallego produce un número mayor de estas formas. Para acabar, la autora comenta muchas de las limitaciones que impone la metodología de la disponibilidad léxica para el estudio de estos fenómenos y declara que los datos muestran la “imposibilidad que tienen los bilingües de mantener sus lenguas completamente separadas” (Álvarez de la Granja 2011: 79).

También tiene cabida en este epígrafe el trabajo de Prado *et al.* (2009) sobre la disponibilidad léxica en situación de contacto de lenguas en las zonas limítrofes de Andalucía y Extremadura (España), por un lado, y, por el otro, Algarve y Alentejo (Portugal). El estudio analiza la productividad léxica de dos muestras integradas por alumnos preuniversitarios españoles y portugueses procedentes de municipios del centro y sur de la frontera entre España y Portugal²⁰⁶. Se analizan los dieciséis centros de

²⁰⁴ La autora constata que hay realidades a las que se accede a través del gallego porque tienen relación directa con el mundo tradicional, rural o marinero. Se trata de nombres de animales marinos, de plantas o de herramientas o tareas del campo (Álvarez de la Granja 2011: 60-64).

²⁰⁵ Esto es así ya que la investigadora reconoce que, mediante un listado de palabras, es muy difícil dilucidar si estamos ante un uso consciente o inconsciente del término en cuestión (Álvarez de la Granja 2011: 65-66).

²⁰⁶ Forman parte de la zona conocida como “La Raya”, es decir, ‘la frontera oficial’, “donde las poblaciones portuguesas y españolas, separadas por una frontera política, han compartido a lo largo de la historia elementos culturales y económicos” (Prado *et al.* 2009: 13).

interés clásicos²⁰⁷, y se comparan los totales, los promedios y los porcentajes con respecto a los dos léxicos disponibles, en español y portugués, respectivamente. Se comprueba así que los porcentajes de vocablos por informante y centro de interés son algo superiores en la muestra portuguesa, aunque el porcentaje total de palabras por centro otorga cierta ventaja a los datos españoles. Además, ambas muestras coinciden con respecto a los centros más o menos cohesionados.

El estudio también analiza la influencia de ciertas variables sociales sobre los dos repertorios léxicos. En relación con el *sexo*, los varones aventajan a las mujeres en los dos países, mientras que, tanto en la *ubicación del centro* como en la variable *residencia de los padres*, el entorno urbano resulta significativo, aunque por una pequeña diferencia con respecto al entorno rural. La variable *nivel sociocultural*²⁰⁸ covaría en la muestra portuguesa, como es de esperar, otorgando mejores promedios a los estudiantes de nivel más alto. Por el contrario, en los datos españoles no se observa este paralelismo, y por ejemplo, el nivel más productivo es el elemental, dato que no deja de ser sorprendente²⁰⁹. El estudio se acompaña, finalmente, de los diccionarios de los dos repertorios léxicos.

3.2.1.2. *En contacto con el euskera*

El primer trabajo que se realiza en la Península Ibérica sobre disponibilidad léxica es la tesis doctoral de M^a José Azurmendi (1983) en la que, siguiendo los pasos de Mackey, la autora pretende realizar una descripción del bilingüismo vasco-español en la comarca de San Sebastián. A lo largo de sus tres volúmenes se desarrollan los cuatro apartados que componen la obra: la introducción, el marco socio-demográfico de la población y los análisis cuantitativo y cualitativo de los resultados.

El colectivo estudiado lo constituye el alumnado que cursa 5^o y 8^o de la antigua E.G.B. (que corresponde a las edades de 9-10 y 13-14 años, respectivamente), a los que se les proporciona dos cuestionarios bilingües: uno sociolingüístico, donde se hallan las

²⁰⁷ Más otro centro, que es *Portugal*, para los estudiantes españoles, y *España*, para los portugueses, que no se analizan en este trabajo y que pretenden detectar las connotaciones que despiertan ambos términos.

²⁰⁸ Se calcula, como viene siendo habitual, en función de la profesión y del nivel de instrucción de los padres. Por otro lado, la oposición público/privado no se contempla, puesto que todos los centros de la investigación son públicos.

²⁰⁹ Como los mismos autores afirman convendría matizar estos resultados con estudios estadísticos más exhaustivos (Prado *et al.* 2009: 95).

preguntas referidas a las variables sociolingüísticas y que el alumnado puede rellenar en la lengua que desee, y otro lingüístico, donde se encuentran los distintos centros de interés y que el alumnado debe completar en la lengua que se le indica (Azurmendi 1983: 235-257)²¹⁰. Se concede un tiempo aproximado de diez minutos por cada centro de interés y se cumplimentan tres campos semánticos en cada sesión de clase: en total, la encuesta se realiza, en cada escuela o colegio, en sesiones de media hora durante cinco días. A pesar de que la investigación incluye quince centros²¹¹, la autora solo analiza cuatro, a saber, *Partes del cuerpo* (1), *Prendas de vestir y de calzar* (2), *Alimentos. Bebidas* (4) y *Juegos. Diversiones. Entretenimientos* (10). El criterio de selección de estos centros de interés es el de procurar que sean los más distintos entre sí tanto desde el punto de vista lingüístico como sociológico. Las variables sociolingüísticas que se toman en consideración son: la *procedencia geográfica del encuestado*, la *edad*, la(s) *lengua(s) materna(s)*, el *sexo*, la(s) *lengua(s) más utilizada(s) en casa*, la(s) *lengua(s) más utilizada(s) con los amigos*, y las dos *lenguas en que se realizan las encuestas*: el euskera y el español. Por su parte, las variables lingüísticas o aspectos lingüísticos examinados son la cantidad de producción, la clase o categoría léxica, la distancia interlingüística²¹², la corrección lingüística y las divisiones significativas²¹³. Con respecto a la distancia entre las dos lenguas, la autora constata que, a menor distancia interlingüística mantenida (a través de los préstamos, de las interferencias, de la sustitución, etc.), mayor probabilidad hay de que se utilicen de forma incorrecta ambas lenguas. Además, esta disminución de la distancia entre las dos lenguas se hace siempre a costa del euskera, que acaba por parecerse cada vez más al

²¹⁰ La autora procura que en cada grupo haya el mismo número de encuestas para cada una de las dos lenguas, por tanto, la clase se divide y la mitad de los alumnos completa la encuesta en castellano, y la otra mitad, en euskera. Por otro lado, se encuesta un total de 1224 informantes, es decir, la práctica totalidad del alumnado que cursaba 5º y 8º en las *Ikastolak* de San Sebastián en los cursos escolares 79-80 y 80-81 (Benítez 1991: 356).

²¹¹ Se trata de: 1. *Partes del cuerpo*, 2. *Prendas de vestir y de calzar*, 3. *Partes y objetos que hay en una vivienda (casa o piso)*, 4. *Alimentos. Bebidas*, 5. *Relaciones familiares (madre, tío...) y no familiares (alumno, amigo...)*, 6. *Ciudad o pueblo: sus partes; lo que se ve y se hace en ellos*, 7. *Los transportes: medios y vehículos que se utilizan*, 8. *Enseñanza: clases de centros y sus partes; objetos y material utilizados, asignaturas, etc.*, 9. *Trabajo: clases de trabajos, de oficios, de profesiones*, 10. *Juegos. Diversiones. Entretenimientos*, 11. *Deportes: prendas de vestir y objetos que se utilizan*, 12. *Acciones que normalmente realizamos todos los días*, 13. *La naturaleza. Las vacaciones*, 14. *Agricultura y vegetales. Ganadería y animales. Pesca y peces* y 15. *Dinero. Bancos. Comercio*. (Azurmendi 1983: 145-147).

²¹² Con esta variable, la autora trata de estudiar el grado de separabilidad de las lenguas. Para medirlo, usa cuatro parámetros: las palabras fonológicamente diferentes en las dos lenguas en contacto, las palabras fonológicamente parecidas en las dos lenguas, los casos de sustitución lingüística y los nombres propios (Azurmendi 1983: 159-160 y 899).

²¹³ Son las distintas parcelaciones semánticas que pueden realizarse en cada centro de interés. Así, por ejemplo, en el centro 10, se diferencia entre los juegos, diversiones o entretenimientos que pueden realizarse en casa, en la calle, en ambos lugares o que son deportes (Azurmendi 1983: 262).

español. Por dicha razón, la autora opina que la situación social en San Sebastián puede describirse como de bilingüismo social desequilibrado, a favor del español (Azurmendi s.d.)²¹⁴, al menos en las fechas en las que se realizó el estudio.

El segundo trabajo realizado en el País Vasco es el de Maitena Etxebarria (1996a), quien lleva a cabo una investigación con 245 alumnos preuniversitarios del área metropolitana de Bilbao donde tiene en cuenta, además de las variables comunes al resto de investigaciones —*sexo, nivel sociocultural, tipo de enseñanza, ubicación* de los centros de enseñanza—, otras —*modelo de enseñanza bilingüe cursado, lengua materna y lengua de uso* predominante fuera del recinto escolar—, que se deben a la singularidad del contacto de lenguas en la comunidad analizada. En Etxebarria (1996b: 319) se constata que las diferencias en función del *modelo de enseñanza* cursado no son notables. En otros dos artículos (Etxebarria 1997 y 1999) se detallan las diferencias debidas al *nivel sociocultural*, que se producen una vez más, a favor del alumnado del estrato más alto del espectro social.

Años más tarde, en el marco del Proyecto Panhispánico, empiezan las investigaciones sobre disponibilidad léxica en la comunidad foral de Navarra (Saralegui y Taberner 2008). La situación lingüística de esta comunidad autónoma es llamativa puesto que combina zonas de bilingüismo entre el euskera y el castellano, y otras zonas no vascófonas donde hay monolingüismo²¹⁵. En las zonas vascófonas de Navarra tanto el euskera como el castellano son oficiales. Para estudiar mejor la influencia de todos estos condicionantes, en el análisis del léxico disponible de Navarra se incorporan —además de las variables propias de este tipo de estudios como el *sexo*, la *ubicación* y el *tipo de centro*, la *residencia familiar* y el *nivel sociocultural*— otras como la *lengua materna y de uso habitual*, el *modelo lingüístico*, la *modalidad de bachillerato*, la *zona dialectal* y la *procedencia de los progenitores*.

La investigación se realiza sobre una muestra de 430 informantes de todas las zonas de la región (comarca de Pamplona, zona Media, Montaña y Ribera) y de los

²¹⁴ Benítez (1991) apunta algunas carencias en el trabajo de Azurmendi, como la desorganización interna de ciertos capítulos o la no consideración de algunos de los trabajos fundamentales de López Morales.

²¹⁵ Cabe recordar que en la Edad Media ya convivía el euskera con el romance. Los límites geográficos entre estas dos zonas ya estaban establecidos de forma similar a la actual a finales del siglo XVI (Taberner 2008b: 546-548).

distintos modelos de enseñanza vigentes (A: castellano con euskera como asignatura, G: castellano, D: euskera con castellano como asignatura). Se analizan los dieciséis centros de interés clásicos, a los que se añaden tres centros más: *Inteligencia, Virtudes y defectos* y *Colores* (Saralegui y Taberero 2008: 748)²¹⁶. Algunos de los frutos de esta investigación pueden hallarse en Taberero (2008a) donde la autora da cuenta de los regionalismos hallados en las encuestas, clasificados según su filiación vasca o romance, y de su distribución geográfica por la región²¹⁷.

Otra aportación muy interesante, desde nuestra perspectiva, es sin duda la de Taberero (2008b). En este artículo la estudiosa, tras repasar la situación lingüística de la región y las características de la investigación, analiza las transferencias del euskera en el léxico disponible de Navarra. Para empezar, detecta que hay unos centros más propicios que otros a la aparición de interferencias²¹⁸, y las clasifica en *code-shifting*, *interferencias* y *préstamos culturales*. El primer caso alberga las sustituciones de código o términos del euskera introducidos por la imposibilidad de recordar la voz castellana correspondiente con la que no guardan similitud alguna²¹⁹. Bajo la denominación de *interferencias* se recogen los préstamos puntuales o esporádicos, que guardan una gran similitud con las palabras castellanas equivalentes²²⁰ y, en tercer lugar, los *préstamos culturales* son las voces del euskera que han pasado al castellano (como *kalimotxo*, *txapela* o *chistorra*), algunas de las cuales ya se han adaptado ortográfica y gramaticalmente y se hallan en los diccionarios normativos.

Otro de los objetivos de la autora es estudiar, en la disponibilidad léxica, la influencia del *modelo lingüístico* cursado, para lo cual analiza los resultados cuantitativos de cinco de los dieciséis centros de interés. Tras realizar las pruebas

²¹⁶ Con respecto al tratamiento de los datos, cabe destacar el uso novedoso que las investigadoras obtienen del programa informático *Excel* aplicado a la disponibilidad léxica (Mangado y Areta 2008).

²¹⁷ Una de las conclusiones a las que llega la investigadora es que los términos de filiación vasca se localizan en la Montaña y en Pamplona y su comarca y, en cambio, los términos romances aparecen en cualquier punto del territorio (Taberero 2008a: 822).

²¹⁸ Concretamente *La ropa, Partes de la casa, La comida y bebida, La escuela, El campo, Los trabajos del campo y del jardín* y *Los juegos y distracciones*.

²¹⁹ Estos ejemplos, como *beroki* ('abrigo') o *arbela* ('pizarra') son recabados principalmente en la zona vascofona y mixta, y en hablantes que poseen el euskera como lengua materna y habitual.

²²⁰ Algunos de los ejemplos son *txocolate, bicicleta* o *zine*, que únicamente se diferencian del castellano desde el punto de vista gráfico; el canal por el que se realizan las encuestas impide constatar el alcance de las mismas. Casi un 85% de estos vocablos pertenecen a informantes del modelo D. Gran parte de ellos, además, se ubica en Pamplona y su comarca, lo que lleva a la autora a concluir que se trata de meras interferencias gráficas, puesto que muchos de los informantes ni siquiera son bilingües en sentido estricto (Taberero 2008b: 562).

estadísticas pertinentes, aplica el paquete estadístico SPSS y llega a la conclusión de que, en relación con el número de palabras escritas, los peores resultados los proporciona el alumnado que sigue el modelo D y que procede del área de la Montaña²²¹ lo que lleva a la investigadora a terminar su artículo cuestionándose la competencia léxica y comunicativa en castellano del alumnado que sigue el modelo educativo de inmersión en lengua vasca.

El análisis cuantitativo y cualitativo global de los datos navarros constituye la tesis doctoral de Areta (2009). En ella, además de comprobar tendencias comunes con las de otras sintopías hispánicas —respecto al número de vocablos y promedios de respuestas por centro— se deja constancia de las diferencias significativas estadísticamente que arrojan todas las variables del Proyecto Panhispánico²²², con la excepción del *sexo*. Las variables incluidas como consecuencia del bilingüismo resultan también significativas: tanto la *lengua materna* como la de *uso habitual* y el *modelo lingüístico* seguido denotan una superioridad de los castellanohablantes²²³. De las restantes variables, solo la *zona geográfica* aporta resultados relevantes y muestra cómo los informantes de la Montaña, zona vascófona, obtienen promedios más bajos. Los otros factores —*modalidad de bachillerato* y *procedencia de los padres*— no se muestran significativos desde el punto de vista estadístico. Areta (2009) también analiza la convergencia entre los subgrupos de las distintas variables²²⁴, y observa que el porcentaje mayor de compatibilidad se da según la variable *modalidad de bachillerato*, seguida de la *residencia familiar*, el *tipo de centro*, el *sexo* y la *lengua materna*²²⁵. Del mismo modo, se observan diferencias interesantes en el análisis cualitativo según las variables mencionadas²²⁶. Por último, la investigadora aborda la variación diatópica y

²²¹ Los modelos A y G obtienen resultados similares y las diferencias mayores se dan en los centros de interés relacionados con el ámbito académico. Con respecto a la variable geográfica, los mejores resultados los obtiene Pamplona y su comarca. La autora también comprueba estadísticamente “la coincidencia casi total de los alumnos del modelo D con aquellos que poseen el euskera como lengua materna y de uso, o al menos en una de estas dos posibilidades.” (Taberner 2008b: 560).

²²² Esto es, la variable *tipo de centro*, su *ubicación*, la *residencia familiar* y el *nivel sociocultural* establecen diferencias significativas a favor de los centros privados, de las zonas urbanas y de los estratos más altos, respectivamente.

²²³ Con respecto al *modelo lingüístico*, los promedios que ofrece el alumnado que sigue los modelos A y G son siempre mejores frente al D.

²²⁴ Para ello utiliza las palabras que tienen un índice de disponibilidad mayor o igual a 0.05.

²²⁵ Los porcentajes son: 81.3% para la *modalidad de bachillerato*, 75.3% para la *residencia familiar*, 74.9% para el *tipo de centro*, 72.3% para el *sexo* y 67.8 para la *lengua materna*. En las restantes variables, especialmente aquellas formadas por tres o más subgrupos, el porcentaje es menor.

²²⁶ Por ejemplo, la *lengua materna* influye en las respuestas de los informantes, sobre todo en las áreas relacionadas con la vida cotidiana.

socioestilística llevando a cabo un análisis tanto de las voces dialectales como de los términos de distintos registros que aparecen en los listados²²⁷.

3.2.1.3. En contacto con el catalán

Hace ya un cierto tiempo que se han iniciado los estudios sobre disponibilidad léxica en Cataluña. Lamentablemente, los únicos datos que tenemos hasta la fecha son los de nuestro trabajo sobre la ciudad de Lleida²²⁸. En Serrano Zapata (2003b y 2004), se analiza la disponibilidad de una muestra de 48 informantes distribuidos según el *sexo*, el *tipo de centro*, el *nivel sociocultural* y la *lengua materna*. Los resultados muestran una superioridad cuantitativa por parte de las mujeres, de los informantes de centros privados, de nivel sociocultural medio y catalanohablantes. También se documenta la esperable aparición de interferencias en los listados, que se dan con mucho más frecuencia en los hablantes cuya lengua materna es el catalán; y se clasifican en interferencias morfológicas (préstamos híbridos y diminutivos) y léxicas (préstamos, transferencias semánticas, cambios en el entorno de uso, calcos sintácticos y creaciones) (Serrano Zapata 2003b: 93-100). Más tarde (Serrano Zapata 2006), ofrecimos una nueva clasificación de las interferencias recabadas en seis centros de interés. Esta vez, trabajamos ya con todo el corpus—esto es, con 240 informantes, aquellos en los que se basa la presente investigación— y distinguimos cuatro grandes grupos de fenómenos derivados del contacto lingüístico: los préstamos puros, los préstamos híbridos, los calcos semánticos y los calcos sintácticos.

En otra comunidad autónoma bilingüe como es la Comunidad Valenciana hace ya algunos años que han dado comienzo los estudios de disponibilidad²²⁹. Siguiendo un criterio estrictamente geográfico nos referimos, en primer lugar, a los estudios realizados en Castellón. Dentro de la Comunidad Valenciana es precisamente en esta provincia donde el bilingüismo se halla más extendido. A pesar de que los datos finales aún no han sido publicados, contamos ya con varios estudios realizados sobre ocho de los diecisiete centros de interés, y basados en una muestra de 246 informantes

²²⁷ El estudio de los regionalismos cuenta ya con una cierta tradición (*cf.* Arnal 2008a y 2008b). Sin embargo, la segunda perspectiva, referida a la variación socioestilística, es novedosa en los trabajos de disponibilidad.

²²⁸ Los trabajos sobre las provincias de Barcelona, Tarragona y Girona están todavía en proceso puesto que el equipo de investigación que asumió hace algunos años esta empresa ha tenido dificultades para proseguirla, aunque esperamos poder contar pronto con los resultados y, de este modo, ofrecer un panorama global de la disponibilidad léxica de Cataluña.

²²⁹ Véase una perspectiva general de las características de estas investigaciones en Casanova (2006).

estratificados según la *lengua materna*, la *titularidad del centro educativo*, el *nivel cultural de la familia* y el *lugar de residencia*²³⁰. En un primer trabajo (Blas y Casanova 2003a) se expone la poca relevancia que el modelo de enseñanza tiene en los resultados recabados sobre el léxico disponible de esa provincia. Al mismo tiempo se observa que la variable entorno sociocultural sí parece tener más influencia, aunque solo entre los grupos más extremos, esto es, entre el alumnado cuyos padres tienen estudios superiores y aquellos cuyos padres apenas han recibido instrucción escolar.

Más interesante para la perspectiva que nos ocupa, es la relevancia de la lengua materna. En Blas y Casanova (2003b) se observa cómo esta variable incide leve pero significativamente en los resultados a favor del alumnado que se declara castellanohablante o que considera como maternas ambas lenguas²³¹. Los mismos autores profundizan en esta cuestión (Blas y Casanova 2001-2003) y muestran la correlación de esta variable con otros factores. De este modo, al cruzar las variables de forma simultánea llegan a la conclusión de que algunos factores hacen aumentar las diferencias, mientras que otros las neutralizan²³². En época reciente, Casanova (2013) emprende el análisis de los valencianismos detectados en su corpus²³³. En este estudio, la autora nos ofrece los índices relativos a la cantidad de vocablos valencianos hallados, tanto con respecto al total de palabras como con respecto al total de vocablos, y ofrece el porcentaje de valencianismos que presenta cada centro de interés. También analiza, en concreto, las interferencias recabadas en *Trabajos del campo y del jardín* y en *El campo*, tras comprobar que son los centros más productivos en este sentido. Además, compara la cantidad de interferencias hallada en distintos territorios donde también conviven el catalán y el castellano, como Alicante, Valencia o Lleida y señala algunas tendencias comunes, como la mayor producción de catalanismos en torno a *Trabajos*

²³⁰ La muestra abarca toda la provincia, exceptuando alguna comarca interior monolingüe. Cabe señalar que la variable nivel cultural se basa únicamente en el nivel de instrucción de los progenitores y que la muestra global se compone de 402 informantes (Blas y Casanova 2003a).

²³¹ Para comprobarlo usan la prueba estadística T-student. Las diferencias existentes, con todo, no superan los 3 puntos y son mayores en los resultados cuantitativos que con respecto al número de palabras diferentes. Además, los autores comparan sus datos con los de Etxebarria (1996b) donde se analiza la incidencia del modelo lingüístico escolar seguido en el País Vasco.

²³² Con respecto al *sexo*, por ejemplo, la mayor distancia se establece entre las mujeres castellanohablantes y los hombres valencianohablantes, en cambio, las diferencias se reducen drásticamente entre los informantes de nivel cultural alto, independientemente de su lengua materna. Obviamente, usamos el término valenciano para referirnos a la variante dialectal del catalán.

²³³ Esta vez la autora maneja una muestra mayor, de 354 informantes, de los cuales 158 tienen el valenciano como lengua materna.

*del campo y del jardín, El campo, Comidas y bebidas o Animales*²³⁴. La investigadora, en último lugar, relaciona sus datos con las cinco variables contempladas en su estudio y concluye que “el hablante tipo que presenta más interferencias se especifica como proveniente de un estrato o nivel cultural bajo, estudiante de un centro público ubicado en una zona rural y cuya lengua materna es el valenciano” (Casanova 2013: 542).

Sin duda alguna, es en Valencia donde los estudios de disponibilidad han dado algunos de sus mejores frutos, puesto que ya contamos con los resultados de las investigaciones referidas no solo al castellano, sino también al catalán —como veremos más adelante (*vid.* 3.3.)—. Con respecto al léxico disponible en castellano del alumnado de Valencia, la rigurosa obra de Gómez Molina y Gómez Devís (2004) da cuenta de sus principales características desde una triple perspectiva²³⁵. Los autores emprenden el estudio de su comunidad, donde se da un bilingüismo de carácter no diglósico con cierto desequilibrio a favor del castellano y donde se cuenta con un alto porcentaje de bilingüismo pasivo²³⁶. Teniendo en cuenta esta premisa, la muestra se compone de 465 informantes que se distribuyen según las siguientes variables: *sexo*, *tipo de centro*, *nivel sociocultural*²³⁷, *lugar de residencia* (Valencia ciudad, área metropolitana, zona semiurbana y ámbito rural²³⁸), y *lengua habitual* (monolingües o bilingües)²³⁹. En el análisis, se aportan los totales de palabras y vocablos, sus promedios y los índices de cohesión, y se comparan estos datos con los de otras sintopías del mundo hispánico, observándose así tendencias similares.

Además, los investigadores valencianos llevan a cabo la covariación de las variables dependientes (los centros de interés) con las variables independientes (las sociales) y llegan a la conclusión de que el factor que más incide en el léxico disponible

²³⁴ La estudiosa maneja nuestros primeros datos de 2006 (Serrano Zapata 2006), relativos únicamente a seis centros de interés. De otra parte, en Casanova (2008) se analiza los anglicismos presentes en el mismo corpus.

²³⁵ Las conclusiones se recogen desde la léxico-estadística, la sociolingüística y la psicolingüística, aunque la perspectiva sociolingüística es la predominante.

²³⁶ Aproximadamente el 87% de la población habla castellano y entiende el valenciano. Por tanto, la inmensa mayoría de la comunidad se puede clasificar como monolingüe activa en castellano y bilingüe pasiva en valenciano o como bilingüe activa.

²³⁷ Se determina según el grado de instrucción y la profesión de los progenitores.

²³⁸ Estas zonas se distinguen según la residencia del informante. El área metropolitana es la corona que rodea la capital de la Comunidad, mientras que la zona semiurbana comprende las ciudades medianas y capitales comarcales.

²³⁹ La muestra se integra por 343 estudiantes monolingües y 122 bilingües. En la población real los monolingües suponen el 58.3% y los bilingües el 41.7% (Gómez Molina y Gómez Devís 2004: 43-44). Como ya hemos indicado, los monolingües son bilingües pasivos en valenciano.

de los valencianos es el *tipo de centro educativo*, donde el alumnado de los centros privados muestra una productividad mayor²⁴⁰. En segundo lugar, la variable *nivel sociocultural* también influye vehementemente —a favor de los estratos más altos—, mientras que otros factores como el *sexo* o el *lugar de residencia* también muestran, aunque en menor medida, una preponderancia de las mujeres y de los informantes de la zona rural en la productividad léxica²⁴¹. Sin embargo, la *lengua habitual* del informante muestra una influencia mucho menor, apenas apreciable²⁴², tal como los autores ya aventuran en las hipótesis iniciales (Gómez Molina y Gómez Devís 2004: 16), dado que todos los informantes han tenido doce años de escolarización mayoritariamente en lengua castellana.

Por otro lado, los autores también aplican diferentes pruebas estadísticas para constatar qué centros de interés son más propicios a la estratificación mediante las variables, qué factores sociológicos correlacionan más entre sí, etc. (Gómez Molina y Gómez Devís 2004: 123-138). El análisis cuantitativo se complementa con el cualitativo, donde los autores comprueban las coincidencias que se dan entre las cincuenta primeras palabras de cada subgrupo según las variables independientes ya mencionadas. Al respecto, el porcentaje mayor de convergencia se da en la *lengua habitual*, seguida del *tipo de centro*, el *sexo*, el *lugar de residencia* y el *nivel sociocultural*, que arroja el menor porcentaje de coincidencia entre las cincuenta primeras lexías.

Además, también detectan la presencia de voces del catalán en los listados y diferencian entre: préstamos puntuales, en vías de integración o integrados y sustituciones de código. Utilizan los índices de disponibilidad para mostrar el alcance de estos términos comparándolos con sus equivalentes en castellano y reconocen la dificultad de saber qué realiza realmente el hablante: si una sustitución de código (al no recordar el término castellano), o un préstamo (al desconocer la denominación en

²⁴⁰ Los centros privados obtienen mejores promedios en todas las áreas. Además, todas las diferencias, salvo una, resultan significativas estadísticamente.

²⁴¹ Aunque destacan los resultados superiores del estrato rural, los autores achacan el valor discriminatorio medio de esta variable a las buenas comunicaciones de la provincia y a la dinámica sociodemográfica de sus habitantes (Gómez Molina y Gómez Devís 2004: 119-120).

²⁴² Únicamente dos de las diecisiete diferencias son significativas; y además, la distancia entre los promedios es la menor de todas las variables.

castellano). Del mismo modo, según el índice de disponibilidad alcanzado clasifican estas voces en:

- ✓ préstamos culturales: voces del catalán que designan realidades propias del entorno,
- ✓ vocablos con claro predominio del índice de disponibilidad en catalán, que reflejan cierto desconocimiento u olvido del término castellano y podrían ser préstamos o sustituciones de código, y
- ✓ dobles, en los que el mayor índice de disponibilidad es para la palabra castellana y, por tanto, pueden ser casos de transferencias léxicas pero solo para determinados informantes, no para el colectivo.

Finalmente, se presentan las conclusiones desde las diferentes perspectivas mencionadas. La aportación más interesante, desde nuestro punto de vista, es el exhaustivo rigor estadístico con el que los autores comprueban si sus resultados son extrapolables al conjunto del universo estudiado.

Por otro lado, es de justicia reconocer que —con anterioridad al trabajo de Gómez Molina y Gómez Devís (2004)— existe una tesis de licenciatura sobre la disponibilidad léxica del español hablado en Valencia realizada con hablantes de diferentes estratos sociales y generacionales y que se aleja, por tanto, de las directrices del Proyecto Panhispánico (Terrádez 1996). Entre las conclusiones más reseñables de este trabajo, destaca la relación directamente proporcional que existe entre el carácter estructurado y objetivo del centro de interés, por un lado, y, por el otro, el número de palabras que este recibe²⁴³. En cuanto a la relación entre el número de vocablos actualizados y las variables sociales empleadas en su estudio, las variables *sexo* y *conocimiento del valenciano* son las que resultan de menor significatividad, mientras que la *edad* y el *nivel sociocultural* presentan un patrón de distribución regular y se revelan como indicadores sociolingüísticos²⁴⁴.

²⁴³ Es decir, cuando el campo léxico está muy estructurado y presenta un cariz objetivo (como *Alimentos*, *Animales* o *El Cuerpo Humano*), el informante actualiza muchas palabras relacionadas con él. Por el contrario, si se trata de un centro de interés subjetivo, como *Cualidades Humanas* o *Juegos y Distracciones*, aparece un menor número de palabras.

²⁴⁴ Agradecemos, desde aquí, al autor de este trabajo la información facilitada sobre el mismo.

Para finalizar con las aportaciones al Proyecto Panhispánico en la Comunidad Valenciana, debemos referirnos a la tesis doctoral de Martínez Olmos (2008) sobre la disponibilidad léxica en las comunidades de habla alicantinas. El objetivo principal de este trabajo es reflejar la estratificación social de estas comunidades de habla a partir de la covariación de la variable léxica con diversos factores lingüísticos y sociales. Para ello, y siguiendo un enfoque sociolingüístico, se estratifica una muestra, compuesta por 460 informantes, según las siguientes variables sociales: *sexo*, *tipo de centro*, *tipo de hábitat* (rubano o rural), *ubicación geográfica* (litoral o urbana), *nivel sociocultural* (medio-alto, medio-medio o medio-bajo, basados en la profesión y en los estudios de los progenitores), y *comunidad de habla*. Especialmente importante resulta la incorporación de esta última variable, infrecuente en los estudios de disponibilidad²⁴⁵, que Martínez Olmos (2008: 45-70) justifica con vehemencia como el mejor punto de partida para emprender un análisis de carácter sociolingüístico. De esta manera, las comarcas de la provincia de Alicante se clasifican en cuatro comunidades de habla: (a) comarcas con bilingüismo activo con predominio del valenciano, (b) con bilingüismo activo con predominio del castellano, (c) con bilingüismo pasivo con dominio del castellano e históricamente parte del Reino de Valencia y (d) con bilingüismo pasivo con dominio del castellano, pero incorporadas a la provincia de Alicante durante el siglo XIX²⁴⁶.

Tras realizar un exhaustivo y riguroso análisis de los resultados, la autora comenta los datos globales por variables lingüísticas, observando tendencias similares a las obtenidas en otras investigaciones y hace hincapié en el hecho de que en las áreas semánticas contenidas en el currículo escolar se comparte un mayor número de tipos léxicos diferentes²⁴⁷. Con respecto a la incidencia, tanto cuantitativa como cualitativa, de las variables sociales, la que parece tener mayor influencia es la *comunidad de habla*, donde el promedio de palabras es claramente superior en los informantes de la comunidad de habla más castellanizada²⁴⁸. Además, el porcentaje de compatibilidad léxica entre los subgrupos es el más bajo de todos, con lo que la actuación lingüística

²⁴⁵ Aunque Llopis (2009) también la utiliza (*vid.*3.3).

²⁴⁶ Son las siguientes: (a) área de *la Marina (Alta y Baixa)-el Comtat-l'Alcoià*, (b) *l'Alacantí-el Vinalopó*, (c) el Bajo Segura (excepto Guardamar y Barba-roja) y (d) el Alto Vinalopó (Villena y Sax).

²⁴⁷ La autora prefiere el término *tipos léxicos diferentes* al de *vocablos*, así como el de *total de ocurrencias* al de *palabras*. Algunas de estas áreas, que también cuentan con un índice de cohesión alto, son *Partes del cuerpo humano*, *Los animales* o *La ropa* (Martínez Olmos 2008: 168 y 177).

²⁴⁸ Es decir, la del Alto Vinalopó, que comprende las localidades de Villena y Sax.

parece diferir considerablemente según esta variable²⁴⁹. También se observa cómo en las comarcas donde predomina el valenciano aparecen más transferencias léxicas del catalán.

La segunda variable que más parece covariar con los resultados es el *nivel sociocultural*, aunque, contrariamente a lo que cabría esperar, es el estrato medio-medio el que supera los otros dos²⁵⁰. Por otro lado, aunque el porcentaje de convergencia entre los subgrupos es alto, se aprecian diferencias sociolectales considerables entre cada uno de los niveles socioculturales, y el mayor número de transferencias del catalán, considerando las cincuenta primeras palabras, se registra en el nivel medio-bajo. En tercer lugar, según el orden de importancia, hallamos la variable *ubicación geográfica*, que arroja unos resultados mejores para el alumnado de instituciones ubicadas en el litoral²⁵¹; desde la perspectiva cualitativa, no se observan diferencias notables.

El resto de variables parecen tener una influencia menor en los resultados. El *tipo de hábitat*, contrariamente a lo que sucede en la mayoría de las investigaciones, otorga unos promedios más altos al alumnado del ámbito rural, en quienes también se observa un mayor número de transferencias. En cuanto al *tipo de centro*, una vez más las instituciones privadas aventajan a las públicas, aunque con una diferencia escasísima, y con un porcentaje alto de compatibilidad léxica. En último lugar, la variable *sexo* favorece una vez más a las mujeres que, además, responden de forma más cohesionada²⁵² y son más permeables a las transferencias de otras lenguas.

El segundo objetivo del trabajo de Martínez Olmos (2008) es el análisis de las transferencias léxicas detectadas en las comunidades de habla. La investigadora analiza de forma cuantitativa —sirviéndose de los distintos índices que ofrece el programa *Lexidisp*— y de forma cualitativa todas las transferencias halladas en los listados y confirma la hipótesis inicial de que hay más transferencias lingüísticas del vernáculo en aquellas comunidades de habla donde los hablantes poseen un mayor grado de

²⁴⁹ Este porcentaje se calcula sobre las cincuenta primeras palabras ordenadas según el índice de disponibilidad y es concretamente de un 74.1%.

²⁵⁰ Este sociolecto sobrepasa el medio-bajo en todos los centros y el medio-alto en diez de las dieciséis áreas.

²⁵¹ Como ya habían detectado los contados trabajos que dividen esta variable de la misma forma (Martínez Olmos 2008: 81).

²⁵² Resulta curioso este dato, puesto que las mujeres son superiores numéricamente a los varones en la muestra (270 frente a 190).

competencia comunicativa bilingüe²⁵³. Del mismo modo, resulta curioso observar cómo en estas comunidades también se da un mayor uso de transferencias de otros idiomas. En general, el número de catalanismos es escaso en el corpus, aunque se detecta su presencia principalmente en los centros de interés más próximos al ámbito familiar (v.gr. *El campo, Alimentos y bebidas, Los animales, La cocina...*) y, sobre todo, en *Trabajos del campo y del jardín*. Según el perfil sociológico de los informantes, se advierte una mayor tendencia al uso de las transferencias léxicas en las mujeres, en los informantes de centros públicos, en las comunidades con un bilingüismo más activo y en el alumnado ubicado en el nivel sociocultural medio-medio. En cuanto a la clasificación de las transferencias, su propuesta sigue de cerca la tipología presentada por Gimeno y Gimeno (2003) basada en la mayor o menor integración lingüística y social del préstamo. Para valorar la integración social del préstamo se vale de la frecuencia relativa de las transferencias y así determina que, por debajo de un 1%, son interferencias y el cambio de código, en cambio, posee una frecuencia relativa del 3 al 15%. Tras aplicar estos porcentajes, concluye que la mayor parte de catalanismos hallados constituyen cambios de código, aunque también detecta algún caso aislado de calco e híbridos (derivados y compuestos)²⁵⁴.

No podemos acabar este apartado, sin mencionar que en el léxico disponible de Aragón (Arnal *et al.* 2004) también aparecen términos, aunque escasos, de filiación catalana. Esto resulta lógico toda vez que, en su elaboración, se tienen en cuenta las distintas áreas lingüísticas de Aragón. Al incluir el área oriental de habla catalana —también llamada *La Franja*— aparecen interferencias del catalán, “catalanismos” y “aragonesismos”, o términos dialectales comunes en Aragón y Cataluña²⁵⁵. Arnal (2008a, 2008b y 2009) hace un exhaustivo análisis de todos los términos dialectales, donde también se incluyen aquellas voces que muestran la pervivencia de algún rasgo del dialecto aragonés y las voces propias de Aragón que no existen en el español general, o tienen alguna diferencia respecto a él. La investigadora maña llega a la conclusión de que los informantes de las zonas rurales utilizan más términos dialectales y, además, que, de las tres zonas lingüísticas de Aragón, es precisamente el área oriental

²⁵³ Es decir, se detectan más transferencias en *La Marina, el Comtat y l'Alcoià* y en *l'Alacantí-el Vinalopó*.

²⁵⁴ Véase también al respecto Gimeno y Martínez Olmos (2008a) y Gimeno y Martínez Olmos (2008b).

²⁵⁵ Entre el primer grupo (catalanismos) hallamos *barretxa, gualla, palot* o *sindria*, y entre el segundo (aragonesismos), *bróquil, devantal, pichar* o *paleta* (Arnal 2009: 273-274). Nótese los paralelismos entre estas palabras y las halladas en Lleida ciudad (Serrano Zapata 2003b) o en este mismo trabajo.

la que más dialectalismos aporta, sobre todo las poblaciones de Fraga y Valderrobres (Arnal 2009)²⁵⁶.

3.2.1.4. En contacto con otras lenguas

En primer lugar nos referimos a la obra que recoge el léxico disponible de Asturias, región donde hace muchos siglos que conviven el castellano y el asturiano (o bable), lengua románica que, en época medieval, llegó a hablarse en una franja extensa que discurría paralela al territorio de habla gallego-portuguesa y que alcanzó hasta parte de Extremadura²⁵⁷. El trabajo de Carcedo (2001) se lleva a cabo dentro de las directrices del Proyecto Panhispánico, pero a las variables comúnmente analizadas —*sexo*, *tipo de centro*, *tipo de comunidad* y *nivel sociocultural*— se añade la de *zona geográfica* con la que el autor pretende explorar en qué medida influye la variación dialectal sobre el léxico disponible de los asturianos. De esta forma, se establecen cuatro áreas geográficas, coincidentes con las zonas tradicionales que la dialectología reconoce en la región: el bable oriental, el central y el occidental, y la zona de gallego-asturiano. En el análisis cuantitativo de las variables, se observa cómo los resultados favorecen la zona del bable central, aunque las diferencias entre las distintas zonas no parecen ser muy destacables²⁵⁸. Con respecto a las restantes variables, los mejores resultados los aportan las mujeres, el alumnado de centros privados, de entorno rural y de nivel sociocultural alto²⁵⁹. De otra parte, el análisis cualitativo aborda la compatibilidad de los inventarios obtenidos según los factores sociales, comparados bien en su totalidad, bien teniendo solo en cuenta las veinte primeras palabras. En estos dos parámetros, el porcentaje menor de convergencia se da entre las respuestas de los informantes en función de la *titularidad del centro educativo*, hecho que vuelve a destacar la importancia de esta variable en el léxico de los preuniversitarios asturianos. En cambio, el porcentaje mayor de coincidencia se da según la variable *sexo*²⁶⁰, analizando todos los datos, o entre los niveles medio y bajo del

²⁵⁶ También demuestra, con respecto al nivel sociocultural, que el grupo más bajo aporta más voces dialectales (Arnal 2008b).

²⁵⁷ El nombre más adecuado para hablar de la situación lingüística medieval es “asturleonés”, pero esa etiqueta no se ajusta a la realidad diatópica actual puesto que el último reducto de esta lengua románica coincide hoy prácticamente con los límites geográficos de Asturias.

²⁵⁸ Los promedios de respuestas son: 19.1 (bable central), 20.4 (occidental), 19.2 (oriental), 18.7 (gallego-asturiano) y 19.2 (Oviedo). El autor hace una nueva subdivisión al aislar la capital dentro de la zona del bable central.

²⁵⁹ Las diferencias con respecto al *sexo* o al *tipo de comunidad* son menos destacables que las referentes al *tipo de centro escolar* o al *nivel sociocultural*, que se determina tanto en función de la profesión y el nivel de estudios de los progenitores como de los ingresos económicos familiares.

²⁶⁰ Factor que, como ya hemos dicho, no ejerce apenas influencia en el trabajo sobre Asturias.

espectro social, considerando las veinte palabras más disponibles.

Por lo que respecta a la variable *zona geográfica*, el mayor porcentaje de convergencia se da entre el bable de la ciudad de Oviedo y el resto de la zona central. Atendiendo a este factor, el autor concluye que: “las similitudes son muy superiores a las divergencias, lo que pone de manifiesto la unidad esencial del vocabulario de los asturianos, independientemente de la variable dialectal característica de la zona regional en la que se inscriben” (Carcedo 2001: 105). Por otro lado, el autor comenta, en los criterios de edición, la presencia de voces autóctonas en algunos centros de interés como *fabes* (en *Comidas y bebidas*) o *sachar, sallar y semar* (en *El campo y/o en Trabajos del campo y del jardín*).

Otra investigación que trata de establecer el léxico disponible en una situación de contacto de lenguas en la Península, en este caso entre el español y el inglés, es la que se realiza, desde hace años, en Gibraltar. Los objetivos de esta investigación —llevada a cabo desde el Área de Lingüística General de la Universidad de Cádiz— son describir los elementos léxicos más disponibles tanto en español como en inglés y, al mismo tiempo, comprobar el grado de penetración en el uso de elementos léxicos de una lengua en la otra (Escoriza 2007). Para alcanzar estos objetivos, se aplican las encuestas de disponibilidad en cada lengua con un número similar de informantes preuniversitarios²⁶¹. Además, se establecen veintiún centros de interés, es decir, se añaden cinco (*Los colores, La mar, Defectos y cualidades físicas y morales, Acciones y Religión*) a los dieciséis clásicos, con el propósito de recoger otras categorías gramaticales o motivados por el entorno sociocultural de la comunidad.

Tras una primera revisión de los listados en castellano, Escoriza (2007) constata la presencia de numerosos anglicismos, extendidos de forma general en las comunidades de habla hispana, así como una multitud de elementos léxicos que reflejan una interferencia del inglés en el uso del español²⁶². El autor distingue, *grosso modo*,

²⁶¹ Concretamente, 137 estudiantes la completan en inglés y 129, en español. El estudio sigue, en general, las pautas metodológicas del Proyecto (dos minutos, etc.) (Escoriza 2007). La información personal que se recoge de cada informante es el sexo, la lengua materna, la lengua utilizada en casa, la lengua utilizada fuera de la escuela, el nivel de ingresos, la profesión y los estudios de los padres (Escoriza 2003).

²⁶² También observa la presencia de numerosos errores ortográficos, que caracterizan dialectalmente la comunidad de habla. Por otro lado, aunque la comunidad de Gibraltar tiene un carácter multiétnico y

dos grandes grupos de elementos en español donde se percibe la influencia de la lengua inglesa. En primer lugar, se observan lexías en las que se reproducen aspectos estructurales del inglés (“air acondicionado” o “jardinear”) y, en segundo lugar, también hay elementos en los que la interferencia parece causada por motivaciones de carácter fónico o gráfico, tanto si se adopta únicamente la grafía y no la pronunciación (“officinas”, “kanguro”) como si hay una interferencia de carácter fónico (“confession”, “receptionista”). Como ya hemos dicho, esta investigación pretende establecer el léxico disponible en español y en inglés y también indagar en la relación existente entre interferencias y tipos de interferencias y otros factores como el índice de disponibilidad, el centro de interés o las lenguas que maneja el informante (Escoriza 2003, 2007)²⁶³.

3.2.2. El español en contacto con otras lenguas no peninsulares

3.2.2.1. En el norte de África

Para empezar, analizamos las obras que han tratado de la disponibilidad léxica en España, aunque fuera de los límites de la Península Ibérica y, además, en contacto con otra lengua que no tiene el estatus de oficial. Nos referimos, pues, a las ciudades de Ceuta y Melilla, donde se usa de forma mayoritaria el castellano, pero también otras lenguas como el árabe o el bereber.

El primer trabajo referido al léxico disponible de estas ciudades es el de Ceuta (Ayora 2006). En este estudio, que sigue fielmente las pautas acordadas por los estudiosos de la disponibilidad, la autora trabaja con una muestra de 266 informantes²⁶⁴ de los cinco centros educativos de la ciudad²⁶⁵. Las variables sociales que se consideran en el análisis son la *lengua materna*²⁶⁶, el *tipo de centro*, el *sexo* y el *nivel sociocultural*²⁶⁷. Con respecto a la primera de las variables, hay que aclarar que en Ceuta se da un bilingüismo unidireccional puesto que únicamente los hablantes de

multicultural, desde el punto de vista lingüístico lo más relevante es el contacto entre el inglés, la lengua oficial, y el español, a pesar de que en realidad la situación es de multilingüismo social (Escoriza 2007).

²⁶³ Lamentablemente, no contamos aún con los resultados, lo que permitiría establecer interesantes comparaciones, dada la cercanía de sus planteamientos metodológicos con los nuestros. Véase más información al respecto en Fernández Smith *et al.* (2012).

²⁶⁴ Suponen un 45.08% de la población escolar de ese nivel.

²⁶⁵ A pesar de que solo uno de ellos es privado, los resultados cuantitativos otorgan cierta relevancia a esta variable.

²⁶⁶ En la encuesta, además, también se pregunta por la lengua utilizada en casa y la usada fuera del recinto escolar.

²⁶⁷ Para establecer el nivel sociocultural se tiene en cuenta el nivel de instrucción de los padres, su profesión y sus ingresos. Por otro lado, la variable *factor espacial* (rural/urbano) no se incluye porque todos los centros educativos se sitúan en la ciudad.

dariya son bilingües²⁶⁸ y es muy extraño que un hispanohablante sienta la necesidad de aprender la lengua árabe. Además, la coexistencia de estas dos lenguas presenta un carácter diglósico a favor del español, que se considera la única lengua oficial de la ciudad y, por ende, la de la escuela y la administración. En la muestra, únicamente hablan dariya 29 alumnos, frente a los 237 que se declaran hispanohablantes, reflejándose así la distribución social de la ciudad donde la población mayoritaria es de origen europeo²⁶⁹.

Con respecto a los estímulos verbales, además de los dieciséis centros de interés clásicos, se incluyen otros más con el propósito de recabar otras categorías gramaticales como *El mar*, *Defectos y cualidades físicos y morales* y *Acciones*. A la hora de evaluar los resultados cuantitativos, la autora constata que el *sexo* no establece diferencias importantes, ya que existe bastante equilibrio entre la media de palabras por sujeto según esta variable. Sin embargo, un factor muy influyente es la *lengua materna*, ya que los hablantes de árabe manejan un bagaje léxico menor al de los hablantes de español, tanto en el promedio de palabras por individuo como en el número de palabras diferentes. Otro factor que presenta diferencias marcadas es el *tipo de centro* escolar: los estudiantes del centro privado aventajan en disponibilidad a los de los centros públicos²⁷⁰. También en el *nivel sociocultural*, como viene siendo habitual en los trabajos sobre disponibilidad, se constata una estrecha relación entre esta variable y la cantidad de léxico que maneja el individuo.

El estudio cualitativo del corpus comprende el análisis de las asociaciones realizadas en cada estímulo, las categorías gramaticales y los compuestos sintagmáticos recabados, así como un examen de las principales marcas comerciales, siglas y acrónimos hallados. También se estudia el origen de las palabras obtenidas prestando especial atención al léxico de origen dialectal, en su mayoría de influencia andaluza, y se señala la presencia de numerosas voces del habla autóctona o localismos (*alita*,

²⁶⁸ El dariya es un dialecto oral exclusivo de la zona fronteriza que deriva del árabe clásico. Sus hablantes, de religión musulmana, aprenden el español en la escuela.

²⁶⁹ Es decir, el 70%, “[e]l 30% restante “está integrado por las poblaciones árabe-musulmana (22%), judeo-hebraica (4%), hindú-brahmanista (3%) y gitano-cristiana (1%)” (Ayora 2006: 37).

²⁷⁰ Como la investigadora reconoce, hay una relación entre esta variable y la anterior —la *lengua materna*— y el *nivel sociocultural*. Así pues el alumnado del centro privado suele pertenecer al estrato alto y los hablantes de dariya, en cambio, al más bajo.

cusús, harira, Kebab...) algunas de los cuales reflejan, como es lógico, la convivencia en Ceuta de distintas culturas.

Poco tiempo después, ve la luz el léxico disponible de la otra ciudad española en el continente africano, esto es, Melilla (Fernández Smith *et al.* 2008). Esta investigación, que también sigue las directrices metodológicas del Proyecto, se basa en los datos de 315 estudiantes preuniversitarios²⁷¹ estratificados según el *sexo*, la *lengua materna* y el *nivel sociocultural*²⁷². Con respecto a la *lengua materna*, la muestra se divide entre hispanohablantes (290 informantes), hablantes de mazigia, una variedad de bereber²⁷³ (22), y hablantes de otras lenguas (3)²⁷⁴. Tras descartar estos últimos del análisis por la escasa representatividad numérica que suponen, los autores comprueban estadísticamente la existencia de una relación significativa en favor de los hablantes de español, que aventajan a los hablantes de bereber en trece de los dieciséis centros de interés. Cabe aclarar que los hablantes de mazigia son bilingües en distinto grado, mientras que los hablantes de español no lo son en absoluto, puesto que es muy raro que un hispanohablante aprenda bereber, excepto en el caso de que pertenezca a una familia mixta (Fernández *et al.* 2008: 38)²⁷⁵.

En cuanto a las otras variables, las mujeres parecen obtener mejores resultados, aunque las pruebas estadísticas aplicadas no corroboran este dato, y, con respecto al *nivel sociocultural*, que se distribuye en cuatro grupos, los investigadores solo detectan que el nivel más bajo de la escala covaría significativamente con los niveles inmediatamente superiores, es decir, el medio-bajo y el medio-alto²⁷⁶. En otro orden de cosas, al analizar la distribución de los rangos de los centros de interés según las

²⁷¹ Cifra que supone el 66.60% del total de alumnos matriculados en 2º de bachillerato durante el curso académico 2004-2005 y el 0.4690% de la población total de Melilla.

²⁷² Las variables *tipo* o *ubicación del centro escolar* no son tenidas en cuenta en esta investigación por las condiciones particulares de la misma (solo existe un centro concertado, donde también se aplica la encuesta, y todas las instituciones escolares se ubican en la ciudad).

²⁷³ La lengua mazigia o tamazight pertenece al tronco común de las lenguas afroasiáticas, dentro de la familia del bereber.

²⁷⁴ Una distribución similar se observa en la población según su origen, ya que el 60% es de procedencia peninsular, mientras que la población de ascendencia bereber no alcanza el 40%.

²⁷⁵ Por esta razón, se toma la resolución de considerar hablantes de tamazight a todos los informantes que dicen ser bilingües o que consignan las dos lenguas como maternas. Como también sucede en Ceuta, existe una fuerte correlación entre los hablantes de bereber y los niveles más bajos de estratificación social. Por otro lado, Melilla es una comunidad diglósica, donde el español es la lengua de la cultura, las instituciones y los negocios.

²⁷⁶ Para clasificar a los informantes en un nivel u otro se tiene en cuenta el nivel de instrucción y la profesión de los progenitores, como es habitual en los estudios de disponibilidad léxica. Aunque se solicitan en la encuesta los ingresos familiares, estos no se consideran finalmente.

palabras distintas producidas, es decir los vocablos, se aprecian también diferencias según la *lengua materna*. Mientras que los hablantes de lengua materna española siguen el patrón general del resto de investigaciones, los hablantes de bereber manifiestan divergencias que pueden interpretarse, teniendo en cuenta que son bilingües, “como el resultado de sus particulares experiencias vitales y culturales asociadas al uso de una lengua en contextos diferentes” (Fernández Smith *et al.* 2008: 89)²⁷⁷. Por otra parte, con respecto al grado de compatibilidad de los informantes²⁷⁸, se observa que la distribución según la *lengua materna* arroja porcentajes menores de convergencia que al analizar otras variables, como por ejemplo el *sexo*. Cabe destacar la rigurosidad estadística del trabajo sobre el léxico disponible de Melilla que, por otro lado, aporta datos significativos en relación con la producción de palabras, vocablos y en la línea del resto de investigaciones españolas en lo que respecta a la productividad y la cohesión de cada área temática²⁷⁹. Además, el estudio cuantitativo se acompaña de un exhaustivo análisis cualitativo donde se abordan aspectos como las categorías gramaticales, los procesos de formación de palabras, el origen de las palabras, la ortografía, los dialectalismos, los vulgarismos o las marcas comerciales.

Cabe destacar para la perspectiva que nos ocupa, que, en ninguno de estos dos trabajos se menciona la presencia de elementos lingüísticos a caballo entre las dos lenguas, o lo que es lo mismo, la esperable aparición de interferencias. Una razón que explique esta circunstancia puede ser que, como ya hemos dicho, la mayoría de hablantes no son bilingües, y la proporción de los que sí lo son es reducida en ambos casos²⁸⁰.

²⁷⁷ En este sentido, los autores encuentran coincidencias con los datos recabados en Ceuta. En ambas ciudades, los hablantes de lengua materna distinta del español proporcionan más vocablos en *Alimentos y bebidas* y *Los animales*, y menos en *Trabajos del campo y del jardín*. La comparación de vocablos se basa en los rangos que alcanzan los distintos centros y no en cifras absolutas, ya que no cuentan con el mismo número de informantes.

²⁷⁸ Los investigadores lo calculan de distintas formas: basándose en las 100 palabras más disponibles, en las 20 o en las que alcanzan un índice de disponibilidad superior a 0.02, a pesar de que rechazan esta última solución. Véase al respecto Fernández Smith *et al.* (2008: 100-104).

²⁷⁹ Así, por ejemplo, se documenta una producción total de 88 675 palabras y 7364 vocablos, siendo los centros más productivos *Los animales*, *Partes del cuerpo* y *Alimentos y Bebidas*, y los menos, *Trabajos del campo y del jardín*, *Iluminación y medios de airear un recinto* y *Objetos colocados en la mesa para la comida*. Los centros más compactos desde el punto de vista semántico son *Partes del cuerpo humano*, *La ropa* y *Los animales*.

²⁸⁰ A pesar de las palabras de Fernández Smith *et al.* (2006: 772) “no es difícil comprender que estos jóvenes bereberes empleen elementos o estructuras de su lengua nativa para hablar la que adquieren en segundo lugar”, no se dedica ningún apartado en el léxico de Melilla a esta cuestión. En el caso de Ceuta, la autora indica en el marco teórico que sería provechoso aplicar los resultados de la disponibilidad léxica

3.2.2.2. En EEUU

En segundo lugar, analizamos las obras que han estudiado la disponibilidad léxica del español en contacto con el inglés fuera del territorio peninsular. El primer trabajo que se realiza es el de Bailey Victory (1971) basado en el español y el inglés de Texas. El investigador, quien sigue de cerca los planteamientos de Gougenheim *et al.* (1964) y Dimitrijévic (1969), lleva a cabo su estudio entre 1970 y 1971 en las ciudades de Houston (Texas, EEUU) y Monterrey (Nuevo León, México). El propósito que le mueve es conocer el grado de disponibilidad léxica de dos grupos lingüísticos distintos, teniendo en cuenta el *sexo* y la *procedencia* y en este caso, además, la *nacionalidad*.

Los grupos que se establecen para la investigación son tres: un grupo de 33 estudiantes de Monterrey (monolingües en español), otro grupo de 33 estudiantes de Houston (monolingües en inglés) y un grupo de 33 estudiantes de Houston (bilingües español – inglés). Este último grupo completa la encuesta dos veces, primero en español y luego en inglés²⁸¹. Por tanto, la muestra comprende un total de 99 estudiantes de entre 16 y 18 años, de los cuales 53 son mujeres y 46 varones. En relación al *nivel sociocultural*, establecido a partir de la ocupación del cabeza de familia, el autor constata que este es superior en los estudiantes mejicanos, cuyos padres son profesionales o semiprofesionales. Los estudiantes americanos, en cambio, son de clase media o media baja de un área comercial y residencial de Houston²⁸².

Tras realizar un primer experimento con 40 personas de distintas edades, el autor considera que cinco minutos son excesivos y fija el tiempo en dos minutos por centro de interés; asimismo, desecha la idea de realizar las encuestas oralmente²⁸³. Los centros de interés elegidos por el autor son: *Animales/Animals*, *La ciudad/The city*, *Diversión/Entertainment*, *Familia/Family*, *Dios/God*, *Trabajos y profesiones/Jobs and*

a la enseñanza del español como segunda lengua para los niños musulmanes ya que “nos permitiría conocer (...) las interferencias de su lengua al emplear la lengua española” (Ayora 2006: 61).

²⁸¹ El autor afirma que el transcurso ideal entre las dos encuestas sería de dos a tres semanas, aunque reconoce no haber podido comprobar este hecho puesto que este aspecto lo supervisaba cada profesor (Bailey Victory 1971: 16, nota 1).

²⁸² Solo el 20% proseguirá con sus estudios e irá a la universidad, frente a la totalidad de estudiantes mejicanos que realizará estudios universitarios (Bailey Victory 1971: 12).

²⁸³ Concretamente, constata tres problemas: la inhibición, la repetición de palabras y la falta de concentración (Bailey Victory 1971: 14).

*professions, Medicina/Medicine, Música/Music, Espacio/Space y Guerra/War*²⁸⁴. El investigador elige estas áreas temáticas —tres de ellas son las mismas que usa Dimitrijévic (1969)— basándose en la temática cultural y en la “relatividad binacional”. Por un lado, su intención es buscar vocabulario que pueda resultar homogéneo en todos los grupos, y por otro, quiere tener en cuenta las diferencias que puedan surgir según la nacionalidad y la cultura. Las encuestas se completan en sesiones de una hora y se solicitan, al inicio, los datos personales de los informantes (nombre, edad, sexo, lenguas habladas en casa y clasificación ocupacional del cabeza de familia)²⁸⁵. Con relación al bilingüismo de los 33 estudiantes de Houston, el investigador cree que puede haber diferencias en una lengua u otra según el centro de interés y su experiencia lingüística anterior. También se pregunta hasta qué grado son totalmente bilingües estos estudiantes o si mostrarán preferencia por alguna de las dos lenguas; étnicamente, son ciudadanos americanos pero de origen mejicano que tienen un bilingüismo “individual”, puesto que se hallan inmersos en una ciudad preparada para hablantes de inglés únicamente.

Con respecto a los resultados, el estudioso constata que, en la producción de palabras, el grupo preponderante es el de los monolingües en inglés, seguido de (1) los bilingües que completaron la encuesta en español, (2) los bilingües que completaron la encuesta en inglés y (3) los monolingües en español. En la producción de vocablos, solo varían las dos primeras posiciones que, en este caso, son para los bilingües en español, seguidos por los monolingües en inglés. Tanto la producción general de palabras (12012) como la de vocablos (4838) es superior en lengua inglesa, aunque los resultados en español no distan demasiado (10368 para las palabras y 4443 para los vocablos). Por otro lado, las mujeres —superiores numéricamente en la muestra, aunque esta circunstancia ya se tiene en consideración a la hora de valorar los resultados— aventajan a los varones tanto en la cantidad de palabras como en la de vocablos, siendo

²⁸⁴ El autor prefiere enunciados cortos que permitan hacer asociaciones más libres, ya que —en su opinión—, algunos de los enunciados usados por los pioneros franceses son demasiado largos y obligan al informante a pensar en un área pequeña donde el vocabulario por fuerza debe ser muy limitado (Bailey Victory 1971: 19).

²⁸⁵ Con respecto a las instrucciones para completar las encuestas, estas son similares a las usadas en otras investigaciones, a saber, poner un ejemplo al principio que contestan entre todos, advertirles que el resultado de la prueba no afectará a su trabajo en la escuela y restar importancia a la ortografía de las palabras. Se cuenta con una hoja por cada centro de interés (Bailey Victory 1971: 17-18).

mayor la diferencia primera²⁸⁶. La producción léxica general arroja un promedio de 226.06 palabras y 93.75 vocablos por informante.

Algunas de las respuestas que aportan los informantes son tenidas en cuenta bajo un código concreto, aunque se eliminan de la edición final. Estas son las correspondientes a las falsas secuencias (respuestas no apropiadas para los centros, como los nombres propios²⁸⁷), las respuestas indescifrables y las palabras repetidas. A continuación, el investigador analiza la producción de palabras y vocablos en cada área, destacando el lugar que ocupa con relación al resto de centros de interés. De este modo, por ejemplo, el centro que evoca un mayor número de respuestas es el de *Animales*, seguido de *La ciudad* y *La guerra*; en cuanto a los vocablos, las primeras posiciones las ostentan *La guerra*, *Dios* y *Trabajos y profesiones*, respectivamente. También analiza la importancia de cada área temática concreta con respecto a la totalidad de producción del corpus y el porcentaje que representan en cada área los vocablos con ocho menciones o más, cuya lista se incluye por cada centro de interés²⁸⁸. Si se tiene en cuenta este último criterio de las 8 menciones, que constituye el 44.81% del corpus, entonces las posiciones principales en las dos lenguas en cuanto a palabras y vocablos son ocupadas por *La familia* y *Los animales*.

Con relación al estudio comparado de las dos lenguas, el autor aporta dos tablas (Bailey Victory 1971: 80) donde compara las palabras más disponibles en lengua castellana (esto es, las que alcanzan una índice de frecuencia superior o equivalente al 50%), con su correspondiente término correlativo en inglés. La conclusión a la que llega el investigador es que la convergencia léxica entre las dos lenguas es muy alta, ya que el 75% de las palabras más disponibles en castellano (en este caso, 18 sobre 24) aparecen

²⁸⁶ Las mujeres también superan a los varones en todos los centros de interés, en el centro *Dios* se aprecia la mayor distancia y, por el contrario, en *Guerra* la diferencia es menor, siempre a favor del colectivo femenino (Bailey Victory 1971: 99).

²⁸⁷ El criterio con respecto a los nombres propios parece un poco ambiguo puesto que en algunos centros se eliminan por completo, como, por ejemplo, los nombres de ciudades en *La ciudad*, pero, en cambio, sí se mantienen los nombres de ciudades en otros estímulos como *Guerra* o *Dios*. Con todo, parece muy oportuno el hecho de mantener toda esta información bajo un código numérico, para que así sea tenida en cuenta al menos en el recuento final de la producción (Bailey Victory 1971: 27).

²⁸⁸ Al final, además de esta lista de vocablos, también se indica cuántas personas los han mencionado según el *sexo* y el *grupo lingüístico*. Asimismo, se incluyen los totales de producción de palabras y vocablos según el centro de interés en cada idioma y la lista de palabras más disponibles, hasta un porcentaje mínimo de un 15% en cada lengua (Bailey Victory 1971: 67-80).

también como más disponibles en lengua inglesa²⁸⁹.

La única alusión a las consecuencias derivadas del bilingüismo es una tabla (Bailey Victory 1971: 96, tabla 15) donde el estudioso incluye un elenco variado de fenómenos, como analogía, disimilación, epéntesis, haplología o interferencias (como “cientistas” u “obscuridad”), detectadas en los listados de los informantes bilingües. Cabe destacar, finalmente, que es una lástima que el trabajo de este autor, pionero en el estudio de la disponibilidad en grupos monolingües y bilingües, obvie —como es lógico, por otro lado, dada la fecha del mismo— tantas posibilidades de análisis que se podrían realizar desde la perspectiva actual, como la covariación con los factores sociales, por citar solo un ejemplo²⁹⁰.

Un trabajo mucho más reciente y con un enfoque ya moderno (Moreno Fernández 2007) vuelve a abordar la cuestión del léxico disponible en EEUU. En este caso, se analiza la disponibilidad léxica de 48 adolescentes de origen hispano, fundamentalmente nacidos en Méjico o procedentes de allí, de primera y segunda generación con edades comprendidas entre los 15 y los 17 años que estudian en una escuela secundaria de la ciudad de Chicago. Los objetivos que el autor se plantea son, por tanto, conocer el léxico utilizado por estos jóvenes, observar la presencia tanto de palabras inglesas como de formas léxicas mezcladas en las listas de palabras y determinar los usos léxicos según unas variables concretas. Las pautas metodológicas de este breve estudio siguen de cerca las del Proyecto Panhispánico (dos minutos de tiempo por centro), aunque a la lista habitual de centros de interés se añaden: *Los colores*, *La vida*, *Salud y enfermedades* y *Árboles y plantas*. La muestra se post-estratifica según las siguientes variables: *sexo*, *curso* (senior o junior), *generación* de inmigración (primera o segunda), *nivel cultural de los padres* (bajo o medio), *edades* (15-17) y el *nivel* del curso de español en el que están (I, II o III)²⁹¹.

Después de homogeneizar los datos y resolver los problemas que conlleva el

²⁸⁹ Tomando como referencia el porcentaje del 50% de frecuencia antes mencionado. Recordemos que el cálculo de la disponibilidad solo se basa en las menciones, aún no se tiene en cuenta la posición de las palabras en los listados.

²⁹⁰ Otro aspecto que llama la atención, desde la actualidad es que el autor mantenga todas las variantes de los vocablos (singular, plural, derivadas...) en los listados finales.

²⁹¹ Todos los informantes tienen el español como lengua materna, aunque su dominio del castellano es muy variado, por eso se clasifican según el nivel que tienen en el “curso de español para nativos” que siguen en su escuela (Moreno Fernández 2007: 48).

trasvase entre las normas ortográficas de los dos idiomas en contacto²⁹², el investigador recopila el número de palabras por centro de interés y constata —como ya se viene repitiendo en otros trabajos— la supremacía cuantitativa de *Comidas y bebidas*, *La ciudad* y *Animales*²⁹³. Con respecto a la presencia de anglicismos, el promedio general en las listas es del 15%, aunque este porcentaje desciende hasta el 6.5% si se tienen en cuenta las veinte palabras más disponibles de cada centro. Dentro de este último margen, los estímulos que presentan más palabras en inglés son *Juegos y distracciones*, *La ropa* y *Calefacción, iluminación y medios de airear un recinto*. El autor lleva a cabo un análisis detallado de los anglicismos que aparecen en cada área léxica, y advierte la tendencia general a la aparición de dobles, en los que la palabra en español es más disponible que el sinónimo equivalente en inglés.

Con respecto a la covariación entre las variables sociales y los anglicismos presentes en los listados, tras realizar un estudio estadístico, el autor llega a la conclusión de que son los varones de segunda generación quienes propician más el uso del anglicismo, aunque estas variables no parecen perfilar grandes diferencias sino que el factor que parece ser determinante es el *nivel de conocimiento y uso* del español, que se revela como una variable decisiva a la hora de pronosticar la aparición o no de anglicismos. Esto es, al descender el conocimiento del español por parte del alumnado, aumenta el uso de palabras en lengua inglesa. En relación con la presencia de formas híbridas en los niveles morfológico, sintáctico o semántico, hay realmente pocas (un 0.4%) en el español de los estudiantes de Chicago.

El trabajo de Sancho (2006) también se circunscribe a EEUU, aunque en este caso las pruebas de disponibilidad se aplican en ocho estados (Nueva York, Nueva Jersey, Carolina del Norte, Georgia, Mississippi, Florida, Texas y California) a un total de 190 personas hispanas de distintas edades²⁹⁴. Los centros de interés tomados en

²⁹² Moreno Fernández sigue los criterios de edición clásicos de Samper Padilla (1998), reconociendo que están orientados hacia comunidades monolingües y hace hincapié en las dificultades que se plantean en las comunidades bilingües, donde a veces es incluso difícil saber en qué lengua está escribiendo el informante, inconveniente que también constatamos en este trabajo (*vid.* 5.4. y 7.1.)

²⁹³ Desde el punto de vista cualitativo, el autor aporta las tres palabras más disponibles de cada centro, donde ya pueden apreciarse diferencias etnográficas con respecto a las listas peninsulares. Por ejemplo, en *Comidas y bebidas*, “taco”, “soda” y “frijoles” son las tres palabras más disponibles (Moreno Fernández 2007: 52).

²⁹⁴ A pesar de que la fecha de publicación de este artículo es anterior al trabajo de Moreno Fernández (2007), lo glosamos a continuación ya que la investigación de Sancho (2006) es en realidad posterior a la

consideración son diez: *Comidas y bebidas, La cocina y sus utensilios, La ropa, Medios de transporte, Partes de la casa, Muebles de la casa, Profesiones y oficios, La enseñanza, El campo y La ciudad*. Con respecto a las variables, la autora estudia cuatro: *sexo, edad* (mayor o menor de veinte años), *idioma* en el que habla fuera de la familia (inglés, español o ambos) y el *tiempo* de permanencia en EEUU (menos de cinco años, entre cinco y diez y más de diez años). Partiendo de una noción amplia de anglicismo²⁹⁵, la investigadora analiza en cada uno de los campos léxicos los anglicismos recabados y, al mismo tiempo, los compara con los recogidos en los léxicos disponibles de otros lugares del continente americano (Puerto Rico y la República Dominicana) o en España (Cuenca y Zamora). Tras realizar este análisis, Sancho concluye que es en EEUU donde más se utilizan los anglicismos, seguido de la República Dominicana, Puerto Rico y, finalmente, España²⁹⁶; aunque llama la atención que en algunas áreas léxicas concretas (*La ropa y La ciudad*) hay más anglicismos en las provincias españolas que en América. La autora también observa que algunos anglicismos concretos se usan en todos los lugares hispánicos (por ejemplo, “microondas”, “bar”, “gasolinera” o “supermercado”).

Otro aspecto que analiza la investigadora es la presencia de americanismos en los listados y constata que los centros de interés que menos americanismos contienen son *Muebles de la casa, Partes de la casa, Profesiones y oficios y La ciudad*²⁹⁷. Con respecto a la relación entre las variables sociales y la cantidad de palabras aportada por cada informante, las mujeres y las personas mayores de 20 años obtienen los mejores resultados. En relación con las lenguas habladas, las diferencias son mínimas, aunque los que hablan español aportan un mayor número de respuestas. También muestran una superioridad cuantitativa las personas que llevan entre cinco y diez años viviendo en EEUU, seguidas del grupo que lleva menos de cinco años residiendo en el país. Por otro

de Moreno Fernández (2007), cuyos resultados se presentaron en 2004 (Moreno Fernández 2007: 43, nota 1).

²⁹⁵ Sancho (2006: 23) entiende así tanto los términos de etimología inglesa como los que proceden de otras lenguas pero que han entrado en el español a través del inglés. A su vez los divide en anglicismos crudos, adaptación de palabras, préstamos y calcos.

²⁹⁶ La autora detalla el mismo orden con respecto a la presencia de anglicismos crudos (es decir, los que provienen de la influencia directa del inglés y conservan la grafía inglesa). Cabe decir que se encuentran más anglicismos en las listas de Zamora que en las de Cuenca, aunque la diferencia es mínima según la autora (Sancho 2006: 28).

²⁹⁷ Sancho (2006: 28) considera americanismos las palabras originarias de una lengua indígena y aquellas cuyo uso o significado, sin tener en cuenta su origen, es distinto al del español peninsular (como *agujetas, birria, bombilla* o *burrito*).

lado, la autora no observa diferencias cualitativas importantes según cada grupo social sino que el léxico usado es bastante semejante²⁹⁸. En cuanto al uso de anglicismos, las diferencias son poco significativas teniendo en cuenta la *lengua usada fuera de la familia*, y, en cambio, se constata un mayor uso de los mismos entre los informantes que llevan menos tiempo viviendo en EEUU.

Recientemente, Verdeses-Mirabal (2012) analiza el léxico disponible en castellano de un grupo de estudiantes hispanos de la ciudad de Redwood City, en California. Su estudio, que secunda los acuerdos del Proyecto Panhispánico en relación con el tiempo destinado a completar el cuestionario y los centros de interés, se basa en los datos que proporcionan 518 informantes de duodécimo —último grado de la enseñanza general—. Estos se estratifican según el *sexo* (hombres o mujeres), la *lengua habitual* (español o inglés), el *nivel sociocultural* (bajo, medio bajo, medio alto o alto), la *generación de inmigrantes* (primera, segunda o tercera), el *lugar de nacimiento* (en EEUU o en un país hispano) y el *nivel de español* (bajo, medio bajo, medio alto o alto). El investigador ofrece los resultados generales (promedios de palabras, vocablos, cohesión y densidad) y los compara con los obtenidos en otros territorios hispánicos. También compara los promedios obtenidos según cada factor social y estudia el peso que dichos factores ejercen en la producción léxica mediante una serie de análisis estadísticos (análisis factorial simple o de varianza, análisis de regresión lineal múltiple, coeficiente de correlación múltiple...) realizados con el programa SPSS. El autor concluye al respecto que, mientras la variable *nivel sociocultural* no es discriminatoria, la influencia mayor la ejercen las variables *sexo*, *lengua habitual*, *lugar de nacimiento*, *nivel de español* y *generación de inmigrantes*. Con respecto a los trasvases lingüísticos entre las dos lenguas en contacto, el autor afirma que el léxico recabado refleja un español patrimonial y que la incidencia de préstamos lingüísticos en los vocablos recogidos es mínima²⁹⁹.

Para acabar, hemos de mencionar que, en la ciudad texana de San Antonio, y también fuera de los límites del Proyecto, se ha iniciado el estudio de la disponibilidad

²⁹⁸ Únicamente encuentra algunas pequeñas disimilitudes entre hombres y mujeres en el campo léxico de *La ropa* y el de *El campo* (Sancho 2006: 29).

²⁹⁹ Esta conclusión es llamativa puesto que en las instrucciones ofrecidas por escrito antes de cada centro de interés se especifica al informante que, si no recuerda la palabra en castellano, puede escribirla en inglés. También Gómez Molina y Gómez Devís (2004: 49, nota 54) ofrecen esta posibilidad.

léxica y ya se han recopilado los materiales, aunque no se han publicado todavía los resultados (Hernández Cabrera y Samper Padilla, en preparación)³⁰⁰.

3.2.2.3. En el norte de Europa

En último lugar, nos referimos al trabajo de Bernal (2004-2005), quien analiza cuantitativamente la disponibilidad léxica de hablantes bilingües de finés y español³⁰¹. Se trata de un estudio del bilingüismo individual de un grupo de hablantes que viven en Finlandia y, por tanto, están inmersos en una comunidad bilingüe donde el finés y el sueco son lenguas oficiales. Estas circunstancias tan peculiares suponen una nueva perspectiva de estudio de la disponibilidad, puesto que en las referidas hasta ahora se trataba de informantes que dominaban dos lenguas como consecuencia de una situación de bilingüismo social.

Tras un repaso a la definición y a las distintas tipologías de bilingüismo, la autora especifica las características de su muestra y las variables sociales que tiene en cuenta. Para establecer el corpus, recoge los datos de 20 informantes de edades comprendidas entre los 7 y los 18 años. La prueba se aplica dos veces (primero en español y luego en finés) en un intervalo de una a tres semanas y se completa en torno a siete centros de interés: 01. *Partes del cuerpo humano*, 02. *La ropa*, 05. *Las comidas bebidas*, 08. *La escuela (muebles y materiales)*, 10. *La ciudad*, 14. *Los animales* y 15. *Juegos y distracciones*³⁰². Según la autora, la mayoría de jóvenes que participa en su estudio puede clasificarse dentro del bilingüismo infantil en que cada uno de los progenitores tiene una lengua nativa que utiliza para comunicarse con el niño y además, una de las lenguas usadas en casa es también la lengua dominante en la comunidad³⁰³. Las variables sociales tenidas en cuenta son el *sexo*, los *estudios de español*, el *tiempo vivido en un país hispanohablante*, la *edad* y la *procedencia* de los padres (Bernal 2004-2005: 15-18).

³⁰⁰ Información proporcionada por los propios autores y también a través de Samper Padilla y Samper Hernández (2006: 83).

³⁰¹ El artículo es una ampliación del primer estudio que realizó la autora como trabajo final de licenciatura. Partiendo del mismo marco teórico, en aquella ocasión (Bernal 2003) su único informante era una niña bilingüe finés-español. Del mismo modo que en el presente artículo, la autora también constató la preponderancia del finés sobre el español.

³⁰² La investigadora conserva la numeración tradicional de estas áreas temáticas (Bernal 2004-2005: 15).

³⁰³ Concretamente, atienden a esta clasificación de bilingüismo 14 de los 20 informantes. Bernal sigue la clasificación de bilingüismo infantil de Romaine (1995).

Bernal recaba un total de 1374 palabras en español y 1887 en finés. En general, tanto la cantidad de palabras como la de vocablos y el promedio de respuestas por informante son superiores en finés, tal como resulta lógico³⁰⁴. Tras informatizar los datos y analizar la incidencia en ellos de las distintas variables, la autora constata que los mejores porcentajes con relación a la lengua española los alcanzan los niños o adolescentes que asisten regularmente a clases de castellano y que, en algún momento de su vida, han residido en un país hispanohablante. La *edad* también se perfila como una de las variables más importantes, puesto que se comprueba, analizando las distintas franjas de edades, que los informantes aumentan gradualmente su léxico disponible en las dos lenguas a medida que crecen. Por otro lado, los de menor edad registran más palabras y vocablos en finés que en español, mientras que entre los mayores las diferencias son menores.

Con respecto a otros factores, como el *sexo* o la *procedencia de los progenitores*, la investigadora no llega a ninguna conclusión relevante y deja constancia de que otras variables, como la *edad*, resultan claramente más decisivas³⁰⁵. Para acabar, la investigadora incluye el listado de las cinco unidades más disponibles en cada idioma por cada centro de interés.

3.3. ESTUDIOS SOBRE EL CATALÁN (EN CONTACTO CON EL ESPAÑOL)

Los estudios que se han realizado, hasta la fecha, sobre la lengua catalana se hallan circunscritos a la provincia de Valencia, como ya hemos comentado (*vid.* 3.2.1.3.). El primer trabajo que se realiza pretende establecer el nivel de disponibilidad léxica de la población valencianohablante de la ciudad de Valencia y su área metropolitana (Gómez Devís 1995). También intenta determinar el grado de covariación entre la competencia léxica y distintas variables sociológicas, como el *sexo*, la *edad*, el *nivel sociocultural* y la *lengua materna*. Para llevarlo a cabo, se elige a 139 informantes estratificados mediante un muestreo por cuotas. Este estudio inédito mantiene algunas de las pautas metodológicas del Proyecto Panhispánico, como los dieciséis centros de

³⁰⁴ La autora también lleva a cabo otros análisis como el estudio de la densidad, el índice de cohesión o los índices de disponibilidad, aparición y frecuencia acumulada. Llega a algunas conclusiones, ya reiteradas en el mundo de la disponibilidad, como el hecho de que el centro de interés que mayor cohesión alcanza en ambas lenguas es *Partes del cuerpo humano* (Bernal 2004-2005: 23-25).

³⁰⁵ Otro problema con el que se enfrenta para poder llegar a conclusiones válidas es el hecho de que la muestra que usa no es proporcionada, como ella misma reconoce. Del total de 20 informantes, solo 6 son varones y la mayoría (14) tiene una madre hispanohablante y un padre finés, con lo que es difícil valorar las diferencias (Bernal 2005: 25-31).

interés y la limitación temporal, pero se aleja de otras puesto que la *edad*, como acabamos de ver, se contempla como variable³⁰⁶. En el análisis se aportan las palabras que han sido mencionadas un mínimo de diez veces y se muestran los promedios según las variables³⁰⁷. Además, se analizan las interferencias que superan el 0.5 de frecuencia relativa en cada centro de interés según las variables sociológicas³⁰⁸. Tras este análisis, se llega a la conclusión de que hay centros que propician mucho la aparición de interferencias —como *Objetos colocados en la mesa para la cocina* y *Calefacción e iluminación*— y en otros, en cambio, se actualizan muchísimas menos, como *Partes del cuerpo humano*. El prototipo sociológico que más favorece la aparición de este tipo de fenómenos es mujer, mayor de 55 años, de nivel sociocultural bajo y bilingüe³⁰⁹. En conclusión, la autora opina que el nivel de penetración de otras lenguas en la disponibilidad léxica del valenciano no es muy elevado, aunque en algunos campos la presencia sí sea relevante.

En segundo lugar, hallamos la investigación de Llopis (2009)³¹⁰, que se centra en el análisis de la disponibilidad léxica en valenciano de los preuniversitarios de la provincia de Valencia, estudio que, recordamos, ya había sido acometido con anterioridad con respecto al castellano por Gómez Molina y Gómez Devís (2004). En esta ocasión, Llopis (2009) cuenta con una muestra muy parecida numéricamente, integrada por 464 informantes procedentes de 24 centros educativos. Estos informantes se distribuyen según las siguientes variables: *sexo*, *tipo de centro*, *nivel sociocultural*, *lugar de residencia*³¹¹, *lengua habitual* y *lengua materna*³¹². Llopis (2009) incorpora dos variables más, que resultan, en parte, novedosas. La primera es el *programa de educación bilingüe*, que puede ser el PIP (*Programa d'Incorporació Progressiva*),

³⁰⁶ Se divide en menores de 25 años, entre 26 y 55 años y mayores de 55. Otros estudios sobre disponibilidad léxica se han realizado con otras franjas de edad, por ejemplo, Borrego (2002a).

³⁰⁷ Estos favorecen de forma constante a los varones y a los informantes del nivel sociocultural más alto, y de forma bastante menos evidente, a la franja de edad media y a los bilingües o valencianohablantes.

³⁰⁸ No se contemplan, en cambio, las que alcanzan este índice, pero que solo aparecen una vez.

³⁰⁹ Aunque la investigadora advierte de que este perfil solo se da en algunos centros de interés y precisamente en otros, que contienen un número superior de interferencias, como *Objetos colocados en la mesa para la comida*, el patrón cambia (Gómez Devís 1995: 96).

³¹⁰ Puede hallarse un resumen de los puntos más importantes de la misma en Llopis y Gómez Devís (2010).

³¹¹ La manera de determinar el *nivel sociocultural* y los subgrupos con respecto a la variable *residencia*, basada en el alumno, son los mismos que usan Gómez Molina y Gómez Devís (2004), *vid.* 3.2.1.3.

³¹² Estas dos variables se subdividen en hablantes monolingües (que en realidad son bilingües pasivos) y bilingües activos. Los informantes que se declaran bilingües activos son 269 y los que tienen el valenciano como lengua habitual, 253. Es decir, el 58% y el 55% de la muestra. Es importante aclarar que en el trabajo de Llopis (2009) solo se tienen en cuenta las comarcas valencianohablantes de la provincia.

donde la lengua base es el castellano y que es seguido por el 75% de informantes, o el PEV (*Programa d'Educació en Valencià*), donde la mayor parte de áreas no lingüísticas se dan en valenciano). La segunda variable es la de *comunidad de habla*³¹³, establecida, en este caso, en tres grupos a partir del grado de castellanización de las comarcas: bilingüismo activo con predominio del castellano, bilingüismo activo con uso equilibrado y bilingüismo activo con predominio del valenciano³¹⁴.

En el análisis, se aportan los totales de palabras y vocablos, sus promedios, la densidad léxica y los índices de cohesión, y se comparan estos datos con los de otras sintopías del mundo hispánico, especialmente con la investigación sobre el castellano de Valencia, para concluir que las tendencias que se observan son similares.

Para poder determinar la covariación de las áreas temáticas con las variables sociales elegidas, se realizan diversos análisis estadísticos que muestran que los factores más influyentes en el léxico disponible catalán de los valencianos son el *tipo de centro*, la *comunidad de habla* y la *residencia* del alumno, ofreciendo mejores resultados el alumnado de colegios privados, de zonas bilingües con predominio del valenciano y semiurbanas³¹⁵. Otros factores con un valor discriminante menor, y parcial según los datos estadísticos, son: la *lengua materna* —donde los bilingües tienen una mayor productividad—, el *sexo* —a favor de los hombres— y el *nivel sociocultural* —que parece ofrecer mejores promedios a los estratos medios—, a pesar de que la mayoría de estas diferencias no son significativas, como ya hemos afirmado³¹⁶. Por lo que respecta a otras variables contempladas, como el *programa de educación* y la *lengua habitual*, estas muestran una influencia muy escasa en la covariación de los datos. Estos análisis se complementan con el estudio del grado de convergencia conceptual de las cincuenta primeras palabras según las distintas variables, que se realiza con la finalidad de

³¹³ Concepto que sirve de unidad de análisis de un contexto social determinado por unas normas lingüísticas de comportamiento social. Véase al respecto Martínez Olmos (2008).

³¹⁴ El primer caso es el que atañe a la ciudad de Valencia, el uso proporcional de las dos lenguas se da en *la Horta*, *el Camp de Morvedre* y *el Camp de Túria*, y el predominio del valenciano se observa en *Les Riberes*, *la Costera* y *la Vall d'Albaida*. Para establecer esta división, el investigador se basa en los datos de *l'Acadèmia Valenciana de la Llengua*.

³¹⁵ Según las conclusiones generales el *tipo de centro* contribuye a la covariación en quince centros de interés, la *comunidad de habla* en catorce y el *lugar de residencia* en once (Llopis 2009: 317-320).

³¹⁶ La *lengua materna* contribuye a la explicación de cinco áreas y el *sexo* y el *nivel sociocultural* únicamente de cuatro. Llama la atención, especialmente, los resultados con respecto a esta última variable, ya que en la mayoría de trabajos de disponibilidad sí que resulta fundamental, como el mismo autor observa acertadamente (Llopis 2009: 319).

observar qué factores sociales ofrecen una mayor homogeneidad léxica. Con respecto a este análisis, las variables analizadas se ordenan de la siguiente forma: *nivel sociocultural*, *sexo*, *comunidad de habla*, *tipo de centro*, *residencia* y *lengua habitual*³¹⁷. En relación con el análisis cualitativo, cabe destacar que se observa un menor número de transferencias del castellano en las comunidades de habla menos castellanizadas³¹⁸.

Otro punto importante del trabajo de Llopis (2009) es el estudio de las transferencias como resultado del contacto entre el valenciano y el castellano. En primer lugar, el autor repasa, desde una perspectiva epistemológica, los diferentes resultados que pueden darse tras el contacto lingüístico continuado de dos lenguas, esto es, las interferencias, los préstamos, los calcos y los compuestos híbridos, y, por otro lado, el cambio o la sustitución de código. Desde estos presupuestos teóricos, el autor clasifica los fenómenos que encuentra en su corpus valenciano de la siguiente forma³¹⁹:

- a) interferencias (palabras agramaticales, inexistentes en ninguna de las lenguas y con una frecuencia inferior al 6%),
- b) préstamos en vías de integración (con una frecuencia entre el 6 y el 15%),
- c) préstamos integrados (con un índice de frecuencia superior al 15%),
- d) compuestos híbridos (sintagmas donde se combinan palabras de las dos lenguas), y
- e) sustituciones de código (palabras normativas en la otra lengua, actualizadas en muy pocas ocasiones).

El mismo autor reconoce que el test de disponibilidad léxica es limitado y que sería necesario realizar un estudio más exhaustivo para ver el grado de integración lingüística y social de cada palabra, para poder diferenciar mejor las interferencias de los préstamos.

³¹⁷ La primera, el *nivel sociocultural*, es la que presenta el mayor índice de convergencia, concretamente, un 84.76%. No se analizan todas las variables, sino únicamente las que hemos mencionado.

³¹⁸ Estas son *Les Riberes*, *la Costera* y *la Vall d'Albaida*, que ofrecen también —como ya hemos dicho— mejores promedios. Por tanto, se cumple la hipótesis inicial del autor que esperaba hallar en las comunidades de habla con menor competencia comunicativa bilingüe, un mayor número de transferencias del castellano.

³¹⁹ En Llopis (2005) puede hallarse el análisis de estos fenómenos en tres centros de interés con una muestra de menor tamaño. Véase también Llopis (2008) para la clasificación de estos mismos fenómenos lingüísticos en otros tres centros de interés.

Tras analizar los ejemplos concretos para cada centro de interés de cada uno de los fenómenos de contacto lingüístico mencionados³²⁰, se llega a la conclusión de que su presencia en el corpus es bastante notable, aunque distinta según el área temática. Por ejemplo, en algunas áreas estos fenómenos llegan a suponer una tercera o cuarta parte del total de vocablos (como sucede en *Animales*, *Partes de la casa*, *Los muebles*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios* y *Colores*), mientras que otras contienen muchos menos (como *Profesiones y oficios* o *Alimentos y bebidas*). Se corrobora también que, en esta comunidad en concreto, se dan más condicionantes para que estos fenómenos se den en sentido castellano > valenciano y no en sentido inverso.

En trabajos posteriores, el autor vuelve a analizar las transferencias en su corpus desde distintas perspectivas. En Gómez Devís y Llopis (2010b), Llopis (2010) y Llopis (2011), se analiza cuantitativa y cualitativamente la convergencia de las cincuenta primeras palabras según la variable *comunidad de habla*. También se analizan los fenómenos de contacto lingüístico concretos, clasificados según Llopis (2009), que aparecen en estas posiciones y se concluye que la comunidad de habla que presenta un mayor número de estos fenómenos es aquella en la que existe un bilingüismo activo con un uso equilibrado de las dos lenguas³²¹. En Llopis y López Río (2013) se analizan también las transferencias en las primeras cincuenta palabras según la variable *lengua habitual* y se observa cómo los monolingües actualizan más transferencias en sus listados que los bilingües³²².

Otro estudio muy interesante es el que realizan Gómez Devís y Llopis (2010a) al comparar sus corpus respectivos y analizar el grado de convergencia semántica entre los léxicos disponibles de las dos lenguas de la provincia de Valencia. Tras ver las coincidencias entre las treinta palabras más disponibles en castellano y en catalán, repasan la compatibilidad de cada centro de interés y la relacionan con el índice de cohesión respectivo. Llegan a la conclusión de que el grado de convergencia entre las

³²⁰ También se aportan los dobles más importantes con sus índices de disponibilidad respectivos.

³²¹ Por otro lado, la zona de bilingüismo con predominio del catalán es la que obtiene menor presencia de estos fenómenos, como es lógico.

³²² Recuérdese que, en realidad, los informantes monolingües son bilingües pasivos.

lenguas es realmente elevado (concretamente, un 82.3%)³²³, aunque se observan diferencias entre las distintas áreas. Por otro lado, y desde una perspectiva ya piscolingüística, Gómez Devís (2010) analiza las conexiones entre las diversas unidades léxicas, centrándose en las transferencias halladas en la muestra valenciana y viendo la relación que estas tienen con las formas normativas correspondientes. Para llevar a cabo este análisis, confecciona varios “grafos”³²⁴ que detallan las redes semánticas concretas que reflejan las asociaciones que han realizado los encuestados. Al comparar las redes de los informantes según su *lengua de uso habitual* (castellano o valenciano) se observa que tanto la organización de las transferencias como las interrelaciones de estas con las soluciones normativas son bastante diferentes; las redes de los informantes bilingües tienen mayor riqueza.

Los mismos autores han trabajado también en la línea de la aplicación didáctica de los resultados de la disponibilidad léxica, con aportaciones concretas muy interesantes (Gómez Devís 2005, Gómez Devís y Llopis 2011 y Gómez Devís y Llopis 2012)³²⁵.

3.4. CONCLUSIONES PARCIALES

Como colofón final a este apartado, repasamos, en primer lugar, las distintas perspectivas que se han aplicado en el estudio de la disponibilidad en comunidades donde coexisten dos sistemas lingüísticos. A continuación, enumeramos los trabajos que siguen el mismo enfoque comparativo que aplicamos en este trabajo, precisando las diferencias metodológicas existentes. Para finalizar, revisamos las dos ópticas actuales de análisis de la disponibilidad léxica en comunidades bilingües, ya sea de forma comparada o analizando solo una de las dos lenguas en contacto: el estudio de la incidencia de algunas variables propias de estas comunidades y el análisis de los elementos que se transfieren de una lengua a otra.

3.4.1. Perspectivas aplicadas

Como hemos visto, son muchas —y muy distintas a su vez— las perspectivas desde las cuales se ha abordado el estudio de la disponibilidad léxica en hablantes o

³²³ Los profesores valencianos reconocen que el porcentaje aún sería mayor si no fuera por la aparición de dobles en el léxico disponible valenciano (Gómez Devís y Llopis 2010a).

³²⁴ Para establecerlos, la autora utiliza el programa computacional Dispógrafo, creado por un grupo de investigadores de la Universidad de Chile (Echeverría *et al.* 2008).

³²⁵ También es digno de mención el análisis que realiza sobre la variable *sexo* otro miembro del equipo valenciano (Gómez Molina 2006).

comunidades bilingües. El enfoque general y mayoritario —que seguimos también en este estudio— es el que inicia Mackey (1971a, 1976) y sigue Knock (1978), es decir, el estudio del léxico disponible en un territorio donde la mayoría de sus hablantes son bilingües y donde las dos lenguas tienen un estatus similar, con las diferencias de cada caso particular. Esta es la perspectiva desde la que se acomete en la actualidad el estudio de la disponibilidad en las regiones bilingües de España, donde el español es cooficial con la lengua propia de cada comunidad. También es esta la óptica desde la que se estudia la disponibilidad de los hablantes de español en algunas ciudades de EEUU (Moreno Fernández 2007, Sancho 2006, Verdeses-Mirabal 2012), aunque, en este caso, las lenguas en contacto no tienen el mismo estatus político. Por el contrario, no sería el caso de Bernal (2004-2005) puesto que sus informantes responden a un bilingüismo individual (español/finlandés) que no se corresponde con el que se da en su país de forma oficial (sueco/finlandés).

Otra opción distinta, y de carácter minoritario, es la que observamos en el estudio de O'Huallachain (1966) en Irlanda, donde el autor deja claro que sus hablantes —tanto los irlandeses como los ingleses— son monolingües y, por ende, solo dominan un idioma, a pesar de que viven en un territorio donde las dos lenguas se utilizan y gozan de oficialidad.

La tercera perspectiva, que cuenta ya con varios estudios, es la que engloba en una misma investigación la comparación entre hablantes bilingües y monolingües que comparten una de las dos lenguas y conviven en un mismo lugar. Sigue este enfoque, en primer lugar, Bailey Victory (1971) con su comparación entre estudiantes monolingües de lengua materna inglesa y estudiantes bilingües que emplean con similar competencia el inglés y el español, ubicados todos ellos en Houston. En segundo lugar, Dimitrijevic (1981) también compara en su trabajo sobre Novi Sad (Yugoslavia) los informantes monolingües de serbocroata con los bilingües en húngaro-serbocroata. Situación semejante es la que se da en Navarra, donde también conviven hablantes monolingües con hablantes bilingües (Taberner 2008b). También es esta la situación que se describe, *grosso modo*, en las ciudades españolas de Ceuta y Melilla, donde existe diglosia a favor del español y donde únicamente son bilingües los hablantes de mazigia y dariya, puesto que los castellano hablantes no tienen la necesidad de aprender estas lenguas (Ayora 2006, Fernández Smith *et al.* 2008).

Tampoco es infrecuente —además de comparar estos dos colectivos de hablantes entre sí— extender este análisis contrastivo a otro grupo de hablantes monolingües situado en otro territorio. Es lo que hace Bailey Victory (1971) al cotejar los datos de Houston de los dos colectivos mencionados con los de un grupo de estudiantes monolingües de Monterrey (México) y, también desde esta perspectiva, Dimitrijevic (1981) da cuenta de algunas de las diferencias entre el léxico disponible en Yugoslavia y el que había recogido con anterioridad en Escocia (Dimitrijevic 1969), aunque, en este caso, se trata de dos lenguas distintas. Este investigador constata, al respecto, diferencias debidas a la distancia cultural y social existente entre los dos países. Se abre así otra vía de análisis, también secundada desde el inicio, que consiste en establecer las diferencias culturales que se dan entre dos territorios que utilizan una misma lengua pero que se hallan separados geográficamente. Las investigaciones que toman esta perspectiva normalmente comparan dos territorios distintos donde la misma lengua se da en una situación de bilingüismo social o en un estado de monolingüismo. Es el caso de Mackey (1971a, 1976) al cotejar el francés de Francia con el de Canadá y el de Knock (1978), quien compara el francés de Camerún con el de Francia³²⁶. Esta óptica es muy interesante porque, además de las diferencias culturales, permite observar el grado de penetración de la segunda lengua en la primera y establecer comparaciones con un territorio donde la lengua no está expuesta a esta influencia. Por ejemplo, Mackey (1976) compara los anglicismos presentes en el francés de Canadá y observa si estos aparecen en las listas de Francia. También el mismo análisis está presente en la comparación que lleva a cabo Sancho (2006) entre los anglicismos detectados en EEUU y los índices que estos adquieren en las listas de algunas regiones españolas.

El último enfoque sobre lenguas en contacto aplicado, y que difiere de los anteriores, sería la comparación de dos lenguas distintas con hablantes monolingües en una zona fronteriza de contacto entre ambas (Prado *et al.* 2009).

3.4.2. El enfoque comparado de dos lenguas

La perspectiva que usamos en esta investigación, realizada con hablantes bilingües, y que tiene como objetivo la comparación entre las dos lenguas que conviven en un mismo territorio, cuenta con varios antecedentes. En primer lugar, Mackey

³²⁶ En sus cotejos, los dos autores se sirven de los datos de Gougenheim *et al.* (1964) sobre el francés de Francia.

(1971a, 1976) aplica la encuesta relativa a cinco centros de interés en dos idiomas distintos (inglés y francés) a unos mismos informantes bilingües con un tiempo de reacción de quince minutos y dejando unos meses entre una prueba y otra. Asimismo, también en Bailey Victory (1971) son los mismos estudiantes los que completan dos veces la encuesta, que tiene diez centros de interés y un margen de dos minutos, tras aproximadamente dos o tres semanas de intervalo entre una y otra encuesta³²⁷. Knock (1978) también sigue esta perspectiva y compara los resultados obtenidos al aplicar la encuesta dos veces a los mismos hablantes, con diez centros de interés y quince minutos por centro, y dejando dos semanas de intervalo entre las dos pruebas. También en Bernal (2004-2005) son los mismos informantes quienes completan la encuesta, aunque en una situación especial de bilingüismo que no puede compararse con las demás, como ya hemos indicado.

También se ha dado otra solución metodológica en los estudios que tratan de forma comparada dos lenguas que coexisten en una situación de bilingüismo social. De este modo, en el trabajo pionero de Azurmendi (1983), los informantes se dividen en el momento de aplicar la encuesta y la mitad de la clase la completa en cada lengua. En su caso contaban con diez minutos para quince estímulos. Dentro del Proyecto Panhispánico, la investigación que se lleva a cabo en Gibraltar desde la Universidad de Cádiz (Escoriza 2007, Fernández Smith *et al.* 2012) también opta por esta solución, y por tanto, la comparación entre las dos lenguas se realiza a partir de los datos que aportan distintos informantes. También dentro de esta perspectiva tiene lugar el análisis comparado que realizan de sus corpus Gómez Devís y Llopis (Gómez Devís y Llopis 2010a), sobre la base de sus investigaciones realizadas acerca del alumnado preuniversitario de Valencia en castellano y catalán, respectivamente³²⁸.

Es evidente que necesitaríamos contar con más investigaciones de este tipo para que pudieran contrastarse las diferentes opciones metodológicas y se determinara si es relevante o no, por ejemplo, el hecho de que sean los mismos informantes quienes completen la encuesta por duplicado. Por otro lado, los pocos estudios realizados desde

³²⁷Recuérdese lo dicho en la nota 281.

³²⁸ En el trabajo de Dimitrijevic (1981) no es posible saber si son los mismos informantes los que completan la encuesta dos veces, puesto que el autor no lo especifica. Por otro lado, en Galicia se ha iniciado ya el análisis de la disponibilidad léxica en gallego, lo que permitirá un estudio comparado con los datos de López Meirama (2008) sobre el castellano en Galicia.

esta perspectiva comparada siguen distintas pautas metodológicas en cuanto a número de centros de interés, tiempo de reacción, con lo que los resultados no resultan comparables; constituyen dos excepciones las investigaciones en Gibraltar y en Valencia, por enmarcarse dentro del Proyecto Panhispánico. Merece la pena destacar que, en todos los estudios en que son los mismos informantes los que completan la encuesta, el tiempo transcurrido entre las dos pruebas es, como mínimo, de dos semanas³²⁹, el mismo intervalo que utilizamos en este trabajo. Intuitivamente parece este el transcurso necesario para que el informante haya olvidado las respuestas anteriores y deba realizar de nuevo las asociaciones pertinentes.

3.4.3. Variables analizadas en comunidades bilingües

A la nómina de variables estudiadas tradicionalmente en los trabajos de disponibilidad, se suelen añadir otras en los estudios sobre territorios bilingües³³⁰. La más frecuente, sin duda, es la *lengua materna*, hasta la fecha incluida en la mayor parte de trabajos de este tipo (Dimitrijevic 1981, Azurmendi 1983, Gómez Devís 1995, Etxebarria 1996a, Blas y Casanova 2003b, Serrano Zapata 2003b y 2004, Ayora 2006, Escoriza 2007, Areta 2009 y Fernández Smith *et al.* 2008). Esta variable ha mostrado hasta la fecha, que, como resulta obvio, alcanzan mejores resultados los informantes que completan la encuesta en su propia lengua materna, ya sea el castellano (Blas y Casanova 2003b³³¹, Ayora 2006, Fernández Smith *et al.* 2008, Areta 2009) o el catalán (Gómez Devís 1995, Llopis 2009). La única excepción sería nuestro trabajo sobre la ciudad de Lleida (Serrano Zapata 2003b y 2004) donde los catalanohablantes obtenían mejores promedios en castellano, conclusión que rebatimos en la presente investigación, basada en una muestra considerablemente mayor (*vid.* 6.2.5.)³³².

Otra variable frecuentemente analizada es la *lengua habitual* que —con la excepción de Areta (2009) y Verdeses-Mirabal (2012)— no parece tener una incidencia

³²⁹ En el trabajo de Bernal (2004-2005), concretamente, era de una a tres semanas (*vid.* 3.2.2.3)

³³⁰ Véase una perspectiva de los condicionantes añadidos en estas comunidades en Samper, Bellón y Samper Hernández (2003: 87-89) y Samper Padilla y Samper Hernández (2006: 79-84).

³³¹ Los autores afirman que esta variable incide leve pero significativamente a favor del alumnado castellanohablante o que considera ambas lenguas como maternas. En Blas y Casanova (2001-2003) se examina la incidencia de esta variable junto con otras. También Dimitrijevic (1981) afirma que informantes bilingües y los monolingües no difieren demasiado con respecto a las palabras que alcanzan mayor grado de disponibilidad.

³³² Como ya indicamos en aquella ocasión, el tamaño de la muestra, u otra variable con mayor incidencia, podía estar distorsionando los resultados (Serrano Zapata 2004: 160-161).

importante en la productividad léxica, al menos comparada con otras variables (Gómez Molina y Gómez Devís 2004, Sancho 2006, López Meirama 2008 y Llopis 2009). También ha surgido interés por saber si en las comunidades bilingües en las que hay distintos modelos de enseñanza según la lengua predominante, este influye en el caudal léxico del alumno. En este sentido, mientras Etxebarria (1996b), Blas y Casanova (2003a) y Llopis (2009) constatan que las diferencias según este factor son escasas; Taberero (2008b) y Areta (2009) comprueban estadísticamente cómo el alumnado que sigue el modelo D, con el euskera como lengua vehicular, obtiene peores resultados³³³.

Una variable que ha mostrado tener mucha relevancia en las comunidades bilingües y de reciente incorporación es la *comunidad de habla*. En el trabajo de Martínez Olmos (2008) es el factor que presenta una mayor covariación con los resultados, a favor de las comunidades de habla más castellanizadas, y en Llopis (2009) se observa cómo las comunidades de habla bilingües con predominio del valenciano alcanzan promedios más favorables. De otra parte, las investigaciones que trabajan con población inmigrada en sus orígenes incorporan otras variables de orden distinto: si siguen o no un *curso de español* y el *nivel* en el que están (Bernal 2004-2005, Moreno Fernández 2007, Verdeses-Mirabal 2012), el *tiempo* que llevan viviendo en el país (Sancho 2006) o el que han vivido en un país de habla hispana (Bernal 2004-2005), el *lugar de nacimiento* (Verdeses-Mirabal 2012), la *generación* de inmigración (Moreno Fernández 2007, Verdeses Mirabal 2012), o la *procedencia* de los padres (Bernal 2004-2005). Estas investigaciones han constatado que los informantes que siguen cursos de español de forma regular ostentan resultados superiores (Bernal 2004-2005) y, a su vez, cuanto mayor es el nivel de español adquirido, menos préstamos se toman de la otra lengua (Moreno Fernández 2007). El tiempo residido en el país también parece ser relevante a la hora de incorporar palabras de la segunda lengua (Moreno Fernández 2007, Sancho 2006)³³⁴.

Por último, Carcedo (2001) y Areta (2009) incorporan la variable *zona geográfica*. Mientras que en el caso de Asturias esta variable no ejerce una

³³³ La mayoría de alumnos que sigue este modelo tiene el euskera como lengua materna o habitual. Taberero (2008b) examina únicamente cinco centros de interés, mientras que el estudio de Areta (2009) analiza todos los datos.

³³⁴ Sancho (2006) constata un mayor uso de anglicismos entre los informantes que llevan menos tiempo viviendo en EEUU, en cambio, Moreno Fenández (2007) comprueba que en la segunda generación de inmigración sobresale ligeramente la frecuencia de uso de anglicismos.

discriminación destacable; en Navarra sí resulta relevante: los informantes de la zona vascófona obtienen peores resultados.

3.4.4. Transferencias entre sistemas lingüísticos

En todas las investigaciones que se han llevado a cabo en territorios bilingües se observa, al recoger los datos de una lengua o de las dos, un trasvase de elementos entre las dos lenguas, ya sea en el nivel gráfico, como en el fónico, morfológico e incluso sintáctico. A pesar de que no siempre estas transferencias o elementos a caballo se analizan, desde el primer momento se deja por escrito constancia de su presencia. Así, por ejemplo, Knock (1978) detecta préstamos, aunque no los examina, y Dimitrijevic (1981) también observa que los bilingües proporcionan palabras de su dialecto al completar la encuesta en la segunda lengua. Otros van más allá, como Bailey Victory (1971), que ofrece un elenco de las interferencias halladas, o Mackey (1971a y 1976), quien compara los dobles, indica el porcentaje que suponen las palabras del inglés en las listas francesas e inventa una fórmula para conocer el grado de integración de estas palabras en el francés.

En época más reciente, muchos investigadores relatan los problemas que suponen las interferencias gráficas (Bernal 2003, Escoriza 2007, Moreno Fernández 2007, López Meirama 2008, Tabernero 2008b, Casanova 2013...), sobre todo, respecto a los criterios de edición. La mayoría de estudios llevados a cabo en las comunidades bilingües intenta detallar el peso que estas palabras tienen en el corpus final, o, dicho de otra forma, el porcentaje de vocabulario disponible que suponen. También indican qué centros propician más su aparición, con lo que ya se pueden reconocer algunas tendencias comunes, por ejemplo, al empleo del vernáculo en *El campo y Trabajos del campo y del jardín*. Otras investigaciones van más allá y determinan cuál es el patrón social que más propicia la aparición de estos elementos, es decir, covarian las interferencias con algunas variables sociales como la *comunidad de habla* (Martínez Olmos 2008, Llopis 2009) o la *lengua materna* (Serrano Zapata 2003b y 2004)³³⁵. Más compleja resulta la forma de clasificar estos fenómenos, aunque muchos de los que se detectan se repiten en las distintas comunidades:

³³⁵ Casanova (2013) ofrece la cantidad de informantes que producen interferencias según las cinco variables sociales que contempla en su estudio, esto es, *sexo, ubicación del centro, tipo de centro, nivel educativo de los padres y lengua materna*.

- *dobletes* (Gómez Molina y Gómez Devís 2004, Moreno Fernández 2007 o Llopis 2009),
- *préstamos culturales* (Gómez Molina y Gómez Devís 2004, Taberero 2008b),
- *préstamos en vías de integración o asimilados o consolidados o integrados* (Gómez Molina y Gómez Devís 2004, Serrano Zapata 2006, Sancho 2006, Llopis 2009, Álvarez de la Granja 2011),
- *interferencias* (Taberero 2008b, Llopis 2009, Álvarez de la Granja 2011),
- *sustituciones o cambios de código* (Gómez Molina y Gómez Devís 2004, Taberero 2008b, Martínez Olmos 2008, Llopis 2009, Álvarez de la Granja 2011),
- *calcos* (Serrano Zapata 2006, Sancho 2006, Martínez Olmos 2008, Álvarez de la Granja 2011, Fernández Smith *et al.* 2012),
- *compuestos híbridos* (Serrano Zapata 2006, López Meirama 2008, Martínez Olmos 2008, Llopis 2009, Casanova 2013) o
- *derivados híbridos* (Serrano Zapata 2006, Martínez Olmos 2008, Fernández Smith *et al.* 2012³³⁶, Casanova 2013).

El hecho de que las investigaciones en estas comunidades sean relativamente recientes y que el test de disponibilidad sea insuficiente por sí mismo para analizar el alcance de estos elementos, hacen que no se haya llegado a un acuerdo sobre el valor de su presencia en los listados. A pesar de esto, en el último encuentro sobre disponibilidad léxica³³⁷, Gómez Molina propuso una clasificación basada en la frecuencia de aparición de la palabra para lograr diferenciar entre *interferencias*, *préstamos en vías de integración* y *préstamos integrados*³³⁸.

³³⁶ En realidad, los autores hablan de *hybrid loan words*, etiqueta bajo la que engloban distintos fenómenos como la aféresis, la confusión de vocales, la geminación o las sustituciones morfémicas.

³³⁷ Organizado por la Universidad de Salamanca y celebrado en esta ciudad los días 14, 15 y 16 de abril de 2011.

³³⁸ El investigador valenciano propuso considerar *interferencias* los términos con una frecuencia de aparición inferior al 5%, *préstamos en vías de integración* cuando la frecuencia fuera entre el 5 y el 20% y *préstamos integrados* si la frecuencia superaba el 20%.

Desde esta perspectiva, el único trabajo comparado que existe hasta la fecha es el de Bartol (2011)³³⁹. El estudioso salmantino analiza los datos aportados por distintas investigaciones realizadas en territorios bilingües —concretamente en Galicia, Valencia, Lérida, Vizcaya, Álava y Navarra³⁴⁰— con la finalidad de buscar tendencias comunes. De este modo, por ejemplo, con respecto a los centros de interés, llega a la conclusión de que algunos favorecen más que otros la presencia de elementos de la otra lengua, como ocurre en *El campo*, *Trabajos del campo y del jardín*, *Animales* o *Partes de la casa*. En cambio, la presencia de términos de la lengua vernácula es minoritaria en *La Escuela*, *La ciudad*, *Los muebles de la casa* o *Calefacción e iluminación*. El investigador también mide el peso que suponen las interferencias léxicas en el conjunto de los léxicos disponibles regionales y concluye que Lérida es la zona donde mayor porcentaje de palabras de la otra lengua se registran en el léxico disponible en castellano, seguida, en segundo lugar, de Galicia³⁴¹. Cabe señalar que los datos que usa Bartol (2011) de nuestra provincia corresponden únicamente a seis centros de interés³⁴².

Además, el autor analiza la integración de los préstamos en cada una de las comunidades de habla. Para ello se sirve del índice de disponibilidad y del número de informantes que mencionan la palabra en cuestión. Una vez detectados los vocablos que alcanzan o superan el 0.01 de disponibilidad, los compara con sus respectivos sinónimos correspondientes en la otra lengua de la comunidad³⁴³.

Una vez finalizado el análisis de las obras referidas al léxico disponible en una situación de contacto de lenguas, pasamos ahora a ver las características concretas del marco de nuestra investigación, esto es, el contacto entre el castellano y el catalán en la provincia de Lleida.

³³⁹ Con posterioridad, Casanova (2013) también compara sus datos con los de otros territorios donde conviven el catalán y el castellano.

³⁴⁰ Los datos de los que se sirve son los siguientes: López Meirama (2008), para Galicia, Gómez Molina y Gómez Devís (2004) para Valencia, Serrano Zapata (2006) para Lérida, Martínez Olmos (2008) para Alicante, Areta (2009) para Navarra, y los datos de Vizcaya y Álava los obtiene a partir de su colaboración en el *Léxico disponible de España* (Bartol 2011: 159).

³⁴¹ Para realizar este análisis, el autor utiliza el porcentaje que suponen las interferencias sobre el total de vocablos. De esta manera, en nuestra provincia, la media es de 9.21% y, en Galicia de 2.88%. En la zona catalana, contrastan fuertemente las diferencias halladas, puesto que, mientras en Lérida la influencia del catalán es considerable, en Valencia es media y en Alicante, ínfima (Bartol 2011: 173-174).

³⁴² Los criterios de edición usados en aquella ocasión son distintos a los utilizados en la presente monografía, razón por la cual los datos difieren levemente (*cf.* capítulo 7).

³⁴³ Este análisis se circunscribe a Galicia, Valencia, Alicante, Vizcaya y Navarra (Bartol 2011).

4. MARCO SOCIOLINGÜÍSTICO DE LA INVESTIGACIÓN

En el presente apartado, realizamos una aproximación general a la realidad sociolingüística de la comunidad autónoma catalana, centrándonos en la provincia de Lleida, con la finalidad de poder analizar en su contexto, posteriormente, los datos obtenidos. En primer lugar, tras efectuar un breve repaso histórico del contacto lingüístico entre las dos lenguas³⁴⁴, profundizamos en la situación actual de Cataluña, de una cierta complejidad. A continuación, glosamos las características del español hablado por los catalanohablantes, atendiendo con especial interés a los estudios que se circunscriben a nuestra región y a los dedicados, como la presente investigación, al nivel del léxico. Para finalizar, trazamos un panorama de las características geográficas generales de la provincia, y ahondamos en la realidad social y lingüística de misma.

4.1. EL CONTACTO ENTRE EL CASTELLANO Y EL CATALÁN EN CATALUÑA

El catalán entra en contacto con el castellano a finales de la Edad Media. Hasta ese momento, la lengua catalana goza de prestigio y reconocimiento como lengua propia y única en Cataluña y en el resto de territorios conquistados por la corona catalanoaragonesa, como las Islas Baleares o Valencia³⁴⁵. Con el matrimonio entre Fernando II de Aragón e Isabel de Castilla (1469), la historia de Cataluña se une a la de España. A partir de este momento, el catalán decrece como instrumento de creación literaria, mientras la economía catalana también se deteriora debido a que Sevilla monopoliza el comercio con las Américas. Durante el siglo XVI, se desarrolla un proceso de castellanización que afecta a parte de la nobleza catalana, así como a la Iglesia y su jurisdicción. Sin embargo, el catalán continúa siendo la lengua de las clases populares y de la burguesía mercantil, y predomina en el campo y en la mayoría de las ciudades.

Durante el siglo posterior, el prestigio de la lengua catalana va disminuyendo, mientras crece, de forma simultánea, el uso del castellano entre las capas dominantes de la sociedad y se acentúa la política centralizadora de los monarcas. El castellano es una

³⁴⁴ Para el que nos basamos, sobre todo, en Hoffman (1991: 260-288), Sinner (2004: 9-18), Vila Pujol (2008: 79-87) y el *Institut de Sociolingüística Catalana* (s.d.).

³⁴⁵ Durante este periodo, el catalán convive con el latín, que funciona como lengua de cultura, produciéndose así una situación diglósica que afecta únicamente a los hablantes cultos (Vila Pujol 2008: 80).

lengua de distinción social que la mayoría de la población no utiliza. Es a partir del siglo XVIII cuando el español realmente se asienta en Cataluña, tras la guerra de sucesión (1701-1713), que culmina con el centralismo político, territorial, socioeconómico y lingüístico de Felipe V. Con los Decretos de Nueva Planta, promulgados por este monarca, el castellano se impone como única lengua de toda España y se prohíbe el uso del catalán en cualquier ámbito. Pese a todas las medidas que se adoptan, la lengua catalana no desaparece sino que se reserva para usos informales o familiares, lo que acarrea una clara situación de diglosia entre el castellano y el catalán durante este periodo.

La industrialización y el nacimiento del catalanismo político son los hechos más decisivos relacionados con los avatares lingüísticos del siglo XIX. Mientras la nobleza y parte de la alta burguesía continúan castellanizadas, surge con fuerza una burguesía de clase media que aboga por la defensa y recuperación del idioma, a la vez que lleva a cabo una nueva política económica, basada en el comercio y en la industria. La lengua catalana vuelve a convertirse en vehículo de la literatura y vive un periodo de esplendor cultural conocido como la *Renaixença*. No obstante, la castellanización también aumenta, y gran parte de la población se convierte en bilingüe durante este siglo (Badia Capdevila 2002: 78)³⁴⁶. En las últimas décadas, empiezan a llegar las primeras oleadas migratorias de obreros a Cataluña, que no cesarán a lo largo del siglo XX, lo que ocasiona que, por vez primera, haya castellanohablantes entre las clases populares.

En el siglo XX suceden algunos hechos históricos que incrementan la vitalidad y el reconocimiento de la lengua catalana, como la creación de la *Mancomunitat* (1914) o el periodo de la Segunda República (1931-1936)³⁴⁷; aunque también acaecen otros que resultan claramente negativos para el idioma vernáculo, como la dictadura de Primo de Rivera (1923-1930) y el largo periodo franquista (1939-1975). Será este último el que resulte decisivo para entender la situación del catalán en la actualidad.

Tras la victoria del bando nacional en la guerra civil (1936-1939), se veda tajantemente cualquier uso público de la lengua catalana. Su empleo se entiende, a partir

³⁴⁶ Marsà (1986: 98) dice, al respecto: “[l]a restauración de la lengua catalana, iniciada con la *Renaixença*, logra óptimos frutos en el plano literario, pero actúa con menos eficacia en el ámbito popular”.

³⁴⁷ Durante estos dos periodos fue posible el uso oficial del catalán, la extensión en su conocimiento y la creación incipiente de escuelas en catalán (Boix y Vila 1998: 35).

de ese momento, como una provocación en contra del Régimen. El catalán, tras ser desterrado de las escuelas, de los medios de comunicación, de las instituciones y de cualquier situación formal, queda relegado, una vez más, al contexto familiar. La situación de la lengua catalana solo mejora ostensiblemente tras la muerte de Franco (1975) y la posterior restauración de las instituciones catalanas como la *Generalitat* (1977). Al mismo tiempo, la llegada continua de inmigrantes de origen meridional, sobre todo en la última mitad del siglo, conlleva una mayor extensión social del castellano³⁴⁸. Desde el gobierno autónomo de Cataluña se inicia, a partir de ese momento, un proceso de normalización lingüística con el que se pretende restablecer el uso del catalán en la sociedad. Para llevarlo a cabo, se aprueban distintas leyes —como la *Llei de normalització lingüística* del año 83 o la *Llei de Política Lingüística* del 98— con las que se normaliza el uso de la lengua catalana, especialmente, en la enseñanza, en la administración y en los medios de comunicación³⁴⁹.

Gracias a las medidas adoptadas en política lingüística, Cataluña se encuentra, actualmente, en el umbral de la normalización de su lengua vernácula (*Institut de Sociolingüística, s.d.*), aunque en este punto las opiniones divergen, como veremos a continuación. Si bien es cierto que el catalán se ha normalizado en la administración, en el sistema educativo y en los medios de comunicación³⁵⁰; aún quedan algunos ámbitos, como el judicial, donde su presencia sigue siendo excepcional³⁵¹. En la industria cultural, la situación oscila entre la vitalidad que tiene el teatro en lengua vernácula, por un lado, y, por el otro, el bajo índice que alcanzan las películas dobladas al catalán. También en el terreno de las nuevas tecnologías, a pesar de la existencia de oferta en lengua catalana, la del castellano es claramente superior.

La implantación del catalán como lengua vehicular de la enseñanza ha hecho que el nivel de conocimiento de esta lengua aumente considerablemente. Por ejemplo, en el año 1991, el porcentaje de personas que eran capaces de escribir la lengua propia

³⁴⁸ Alrededor de 1975 casi la mitad (el 49%) de los habitantes del municipio de Barcelona eran de origen no catalán. Entre 1950 y 1975 se asentaron, aproximadamente, dos millones de inmigrantes en Cataluña (Hoffman 1991: 271).

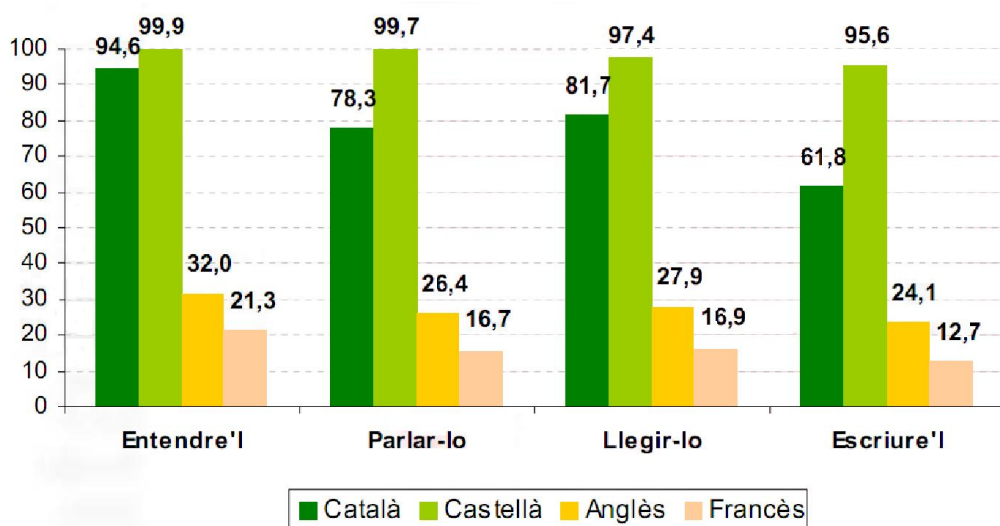
³⁴⁹ La ley 7/1983 de 18 de abril, aprobada por unanimidad en el *Parlament*, establece que el catalán es la lengua propia de la enseñanza y del gobierno de la *Generalitat* y de la administración local.

³⁵⁰ Aunque los últimos datos sobre oferta lingüística (*Ofercat*) otorgan una posición baja a esta esfera de la vida pública (Fabà 2008: 42).

³⁵¹ Esto es debido a que la mayoría de los jueces son de origen castellano hablante y sus destinos en Cataluña son, muchas veces, provisionales (*Institut de Sociolingüística, s.d.*).

de la comunidad suponía solamente el 39.94%, mientras que el porcentaje de personas con capacidad de hablarlo era de 68.3%; en 2001, los mismos índices se sitúan en el 49.82% y el 74.5%, respectivamente. Los datos de 2008 son aún más positivos con el 61.8% y el 78.3%, respectivamente³⁵². En la actualidad, la mayoría de los jóvenes sabe hablar y escribir en catalán, pero aún quedan personas en la franja de más edad que se consideran incapaces de escribir en catalán, a pesar de ser catalanohablantes, como consecuencia de que su escolarización fue exclusivamente en lengua castellana³⁵³. Sin embargo, la última encuesta sobre los usos lingüísticos de la población, llevada a cabo en el 2008 por la *Secretaria de Política Lingüística*³⁵⁴, señala que el conocimiento del castellano sigue siendo superior al del catalán, especialmente en la escritura.

Gráfico 4.1. Conocimiento de lenguas de la población de 15 años y más, en porcentajes.



Fuente: Encuesta d'usos lingüístics de la població (EULP) 2008.

También son más los hablantes que tienen el castellano como lengua inicial o materna (la L1, la que primero habló en casa), como lengua de identificación (la que el informante considera suya y con la que se identifica) y como lengua de uso habitual (la

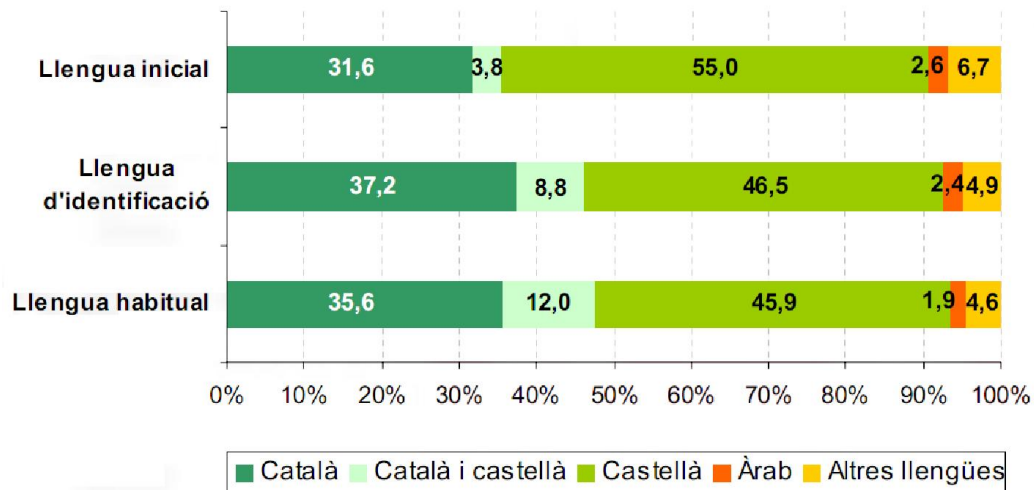
³⁵² Datos del *Institut d'Estadística de Catalunya* (años 1991 y 2001) y de la *Encuesta d'usos lingüístics de la població* (2008). Es importante señalar que los datos se refieren a competencias lingüísticas y no a usos.

³⁵³ Según los datos más recientes de la *Encuesta d'usos lingüístics de la població* de 2008, habla catalán un 84.9% de los jóvenes (entre 15 y 29 años) y lo escribe un 80.5%. Este último dato, contrasta fuertemente con el porcentaje global, que se sitúa en un 61.8% para la expresión escrita. Los jóvenes son la franja poblacional que más domina esta destreza mientras que entre los mayores de 65 años, el índice no llega al 43%.

³⁵⁴ Frente a las encuestas de uso y censos lingüísticos anteriores —que solo obtenían datos relativos al conocimiento del catalán—, la de 2008 incorpora los de otros idiomas como castellano, aranés, inglés y francés.

que utiliza con mayor asiduidad). No obstante, es muy significativo el incremento que se observa con respecto al catalán entre el índice de lengua materna y el de identificación, lo que demuestra que un porcentaje de personas hacen suyo el catalán con el paso de los años, a pesar de nacer en familias castellanohablantes.

Gráfico 4.2. Lengua inicial, de identificación y habitual de la población de 15 años y más, en porcentajes.



Fuente: Enquesta d'usos lingüístics de la població (EULP) 2008.

El hecho de que el castellano supere el catalán se explica por la importancia demográfica que adquiere el contingente de inmigración, principalmente meridional, llegado en las últimas décadas y porque muchos de sus descendientes, a pesar de conocer el catalán, prefieren emplear el español en la mayoría de sus intercambios lingüísticos (Vila i Moreno³⁵⁵ 2004: 34)³⁵⁶. Según Boix (2004: 21) hay aún un segmento de la población, de origen castellanohablante, —tanto de clases populares como de clases medias y altas— que continúa muy separado de la lengua y de la identidad catalanas autóctonas; mientras esta minoría es reticente o contraria a una mayor extensión social del catalán, la mayoría acepta su presencia institucional prioritaria. Es decir, existe un amplio sector de origen castellanohablante que se identifica con el catalán, aunque en distintos grados. Vila Pujol (2008: 88), con respecto a los descendientes de esta inmigración, comenta, acertadamente, que “este conjunto de la

³⁵⁵ Es costumbre tradicional catalana enlazar los dos apellidos mediante la conjunción copulativa, tal como aquí hacemos, respetando la voluntad del propio autor.

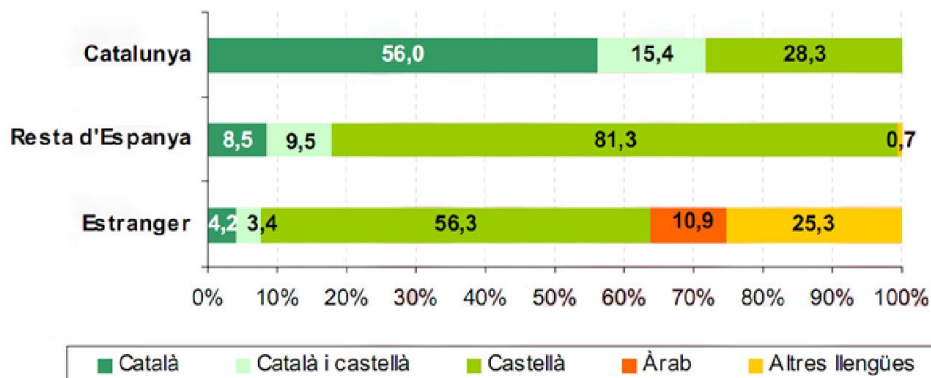
³⁵⁶ También Galindo (2006: 395) detecta esta tendencia entre la población escolar que estudia: los castellanohablantes muestran mayor fidelidad al uso de su lengua familiar o a la habitual en sus redes sociales que los catalanohablantes.

población catalana se debate entre mantener el español como lengua propia, [...] adoptar el catalán como lengua propia o tener las dos lenguas como propias”.

Por otro lado, tal como se ilustra en el gráfico 4.2., no es nada desdeñable el papel que están empezando a ejercer otras lenguas, como el árabe, en la realidad sociolingüística de Cataluña. Estas nuevas oleadas migratorias procedentes, en su mayoría, del Magreb, de Latinoamérica y de la Europa del Este, desequilibran la situación lingüística a favor del castellano y en detrimento claro del catalán, puesto que prácticamente un 75% de este grupo no sabe hablarlo (Fabà 2008: 47)³⁵⁷.

Otro factor digno de mención es la correlación tan nítida que hay entre el origen de la población y la lengua habitualmente empleada en los intercambios lingüísticos. Así pues, entre las personas nacidas en Cataluña, el catalán suele ser la lengua habitual de uso, por el contrario, está mucho más extendido el empleo del castellano entre las personas nacidas en el resto del Estado y en el extranjero.

Gráfico 4.3. Lengua habitual según lugar de nacimiento de la población de 15 años y más, en porcentajes.



Fuente: Enquesta d'usos lingüístics de la població (EULP) 2008.

La situación del contacto lingüístico entre el catalán y el castellano ha sido tildada, en no pocas ocasiones, de diglósica (Ninyoles 1969), sobre todo desde la óptica de la sociolingüística catalana, que ha puesto verdadero empeño en calificarla de este

³⁵⁷ Aunque el 80% lo entiende, y en cambio, únicamente un 12.7% sabe escribir en lengua catalana. Fabà (2008) se basa en los datos de la *Enquesta sobre les condicions de vida i hàbits de la població de Catalunya* del año 2006.

modo (Pueyo 1991: 78)³⁵⁸. Si bien es cierto que, durante la época franquista, podían reconocerse algunos de los patrones de este concepto, tras esa etapa, la mayoría de los autores rechaza la noción de diglosia como ilustrativa de la situación de Cataluña (Sanz 1991, Hoffman 1991, Pueyo 1991). El argumento principalmente argüido es que no hay una distribución rígida de los ámbitos de uso que debe ocupar cada lengua³⁵⁹. Hoffman (1991: 272-273), por ejemplo, constata que “[t]oday the language choice of a Catalan-Castilian bilingual is based rather more on the language of her or his interlocutor, and not so much on the formality of the situation or on the medium (i.e. written or spoken) employed” y Hernández García (1998a: 75-76), en su investigación sobre el Área Metropolitana de Barcelona, precisa: “los factores que explican una mayor o menor motivación para usar habitualmente el catalán o el español, son bastante heterogéneos en función de los campos de actividad y en función del grado y del tipo de contacto con el otro grupo lingüístico”³⁶⁰.

Por estas razones, se considera más oportuno sustituir la noción de diglosia por otra, de naturaleza más dinámica, como es la de *conflicto lingüístico* (Pueyo 1991: 86). Este concepto, acuñado por Lluís V. Aracil (Aracil 1986), refleja mejor la realidad contemporánea catalana y establece que la supervivencia de un idioma se basa en su capacidad de acometer las funciones sociales de una comunidad. Cuando dos variedades lingüísticas compiten entre sí por abarcar los distintos ámbitos de uso, la situación solo puede desembocar en la normalización de la lengua inferior, o en su desaparición porque se haya dado un proceso de sustitución lingüística (Pueyo 1991: 81-86, Boix y Vila 1998: 36). Por otra parte, destacando la situación de inferioridad en la que se halla el catalán en Cataluña, también se ha considerado el contexto actual de *bilingüismo asimétrico*, por cuanto mucha población desconoce el catalán oral mientras que prácticamente nadie desconoce la lengua castellana, y el dominio escrito y formal de la lengua catalana aún es deficitario (Boix 2004: 22, 2008: 16-17). Concretamente, Boix (2004: 24) opina que la situación sociolingüística es compleja y frágil y deja claro que “si per normalitat entenem que la majoria de catalans adopten el català com a llengua

³⁵⁸ El término, además, ha servido para etiquetar un sinfín de situaciones de contacto lingüístico, muy alejadas de los parámetros en los que lo definió Ferguson (Pueyo 1991: 73).

³⁵⁹ Así lo argumenta F. Vallverdú, figura importantísima dentro de la sociolingüística catalana, quien inicialmente aplica el concepto de *diglosia* al caso catalán, para más tarde revisarlo y rechazarlo (Pueyo 1991: 63-79).

³⁶⁰ A pesar de esta afirmación, la autora diferencia dos contextos que favorecen el uso del catalán: “los relacionados con el mundo institucional y laboral y los relacionados con el mundo de las relaciones interpersonales cotidianas” (Hernández García 1998a: 76).

pública comuna, aquesta encara no s'ha assolit". De manera similar, Vila i Moreno (2004: 46) es contundente al expresar que el catalán en Cataluña está "[f]orçat a sobreviure en situació de minorització".

Como acabamos de ver, en todas estas concepciones subyace la consideración del castellano como lengua dominante, fuerte, dotada de reconocimiento político y prestigio, mientras que el catalán, por el contrario, es la lengua que ha sido minorizada, como consecuencia de los avatares sociales e históricos sufridos. Las principales razones que avalan esta concepción ya han sido expuestas: un largo periodo de subordinación política y la llegada masiva de inmigrantes monolingües en un entorno en que los hablantes o ya eran bilingües o acabaron siéndolo (Boix 2004)³⁶¹.

Además, Hernández García (1998a: 70) cita otros condicionantes externos más recientes que han facilitado la expansión del castellano: la existencia de un grupo muy importante de personas que, a pesar de aprender el catalán, han seguido transmitiendo el español a sus hijos; el hecho de que la tasa de natalidad fuera superior entre los inmigrados que entre los autóctonos durante los primeros años de la migración, y que, durante años, se ha favorecido el paso al español como única lengua de transmisión de los matrimonios mixtos³⁶². Todas estas circunstancias han influido en los hábitos lingüísticos de la población, caracterizados por una convergencia hacia el español, es decir, se tiende a emplear el castellano frente a un interlocutor desconocido o castellano hablante. Este comportamiento se desarrolló, sobre todo, tras la guerra civil, ya que los condicionantes sociopolíticos llevaron a los catalanohablantes a emplear su lengua vernácula únicamente con otros catalanes y en situaciones privadas e informales³⁶³. Este mismo hábito, impidió, a su vez, la integración lingüística de los inmigrantes, que no tenían ocasión así de aprender o practicar el otro idioma (Sinner 2004: 19).

³⁶¹ De hecho, esta es una de las características del catalán en la actualidad: la inexistencia de hablantes monolingües (Mollà y Palanca 1987: 209).

³⁶² En este último punto, la autora precisa que este hecho tiene lugar, principalmente, en Barcelona y en su Área Metropolitana (Hernández García 1998a: 70).

³⁶³ Badia i Margarit (1979: 138) ya da fe de esta realidad sociolingüística: "...la convivència amb els no catalans (companys de feina o d'esbarjo, veïns, minyones de servei, superiors i subordinats) es fa, gairebé sempre, sota el signe de cedir l'expressió vernacla a la castellana." Recuérdese lo dicho en la nota 355 respecto a la unión de los apellidos catalanes mediante la conjunción.

A pesar de todo lo que acabamos de exponer, la situación sociolingüística de Cataluña tiene unos condicionantes tan singulares que no falta quien ha aplicado, más recientemente y de forma mucho más minoritaria, los mismos conceptos de *diglosia* y *conflicto lingüístico* considerando esta vez el castellano como lengua inferior. Esta perspectiva hace hincapié en un hecho particular del catalán en Cataluña como es el prestigio del que siempre ha gozado entre la población nativa (Blas Arroyo 2005: 407). Hernández García (1998a: 71) recoge una serie de factores que han hecho que la situación entre el castellano y el catalán cambie, en los últimos tiempos, favoreciendo el crecimiento de esta última lengua, como son: la cooficialidad del catalán y su reconocimiento como lengua propia, la promulgación de decretos que regulan su uso, el traspaso de competencias del Gobierno Central al Gobierno de la *Generalitat* y el cese prácticamente total de la inmigración procedente de otras zonas del Estado español³⁶⁴.

El catalán es la lengua propia del gobierno autonómico, de la educación y de cualquier institución pública de Cataluña. Se ha señalado, además, que en algunas ocasiones, el catalán parece ser la lengua más adecuada para los usos públicos mientras que el castellano queda restringido al uso interno de la comunidad castellanohablante (Boix 1993: 113). Por otro lado, no hay que olvidar que es también la lengua que habla gran parte de las clases sociales medias y acomodadas de origen autóctono, mientras que el castellano se asocia al fuerte contingente obrero que llegó durante el siglo XX a Cataluña y, por ende, a los estratos más populares³⁶⁵. Además, la política lingüística llevada a cabo por el gobierno autonómico ha hecho que el dominio del catalán sea imprescindible para una inserción adecuada en el mundo laboral. Todos estos hechos son la causa de que el catalán avance, en la sociedad, desde el punto de vista intergeneracional, lo que equivale a decir que hay un porcentaje importante de personas que hablaban castellano con sus progenitores y que deciden hablar en catalán a sus hijos (EULP³⁶⁶ 2008: 85), así mismo el índice de identificación con el catalán es superior al del catalán como lengua inicial, como ya hemos observado en el gráfico 4.2. Como consecuencia de todo ello, los usos lingüísticos de la población también parecen estar

³⁶⁴ Hernández García (1998a: 70-71), a su vez, cita una ponencia de Payrató y Tusón realizada en 1991 titulada “El castellano en Cataluña”. Los autores ratifican en ella: “La lengua catalana es, nunca ha dejado de serlo, una lengua de prestigio”.

³⁶⁵ Esta situación se aprecia más en las ciudades que en el campo y, por supuesto, incluye un gran número de excepciones.

³⁶⁶ Con estas siglas nos referimos, a partir de este momento, a *l'Enquesta d'Usos Lingüístics de la Població*, reseñada adecuadamente en el apartado bibliográfico. Citaremos siempre la edición a la que nos referimos, puesto que se llevó a cabo en distintos años: en 2003 (*EULC 2003*) y en 2008 (*EULP 2008*).

cambiando y algunos estudios realizados con jóvenes evidencian que la tendencia anterior hacia la convergencia al castellano está siendo superada o, por lo menos, convive con otras normas de uso que favorecerían más el uso del catalán en contextos que le son favorables³⁶⁷.

En resumidas cuentas, las consecuencias de la política lingüística llevada a cabo en Cataluña se evalúan de forma muy distinta según la perspectiva que se adopte, no siempre exenta de connotaciones ideológicas. Como ya hemos visto, para algunos los resultados son claramente insuficientes y argumentan que la lengua catalana sigue lejos de ser la lengua común de la sociedad y el instrumento de comunicación normal entre sus miembros³⁶⁸. Para otros, las leyes que amparan la normalización lingüística han ido demasiado lejos, en un intento de emplear exclusivamente el catalán en muchos ámbitos y no de potenciar realmente el bilingüismo, con lo que denuncian que sus derechos lingüísticos individuales han sido mermados³⁶⁹. Así que, a pesar de la cooficialidad de la que goza en el territorio bajo el amparo de la Constitución y del importante porcentaje demográfico de origen castellanohablante, para algunos colectivos, “el castellano ha adquirido, paradójicamente, la condición de lengua minoritaria en la comunidad” (Blas Arroyo 2005: 419-526)³⁷⁰.

Con todo, no puede obviarse el hecho de que, fuera de los límites de la comunidad autónoma catalana, la vitalidad y el futuro del español como lengua en expansión está fuera de toda duda, mientras que el catalán únicamente goza de esta posición —que podríamos llamar “privilegiada”— en Cataluña. En el resto de territorios donde el catalán convive con otra lengua, ya sea el español, el francés o el italiano, la situación sociolingüística es de índole muy distinta, cuando no está en claro

³⁶⁷ Así lo determina, por ejemplo, Boix (1993), en su estudio sobre los usos lingüísticos de jóvenes del Área Metropolitana de Barcelona que participan en cursos de monitores de tiempo libre. Más recientemente, Galindo (2006: 394-395) observa también distintos comportamientos lingüísticos entre el alumnado de primaria que no dependen ya tanto de la identidad lingüística del interlocutor como de la situación sociolingüística del entorno donde se está hablando. Véase al respecto también Vila i Moreno (1996: 187-190 y 466-467). De otra parte, en Pujolar *et al.* (2010) se ofrece una visión actual de los usos lingüísticos y de las percepciones de los jóvenes catalanes y se desmienten algunos de los tópicos comúnmente asumidos, como que el catalán es poco popular en este grupo generacional o que la juventud renuncia con facilidad a su uso.

³⁶⁸ Así lo indica Badia Capdevila (2002: 79) en su *Diccionari de les llengües europees*, por poner solo un ejemplo.

³⁶⁹ Sobre decir que estas minorías castellanohablantes son las que no se identifican con el catalanismo dominante. Sus ideas se han vehiculado, en las últimas elecciones, a través de la agrupación política *Ciutadans*-Partido de la Ciudadanía.

³⁷⁰ La cita entrecomillada corresponde a la página 509.

retroceso, como sucede en el *Alguer* o en la *Catalunya Nord* (Querol 2007). De otro lado, a pesar de que el conocimiento del catalán en Cataluña alcanza cotas inimaginables con respecto a otros tiempos —los últimos datos sitúan en 94.6 el porcentaje de personas que entiende la lengua vernácula (EULP 2008: 36)—, la realidad con respecto a su uso es otra. Tal como concluye Vila Pujol (2008: 85): “se conoce y se usa el español en muchas de las situaciones comunicativas de la vida diaria; y son muchos los que entienden el catalán e, incluso, conocen sus normas y son capaces de usarla [sic], pero son muchos menos los que lo usan habitualmente en la vida cotidiana, sobre todo en las grandes ciudades industriales”. De un modo similar, Hernández García (1998a: 81-82) reconoce: “si es cierto que el catalán ha conseguido grandes avances en el terreno institucional y público, los usos privados interpersonales siguen siendo dominio del español, siempre que los interlocutores —y es suficiente con que lo sea uno de ellos— sean de origen no catalán”. Con respecto a los usos en el terreno de la enseñanza, Galindo (2006: 397) postula que, en general: “als patis de les escoles catalanes es parla més castellà que català”, aunque en la relación con los maestros, los escolares usen más el catalán que el castellano. La misma autora, manejando datos de la *Conselleria d’Educació* del curso 2004-2005, afirma que entre el 30 y el 40% de las clases que se imparten en los institutos catalanes se realizan en lengua castellana (Galindo 2006: 402).

Estando así las cosas, en una situación que podemos denominar de bilingüismo social extenso —dado el número elevado de bilingües—, estable y no diglósico, los fenómenos de transferencia lingüística entre las dos lenguas, con sistemas lingüísticos además emparentados, son más que esperables³⁷¹. Es evidente que “en una situación de contacto estable en el tiempo, la interferencia puede ser mutua y encontrarse en ambas lenguas” (Thomason 1986 *apud* Blas Arroyo 1993: 40). De hecho, en la comunidad concreta que analizamos sobran razones para que esto se produzca. Por un lado, y tal como acabamos de ver, la influencia del castellano sobre el catalán ha sido constante e innegable a lo largo de los siglos y ha contado con una supremacía política que le ha permitido regular e imponer el uso del castellano entre los catalanes en diferentes periodos históricos. El otro factor fundamental, al que ya hemos hecho referencia, es el fuerte contingente inmigratorio llegado a las tierras catalanas desde finales del XIX, que

³⁷¹ Véase Silva-Corvalán (1989: 179).

si bien tuvo una cierta integración lingüística en un primer momento, los condicionantes sociopolíticos impidieron que esta se produjera adecuadamente tras la guerra civil (Bastardas 1986: 14-15).

Por otro lado, la situación actual se muestra, también, muy favorable a la influencia del catalán sobre el castellano. De un lado, el catalán es la lengua de la administración y del poder político autonómico y local; su conocimiento es condición *sine qua non* para acceder a una plaza en la administración, y es un mérito importante en las empresas privadas, sobre todo para ocupar un puesto de trabajo de cara al público³⁷². De otro, la lengua catalana goza de un estatus jurídico importante como única lengua propia de la comunidad, amparada en distintas leyes que protegen y regulan su uso. De hecho, en el último *Estatut d'Autonomia* (refrendado por la población el 18 de junio de 2006) se llegó a establecer que el catalán era la lengua de uso normal y *preferente* de las administraciones y de los medios de comunicación públicos, y que los ciudadanos de Cataluña tenían el derecho y el *deber* de conocer las dos lenguas oficiales, lo que fue fuente de innumerables trifulcas políticas³⁷³. La misma ley también precisa que el catalán es la lengua normalmente empleada como vehicular y de aprendizaje en la enseñanza en todos los niveles de instrucción.

De esta última afirmación se deduce que el conocimiento del catalán también resulta hoy imprescindible para poder recibir una formación académica adecuada en Cataluña. Esta realidad no ha estado exenta de polémica, porque algunos sectores consideran que el hecho de que toda la escolaridad se desarrolle en lengua catalana puede hacer mermar el nivel de castellano del alumnado. Además, tal como se ha señalado, en muchos centros educativos, el catalán se considera la lengua “alta”, apropiada para objetivos formales y académicos (Boix 1993: 122) lo que, según Hernández García (1998a: 77), puede ser la causa de la aparición de interferencias³⁷⁴. Tal como sigue la misma autora, al cursar toda la enseñanza en lengua catalana, los estudiantes tienen algunas convenciones gráficas mucho más interiorizadas en esa

³⁷² Casi la mitad de las empresas valora como criterio la competencia oral en catalán. Todo ello ha contribuido a que, actualmente, se considere difícil entre la población obtener un trabajo si no se tienen unos ciertos conocimientos de la lengua catalana (*Institut de Sociolingüística, s.d.*).

³⁷³ Estas terminaron, por el momento, con la sentencia 31/2010 del Tribunal Constitucional, emitida el 28 de junio de 2010, que obligó a modificar algunos artículos de la ley.

³⁷⁴ Galindo (2006: 392) también constata en sus conclusiones: “el català s’associa amb la llengua de les institucions educatives formals i dels membres que les representen, els mestres”.

lengua. No obstante, los estudios realizados hasta la fecha no han demostrado que el nivel de lengua española haya disminuido desde que se introdujo el catalán en la enseñanza a través de los programas de inmersión lingüística (Sinner 2004: 18)³⁷⁵. Sin ir más lejos, en nuestro trabajo sobre disponibilidad léxica en la ciudad de Lleida dejamos constancia de que el promedio de palabras por centro de interés que alcanzaba nuestros informantes, si bien era algo bajo, no distaba en mucho del índice alcanzado por el resto de comunidades hispánicas, y superaba al de algún otro territorio bilingüe (Serrano Zapata 2003b: 82-83, 2004: 154-156). Con todo, en la presente monografía, al manejar una muestra numéricamente superior y abarcar todo el territorio provincial, tendremos ocasión de corroborar esta tendencia y de compararla con la que observemos, de forma paralela, en lengua catalana.

Como ya hemos visto, es esperable que, en nuestra comunidad de habla, se produzcan interferencias en ambas direcciones. No existen demasiados estudios que hayan analizado, concretamente, la presencia de fenómenos de contacto lingüístico de forma simultánea en las dos lenguas analizadas, tal como realizamos en esta tesis. Los que se han dedicado, al menos en parte, a ello, ya han constatado que la proporción de fenómenos de contacto que aparecen en conjunto en el habla de las dos lenguas es baja, pero que el porcentaje de interferencias detectadas en catalán es superior al que se obtiene en castellano (Galindo 2006: 78).

Bastardas (1986), por ejemplo, lleva a cabo un interesante estudio en Vilafranca del Penedés (Barcelona), donde analiza la interrelación entre el contexto, el comportamiento y la competencia lingüística en el proceso de adquisición de una segunda lengua (el catalán) por parte de la segunda generación de inmigrantes castellanohablantes. Nos detendremos en dicho estudio, ya que el autor se sirve de la metodología de la disponibilidad léxica —aunque en ningún lugar la denomine así— para calcular tanto el coeficiente de bilingüismo como los coeficientes de interferencia en catalán y en castellano, hecho que presenta cierta similitud, aunque con salvedades importantes, con nuestra investigación³⁷⁶. Para el autor, estos coeficientes, junto con

³⁷⁵ Véase en Vila i Moreno (1996: 134-135) un resumen de los resultados obtenidos en los estudios que han evaluado la competencia en catalán y en castellano tras aplicar el modelo de inmersión lingüística en las escuelas.

³⁷⁶ Reseñamos esta obra en el presente apartado, y no en el estado de la cuestión anterior, porque el estudio de Bastardas no pretende establecer la disponibilidad léxica en sí, como sí lo pretenden los

otras variables como la capacidad de saber hablar en catalán y el nivel de coloquialización de esta lengua, forman parte de la competencia lingüística (Bastardas 1986: 59). En su investigación, el sociolingüista catalán utiliza una muestra de hijos de inmigrantes castellanohablantes, de donde obtiene la mayor parte de sus análisis, y otra, de descendientes de población catalana autóctona, que le sirve para realizar algunas comparaciones³⁷⁷. Los instrumentos de medida de la investigación son el cuestionario sociolingüístico subjetivo, por un lado, y las pruebas de enumeración léxica contextualizada, por otro. En el primero de ellos, se incluyen preguntas sobre los datos personales, las autoevaluaciones sobre el contexto y la calificación subjetiva de la propia competencia lingüística del individuo. Las pruebas asociativas, realizadas en las dos lenguas con un intervalo mínimo de una semana, versan sobre seis centros de interés (*1. Aliments i begudes, 2. La cuina i el menjador, 3. El carrer, 4. L'escola o l'institut, 5. Jocs, esports i distraccions y 6. El cos humà*) que se completan por escrito en un tiempo máximo de tres minutos por centro. Por otro lado, tal como el mismo autor advierte en diversas ocasiones (Bastardas 1986: 22-23), los sujetos de su estudio tienen una menor práctica en la expresión escrita en catalán, resultado de haber realizado el aprendizaje de la lectoescritura en castellano, aspecto que puede influir en los datos.

Con respecto a los resultados del coeficiente de bilingüismo, el investigador comprueba que, en el grupo inmigrante, este es muy cercano a -1 (-.096), lo que representa un predominio máximo del castellano, mientras que, en el grupo de los descendientes de la población autóctona, este es cercano a 0 (-.005), es decir, al equilibrio ideal entre las dos lenguas (Bastardas 1986: 50)³⁷⁸. El autor también detecta diferencias en función del centro de interés analizado. Así, por ejemplo, en el grupo inmigrante, los coeficientes que denotan un mayor uso del castellano son los correspondientes a *Alimentos y bebidas* y *La cocina y el comedor* —áreas estrechamente

trabajos incluidos en el capítulo 3, sino que se sirve de esta herramienta, entre otras, para el cálculo de la interrelación entre las tres variables —contexto, comportamiento y competencia— ya señaladas.

³⁷⁷ La primera muestra está compuesta por 163 individuos y la segunda por 178. Estos proceden de tres franjas de edad y nivel distintas: 9-10 años (4º de EGB), 12-13 (7º de EGB) y 15-16 (2º de BUP o de FP) (Bastardas 1986: 26). A partir de este momento, llamaremos a estas dos muestras, grupo inmigrante y grupo catalán, entendiendo estas variables como referidas a sus progenitores, ya que la mayoría de los descendientes de inmigrantes ya ha nacido en Cataluña, concretamente el 78% (Bastardas 1986: 37).

³⁷⁸ La fórmula del cálculo del coeficiente de bilingüismo es la siguiente: Total de palabras en catalán/Total de palabras en castellano y en catalán – Total de palabras en castellano/Total de palabras en castellano y en catalán (Bastardas 1986: 28). Los resultados se interpretan dentro de una escala donde el -1 representa el predominio máximo del castellano y el +1, el del catalán; el 0 indicaría una competencia perfectamente equilibrada entre las dos lenguas (Bastardas 1986: 50).

ligadas al medio familiar donde la primera lengua está más presente—; mientras que el único ámbito donde domina el catalán es *Juegos, deportes y distracciones*. En este mismo centro de interés el grupo catalán, por su lado, obtiene los índices más favorables al uso de su lengua materna, lo que nos señala la gran presencia del catalán en este ámbito; aunque es en *El cos humà*, que el autor añade por considerarlo teóricamente neutro, donde los resultados son aún más favorables al catalán por parte de este grupo³⁷⁹. Al respecto, el investigador concluye que todo parece indicar “la variació de la bilingüitat dels individus d’acord amb l’experiència resultant del divers ús lingüístic que tingui lloc als distints àmbits de la seva esfera social” (Bastardas 1986: 51).

Con respecto a los coeficientes de interferencias en las dos lenguas, se constata que entre los inmigrantes, el porcentaje de palabras del castellano presentes en la prueba en catalán asciende a 13.7%, mientras que este índice es únicamente del 5% en el grupo de procedencia catalana. El investigador justifica el grado de interferencia de la población inmigrante como un fenómeno de influencia de su primera lengua sobre la segunda, causado por una competencia menor en este segundo idioma. Con respecto al índice de interferencias detectadas entre el colectivo catalán, considera que muy probablemente se trate de interferencias ya consolidadas en la lengua, resultado de la presión del castellano en la sociedad catalana. En la prueba realizada sobre la lengua castellana, el porcentaje de interferencias del catalán es de 0.7% en el conjunto inmigrante y del 2% entre los autóctonos (Bastardas 1986: 52-53 y 194, cuadro 252), con lo que se confirma que la interferencia del castellano sobre el catalán es mayor que en el sentido inverso³⁸⁰.

³⁷⁹ Añadimos la siguiente tabla comparativa con los coeficientes de bilingüismo de cada centro de interés por si el lector desea ampliar la información (Bastardas 1986: 193, cuadro 253):

Centros de interés	Grupo Inmigrante	Grupo catalán
1. Alimentos y bebidas	-.16	-.01
2. La cocina y el comedor	-.17	-.03
3. La calle	-.11	-.09
4. La escuela o el instituto	-.13	-.08
5. Juegos, deportes y distracciones	+01	+07
6. El cuerpo humano	-.01	+11

³⁸⁰ El autor comenta que, respecto a la interferencia del catalán en el castellano, el centro de interés con el menor grado de interferencia registrado es *La escuela o el instituto* (Bastardas 1986: 53). Finalmente, el sociolingüista catalán comprueba estadísticamente, mediante los paquetes de programas básicos SPSS y SAS, la existencia de una relación significativa entre la competencia lingüística emisora, el contexto social y el comportamiento lingüístico (Bastardas 1986: 69).

Siguiendo los pasos de Bastardas (1986), Boix (1993: 141-142), en su estudio sobre los usos lingüísticos de los jóvenes de Barcelona, también aplica, entre otros procedimientos, lo que denomina “pruebas de enumeración léxica contextualizada” para poder cuantificar la competencia lingüística de los sujetos que integran su muestra. Utiliza tres campos semánticos —*Objectes habituals d’una cuina o menjador, Parts i òrgans del cos humà y Coses per jugar, fer esport o distreure’s*— que los informantes completan tanto en castellano como en catalán³⁸¹. A partir de estos datos, el autor obtiene los coeficientes de bilingüidad y de interferencia en castellano y en catalán, parámetros que forman parte de la competencia lingüística declarada³⁸². El autor detecta que, sea cual sea el grupo lingüístico de los sujetos, el coeficiente de bilingüidad obtiene siempre un resultado positivo, es decir, siempre favorece la lengua castellana (Boix 1993: 151).

Por su parte, Vila i Moreno (1996), en su tesis doctoral inédita, investiga los usos lingüísticos en un colegio público de Santa Coloma de Gramenet. El sociolingüista catalán, que se dedica al estudio de las consecuencias del modelo lingüístico educativo, constata que el porcentaje de marcas transcódicas detectadas en el discurso en catalán es mayor que el respectivo en lengua castellana (3.57% frente a 0.53%) y, además, comprueba estadísticamente que esta diferencia no se debe al azar (Vila i Moreno 1996: 416-417)³⁸³.

Por otro lado, en Baldaquí (2002) se presentan los datos procedentes de una investigación sobre los resultados sociolingüísticos de los programas educativos bilingües y el modelo de lengua de los escolares. En este estudio —realizado con una muestra de 202 alumnos de entre 13 y 14 años de la comarca de *l’Alacantí*, en Alicante,

³⁸¹ Dichos campos semánticos son completados por los informantes en sesiones diferentes de corta duración, con una semana como mínimo de margen entre las dos lenguas. Boix (1993) no ofrece más información sobre cuánto tiempo se daba para completar cada área léxica. Por otro lado, es sorprendente que el autor diera ejemplos de cada centro de interés antes de iniciar la prueba, ya que este hecho, como es lógico, es esperable que interfiera en los resultados (Boix 1993: 216-217).

³⁸² El cálculo de los distintos coeficientes es igual a la fórmula empleada por Bastardas (1986). Boix (1993) se sirve de estas pruebas, idénticas a las de la disponibilidad léxica, para calcular la competencia del individuo. Ya hemos dicho que su estudio no pretende, como todos los que reseñamos en este apartado, establecer la disponibilidad léxica en ambas lenguas.

³⁸³ La misma tendencia es observada por el investigador en el análisis del nivel morfosintáctico (Vila i Moreno 1996: 435-438). El autor, que trabaja con muestras de discurso oral espontáneo no controlado, lleva a cabo una serie de pruebas estadísticas para comprobar la fiabilidad de sus datos, que también son analizados cualitativamente. Por otro lado, una de las principales conclusiones de su estudio es que “unless the out-of-school factors are transformed, the school is not able to substantially modify sociolinguistic reality.” (Vila i Moreno 1996: 469).

provincia muy castellanizada desde hace siglos— entre otras magnitudes, se analiza cuantitativamente la interferencia léxica y se relaciona esta con algunas variables sociales. Con una metodología como la de la disponibilidad léxica, que el autor denomina “test de enumeración léxica”, se mide el vocabulario de los sujetos en dos campos semánticos: *Peces de vestir i objectes d’ús personal y aliments i begudes*³⁸⁴. El investigador concluye, al respecto, que hay una gran diferencia entre las interferencias recabadas en uno y otro sentido: mientras que en valenciano las interferencias del castellano suponen el 17.89%, en el castellano las interferencias del valenciano suponen únicamente el 0.69%. Con respecto a las variables analizadas, se comprueba cómo el alumnado que sigue el modelo educativo bilingüe comete menos interferencias del castellano en el valenciano, aunque, en contrapartida, también actualiza más interferencias del valenciano en el castellano, aunque los porcentajes en este sentido son muy inferiores. Asimismo, el alumnado que tiene el catalán como primera lengua realiza menos interferencias del castellano en el catalán. Por otro lado, el autor también detecta cómo el alumnado de las clases sociales más altas produce menos interferencias que el resto en el sentido del castellano al valenciano³⁸⁵. También las mujeres presentan porcentajes de interferencias inferiores a los de los varones³⁸⁶.

Más recientemente, Galindo (2006), en su investigación sobre los usos lingüísticos de los escolares de primaria a la hora del patio, concluye que en castellano las interferencias constituyen el 0.3% de las palabras emitidas en el discurso en esta lengua, mientras que en catalán el mismo porcentaje alcanza el 3.5%³⁸⁷. De otra parte, como bien afirma la investigadora en una interesante nota a pie de página (Galindo 2006: 79, nota 40), la comparación entre los estudios que analizan la presencia de

³⁸⁴ Lamentablemente, el autor no da más datos respecto a las características metodológicas de su investigación. De este modo, desconocemos el tiempo límite para cada campo semántico o el tiempo transcurrido entre la encuesta de cada lengua.

³⁸⁵ Esta situación atañe únicamente al alumnado de los programas educativos bilingües y no se da en sentido contrario (del valenciano al castellano) (Baldaquí 2002: 53).

³⁸⁶ Aunque esta diferencia solo se advierte en los programas educativos monolingües. Como conclusión, Baldaquí (2002: 53) constata cómo los resultados llaman la atención sobre el empobrecimiento cualitativo del modelo de lengua de las generaciones más jóvenes, hecho que, además, es mucho más evidente en el alumnado que sigue un programa educativo monolingüe. Baldaquí (2002: 55-63) también analiza los distintos tipos de interferencias en el modelo de lengua de los estudiantes en cada uno de los subsistemas lingüísticos. El estudio de Baldaquí (2002), al igual que el de Bastardas (1986), no pretende establecer el léxico disponible del alumnado.

³⁸⁷ Del mismo modo, los turnos con interferencias en castellano representan el 1.8% del total de turnos de habla emitidos en esta lengua, en cambio, en catalán representan el 18.2%. La autora también constata, tanto en castellano como en catalán, más interferencias en el nivel léxico que en el morfosintáctico, y más que son resultado de un proceso de importación que de sustitución (Galindo 2006: 392).

fenómenos de contacto lingüístico entre el castellano y el catalán no siempre es posible, ya que, por un lado, la mayoría de los trabajos sobre el tema se centra en una sola lengua y, por otro, las metodologías y los corpus no son homogéneos, con lo que los resultados obtenidos no son, para nada, comparables. En el presente estudio pretendemos salvar estos inconvenientes puesto que, como veremos en el próximo capítulo, tanto la metodología como los criterios de edición de los corpus, e incluso, los informantes, son exactamente los mismos, con lo que la comparación entre los resultados tiene la máxima fiabilidad.

4.2. RASGOS DEL CASTELLANO EN TERRITORIOS CATALANOHABLANTES

Si bien se ha estudiado, con detenimiento, la influencia del castellano sobre el catalán³⁸⁸ —constante a lo largo de los siglos—; son muchas menos las páginas dedicadas al análisis de la situación contraria, es decir, al influjo del catalán sobre el castellano (Marsà 1986: 98-99), aunque, como ya hemos mencionado, en una situación estable de contacto de lenguas es esperable que se produzcan trasvases que afecten a los dos sistemas lingüísticos implicados.

El interés por este tipo de investigaciones³⁸⁹ se desarrolla, sobre todo, a partir de la década de los 90, siendo minoritarios —aunque fundamentales— los estudios relativos a esta cuestión publicados con anterioridad. A partir de ese momento, se supera, por fin, el tópico de calificar las peculiaridades del castellano en boca de catalanohablantes como un conjunto de errores o defectos sancionables³⁹⁰ que, sobre todo, se asocian a los estratos menos cultivados de la sociedad o a las personas que viven en medios eminentemente rurales (Sinner 2004: 31)³⁹¹.

En los años 90, empiezan a fructificar las investigaciones sobre el castellano en tierras catalanas, basadas en la creencia de que “debe de existir un modelo del castellano

³⁸⁸ Algunos ejemplos son López del Castillo (1975), Cerdà (1984) o la importante obra de Payrató (1985), donde se analiza la interferencia catalán-castellano en las dos direcciones.

³⁸⁹ Nos basamos, a partir de este momento, en el exhaustivo estado de la cuestión de Sinner (2004: 21-36). Puede consultarse también, al respecto, Wesch (2002: 1857-1861) y Sinner y Wesch (2008: 11-55).

³⁹⁰ Por ejemplo, Colón (1967: 203) afirma que “se vería en un aprieto quien tuviera que dar a una de las tres grandes regiones catalanas el premio del mejor castellano”.

³⁹¹ El investigador argumenta cómo esta situación pudo darse, efectivamente, en el pasado, cuando solo las clases altas y urbanas de Cataluña dominaban el castellano. También recuerda que “es una práctica extendida atribuir fenómenos lingüísticos considerados como feos o incorrectos a las clases más bajas o faltas de cultura” (Sinner 2004: 31). A pesar de que el tópico se ha superado entre los lingüistas, por nuestra propia experiencia constatamos que sigue vivo entre las personas ajenas a este ámbito.

hablado por catalanes” (Sinner 2004: 28) y, por tanto, el castellano de Cataluña pasa a considerarse una variedad regional o dialectal. No obstante, se ha subrayado, en varias ocasiones, que la mayor parte de los trabajos se limita a describir los rasgos lingüísticos más típicos —ya comentados anteriormente por otros autores— mientras que escasean las publicaciones basadas en corpus y, por ende, en investigaciones con valor empírico (Sinner 2004: 32). Al tardío interés de la investigación hispánica sobre el tema, hay que sumar que, desde la sociolingüística catalana, se han mostrado reticencias a la existencia de una variedad del castellano peculiar de Cataluña (Vila i Moreno 2004: 43-45)³⁹².

Sinner (2004: 23-24), además, lamenta que muchas de las conclusiones a las que llegaron los autores pioneros en el estudio del contacto castellano-catalán no hayan sido tenidas en cuenta por los investigadores que se han dedicado a ello con posterioridad. Entre ellas, el hispanista alemán advierte que es necesario analizar con profundidad los rasgos propios de esta variedad, y no clasificarlos como interferencias provenientes del catalán sin más. De hecho, Sinner (2004: 23-31, 85-86) critica este hecho y constata que, en no pocas ocasiones, las supuestas interferencias del catalán no son más que elementos perpetuados ya en el sistema —algunos incluso documentados desde hace siglos— aunque también reconoce que es difícil distinguir estas dos circunstancias, a falta de estudios diacrónicos y comparativos. El mismo autor comenta cómo en la bibliografía relativa al tema, “[p]rogresivamente se impone la opinión de que también el desarrollo interno (es decir, el desarrollo lingüístico no explicable con impulsos de otro idioma, sino con tendencias de la propia lengua) así como el papel de castellano como segunda lengua tienen que ver con el desarrollo del castellano de Cataluña” (Sinner 2004: 28-29)³⁹³.

Por otro lado, ya en época temprana, tanto Moll (1961: 473) como Badia i Margarit (1979: 146), reconocen que los rasgos propios del castellano en Cataluña se

³⁹² Con respecto a las marcas transcódicas en el castellano del tipo *aconseguir*, *conversa* o *desenvolupar*, el sociolingüista catalán opina que “tot fa pensar que sovintegen més entre els sectors catalanoparlants joves, escolaritzats en català i que viuen instal·lats molt predominantment en aquesta llengua, uns sectors més aviat escassos en termes quantitius” (Vila i Moreno 2004: 46). Como veremos en el próximo apartado (*vid.* 4.3.2), este sector de población no es escaso cuantitativamente en la provincia de Lleida.

³⁹³ Véanse, por ejemplo, las conclusiones de Casanovas (2000a: 208) o los comentarios de Romera (2003: 370)

encuentran en el habla de personas cultas³⁹⁴. Badia i Margarit (1979: 153), tras analizar las características del castellano hablado por catalanes en los distintos niveles del lenguaje, destaca la falta de naturalidad de los catalanohablantes al expresarse en castellano, debida, en su opinión, a las vacilaciones sintomáticas y a los calcos que se producen entre las dos lenguas.

Otro aspecto interesante, a la hora de abordar el estudio de esta variedad, es la diferencia que ya establece Colón (1967: 203) entre el español hablado por los nativos, “como lengua ocasional o impuesta en los más” y “el español hablado por los inmigrantes y por sus descendientes”. El mismo autor también hace hincapié en la necesidad de distinguir entre la situación lingüística de las distintas comarcas, así como de tener en cuenta la clase social de los inmigrantes³⁹⁵. Este ha sido un aspecto bastante olvidado, ya que, salvo excepciones —como, por ejemplo, Moyer (1991) o Báez (1997)— apenas se ha investigado acerca de las características del castellano hablado por este colectivo o sobre las consecuencias que tiene esta inmigración para el castellano de la población autóctona (Sinner 2004: 124).

Otra carencia dentro de estos estudios³⁹⁶ es la relativa al ámbito geográfico, ya que, hasta la fecha, la mayoría de las investigaciones realizadas en Cataluña se circunscribe a Barcelona y su área de influencia³⁹⁷, mientras que los estudios sobre otras áreas son, cuantitativamente, inferiores, siendo incluso inexistentes los dedicados a algunas provincias como Tarragona o Girona o a áreas tan peculiares lingüísticamente como Andorra, la *Vall d’Aran*, el Delta del Ebro o la *Franja de Ponent* (Vila i Moreno 1996: 476, Sinner 2004: 56, Sinner y Wesch 2008: 33). La situación es distinta en Valencia, donde existen algunos estudios importantes y tempranos sobre el tema, como, por ejemplo, el de Gómez Molina (1986) sobre el contacto castellano-catalán en Sagunto o el de Blas Arroyo (1993a) sobre la interferencia lingüística en Valencia. El español de las Islas Baleares tampoco ha recibido demasiada atención hasta el

³⁹⁴ Concretamente, afirman: “tots els casos que adduïrem han estat recollits del parlar de persones cultes” (Badia i Margarit 1979: 146) y “pueden oírse pronunciaciones y construcciones incorrectas brotar de labios de hombres de carrera, profesores, oradores y literatos” (Moll 1961: 473).

³⁹⁵ “El funcionario público, llegado de tierras castellanias, posee una elevada opinión de su idioma y difícilmente se deja influir por el ambiente (en sus descendientes ya es otra cosa), mientras que el simple obrero, a quien el capataz y los patronos hablan el catalán, está mucho más expuesto” (Colón 1967: 203).

³⁹⁶ Sinner y Wesch (2008: 28-36) señalan las lagunas, a su modo de ver, más importantes en la investigación en este ámbito, algunas de las cuales recogemos en el presente apartado.

³⁹⁷ Algunos ejemplos son Español (1996), Wesch (1997), Hernández García (1998a) o Vann (2009).

momento, a pesar de la existencia de algunos breves estudios sobre el mismo (Moll 1961, Serrano Vázquez 1996 o, el más completo, Romera 2003)³⁹⁸. Si bien es cierto que la mayor parte de rasgos del castellano en zonas catalanohablantes son generalizables a las otras regiones —como ya apunta Badia i Margarit (1979: 145-146)—, también se han advertido diferencias³⁹⁹, con lo que sería deseable que se realizaran estudios en los distintos ámbitos geográficos.

En la zona que nos ocupa, la provincia de Lleida, son pioneros los trabajos de Casanovas. En sus primeras publicaciones (Casanovas 1995, 1996a, 1996b, 1996c, 1996d) la autora describe las características más relevantes del castellano hablado en la ciudad de Lleida, según los distintos niveles del lenguaje⁴⁰⁰. Cabe señalar que las peculiaridades recogidas en estos trabajos no se basan en un corpus concreto, sino que la autora las recaba mediante la “observación directa” (1996b: 150, 1996c: 58)⁴⁰¹. En cambio, en su tesis doctoral (Casanovas 2000a), la investigadora analiza el español hablado en la ciudad de Lleida a partir de un corpus de lengua oral integrado por una serie de entrevistas semidirigidas realizadas a 36 informantes⁴⁰² nacidos en Lleida o en su comarca y cuya lengua materna es la catalana⁴⁰³. La autora centra su análisis en el nivel morfosintáctico y analiza cualitativamente y, en parte, cuantitativamente sus datos⁴⁰⁴. Es especialmente interesante la clasificación cualitativa de los fenómenos léxicos recabados en su corpus, que comentamos seguidamente.

³⁹⁸ La situación parece estar cambiando en la actualidad, *vid.* las novedades bibliográficas sobre el tema en la completísima página de Carsten Sinner: <http://www.carstensinner.de/castellano/bibliografia.htm>

³⁹⁹ Por ejemplo, Blas Arroyo (2004: 1082) comenta que la reducción típica a dos miembros del paradigma deíctico (*este/ese/aquel* y *aquí/ahí/allí*) no se produce en tierras valencianas, mientras que, por el contrario, la confluencia entre *bajo* y *abajo* se da en esa zona, pero no en Cataluña o Baleares. Véase algún otro rasgo distintivo en Blas Arroyo (2004: 1079).

⁴⁰⁰ De este modo, Casanovas (1995) se dedica a las peculiaridades fonéticas; Casanovas (1996a), a las características morfosintácticas y Casanovas (1996c) a las léxico-semánticas. En Casanovas (1996b) la autora aúna las características referidas en 1995, 1996a y 1996c, refiriéndose así a los tres niveles de análisis. No reseñamos aquí estas características, ya que las glosamos a continuación, junto con las señaladas por otros autores.

⁴⁰¹ Véase la opinión de Sinner (2004: 177 y nota 57) al respecto.

⁴⁰² Este conjunto de 36 entrevistas forman parte de un corpus mayor recogido en la ciudad de Lleida bajo la dirección de M^a Ángeles Calero Fernández en el marco del PRESEEA (*Proyecto de Estudio Sociolingüístico del Español de España y América*).

⁴⁰³ Dichos informantes se distribuyen por cuotas uniformes según el sexo (varón/mujer), la edad (primera generación: entre 20 y 34 años, segunda generación: entre 35 y 54 años y tercera generación: más de 55 años) y el nivel de instrucción (estudios primarios, secundarios o universitarios), siguiendo los parámetros metodológicos establecidos en el PRESEEA.

⁴⁰⁴ El análisis cuantitativo, mediante el programa estadístico GoldVarb 2.0, solo abarca algunos rasgos relacionados con la expresión deíctica, ya que las estructuras restantes no son suficientemente frecuentes en el corpus o no covarían, según el análisis de la regla variable, con los factores explicativos (Casanovas 2000a: 30). Para el análisis de la expresión deíctica, puede verse Casanovas (2005).

Como conclusión a su investigación, la autora detecta dos patrones morfosintácticos de los que se sirven los informantes bilingües de la ciudad de Lleida: de un lado, una redistribución semántico-prágmática de estructuras ya existentes en las dos lenguas y, de otro, una importación de estructuras de la lengua autóctona a la lengua de destino⁴⁰⁵. En cuanto a la distribución social de las construcciones lingüísticas, Casanovas (2000a: 209) halla que “la generación mayor y los informantes con estudios elementales son los sectores más tendentes al empleo de los rasgos modificados por el contacto de lenguas”.

Un punto muy interesante, desde nuestra óptica, es la caracterización que realiza Casanovas (1996a: 407, 1996b: 149-150, 1996c: 54) de la variedad del español hablado en la ciudad de Lleida. Según esta investigadora, sus rasgos pueden explicarse a través de dos factores: la interferencia lingüística que el catalán ejerce sobre el castellano y la falta de comunicación oral en esta lengua, hecho que implica que en Lleida no se den algunas de las características que están presentes en otras zonas hispanohablantes como la relajación fonética o el uso de elementos fraseológicos. De este modo, la población catalanohablante tiene contacto principalmente con el castellano a través de los medios de comunicación, donde se difunde el español estándar. Además, también ha de tenerse en cuenta que el castellano no es la lengua materna de buena parte de la población, hecho que ocasiona que los hablantes posean un nivel de competencia desigual, según su relación personal con las dos lenguas.

Por otro lado, la autora también estudia (Casanovas 1998) la interferencia lingüística en la lengua escrita a través de un corpus integrado por textos (exámenes y trabajos de curso) de estudiantes universitarios⁴⁰⁶. La autora clasifica cualitativamente los errores detectados partiendo de la clasificación de Atienza *et al.* (1997)⁴⁰⁷. Merece la pena destacar que una diferencia importante de los trabajos de esta autora con respecto al nuestro, es el ámbito de estudio, toda vez que en la presente investigación se abarca el

⁴⁰⁵ Un ejemplo del primer caso sería el uso de *cada* sin valor distributivo o la presencia de la perífrasis obligativa *tener de*, mientras que en el segundo grupo hallaríamos el uso del partitivo o la doble negación *tampoco no* (Casanovas 2000a: 207-209).

⁴⁰⁶ Estos exámenes le fueron proporcionados por la profesora M^a Ángeles Calero Fernández y habían sido redactados por alumnado de Filología Hispánica de la Universidad de Lleida en diversas asignaturas impartidas por ella.

⁴⁰⁷ Concluye que, una de las principales causas de los errores producidos es la ausencia de competencia en español en un registro formal (Casanovas 1998: 15).

conjunto de la provincia, mientras que Casanovas se refiere siempre al ámbito estricto de la ciudad de Lleida⁴⁰⁸.

Otra carencia advertida (Sinner 2004: 57, Sinner y Wesch 2008: 28-32), como ya hemos comentado, es la ausencia de corpus en los que basar las investigaciones empíricas. La situación es un tanto mejor con respecto a la lengua oral, donde contamos con los datos de Vila Pujol (2001), de Sinner (2001) o de Vann (2009), no obstante; apenas hay corpus escritos sobre los que trabajar. Por otro lado, aunque hace ya mucho que se habla de interferencias en Cataluña⁴⁰⁹, se echan de menos trabajos cuantitativos que analicen la difusión social de las mismas (Vila i Moreno 2004: 43-45), hecho que permitiría conocer su grado de integración en el sistema y su distribución social.

A pesar de que es bastante difícil clasificar los fenómenos de contacto lingüístico según los distintos niveles de análisis (Español 1996, Hernández García 1998a: *passim*), procedemos, a continuación, a detallar —de esta forma— las principales características del castellano en Cataluña. Abordamos, con mayor detenimiento, el nivel léxico-semántico, por ser al que dedicamos la presente monografía. Para esta caracterización nos basamos en los siguientes trabajos: Moll (1961), Badia i Margarit (1979), Solà (1980), Payrató (1985), Casanovas (1996b), Español (1996), Serrano Vázquez (1996), Vila Pujol (1996), Atienza *et al.* (1997), Wesch (1997, 2002), Grupo Interalia (1998), Seib (2001), Blas Arroyo (2004: 1065-1086) y Sinner (2004)⁴¹⁰.

En el nivel fonético, es posible hallar —en esta variedad del español— una ausencia general de relajación fonética, apoyada tanto en el poco uso oral del castellano en esta zona como en la influencia del sistema fonético catalán (Casanovas 1996b: 151). Esta mayor tensión articulatoria es la que lleva a pronunciar la /-d-/ en posición implosiva (**ciudad**) y a conservar sistemáticamente los grupos consonánticos cultos del tipo /ks/ o /ts/. Otros rasgos típicos del español en contacto con el catalán son la

⁴⁰⁸ En alguno de sus artículos, por ejemplo, Casanovas (1996b), la investigadora se refiere al área general de Cataluña, aunque los fenómenos reseñados son exactamente los mismos que cuando alude, en concreto, a la ciudad de Lleida.

⁴⁰⁹ El primer trabajo en el que se habla de interferencias, en el marco del contacto castellano-catalán, es una publicación de la *Assessoria de Didàctica del Català* del año 1973 (Sinner 2004: 25).

⁴¹⁰ No tenemos afán de exhaustividad en esta presentación ni tampoco detallamos qué rasgos constata cada autor, entre otras cosas porque, como ya hemos dicho, se suelen citar casi siempre las mismas peculiaridades.

velarización de la /l/, la sonorización de la /s/ por fonética sintáctica, la pronunciación de /y/ como [j] o el seseo catalán, que es posible hallar aún en medios rurales que no tengan apenas contacto con el castellano. También tienden a conservarse las consonantes finales en determinados sustantivos como *carnet* o *chalet*. En cuanto al vocalismo, ya Badia i Margarit (1979: 149-150) señala una tendencia a la abertura de las vocales intermedias —por asimilación con el sistema vocálico catalán que cuenta con dos fonemas más—, la reducción de las vocales [e] y [a] átonas a una vocal central y el cierre de [o] átona en [u]⁴¹¹.

La influencia del catalán es también tangible en el nivel morfosintáctico. Se registran algunos usos incorrectos en el español estándar como son los frecuentes desajustes con respecto al género por interferencia del catalán (*la* olor o *la* calor), la construcción partitiva con *de* (“*de* coches, arreglo muchos”) o la presencia del artículo determinado ante nombres propios de persona. En el paradigma verbal, también hallamos algunas incorrecciones en los verbos, tanto en la conjugación de formas irregulares como en la concordancia, habitual en toda Cataluña, del verbo *haber* con el complemento directo en oraciones impersonales (“habían tres sillas allí”). Otras características son el mayor uso del futuro flexivo (“haré”) que de la perífrasis *ir a* + infinitivo (“voy a hacer”), contrariamente a lo que ocurre en otros lugares hispanohablantes⁴¹², o el uso del futuro en algunas subordinadas donde sería preceptivo el subjuntivo (“cuando llegará” en lugar de “cuando llegue”).

Además, se constata un uso anormal de las preposiciones (estar **a** un sitio, ir **con** coche), un mayor uso de la perífrasis *haber de* —en detrimento de *tener que*— o el uso del posesivo pospuesto en vez del pronombre personal (delante **tuyo**). Por influencia directa del catalán, se registra una frecuente incorporación de un *que* expletivo ante cualquier oración interrogativa, innecesario en el castellano comúnmente empleado en el resto de regiones castellanohablantes y que da lugar a oraciones del tipo: “¿*Que* vas a llegar tarde hoy?”. Otro asunto largamente tratado en la bibliografía es la reducción de los deícticos *este/ese/aquel* y *aquí/ahí/allí*, a un sistema de dos combinaciones (*este/aquel*, *aquí/allí*), fenómeno a medio camino entre el nivel morfológico y el léxico-semántico. Dentro de las unidades que rebasan el límite de la oración, es decir,

⁴¹¹ Véase más información sobre este nivel lingüístico en Wesch (1997: 295-298).

⁴¹² En Illamola (2008) puede obtenerse más información sobre este fenómeno.

extraoracionales o discursivas, destaca el uso de *pero*, que se extiende a contextos donde el castellano normativo optaría por *sin embargo*, o el de la unidad *por eso*, que funciona como conector adversativo, y no con el valor anafórico con el que lo emplean los castellanos no bilingües.

En cuanto al nivel léxico-semántico⁴¹³, son varios los intentos de descripción y clasificación de los fenómenos propios del castellano en Cataluña. En primer lugar, recogemos la clasificación de Payrató (1985) —basada en Weinreich (1953)— que ampliamos con los ejemplos que recoge Blas Arroyo (2004), procedentes de diversos autores⁴¹⁴. Payrató (1985: 87-88) distingue los siguientes grupos:

- (1) *Subestimación de una distinción* (cuando desaparece una oposición funcional que existía en la lengua por semejanza con otra lengua donde dicha oposición no existe). Serían los casos de las siguientes parejas *hocico/morro*, *cristales/vidrios*, *chuleta/costilla*, *pescado/pez*, *guapo/bonito*, *llevar/traer*, *doler/hacer daño...* siempre resueltas a favor del segundo término, por paralelismo con el término correspondiente en lengua catalana⁴¹⁵.
- (2) *Sobreestimación de una distinción* (el caso inverso, esto es, cuando se realiza una distinción, existente en la lengua modelo, de la que carecía la lengua receptora). Por ejemplo, en el castellano de Cataluña se diferencia entre *fuelle* y *plata*, entre *trabajo* y *faena*, o entre *poner* y *parar*.
- (3) *Reinterpretación de una distinción* (fenómeno por el cual dos elementos dejan de oponerse por unos determinados rasgos y se oponen por otros por similitud con la estructura de la lengua modelo). Aquí entraría la distinción creada entre *puesto* (como equivalente a “sitio/lugar”) y *parada* (como equivalente a “puesto”), o la especial distribución de los verbos *ser/estar* o *ir/venir*.
- (4) *Sustitución* (cuando una unidad de la lengua modelo sustituye a la lexía correspondiente de la lengua fuente). Payrató (1985: 87) recaba aquí los casos de *engegar*, *paleta*, *pica* o *desafavorecer...* entre otros. Blas Arroyo (2004) cita en este punto todos los calcos con el verbo *hacer* (*hacer bondad*, *hacer*

⁴¹³ Solà (1980) ofrece un elenco en el que recaba catalanismos desde 1505 en distintas obras gramaticales y lexicográficas. Sobre términos de origen catalán que se han transmitido al español general, puede verse Colón (1967), y para una visión diacrónica de este mismo fenómeno, consúltese Prat (2003 y 2005).

⁴¹⁴ Principalmente de los trabajos de Casanovas (1995, 1996a, 1996c, y 2000b), de Serrano Vázquez (1996), de Wesch (1997) y de sus propios trabajos (Blas Arroyo 1993a).

⁴¹⁵ También Colón (1967: 204) apunta que, entre dos posibilidades léxicas del castellano, el hablante elige siempre la que tiene correspondencia en catalán y cita los casos de *mortero/almirez* o *concha/pechina*.

buena cara, hacer olor, hacer mal, hacer miedo, hacer cara de, hacer(se) un café, hacer tarde, etc.) y los calcos semánticos del tipo *dejar de pasta de boniato, los días de cada día o a más a más*.

- (5) *Importación* (se produce cuando se transfiere a la lengua receptora un elemento de la lengua modelo que pasa a cubrir un espacio nuevo en la lengua receptora). Este es el caso de los términos *plegar, rajola, trona, caldre* o *chafardear*.
- (6) *Pérdida* (fenómeno consistente en la desaparición de un término de la lengua receptora por carecer de equivalente en la lengua fuente). Payrató (1985: 88) no cita ningún caso en la dirección catalán>castellano, aunque reconoce que pueden interpretarse de este modo los casos referidos en el primer grupo (1).

En segundo lugar, Wesch (1997: 302), siguiendo a Cerdà (1984), distingue entre *préstamos sintagmáticos parciales, préstamos sintagmáticos innovativos* y *préstamos paradigmáticos* (o *calcos semánticos*). Los préstamos sintagmáticos coinciden con los préstamos *in praesentia* de Cerdà, donde se traslada tanto la forma como el contenido, mientras que en los préstamos paradigmáticos, llamados préstamos *in absentia* por Cerdà (1984: 278), solo se manifiesta el influjo en el nivel del contenido (Wesch 1997: 299). Para Wesch (1997: 299) una interferencia sintagmática sería “el uso de elementos catalanes en el castellano que mantienen su forma catalana o aparecen de forma más o menos asimilada”, mientras que la interferencia paradigmática “se basa en la identificación de morfemas o lexemas castellanos y de morfemas o lexemas catalanes equivalentes solamente en parte, y tiene como consecuencia un cambio semántico [...] en los elementos castellanos”. En el castellano de Barcelona, el autor distingue como *préstamos sintagmáticos parciales*⁴¹⁶ la palabra catalana *prou*, variante de “bastante”, *hacer bondad* (“portarse bien”), *pica* (“fregadero”), *hacer servir* (“utilizar”), *embolicar* (“envolver”) o las numerosas locuciones con el verbo *hacer* (“hacer un paseo”, “hacer tarde”, etc.).

⁴¹⁶ Según Cerdà (1984: 278-279), “los préstamos totales aluden a los castellanismos adoptados a costa de formas autóctonas de uso debilitado o arcaizante”, mientras que “[l]os préstamos parciales, a su vez, se refieren a una implantación de castellanismos no sobre un vacío léxico, sino sobre un elemento autóctono equivalente con el que se produce por desplazamiento una cierta complementariedad semántica”. En último lugar, “los préstamos innovativos se refieren a vocablos que se han incorporado, por lo general recientemente, junto con el objeto o la noción que designan. Constituyen por sí mismos una aportación tanto léxica como semántica.”

Como *préstamos sintagmáticos innovativos*, y por tanto, que suponen un enriquecimiento del léxico castellano, el autor detecta *plegar* (“terminar el trabajo”), *chafardear* (“chismorrear”), *enchegar* (“poner en marcha”) y *sucar* (“mojar el pan en una salsa para aprovecharla mejor”). Como interferencia paradigmática el autor da el caso de *parada*, que reúne dos significados (“parada de autobús” y “puesto en el mercado”) en un solo significante. Por otro lado, como influencias léxicas y semánticas paradigmáticas del catalán, el autor constata un uso más habitual del verbo *agradar* (en la pareja *gustar/agradar*), y las peculiaridades semánticas que afectan a las parejas de verbos *ir/venir*, y *traer/llevar*, caracterizadas por utilizarse en contextos no prescriptivos, ya que se asimilan con formas verbales del catalán, que no se distinguen semánticamente del mismo modo⁴¹⁷.

Por otro lado, nos parece muy clarificadora la distinción que elabora el mismo autor más recientemente (Wesch 2008: 57-58) donde hace hincapié en la necesidad de tener en cuenta la frecuencia de las unidades léxicas para poder diferenciar correctamente entre:

- (a) una lexía o construcción, común al castellano y al catalán, que sea más frecuente en el castellano de las tierras de habla catalana por influjo del catalán;
- (b) una lexía o construcción de origen catalán y, sin embargo, que se utilice en toda la comunidad lingüística hispánica, o en gran parte de ella, pero que en el castellano de tierras catalanas se caracterice por una mayor frecuencia; y,
- (c) una lexía o construcción de origen catalán cuyo uso se limite a las tierras de habla catalana⁴¹⁸.

En tercer lugar, Casanovas (2000a: 75-97) establece una tipología basada en los distintos mecanismos de adaptación léxica⁴¹⁹. De este modo, la autora distingue entre *préstamos*, *transferencias semánticas*, *cambios en el registro en el entorno de uso y calcos*. Dentro del primer grupo, la autora documenta formas como *engegar/enchegar*

⁴¹⁷ El investigador (Wesch 1997) ofrece también información sobre la marcación diastrática o diafásica de las unidades mencionadas en su corpus.

⁴¹⁸ Los ejemplos que el autor proporciona de cada categoría son los siguientes: “La Silvia *hace de profesora*” (a), “Tengo el ordenador *enchegado*” (b) y “Se precisa *mujer de hacer faneas*” (c). La cursiva en los ejemplos es nuestra.

⁴¹⁹ La autora parte de una clasificación ya existente a la que añade la categoría: *cambios en el entorno de uso o registro* (Casanovas 2002: 263). Por otro lado, la misma clasificación que reseñamos aquí puede hallarse en Casanovas (2000b, 2002, 2004 y 2008).

(“poner en marcha”), *chafardear* (“chismorrear”) o *girarse* (“volverse”). Todas estas unidades se caracterizan por estar adaptadas fonológicamente, por su integración morfosintáctica en la lengua que las acoge —en este caso, el español—, y por transferirse con el significado que tenían en la lengua de origen. En el segundo grupo, las *transferencias semánticas*, se engloban todas aquellas palabras que, perteneciendo al sistema español, adquieren distintos significados por interferencia con el catalán. Estas voces tienden a confundirse porque la mayor parte de las veces proceden de un mismo étimo y se asemejan en el nivel formal, a pesar de que las restricciones semánticas son muy distintas en cada una de las dos lenguas del territorio. La autora cita como ejemplos de transferencias semánticas los vocablos *avanzar* (“adelantar”), *lampista* (“fontanero”) o *paleta* (“albañil”).

El tercer grupo, *los cambios en el entorno de uso o registro*, está formado por aquellas unidades que existen en español con el mismo significado que aquí se les otorga —no son, por lo tanto, transferencias semánticas—, pero que cuentan en castellano con algún tipo de marcación lexicográfica, ya sea diacrónica, diafásica, diatópica o diastrática. Pertenecen a este grupo *defensar* (marcado como antiguo en los diccionarios), *descambiar* (marcado como propio de América) o *toñina* (marcado como andalucismo) entre otros términos. En último lugar, se consignan los denominados *calcos*, término con el que se alude a la implantación de una expresión catalana en español por traducción directa de cada uno de los términos que conforman el sintagma. Algunos ejemplos serían las siguientes construcciones, habituales en el español hablado en Lleida: *dejar de pasta de boniato*, *encontrar a faltar*, *los días de cada día* o las numerosas combinaciones del verbo *hacer*, como por ejemplo, *hacer un café* (“tomar un café”), *hacer mal* (“doler”), *hacer servir* (“utilizar”) y un largo etcétera.

A continuación, y para finalizar este apartado, ofrecemos una breve síntesis de las tres obras que, a nuestro modo de ver, mejor retratan las peculiaridades del castellano de Cataluña en lo relativo al nivel léxico-semántico, esto es, Szigetvári (1994), Hernández García (1998a) y Sinner (2004). Nos extenderemos más en la tesis doctoral de Hernández García (1998a) debido a que la perspectiva que adopta guarda cierto paralelismo con la nuestra, como veremos a continuación.

El trabajo de Mónica Szigetvári (1994) supone un primer intento de recopilación de todos los fenómenos léxicos y, en parte, gramaticales que caracterizan el español en Cataluña. En la primera parte de la obra (Szigetvári 1994: 1-61), la autora reúne distintas unidades léxicas, ordenadas alfabéticamente, y comenta el uso peculiar de las mismas en el castellano de Cataluña, basándose en la información recogida en distintas obras lexicográficas (de la lengua española y catalana) y también en su propia experiencia, e incluso aporta, en ocasiones, el punto de vista de catalanohablantes sobre estos fenómenos. En la segunda parte de su trabajo (Szigetvári 1994: I-XII), se realizan algunas consideraciones gramaticales y se amplía la información relativa a algunas entradas de la primera parte. También se recogen alfabéticamente los distintos fenómenos analizados (preposiciones, categorías gramaticales, tiempos verbales...). A pesar de que carece de un estudio cuantitativo que dé cuenta de la frecuencia de los elementos recabados, el estudio de Szigetvári (1994) tiene mucho mérito, desde el punto de vista cualitativo, ya que recoge más de 100 entradas⁴²⁰.

La tesis doctoral de Carmen Hernández García (1998a) es de sumo interés, ya que se dedica al estudio de las interferencias en el español de Cataluña. La autora parte del concepto de interferencia como desviación de la norma de cualquiera de las dos lenguas de una comunidad bilingüe debido al contacto lingüístico (Hernández García 1998a: 32-33) y basa su trabajo en dos corpus orales registrados en la Región Metropolitana de Barcelona. En la primera parte de la tesis (Hernández García 1998a: 175-358), se presenta una propuesta de clasificación de interferencias según los distintos niveles (fónico, gramatical, léxico-semántico, discursivo, pragmático y gráfico)⁴²¹ que la investigadora ilustra con varios ejemplos recabados durante las interacciones orales de 20 informantes que tienen las dos lenguas, catalán y castellano, como propias.

En la segunda parte de la tesis (Hernández García 1998a: 359-490), la investigadora analiza cualitativa y cuantitativamente las interferencias realizadas por un grupo de 40 estudiantes universitarios estratificados según su origen familiar y su

⁴²⁰ Suponemos que la autora amplía la información recopilada en este trabajo, en su tesis doctoral del año 2002 (Szigetvári 2002) a la que, lamentablemente, no hemos podido tener acceso.

⁴²¹ A pesar de que su propuesta se efectúa según los niveles lingüísticos, la investigadora hace hincapié en que, frecuentemente, el análisis sobrepasa estos límites puesto que hay muchas interferencias que afectan, simultáneamente, a varios niveles del lenguaje.

contexto de socialización⁴²². Cabe destacar que, en ambas partes de la tesis, la autora ratifica las supuestas interferencias halladas en su corpus con distintas fuentes lexicográficas del catalán y del castellano. Además, contrasta esta información con la opinión de expertos, tanto de Filología Catalana como de Filología Hispánica, y con test de uso llevados a cabo en distintas zonas del mundo hispanohablante⁴²³. Con esto pretende comprobar si las estructuras etiquetadas como interferencias en el castellano a raíz del contacto con el catalán lo son realmente.

En esta segunda parte, la perspectiva que adopta la investigadora es similar a la nuestra —por lo que respecta a la covariación de unos datos lingüísticos con unas variables sociales— en el análisis, concretamente, de las interferencias del español de Cataluña⁴²⁴. Hernández García estratifica la muestra en tres grupos que van desde los informantes que tienen una familia de origen catalán y un contexto de socialización totalmente en esa lengua (grupo A) hasta el caso completamente contrario (grupo C), estableciendo un grupo intermedio de informantes cuya familia es de origen catalán pero cuyo contexto de socialización ha tenido etapas en las que el español era la lengua principal (grupo B). En el análisis global de los datos, se detecta cómo los dos grupos más catalanizados producen más interferencias en general que el grupo más castellanizado. Por otro lado, son muy interesantes las diferencias que la investigadora detecta según los distintos niveles de análisis: al parecer, los sujetos de origen catalán y con un contexto de socialización en esa lengua, producen más interferencias gramaticales, principalmente sintácticas. Por el contrario, los informantes con una historia lingüística más castellanizada incurren más frecuentemente en interferencias léxico-semánticas⁴²⁵. Además, dentro de este nivel de análisis, se observa cómo los

⁴²² En realidad, la muestra queda reducida, finalmente, a 38, ya que se descartan dos informantes cuya familia era de origen castellanohablante, pero que habían vivido un contexto de socialización totalmente en catalán; la autora considera que esta cifra es insuficiente para poder extraer conclusiones más generales. Todos ellos son estudiantes de 2º de Traducción e Interpretación de la Universidad Pompeu Fabra.

⁴²³ Sinner (2004: 36) plantea algún inconveniente en este punto al trabajo de Hernández García.

⁴²⁴ Somos conscientes de que existen otros trabajos —Gómez Molina (1986), Blas Arroyo (1993a) o más recientemente Gómez Devís (2013)— que también relacionan la interferencia lingüística con determinadas variables sociales. Sin embargo, no los reseñamos aquí, porque nos ceñimos a las investigaciones realizadas en Cataluña. La situación sociolingüística en la Comunidad Valenciana es tan distinta que consideramos que sus conclusiones no podrían servirnos para compararlas con nuestros resultados.

⁴²⁵ Hernández García (1998a: 572) justifica este hecho de la siguiente forma: “parecería que existe una cierta independencia entre niveles [...], los aspectos morfosintácticos de la lengua materna quedarían impresos en la mente del individuo de tal modo que difícilmente serían borrados por otros modelos morfosintácticos superpuestos como consecuencia del aprendizaje de nuevas lenguas”.

catalanohablantes tienden a realizar más calcos semánticos, mientras que los castellano hablantes realizan más calcos formales.

En cuanto al análisis cualitativo de las interferencias del nivel léxico-semántico⁴²⁶, la estudiosa documenta tanto calcos formales (*echar a faltar, alcinas, gabias, mandra, degana, en contes de, suerte que, aclapara, se nos acude, de tanto en tanto, comenzamiento, día de cada día, venir de gusto...*) como semánticos (*grande, te encuentras, te pones, tirar adelante, no venir de aquí, bien bien, hacer clase, hacer campana, hacer una conferencia, descuidarse, mirar de + infinitivo...*). A pesar de que muchas de las interferencias que recaba ya han sido consignadas con anterioridad en la bibliografía (como *remenar, embolicar, paradita, plegar*, los numerosos calcos con el verbo *hacer* o la confusión entre *ir/venir* y *traer/llevar*), muchas otras son documentadas por la autora por vez primera hasta ese momento (*soterrano, mediano, arrancar a correr, estada, panera, aire condicionado, aprofundizar, llevarse*, etc.)⁴²⁷.

En último lugar, glosamos el trabajo de Sinner (2004), hasta la fecha el mejor intento de análisis de las particularidades del castellano de Cataluña. Tras esbozar un sólido y amplio marco teórico sobre el tema —que da luz a muchas carencias hasta el momento enquistadas en este tipo de estudios (*cf.* con el principio de 4.2)— el autor emprende la ardua tarea de determinar qué elementos están integrados y, por tanto, forman parte de la norma de uso del castellano de Cataluña⁴²⁸. Para ello, realiza una serie de rigurosos análisis de las frecuencias de uso y de la aceptabilidad formal e informal de un total de 84 elementos léxicos, morfosintácticos y pragmáticos⁴²⁹. Con la finalidad de comprobar si existe una norma regional del castellano catalán, realiza 40 entrevistas a distintos profesionales de la lengua —la mitad de ellas en Barcelona y la otra mitad en Madrid— que constituyen el corpus en el que basa su investigación. En el nivel del léxico, el autor analiza algunos de los ítems más frecuentemente mencionados en los estudios sobre el tema como *lampista, pica, torre, avanzar, chafardear* o *guapo*.

⁴²⁶ La autora analiza todos los niveles lingüísticos —especialmente el gramatical y el discursivo, además del léxico-semántico—.

⁴²⁷ Es muy interesante la perspectiva didáctica que aborda la autora para el tratamiento de las interferencias en el ámbito universitario (Hernández García 1998a: 513-547).

⁴²⁸ El autor parte “de un núcleo mínimo de elementos empleados y considerados habituales y perfectamente aceptables por la mayoría de los catalanes [...] y ya perpetuados por la transmisión generacional de la lengua” (Sinner 2004: 139-140).

⁴²⁹ Concretamente, analiza 6 rasgos morfológicos, 20 sintácticos, 55 elementos léxicos y 3 elementos deícticos (Sinner 2004: 174).

En sus conclusiones, Sinner (2004: 563) constata que el núcleo mínimo de los fenómenos constitutivos del castellano de Cataluña se compone de: (a) elementos inexistentes en las variedades de regiones no catalanohablantes, (b) elementos con significados diferentes de los que tienen en otras variedades del castellano y (c) elementos empleados con otra frecuencia, con otras connotaciones o en otros registros que en otras variedades del castellano.

4.3. CARACTERIZACIÓN DE LA PROVINCIA DE LLEIDA

Abordamos ahora la concreción de las características del que es nuestro marco de estudio: la provincia de Lleida. En primer lugar, detallamos las características geográficas generales, y, en segundo, profundizamos en la situación social y lingüística.

4.3.1. Contexto geográfico del estudio

La provincia de Lleida está ubicada en el nordeste de la Península Ibérica; limita al norte con Andorra y Francia —con los departamentos de *Ariège* y Alto Garona—; al este, con Girona y, principalmente, con la provincia de Barcelona; al sur, con Tarragona, y, al oeste, con Zaragoza y, en mayor amplitud, con Huesca. Abarca una extensión de 12 149.62 Km², lo que representa un 2.40% de la superficie total española. Comprende un total de 231 municipios y su población alcanza los 442 308 habitantes. Pertenece a la comunidad autónoma de Cataluña y es la única, de las cuatro provincias, que carece de costa.

Mapa 4.1. Ubicación de la provincia de Lleida y de su capital en Cataluña y España.

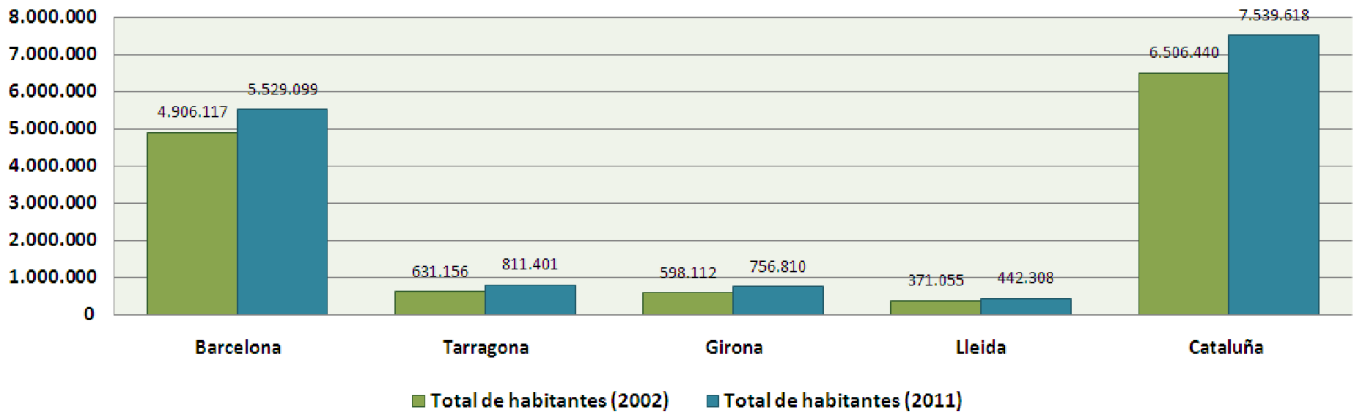


Fuente: Elaboración propia a partir del mapa base del Instituto Geográfico Nacional.

Por otro lado, es la provincia que tiene una mayor extensión territorial —concretamente, ocupa el 37.9% de la superficie catalana— pero, sin embargo, es la menos poblada, con una densidad de 36.41 personas por Km², hecho que contrasta fuertemente con la densidad del resto de provincias catalanas, y que supone que únicamente el 5.87% del total de habitantes de Cataluña se ubica en alguna comarca leridana⁴³⁰. Además, según se aprecia en los siguientes gráficos, Lleida ha tenido el menor incremento demográfico en los últimos años, en comparación con el resto de provincias de la misma comunidad autónoma.

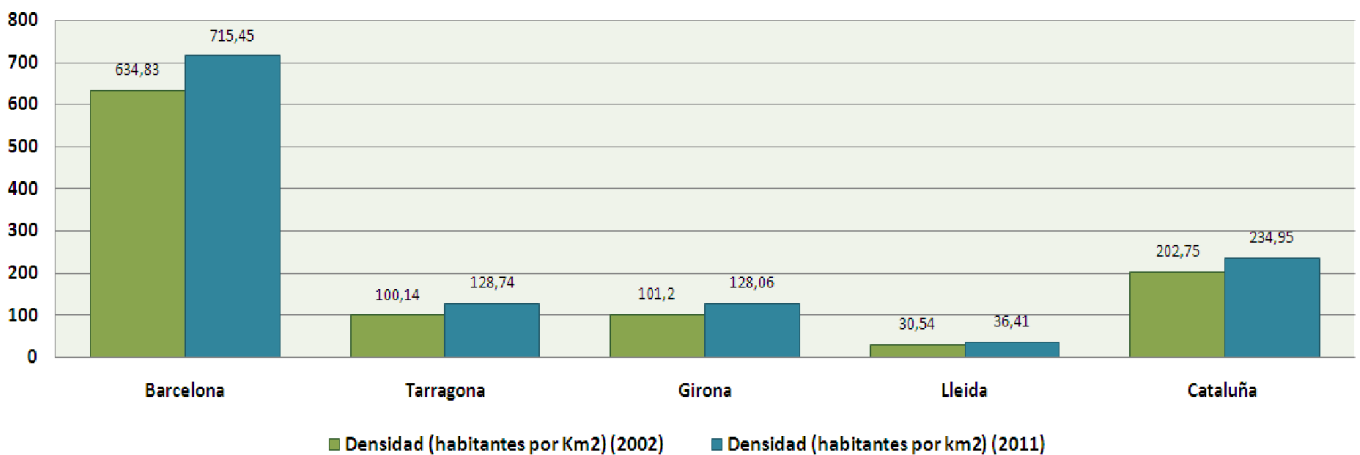
⁴³⁰ Datos del Instituto Nacional de Estadística a 1 de enero de 2011.

Gráfico 4.4. Total de habitantes de las cuatro provincias catalanas en el año 2002 (año de aplicación del estudio) y en el 2011.



Fuente: elaboración propia a partir de los datos de Idescat.

Gráfico 4.5. Densidad de población de las cuatro provincias catalanas en el año 2002 (año de aplicación del estudio) y en el 2011.



Fuente: elaboración propia a partir de los datos de Idescat.

La provincia consta de trece comarcas⁴³¹: *la Vall d’Aran, el Pallars Sobirà, l’Alt Urgell, el Pallars Jussà, l’Alta Ribargorça, la Cerdanya, el Solsonés, la Noguera, la*

⁴³¹ Incluimos, a la hora de citar las comarcas de Cataluña, el artículo que acompaña el nombre propio en la misma lengua, esto es, en catalán, por ser esta la forma más frecuente de nombrarlas en nuestra comunidad. La única excepción al respecto es el caso del artículo contracto “del”, que usaremos, en este caso, a lo largo de estas páginas, evitando de este modo la construcción “de *el*”, que consideramos absurda siendo que la contracción existe en las dos lenguas, aunque reconocemos que esta es una decisión arbitraria. De otra parte, utilizamos el catalán en los topónimos referidos a comarcas, ciudades, pueblos y ríos de la provincia, pero no usamos la cursiva en los nombres que son iguales en las dos lenguas, como Solsona o Viella.

*Segarra, l'Urgell, el Pla d'Urgell, les Garrigues y el Segrià*⁴³². En esta última comarca, situada al sudoeste, se halla la ciudad de Lleida, capital de la provincia, en unas coordenadas geográficas de 0° 38' de longitud este y 41° 36' de latitud norte. Su altitud media es de 155 m sobre el nivel del mar, y el municipio cuenta con una extensión de 211.71 km². Esta ciudad histórica romana —fundada sobre un asentamiento íbero— se alza sobre un pequeño montículo a orillas del río Segre. El fenómeno meteorológico que más la identifica es, sin duda alguna, la niebla, consecuencia de su situación en la depresión del río Ebro y de sus características orográficas.

Mapa 4.2. Ubicación de las comarcas de la provincia de Lleida.



Fuente: Página web de la Diputació de Lleida.

La división administrativa, realizada en 1833, potenció el crecimiento de la ciudad en todos los ámbitos como capital. Desde entonces, es el centro administrativo y comercial de una amplia área territorial, sin que haya otro núcleo de magnitud e influencia similar en el resto de la provincia⁴³³. Su radio de influjo comprende una gran parte de la Depresión Central Catalana, parte del Pirineo Central —el que se articula alrededor de los valles del río Segre y sus afluentes, el *Noguera Pallaresa* y el *Noguera Ribargorçana*— y también, el conjunto de la *Franja* catalana que, desde el Pirineo hasta el río Cinca, abarca una serie de territorios que pertenecen administrativamente a Aragón. Estos límites de influencia, por tanto, no corresponden exactamente con los de

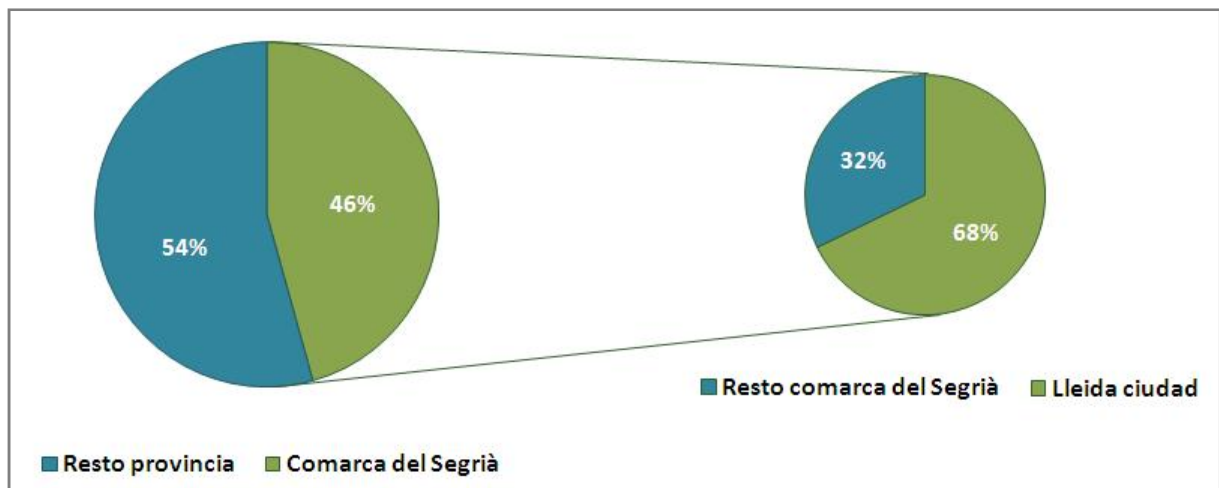
⁴³² También incluye el municipio de *Gòssol*, en la comarca del *Berguedà*, perteneciente a Barcelona. La comarca de *la Cerdanya* se divide entre las provincias de Girona y Lleida.

⁴³³ Además, más de la mitad de los municipios de Lleida tienen menos de 500 habitantes, hecho que acentúa las diferencias entre la ciudad y el territorio (Samper *et al.* 2000: 45).

la provincia, dado que, por ejemplo, el término municipal de Solsona orienta sus relaciones funcionales hacia la ciudad de Manresa, en la provincia de Barcelona, mientras que muchos de los municipios leridanos de *la Cerdanya* y de *l'Alt Urgell* establecen lazos muy directos y lógicos con Andorra o Barcelona, aunque mantienen con la capital, obviamente, una relación administrativa (Vilagrassa 2003: 281-302). Esto es debido, en parte, a la debilidad de la infraestructura de las comunicaciones pero también a la geografía de la región, que no facilita una comunicación rápida y eficaz entre algunos puntos de la misma y la capital.

La población no se reparte de forma equitativa en el territorio. De hecho, durante el siglo XX, la población de la capital se multiplicó por cinco, mientras que el resto de la provincia perdió residentes de forma permanente (Samper *et al.* 2000: 44-45). Concretamente, en el año 2002, año de aplicación de nuestro estudio, casi la mitad de la población de la provincia se concentra en la comarca del *Segrià*, y la mayoría, además, reside en su capital, tal como se observa en el siguiente gráfico:

Gráfico 4.6. Reparto de la población en la provincia de Lleida en el año 2002 (año de aplicación del estudio).

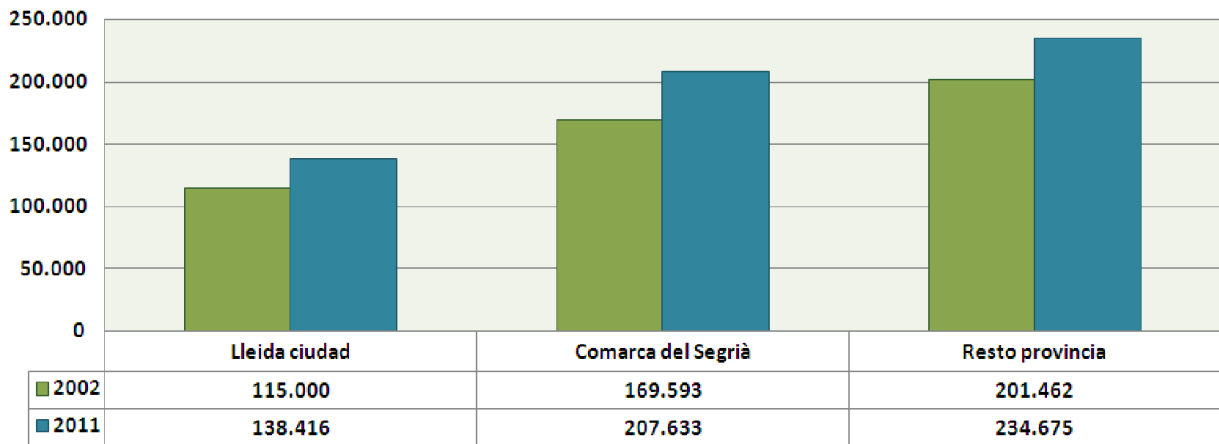


Fuente: elaboración propia a partir de los datos de Idescat.

En la actualidad, la situación sigue siendo similar puesto que la ciudad de Lleida aglutina el 31.3% del total de habitantes y, en su comarca, *el Segrià*, se concentra el 47% de la población provincial. Cabe destacar el desarrollo que han tenido, en época reciente, algunas capitales comarcales, como *Mollerussa*, *Balaguer* o *Tàrrega*, que

reúnen unos 15000 habitantes, aunque su nivel de influencia y actividad económica sigue sin poder compararse con el de la capital.

Gráfico 4.7. Reparto de la población en la provincia de Lleida en el año 2002 (año de aplicación del estudio) y en el 2011.



Fuente: elaboración propia a partir de los datos de Idescat.

La región, como hemos apuntado, presenta grandes diferencias orográficas, que pueden apreciarse en el contraste de los paisajes, y que se deben a que la cordillera de los Pirineos, y el Prepirineo, atraviesan la parte septentrional y correspondiente a las comarcas de *la Vall d'Aran*, *l'Alta Ribagorça*, *el Pallars Jussà*, *el Pallars Sobirà*, *l'Alt Urgell*, *el Solsonés* y *la Cerdanya*, siendo su punto más alto la cima de *la Pica d'Estats*, sita entre la comarca del *Pallars Sobirà* y el departamento francés de *Ariège*. Por otro lado, en la parte más meridional, destaca la llanura formada por la depresión del río Ebro, donde se ubica la capital. La temperatura media anual es de 16°, aunque la provincia se caracteriza por los contrastes entre el clima de montaña del sector pirenaico, y el continental de la depresión del Ebro⁴³⁴.

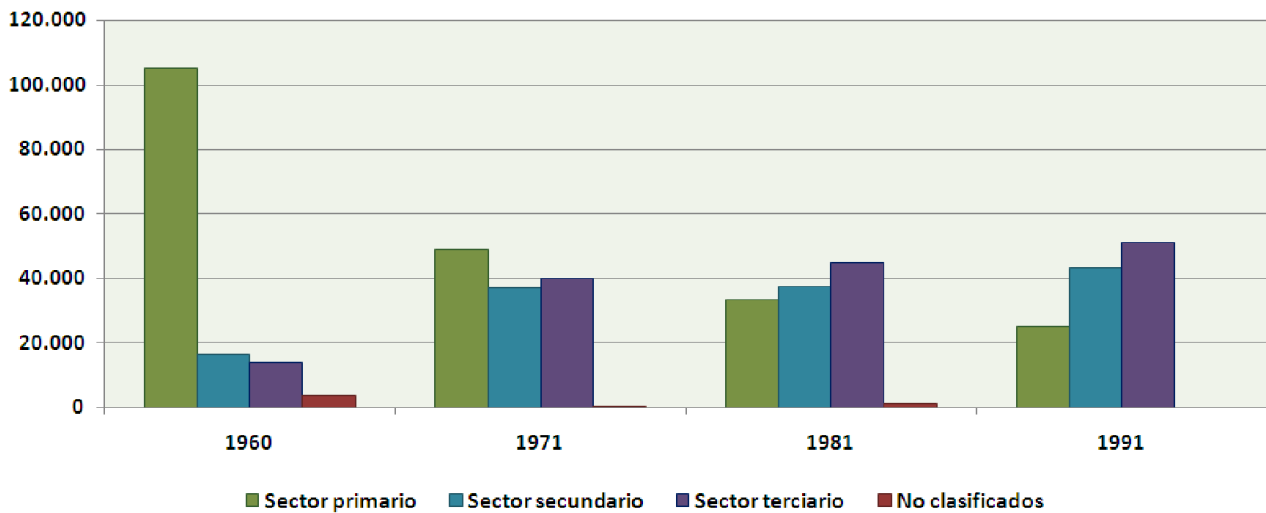
La base del desarrollo económico de la región ha sido, tradicionalmente, la agricultura. La singular orografía y distribución del territorio propician este hecho, ya que la zona montañosa del norte acumula, gracias a las precipitaciones, grandes cantidades de agua que es convenientemente regulada y almacenada en pantanos existentes en las cuencas de los ríos. De este modo, es posible regar la parte sur provincial, de orografía plana y clima templado, y desarrollar distintos cultivos, entre

⁴³⁴ Datos del *Instituto Nacional de Estadística* correspondientes al año 2009.

los que destacan los de cereales, forrajes y árboles frutales. En la actualidad, destacan las zonas de riego de los Canales de *Urgell*, el Canal de Aragón y Cataluña, el Canal de Piñana y el de *Segarra-Garrigues*. El cultivo del cereal, a su vez, contribuye al desarrollo de la ganadería, sobre todo dedicada al porcino en los llanos, en la que la provincia se perfila como una de las más importantes de España. También cabe destacar que las comarcas del sureste, principalmente *les Garrigues*, se dedican, de forma casi exclusiva, al cultivo de olivos y almendros, produciendo un aceite de excelente y reconocida calidad cuya rentabilidad económica se cuestiona hoy en día debido al incremento en los costes de producción. Por otro lado, en la parte norte de la provincia, se ha desarrollado, desde principios del siglo XX, una importante ganadería vacuna y ovina, aunque esta actividad está siendo progresivamente abandonada por razones sociológicas y económicas. Esta parte del territorio basa su desarrollo actual en el turismo de montaña, fomentado por la práctica de numerosos deportes de invierno.

El peso del sector agrario ha sido fundamental en el desarrollo económico de la provincia, sin embargo, tal como se aprecia en el gráfico 4.8, su importancia ha ido decayendo desde finales de los años 60, en favor del sector secundario y terciario, siguiendo la misma tendencia que se observa en otras regiones. Mientras que las áreas dedicadas al sector primario se extienden por la provincia, el sector secundario y terciario se concentran, fundamentalmente, en la capital. La rama de la industria más desarrollada, en los últimos años, es la agroalimentaria, aunque también destaca la rama de la construcción y la maderera, la industria textil y la producción de energía hidroeléctrica. En cuanto al sector terciario, se aprecia una cierta especialización en el funcionariado, un contingente laboral en alza desde que la ciudad fue elegida capital de provincia (Vilagrasa 2003: 350).

Gráfico 4.8. Evolución de los diferentes sectores económicos en la provincia entre 1960 y 1991.



Fuente: elaboración propia según los datos de Samper et al. (2000: 110-111).

En cuanto a la infraestructura de las comunicaciones de la provincia, cabe señalar que la capital se encuentra en el eje transversal que une, principalmente, Madrid con Barcelona. De este modo, se relaciona con estas ciudades gracias a una red ferroviaria de alta velocidad, de creación reciente, y mediante autovías o autopistas. La comunicación interprovincial es, fundamentalmente, de norte a sur, y se basa en una red de carreteras que desemboca en la capital y que discurre por los valles de los ríos pirenaicos. En este caso, no hay prácticamente conexiones ferroviarias, puesto que su uso en términos económicos es claramente deficitario.

4.3.2. Situación social y lingüística de la provincia

Tras haber trazado, en el primer apartado de este capítulo, un panorama general de la situación sociolingüística de Cataluña, pasamos ahora a abordar los aspectos concretos de la misma en la provincia de Lleida. Cabe señalar que las características que señalamos en ese momento (*vid.* 4.1) son, en su inmensa mayoría, extrapolables a nuestra comunidad. Sin embargo, también queremos dejar constancia de que en la bibliografía se hace referencia, en no pocas ocasiones, a circunstancias derivadas de situaciones sociolingüísticas características de Barcelona y su Área Metropolitana, muy alejadas de las realidades de otras provincias como Girona o Lleida. Este hecho se explica porque la mayor parte de las investigaciones realizadas tanto sobre el castellano como sobre el catalán en Cataluña, se ha circunscrito al área mencionada de la capital, mientras que los estudios sobre otras zonas son más bien excepcionales (Vila i Moreno

1996: 476, Sinner 2004: 56). Por eso algunas afirmaciones no dejan de ser sorprendentes a nuestros ojos, por no corresponderse en nada con la realidad leridana, ni de su capital ni de ningún pueblo de su provincia. Por ejemplo, Hoffman (1991: 271), refiriéndose a los inmigrantes llegados de otros puntos de la Península a partir de los años 50, afirma “[c]ontact with the Catalan-speaking population has been, and still is, very limited”⁴³⁵. De igual modo, Sinner (2004: 19) constata: “la concentración de la población inmigrante contribuye a que los contactos con la lengua catalana constituyan, para muchos de los inmigrantes de segunda generación, una excepción en su vida cotidiana”. Las dos afirmaciones son ciertas, sin duda alguna, por lo que respecta a Barcelona y las poblaciones colindantes que la rodean, donde las concentraciones de población inmigrante de origen peninsular alcanzan cotas muy altas. No obstante, esta situación no se produce en Lleida, donde la población de origen castellanohablante es mucho menor, y por esta razón, es un caso excepcional precisamente el contrario, que un inmigrante de segunda generación, no tenga un contacto social continuo con el catalán, como veremos a continuación.

A nuestro modo de ver, el quid de la cuestión radica en que la situación sociolingüística del catalán y del español en Cataluña varía muchísimo de una provincia a otra y de una comarca a otra, por no hablar ya de otras comunidades donde conviven también las dos lenguas como Valencia o Andorra. En el ámbito que estudiamos concretamente, el contacto que puede tener un adolescente de Lleida de un pueblo de *l’Alt Urgell* con el castellano puede limitarse a la hora de la asignatura de lengua y literatura castellanas del instituto, mientras que un joven de la misma edad, tomado al azar, del *Barcelonés* puede, por el contrario, escuchar únicamente el catalán en la escuela⁴³⁶.

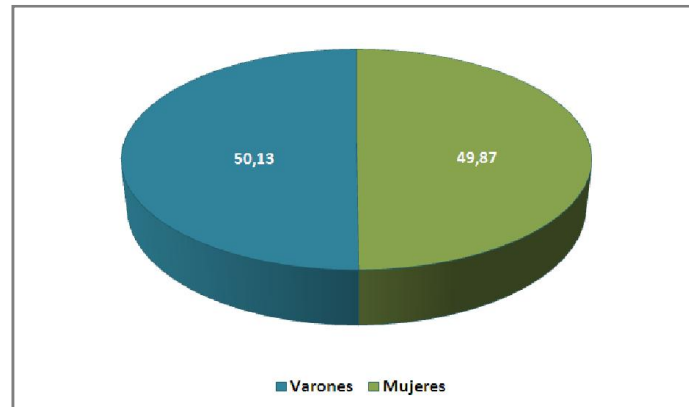
⁴³⁵ Consideramos que la autora se aleja demasiado de la realidad de cualquier lugar de Cataluña, al hablar de delincuencia juvenil y de signos de desintegración social por parte de estas generaciones de inmigrantes y al dibujar unas condiciones casi paupérrimas por lo que respecta a las ciudades dormitorio que rodean la capital: “The provision of schools, hospital and social services was totally inadequate in the dormitory towns. The resulting social and linguistic divisions are marked today by resistance to linguistic and cultural integration, by juvenile delinquency and by other signs of social disintegration which are aggravated further by the level of unemployed (the highest in western Europe) that exists everywhere in Spain” (Hoffman 1991: 271). Con respecto a la desintegración social, muchas veces se ha señalado que, por el contrario, en la sociedad catalana, la lengua no implica ninguna división en grupos etnolingüísticos (Sinner 2004: 21).

⁴³⁶ Hernández García (1998a: 78) presenta ejemplos similares argumentando las causas de la aparición de interferencias lingüísticas en las dos direcciones.

Por esta razón, realizamos, a continuación, una fotografía de la realidad social y lingüística de nuestra provincia en el momento de aplicación del estudio para la que nos basamos en el censo general de población del año 2002, en el censo lingüístico de 2001 y en la encuesta de usos lingüísticos llevada a cabo en 2003⁴³⁷, por la proximidad de los datos con el año de aplicación de nuestro estudio.

Como puede observarse en el siguiente gráfico (4.9), la proporción entre mujeres y hombres, en el año 2002, está muy equilibrada en el ámbito territorial de nuestra investigación. El número de varones, 186 026, sobrepasa —aunque muy ligeramente— la cantidad de mujeres, 185 029.

Gráfico 4.9. Distribución de la población de la provincia de Lleida según sexo en porcentajes. Año 2002.

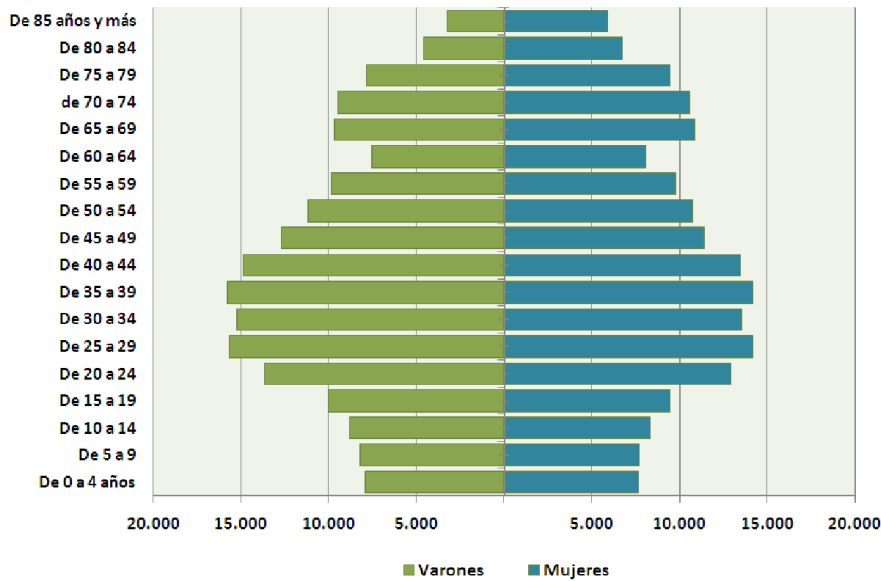


Fuente: elaboración propia a partir de los datos de Idescat.

Esta preponderancia, aunque pequeña, es constante en todas las franjas generacionales y la tendencia solo se invierte al superar los 60 años. La pirámide de población que se perfila, teniendo en cuenta el sexo y los grupos de edad, es la propia de una sociedad desarrollada, con un índice de natalidad bajo y una larga esperanza de vida (gráfico 4.10).

⁴³⁷ Todos los datos han sido obtenidos mediante la consulta de la página electrónica del *Institut d'Estadística de Catalunya* (www.idescat.com), excepto los procedentes de la encuesta de usos lingüísticos de 2003 (*Estadística d'usos lingüístics de Catalunya*, de ahora en adelante *EULC 2003*), que han sido extraídos de *Institut d'Estadística de Catalunya* (2004), cuya referencia completa se halla en el apartado bibliográfico.

Gráfico 4.10. Población según sexo y edad en la provincia de Lleida. Año 2002.

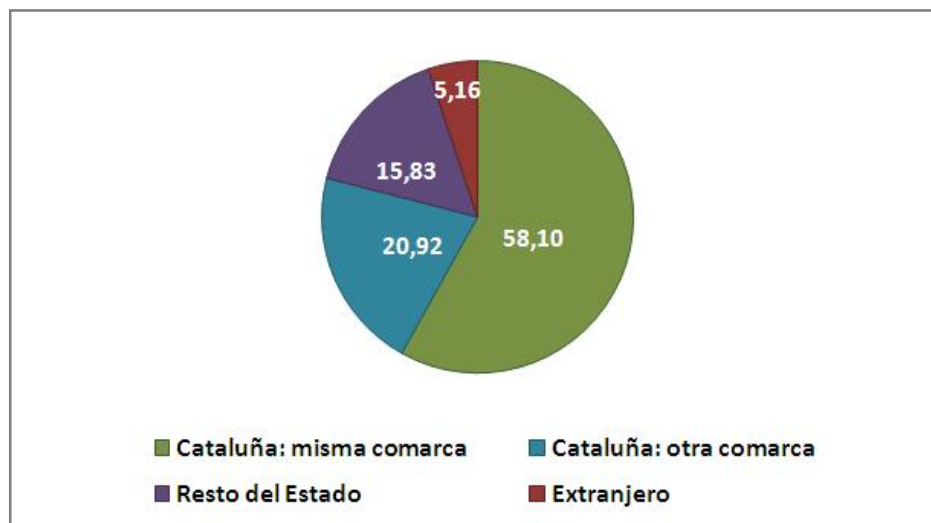


Fuente: elaboración propia a partir de los datos de Idescat.

Como ya mostramos (*vid.* 4.1, gráfico 4.3.⁴³⁸), en Cataluña se da una correlación clara entre la lengua de uso habitual y el origen de la población; de modo que, en términos generales, la población autóctona prefiere utilizar el catalán, mientras que los nacidos en el resto del Estado se decantan, en mayor grado aún, por el uso habitual del castellano. Esta es la causa de que, seguidamente, analicemos con detalle la procedencia de los habitantes de la provincia de Lleida.

⁴³⁸ Recordamos al lector que puede usar el índice de tablas, gráficos y mapas (tras el índice general) para localizar rápidamente cualquier figura o tabla.

Gráfico 4.11. Distribución de la población de la provincia de Lleida según el lugar de nacimiento en porcentajes. Año 2002.

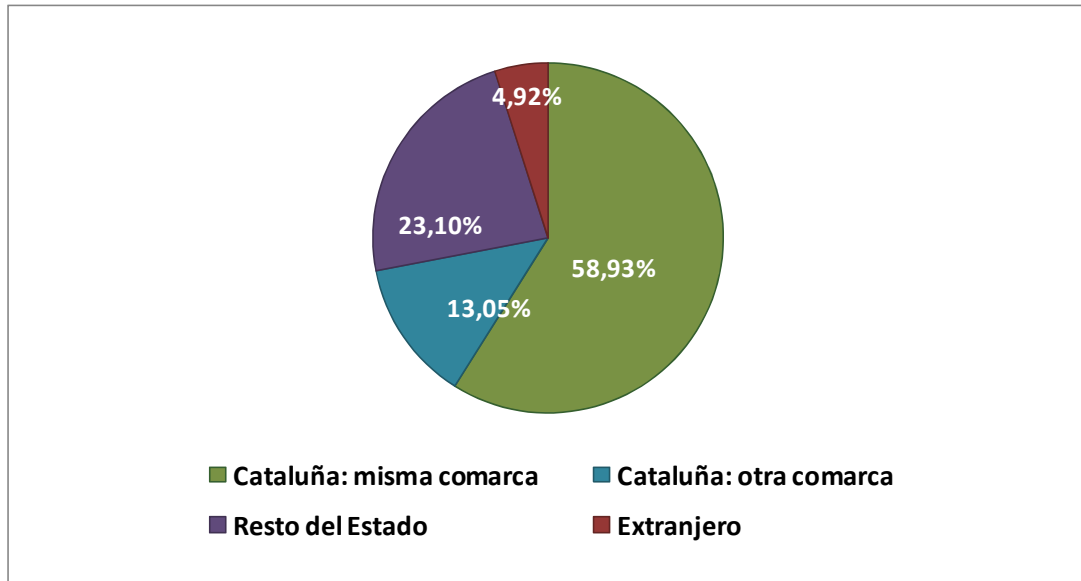


Fuente: elaboración propia a partir de los datos de Idescat.

Como se deduce del anterior gráfico, casi el 80% de los habitantes de Lleida (concretamente, el 79.02%) ha nacido en tierras catalanas, y, además, casi el 60% incluso en la misma comarca en la que reside en la actualidad, lo que nos da un porcentaje global de población autóctona verdaderamente alto. El porcentaje de población inmigrada originaria de otras regiones españolas y, por ende, susceptible de hacer un uso habitual del castellano, se sitúa en torno al 16%. Finalmente, los inmigrantes extranjeros alcanzan cotas bajas, en comparación con los mismos datos en la actualidad, donde alcanzan ya el 18.7% del total de la población de la provincia de Lleida (año 2011).

A continuación (gráfico 4.12), observamos la misma distribución de la población pero teniendo en cuenta, únicamente, el municipio de Lleida. Se constata fácilmente que, mientras la población procedente de la misma comarca del *Segrià* se mantiene, más o menos, en la misma cantidad (58.10% en la provincia y 58.93 en la ciudad); descienden los nacidos en otras comarcas catalanas (20.92% frente a 13.05) y aumentan —notablemente— las personas procedentes del resto del Estado (el 15.83% se convierte en 23.1%). Esta distribución refleja, como ya explicamos, la preferencia por el asentamiento en las ciudades de los inmigrantes castellanohablantes, que llegaron a Cataluña tras la posguerra.

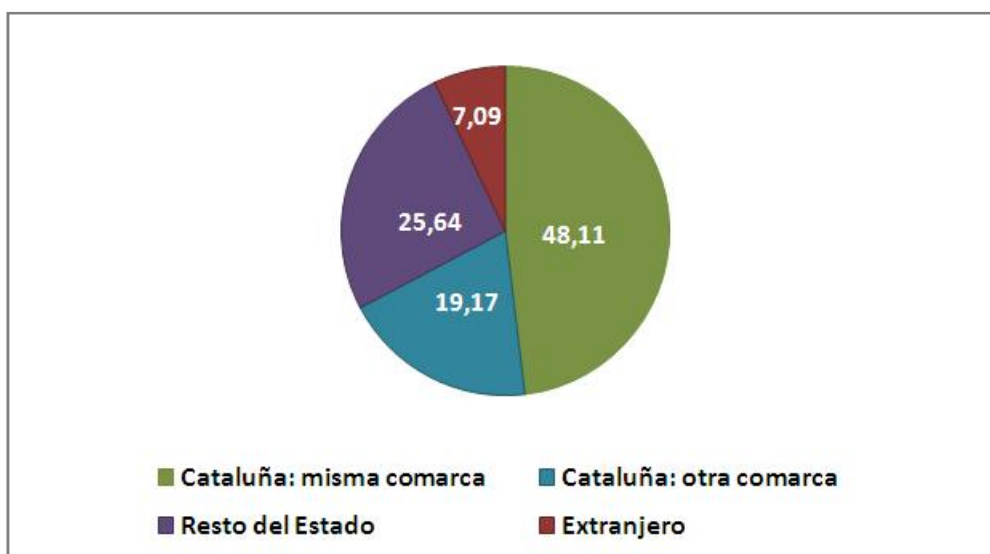
Gráfico 4.12. Distribución de la población del municipio de Lleida según el lugar de nacimiento en porcentajes. Año 2002.



Fuente: elaboración propia a partir de los datos de Idescat.

En el gráfico siguiente, observamos la misma distribución de la población pero respecto al total de población de Cataluña, con la finalidad de observar las diferencias en relación con nuestra provincia:

Gráfico 4.13. Distribución de la población de Cataluña según el lugar de nacimiento en porcentajes. Año 2002.

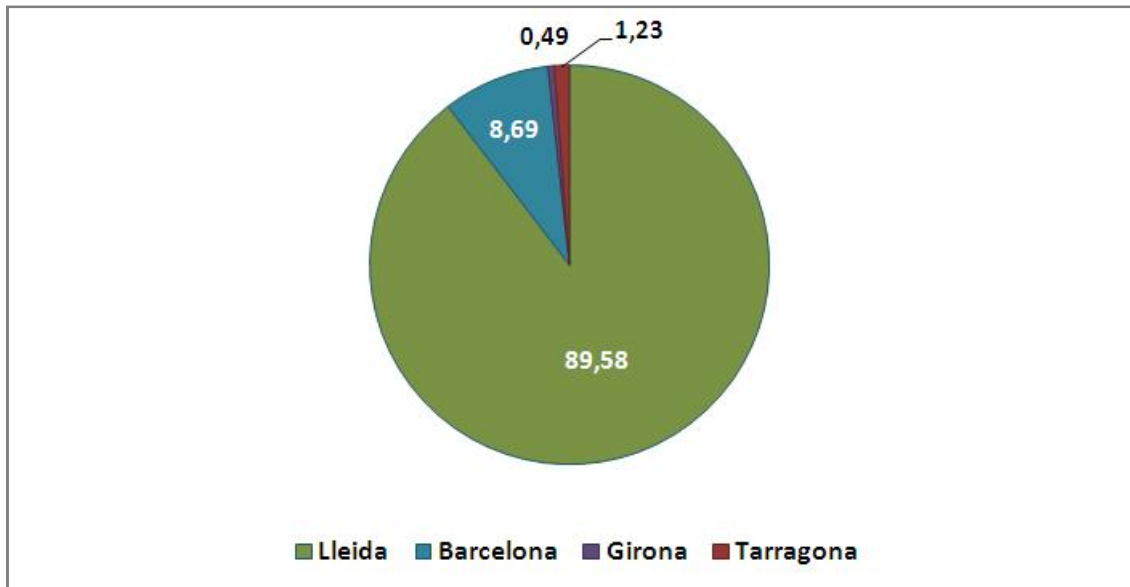


Fuente: elaboración propia a partir de los datos de Idescat.

Como se puede constatar, el porcentaje de población nacida en la misma comarca desciende en el conjunto global de Cataluña, mientras que la población inmigrada, tanto la procedente del resto del Estado como de otros países, asciende. La diferencia mayor se observa respecto a la inmigración de origen castellanohablante, que es, prácticamente, un 10% superior en Cataluña al promedio registrado en la provincia de Lleida (gráfico 4.11), hecho que confirma que este tipo de inmigración no alcanza en nuestro ámbito territorial cotas tan altas como las que se observan en el resto de Cataluña.

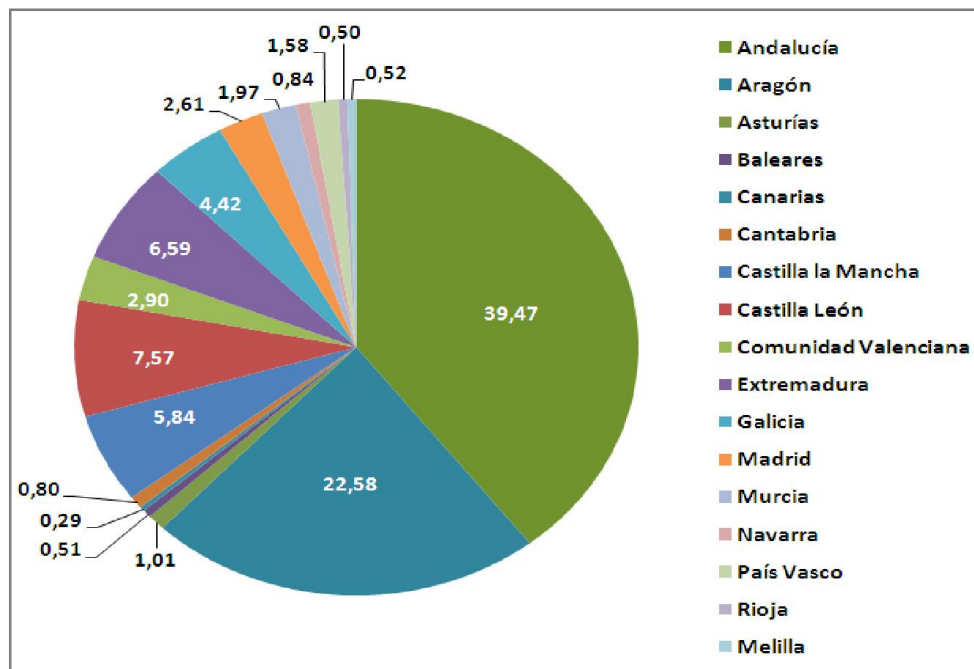
Como se ve en el siguiente gráfico (4.14), la población de la provincia de Lleida originaria de otras provincias catalanas es escasa en relación con la población autóctona, aunque se ha de señalar que mayoritariamente procede de Barcelona. Asimismo, la población procedente del resto de España, que como hemos visto, alcanza casi un 16% del total, no se reparte de forma equitativa sino que la mayoría proviene de tierras andaluzas y aragonesas (gráfico 4.15). La presencia de población originaria de Aragón se explica por la vecindad con la provincia de Huesca, mientras que es larga la tradición de Andalucía como tierra de emigrantes que buscaban en el norte, y más concretamente en Cataluña, las oportunidades laborales que escaseaban en su entorno. Otras comunidades que aportan una cantidad considerable de población son, por este orden, Castilla León, Extremadura, Castilla-La Mancha y Galicia. Estas seis regiones reúnen más del 80% de la población de Lleida nacida en otros lugares de España; mientras que las restantes comunidades tienen una presencia minoritaria dentro de esta inmigración, y más aún en el conjunto global de la población leridana.

Gráfico 4.14. Población de la provincia de Lleida nacida en Cataluña según la provincia de nacimiento en porcentajes. Año 2002.



Fuente: elaboración propia a partir de los datos de Idescat.

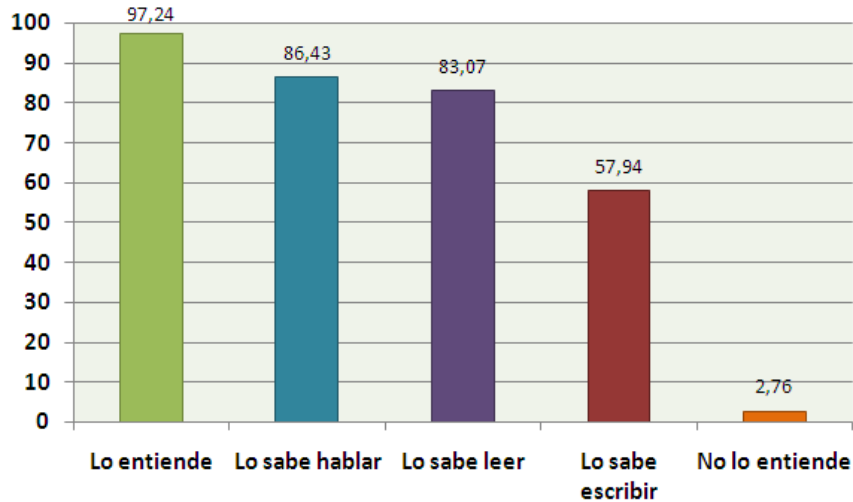
Gráfico 4.15. Población de la provincia de Lleida nacida en el resto del Estado según la comunidad autónoma en porcentajes. Año 2002.



Fuente: elaboración propia a partir de los datos de Idescat.

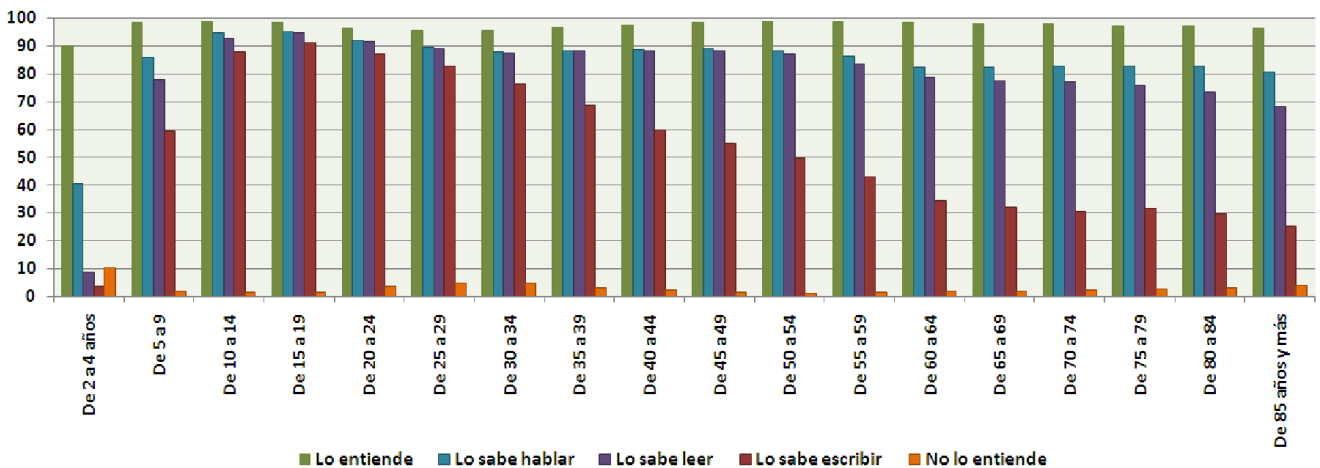
Otra fuente que nos puede ser útil—para reflejar con fidelidad la situación sociolingüística— son los datos que nos proporciona el censo lingüístico de 2001, que versa sobre el conocimiento del catalán⁴³⁹. La situación general para la provincia es la siguiente según los datos generales (gráfico 4.16) y el desglose por franjas de edad (gráfico 4.17):

Gráfico 4.16. Conocimiento del catalán en la provincia de Lleida en porcentajes. Año 2001.



Fuente: elaboración propia a partir de los datos de Idescat.

Gráfico 4.17. Conocimiento del catalán en la provincia de Lleida según la franja de edad en porcentajes. Año 2001.



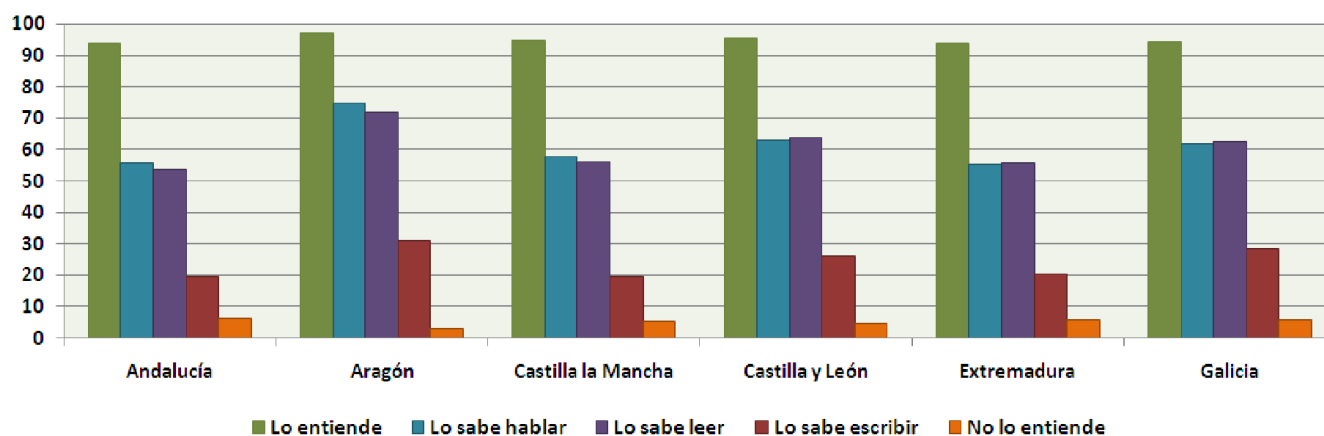
Fuente: elaboración propia a partir de los datos de Idescat.

⁴³⁹ Hasta la encuesta demográfica de 2007 no se solicitan datos referidos al conocimiento de otros idiomas, como el castellano. La razón de ello es que el conocimiento del castellano se supone a toda la población mientras que el del catalán, por el contrario, no. La inclusión de este nuevo parámetro se debe al cambio social que ha supuesto la llegada de población extranjera que se ha instalado en la comunidad autónoma.

Tal como puede observarse, la práctica totalidad de habitantes (97.2%) se declara capaz de entender la lengua catalana y una amplia mayoría es capaz de hablar y leer en catalán. Estas dos destrezas se mantienen constantes en las diferentes franjas de edad, rodeando siempre el 80%, aunque el porcentaje es aún mayor entre la población menor de 30 años, que se incorporaron al sistema educativo cuando ya estaba implantada la normalización lingüística (*vid.* 4.1). La destreza lingüística que cuenta con índices más bajos, en todas las edades, es la escritura —que desciende claramente a medida que aumenta la edad— como consecuencia del escaso índice de alfabetización en lengua catalana que hay, sobre todo, entre la población mayor, fruto de que su escolarización fue, exclusivamente, en lengua castellana.

En el siguiente gráfico, se observa cuál es el conocimiento del catalán que declaran tener las personas procedentes de otros lugares de España. Como ya hemos visto, la cifra total que alcanza este tipo de inmigración dentro del conjunto de población provincial es del 15.83%. Toda vez que, dentro de este grupo, hay un predominio claro de algunas comunidades autónomas, analizamos a continuación solo los datos de las seis comunidades que tienen un mayor peso demográfico en el cómputo global de la provincia.

Gráfico 4.18. Conocimiento del catalán en la provincia de Lleida según la comunidad autónoma de nacimiento en porcentajes. Año 2001.



Fuente: elaboración propia a partir de los datos de Idescat.

La capacidad de comprensión oral de la lengua es muy alta en todos los grupos, aunque la situación es distinta respecto a la expresión oral y a la comprensión lectora.

La escritura en lengua catalana registra, con mucha diferencia sobre las otras capacidades, los peores índices. Estos datos parecen indicar que la integración lingüística de estos colectivos, al menos en el nivel oral pasivo, ha sido efectiva. Los mejores porcentajes, en relación con todas las capacidades, los consiguen los aragoneses, seguidos de la población procedente de Castilla León y de Galicia. Una parte de la provincia de Aragón, donde se habla catalán, colinda con Lleida, lo que podría explicar los altos índices que registra esta población —cuyo peso es importante demográficamente en la provincia— en todas las destrezas lingüísticas, si se comprobara que, efectivamente, la mayoría de aragoneses procede de esa área lingüística, denominada *la Franja de Ponent*. No obstante, es importante señalar, también, que, para poder valorar la información que nos proporciona este gráfico adecuadamente, tendríamos que tener en cuenta muchos otros matices como el año de llegada a la provincia, la edad o la posición socioeconómica de cada grupo poblacional, factores que pueden influir en la necesidad de aprender o no la lengua autóctona.

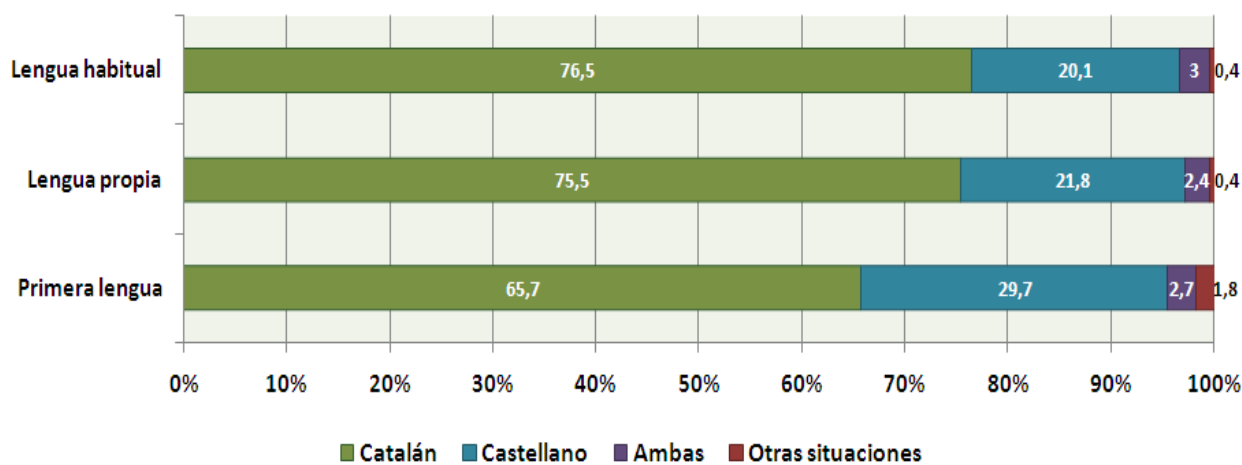
Con respecto a los usos de la lengua, la fuente más próxima al año en el que se construyó el corpus de nuestro estudio es la recopilación estadística de usos lingüísticos de Cataluña (*EULC*) que llevó a cabo el *Institut d'Estadística de Catalunya* en el año 2003⁴⁴⁰. Los datos de la encuesta proceden de una división territorial de Cataluña en siete ámbitos no coincidentes, obviamente, con las cuatro provincias⁴⁴¹. Por esta razón, presentamos a continuación, en diferentes gráficos, los resultados correspondientes al *Àmbit de Ponent*, de un lado, y al *Alt Pirineu i Aran*, de otro, que son los dos ámbitos que aglutina nuestra provincia. En el *Àmbit de Ponent* se engloban las comarcas de *la Noguera, la Segarra, l'Urgell, el Pla d'Urgell, el Segrià y les Garrigues* (esto es, de la 8 a la 13 en el mapa 4.2), y en *Alt Pirineu i Aran*, las de *la Vall d'Aran, el Pallars Sobirà, el Pallars Jussà, l'Alt Urgell, la Cerdanya y l'Alta Ribagorça* (de la 1 a la 6 en el mismo mapa). La comarca del *Solsonès* queda fuera de estos dos ámbitos y se enmarca dentro de las comarcas centrales, con lo que sus datos no se verán reflejados en los siguientes gráficos⁴⁴².

⁴⁴⁰ Con fecha anterior a este año, únicamente contamos con los datos del censo lingüístico que son sobre capacidades, no sobre usos del idioma y que, además, solo se obtenían sobre la lengua catalana.

⁴⁴¹ La ley 24/2001 de 31 de diciembre establecía dicha división territorial.

⁴⁴² Recuérdese lo dicho en el apartado anterior (*vid.* 4.3.1) respecto a que el municipio de Solsona orienta sus relaciones hacia Manresa, en la provincia de Barcelona. Por otro lado, en los siguientes gráficos, el lector atento reparará en que algunos porcentajes suman 99.9 o 100.1. Este mismo hecho se detecta en la

Gráfico 4.19. Población del Àmbit de Ponent según primera lengua, lengua propia y lengua habitual, en porcentajes. Año 2003.



Fuente: elaboración propia a partir de los datos de la EULC 2003.

En relación a cuál es la primera lengua, la lengua propia y la lengua de uso habitual⁴⁴³, observamos que en el *Àmbit de Ponent* hay una preponderancia clara de la lengua catalana, que alcanza índices bastante altos en las tres opciones mencionadas, hecho que demuestra la amplia presencia del catalán en estas comarcas y que contrasta, fuertemente, con la situación general y actual de la comunidad autónoma, ya que como vimos en el primer apartado (*vid.* 4.1, gráfico 4.2), en Cataluña, el castellano sobrepasa al catalán en estas tres magnitudes⁴⁴⁴. Este hecho se relaciona, claramente, con la mayor presencia en esta provincia de población autóctona, como acabamos de ver en los gráficos anteriores. La situación más favorable al catalán se da en relación con la lengua propia del informante y con la lengua habitual de uso, donde más de tres, de cada cuatro informantes, prefieren el catalán antes que el castellano. Cabe destacar un hecho que también se observaba en el gráfico referido al total de la comunidad (gráfico 4.2), a saber, que existe un cierto número de personas que, a pesar de nacer en familias castellanohablantes, acaban identificándose con el catalán y utilizándolo en sus interacciones cotidianas, tal como se aprecia en el gráfico 4.19 al comparar los

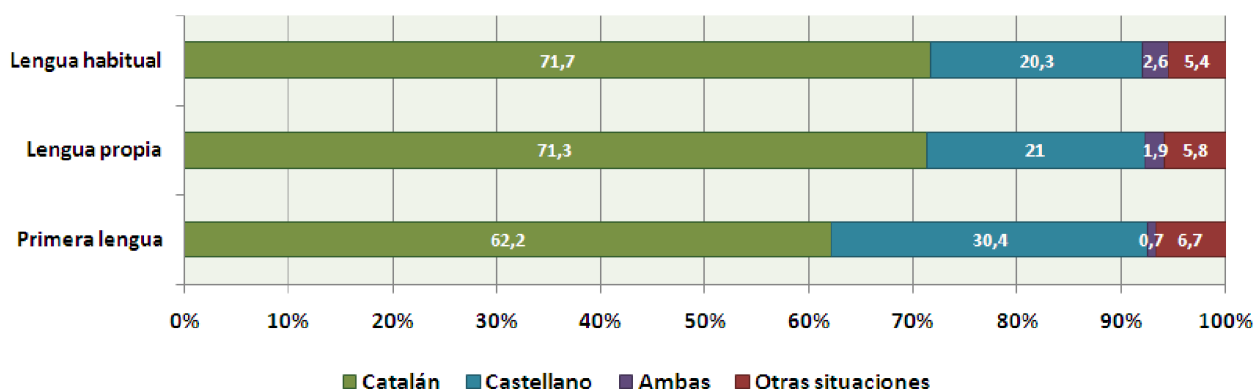
fuente de donde hemos extraído los datos (*Institut d'Estadística de Catalunya* 2004) y se debe a que se consigna solo un decimal en los tantos por ciento.

⁴⁴³ La primera lengua es la que la persona entrevistada habló primero en su casa; la lengua propia es aquella con la que se identifica y considera suya. Finalmente, la lengua habitual es la que el informante usa con más frecuencia.

⁴⁴⁴ Esto es debido a que el predominio del castellano se da en las áreas circundantes a la capital, donde el peso demográfico es también mayor, como también constatamos en 4.3.1. Véase también la nota 448 al respecto.

porcentajes que obtienen las dos lenguas en las tres variables analizadas. Por otra parte, son muy pocos los informantes que se decantan por el uso de las dos lenguas por igual, la identificación con ambas o que reconocen haber empezado a hablar las dos simultáneamente, y menos aún los que optan por otras lenguas en las tres magnitudes estudiadas⁴⁴⁵.

Gráfico 4.20. Población de l'Alt Pirineu i Aran según primera lengua, lengua propia y lengua habitual, en porcentajes. Año 2003.



Fuente: elaboración propia a partir de los datos de la EULC 2003

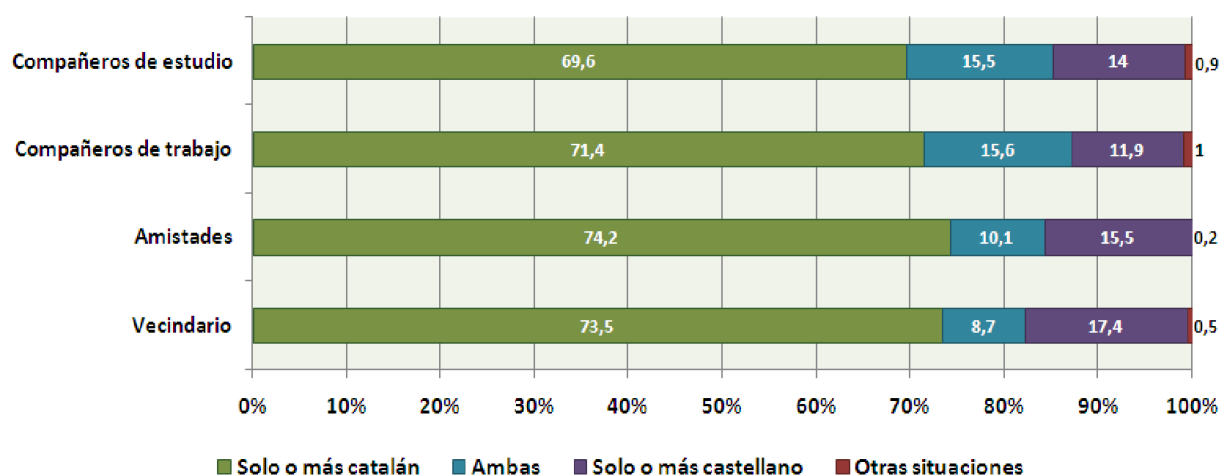
En las comarcas del norte de la provincia, la situación es bastante similar a la que acabamos de ver, con un predominio claro de la lengua catalana. Sin embargo, los porcentajes menores a favor del catalán en los tres parámetros no se corresponden aquí con un crecimiento del castellano —que aumenta muy levemente—, sino que se explican por el mayor uso del aranés, que queda registrado en “otras situaciones”. Se advierte, como en el gráfico anterior, que una parte del contingente migratorio castellano hablante o de sus descendientes se identifica y usa el catalán a lo largo de su vida, mientras que la opción de ambas lenguas en las distintas variables es aún menor que en el gráfico anterior.

En cuanto a la presencia de las dos lenguas en las relaciones personales, se observa un crecimiento del porcentaje de informantes que se decantan por el uso de las dos lenguas, con respecto a los parámetros anteriores. No obstante, continúa dominando el catalán en estos ámbitos, sobre todo en el referido a las amistades, al vecindario y a

⁴⁴⁵ Este es el cambio más importante entre la encuesta de 2003 y la de 2008. En esta última, la proporción de hablantes de otras lenguas crece espectacularmente (*cf.* gráfico 4.2), como consecuencia del peso demográfico que adquiere la inmigración reciente llegada de otros países.

los compañeros de trabajo. El menor índice de uso del catalán, aunque continúa predominando frente al castellano, se observa con respecto a los compañeros de estudio, quizá debido al porcentaje de inmigrantes castellanohablantes o de sus descendientes que, como ya hemos comentado, optan por el catalán a lo largo de su vida. Quizá esta circunstancia no se produzca aún durante los primeros años, cuando se lleva a cabo la formación académica del individuo y, por ende, cuando se inician las amistades con los compañeros de estudio. En el ámbito de *l'Alt Pirineu i Aran*, se observan las mismas tendencias ya señaladas, y se constata el crecimiento del porcentaje de otras situaciones, que engloba el uso del aranés en *la Vall d'Aran*.

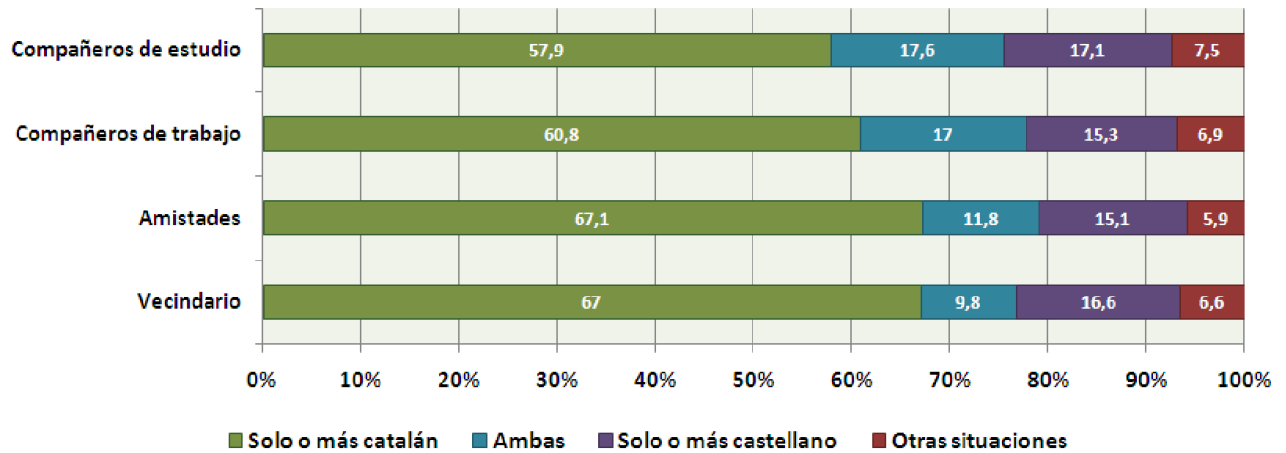
Gráfico 4.21. Usos lingüísticos en ámbitos interpersonales en l'Àmbit de Ponent, en porcentajes. Año 2003.



Fuente: elaboración propia a partir de los datos de la EULC 2003⁴⁴⁶.

⁴⁴⁶ En este gráfico y en los tres siguientes, hemos unido las dos categorías establecidas en la encuesta como “solo catalán” y “más catalán que castellano”, por un lado, y “solo castellano” y “más castellano que catalán”, por otro, en una sola categoría de “solo o más catalán” y “solo o más castellano”.

Gráfico 4.22. Usos lingüísticos en ámbitos interpersonales en l'Alt Pirineu i Aran, en porcentajes. Año 2003.

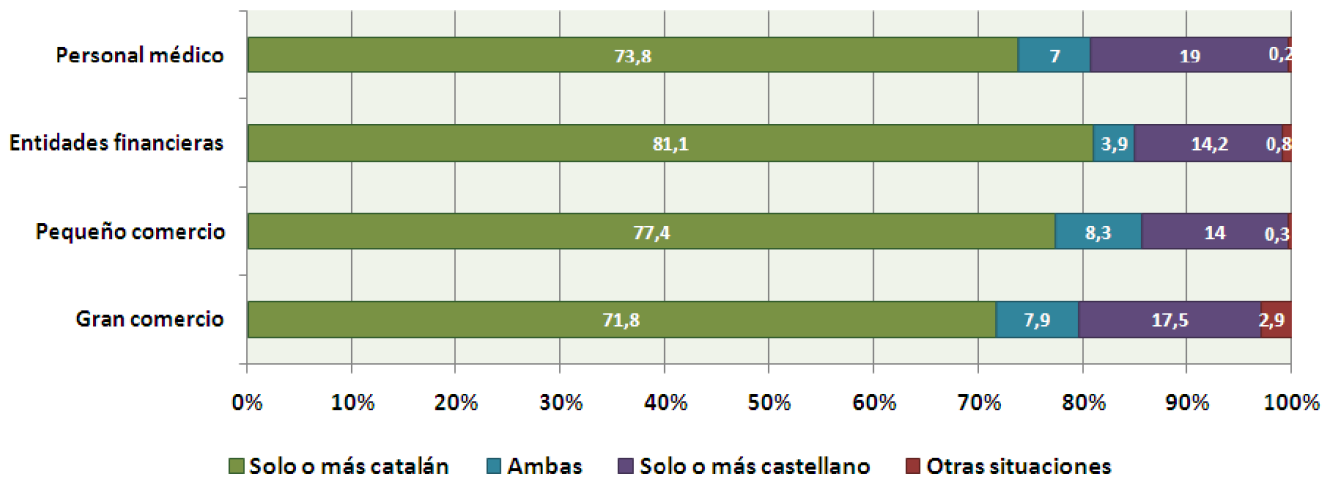


Fuente: elaboración propia a partir de los datos de la EULC 2003.

En las relaciones de servicios, continúa observándose el dominio del catalán, que alcanza su cota más elevada en el terreno de las entidades financieras. En este lugar, además, en los dos ámbitos de la provincia, se da un empleo inferior del uso indistinto de ambas lenguas, lo que equivale a decir que los informantes prefieren decantarse por una u otra, aunque mayoritariamente eligen el catalán. En *l'Alt Pirineu i Aran*, se deduce, en estos contextos comunicativos, un mayor uso del aranés, que, sin embargo, desciende en los restantes espacios. Entre el pequeño y el gran comercio también se observan diferencias interesantes: aunque predomina el catalán en los dos campos, se usa más esta lengua en el pequeño comercio; en el gran comercio, por el contrario, se registra el índice menor de uso del catalán en los dos ámbitos provinciales. El catalán también predomina en las relaciones con el personal médico.

Gráfico 4.23. Usos lingüísticos en ámbitos de consumo y servicio en l'Àmbit de Ponent, en porcentajes.

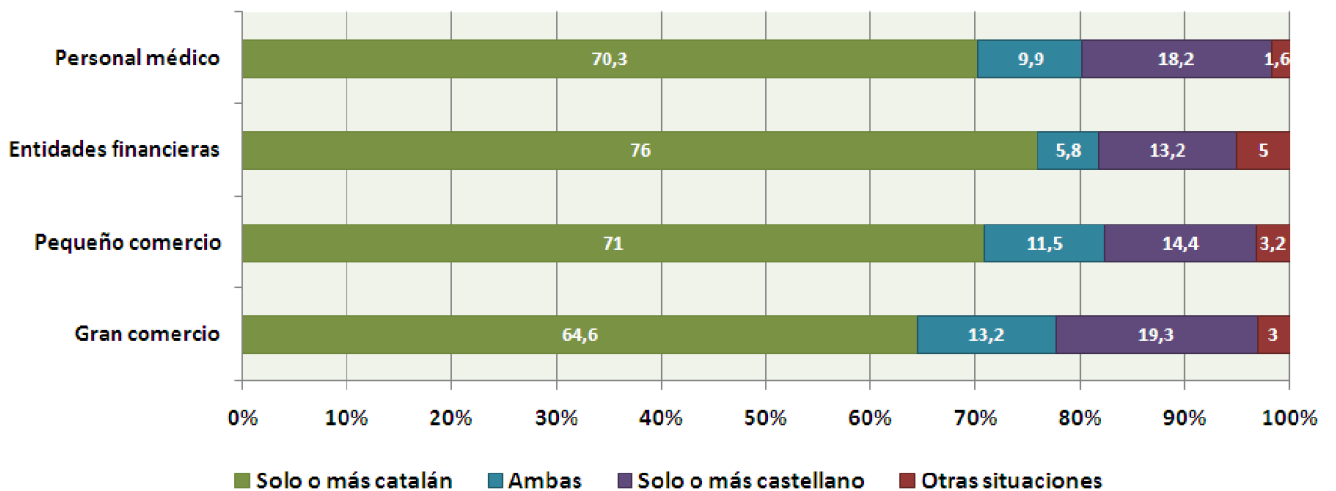
Año 2003.



Fuente: elaboración propia a partir de los datos de la EULC 2003.

Gráfico 4.24. Usos lingüísticos en ámbitos de consumo y servicio en l'Alt Pirineu i Aran, en porcentajes.

Año 2003.

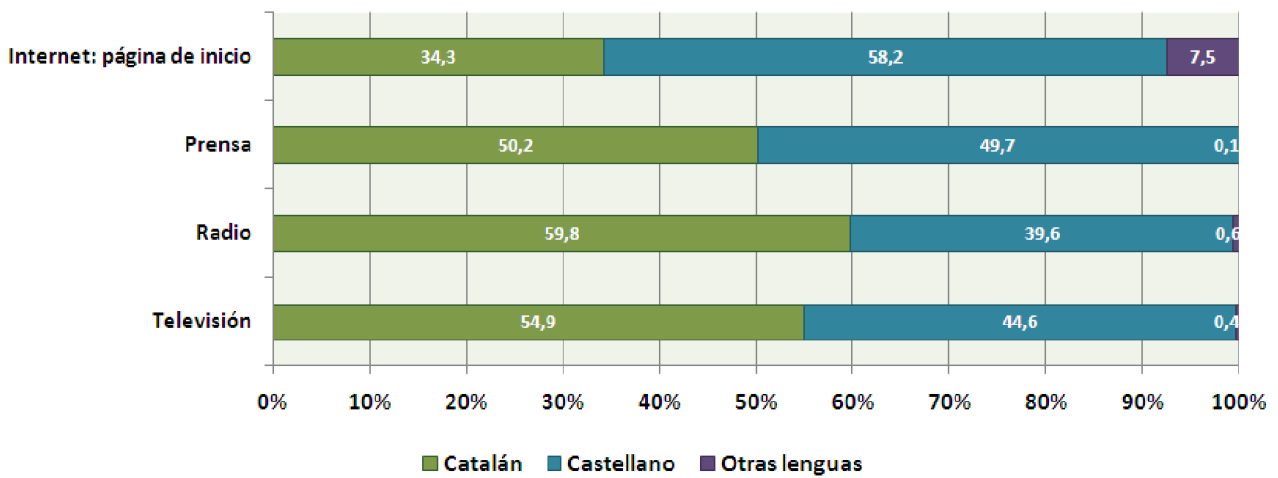


Fuente: elaboración propia a partir de los datos de la EULC 2003

En relación con los medios de comunicación, se constata un uso más proporcional de las dos lenguas, aunque continúa el predominio del catalán en todos los contextos, con la única excepción de internet. La lengua escogida varía significativamente en función del medio del que se trate, de este modo, la radio registra los mejores resultados para el catalán, sobre todo en l'Alt Pirineu i Aran. En la

televisión, aunque los datos apuntan más hacia el equilibrio, también se prefiere la lengua catalana. Con respecto a internet, es notable la escasa presencia del catalán, pero también el crecimiento del uso de otras lenguas, que corresponde, principalmente, al inglés⁴⁴⁷. En la prensa escrita se da el uso más equilibrado de las dos lenguas, aunque en este punto hay que tener en cuenta que parte de la población de más edad no está alfabetizada en catalán.

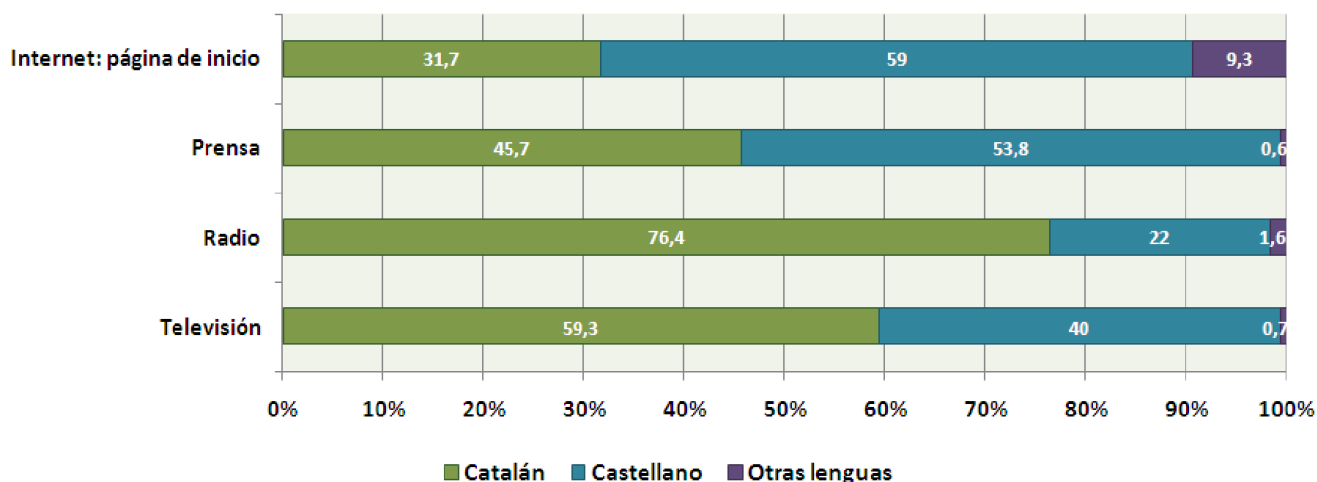
Gráfico 4.25. Usos lingüísticos en el consumo de los medios de comunicación en l'Àmbit de Ponent.
Media de horas. Año 2003.



Fuente: elaboración propia a partir de los datos de la EULC 2003.

⁴⁴⁷ En la tabla original, se diferencia el inglés de otras lenguas, aquí hemos sumado esos dos valores y los hemos considerado como una única categoría.

Gráfico 4.26. Usos lingüísticos en el consumo de los medios de comunicación en l'Alt Pirineu i Aran.
Media de horas. Año 2003.



Fuente: elaboración propia a partir de los datos de la EULC 2003.

Como conclusión a todos los datos que acabamos de aportar, podemos afirmar que el territorio que estudiamos se caracteriza, para empezar, por un alto índice de población de origen catalán, —concretamente el 79.02% declara haber nacido en Cataluña—. Además, dicha población es autóctona, ya que, de este 79.02%, prácticamente el 90% (el 89.58%) es originario de la misma provincia de Lleida e, incluso, en no pocas ocasiones (en el 58.10%), sigue residiendo en la misma comarca donde nació. El porcentaje de inmigración castellanohablante, proveniente sobre todo de Andalucía y Aragón, es menor que el que se registra en el conjunto de Cataluña, y es de apenas un 16% (un 15.83%). La mayor presencia de población autóctona, y, por tanto, catalanohablante, tiene como consecuencia que el catalán registre mejores resultados como lengua propia, como lengua de uso habitual y como primera lengua en el conjunto general de la población. Estos mismos índices, en cambio, otorgan, contrariamente, un resultado favorable al castellano, en la actualidad, en los datos generales de Cataluña⁴⁴⁸. El catalán impera, en la provincia, tanto en los usos interpersonales como en los ámbitos de consumo y servicio. La situación solo se equilibra, en relación con el castellano, respecto al uso lingüístico en los medios de comunicación, donde también se ha de tener en cuenta que la oferta en catalán es siempre inferior a la del español. Todas estas

⁴⁴⁸ Datos de la encuesta de usos lingüísticos de la población de 2008. Sin embargo, los datos de la encuesta de usos lingüísticos de 2003 sobre el global de Cataluña también mostraban un predominio de la lengua catalana, que aventajaba al castellano, aunque tímidamente, como lengua propia y lengua habitual. Como lengua materna o primera lengua, sin embargo, el castellano ya tenía un índice superior al catalán (53.5% frente al 40.4%).

razones nos llevan a ser tajantes al afirmar que la influencia del catalán sobre el castellano será notable en nuestra provincia, como consecuencia de la fuerte presencia, en todos los ámbitos, de la lengua catalana.

Como colofón a este capítulo, pasamos ahora a describir el catalán que se habla en Lleida, toda vez que, cuando entran en contacto dos variedades lingüísticas, no lo hacen sus respectivos estándares, sino dos subvariedades concretas. El castellano hablado en Lleida tiene unas peculiaridades a las que ya nos hemos referido (*vid.* 4.2), y por esta razón, veremos ahora, *grosso modo*, las características del catalán en Lleida, que difieren según el lugar concreto de la provincia al que hagamos alusión.

Como es sabido, el catalán se divide en dos grandes bloques diatópicos, el catalán oriental y el occidental, cuya principal diferencia radica en el diferente tratamiento que se da al vocalismo átono⁴⁴⁹. Los dialectos orientales neutralizan en posición átona las vocales *a/e* dando lugar a la vocal neutra [ə]; mientras que en occidental, en cambio, se mantiene la diferencia de timbre entre las dos vocales. De forma paralela, las vocales *o/u* átonas se neutralizan en [u] en catalán oriental, mientras que en catalán occidental se diferencian. En la mayor parte de la provincia de Lleida predomina una variedad del catalán occidental denominada *nord-occidental*⁴⁵⁰. Este dialecto se extiende por las comarcas de *la Ribargorça*, *el Pallars*, *l'Alt Urgell*, *la Noguera*, la mitad occidental del *Solsonès*, *la Segarra*, *l'Urgell*, *el Segrià*, y *les Garrigues*⁴⁵¹, pero, además, también comprende Andorra, las comarcas aragonesas de La Litera y Bajo Cinca, y la comarca tarraconense del *Priorat* (Veny 1984: 120-121).

⁴⁴⁹ Aunque este es el principal criterio que se usa para discernir entre los dialectos que se engloban en cada uno de los dos bloques, existen también otras diferencias de tipo morfosintáctico y léxico entre las dos variedades.

⁴⁵⁰ El otro dialecto existente del catalán occidental es el *valenciano*, que también cuenta, como el *nord-occidental*, con distintos subdialectos (Veny 1984). A pesar de que la primera documentación en catalán procede, mayoritariamente, de la zona del catalán *nord-occidental*, la dialectología catalana apenas le ha dedicado atención (Casanovas y Creus 1999: 83).

⁴⁵¹ El *Pla d'Urgell* también entra dentro de este dialecto. Veny (1984) no lo cita debido a que la comarca fue creada en fecha posterior. Del mismo modo, cuando el autor habla del *Pallars*, se refiere al *Pallars Sobirà* y al *Pallars Jussà*.

Mapa 4.3. Extensión del catalán nordoccidental y de sus distintos subdialectos.



Fuente: Armengol et al. (2007: 288).

El catalán *nord-occidental* se caracteriza por tener un sistema vocálico tónico de cuatro grados de abertura y siete vocales. El vocalismo átono es el propio del catalán occidental, que mantiene la diferencia de timbre entre *a/e* y *o/u*, y tiene como rasgo eminentemente característico de la variedad la tendencia al cierre de la [a] final de palabra que se convierte en una vocal media, en una [ɛ] abierta. En el nivel morfológico, se usan las formas arcaicas del artículo masculino *lo/los*, y la desinencia *-o* (pronunciada o cerrada) de la primera persona del singular, y la desinencia *-e* (pronunciada e cerrada) de la tercera persona del singular en los verbos de la primera conjugación (Badia Capdevila 2002: 83-83). Además de arcaísmos caídos en desuso en otros dialectos, son propias del *nord-occidental* palabras como *moixó* (“ocell”), *silló* (“càntir”), *sangartalla* (“sargantana”)... Parte del léxico es compartido con el aragonés y con el balear, y, además, hay ciertos significantes que adoptan significados distintos a los de la lengua normativa o general como *popa* (“mamella”) o *padri* (“avi”) (Veny 1984: 137)⁴⁵².

⁴⁵² Para más información sobre el *nord-occidental* puede consultarse Colomina (1999: 223-226). Este autor, además, ofrece un compendio bibliográfico de todos los dialectos del catalán y divide la información en los distintos niveles de análisis lingüístico.

Aunque esta variedad es la que abarca una mayor extensión provincial, es posible hallar otros subdialectos del catalán *nord-occidental* como el *pallarès*, el *ribargorçà* y el *lleidatà*. El primero de todos ellos se localiza en la región natural e histórica del *Pallars*, esto es, en *el Pallars Sobirà* y en *la Vall Fosca del Pallars Jussà*, y es uno de los dialectos más conservadores del catalán, debido a las condiciones geográficas de los valles donde se habla. Se caracteriza por la abertura de la /e/, la no palatalización del grupo -NN-, el artículo plural masculino *les* y los imperfectos verbales en *-eva* e *-iva* (Veny 1984: 138-142).

Por su parte, el *ribargorçà* se habla en los valles de los ríos *Isàvena* y *Noguera Ribargorçana* —en *la Ribargorça*— y en La Litera, dentro ya del territorio oscense. En esta área lingüística, que va desde la población de Benasque hasta el sur de Tamarite de Litera, se extiende este subdialecto caracterizado por la palatalización de los grupos iniciales CL-, PL- y FL-, los arcaísmos fonéticos y el plural femenino en *-as*. Dentro de esta zona se distinguen distintas hablas, como el *benasqués*, consideradas de transición entre el catalán y el aragonés⁴⁵³.

Aunque en un principio *lleidatà* se usó de forma errónea como sinónimo de catalán *nord-occidental* —por asimilación con la provincia que lleva el mismo nombre— hoy en día está claro que se conoce así el catalán que se habla en Lleida capital y su zona de influencia (Veny 1984: 121); es decir, básicamente en las comarcas del *Segrià*, *l'Urgell*, *el Pla d'Urgell*, y buena parte de *les Garrigues*, de *la Noguera* y del Bajo Cinca (Turull 1990: 201). Aún no existe una descripción completa de esta variedad, ni tampoco está clara su delimitación geográfica; no obstante, algunos de sus rasgos más característicos son, en la fonética, la pronunciación [ɛ] de toda /a/ átona y [a] de /e/ inicial trabada, la diptongación en [au] de /o/ inicial, la aparición de una /n/ epentética en palabras como *mangrana*, o la realización africada de las consonantes prepalatales en posición inicial y postconsonántica [tʃokolátɛ]. En la morfosintaxis, son típicos de la zona, además del artículo *lo*, *los*, los posesivos femeninos *meua*, *teua* y *seua*, el plural masculino en *-ns*, el uso del futuro con valor imperativo y la tendencia a

⁴⁵³ La adscripción del *benasqués* al catalán o al aragonés es un tema ampliamente debatido en la bibliografía, sobre el que no hay aún un acuerdo común. También denominada “microlengua”, el *benasqués* tiene, además, ciertas coincidencias con el gascón. En Colomina (1999: 85-86) pueden verse algunas opiniones al respecto.

no realizar la concordancia en el participio verbal (Turull 1990: 203). En cuanto al léxico, Turull (1990: 204-205) ofrece un interesante elenco de términos en función de su grado de vitalidad en el territorio, así por ejemplo *panís, safranòria, xop, carderola, corder, tocino, silló, arena, pastisser, enguany, melic o torre* son vocablos de absoluta vigencia según este autor. Existen algunos recuentos del vocabulario de la zona, como el *Diccionari de Lleidatanismes* (Massana 2004) o el más reciente *Lo nou diccionari lleidatà-català* (Masip *et al.* 2010), si bien ambas iniciativas cuentan con el inconveniente de carecer de un estudio filológico serio que las sustente⁴⁵⁴. Con respecto a la delimitación territorial, Massana (2004) no indica de forma precisa qué área abarca su diccionario, por lo que se desconoce si usa *lleidatà* con un sentido restrictivo o como sinónimo de toda la provincia⁴⁵⁵. Por su parte, Masip *et al.* (2010) sí que indican los límites geográficos de su investigación, que coinciden, más o menos, con Lleida capital y su zona de influencia⁴⁵⁶. Por otro lado, también existen vocabularios concretos de las hablas y de los subdialectos de la provincia que acabamos de mencionar⁴⁵⁷.

Una vez trazado el panorama dialectológico general, hemos de llamar la atención sobre el hecho de que la provincia no limita al oeste directamente con el castellano, sino con el área lingüística extendida verticalmente entre las comunidades autónomas de Cataluña y Aragón denominada *la Franja de Ponent*. Esta área comprende cinco comarcas, cuatro de ellas aragonesas —la Ribargorza, la Litera, el Bajo Cinca y el Matarraña—, y una —*l'Alta Ribagorça*— dividida entre las dos

⁴⁵⁴ *El diccionari de lleidatanismes* es fruto de la experiencia personal de su autor. Massana (2004: 4-5) afirma que no pretende ser exhaustivo y que algunas palabras solo las ha oído nombrar a una persona, como, por ejemplo, su abuela o su padre. Masip *et al.* (2010: 14-15) aclaran que su diccionario no pretende ser un tratado serio de lexicología, sino una herramienta de entretenimiento. Los autores, además, explican, con cierta ironía, el “ampli procés d’investigació i documentació” consistente en ir en coche a distintas poblaciones, oír a la gente hablar y anotar las palabras en unos papeles que, a su vez, eran introducidos en bolsas de supermercado. Si la palabra en cuestión —con la acepción correspondiente— no estaba aceptada en el catalán normativo, la rescataban, si no, la dejaban en la bolsa de plástico. A pesar de lo rudimentario del método, su diccionario, con más de 3000 palabras, es el intento más exhaustivo que existe en la actualidad de recopilación del léxico de la zona. Antes de ver la luz, el *diccionari lleidatà-català*, de los mismos autores, se difundió con mucho éxito a través de internet mediante la web: www.tremendos.net/loDiccionari.htm.

⁴⁵⁵ El autor recoge las palabras que oyó de pequeño cuando recorrió “una bona part de la geografia lleidatana” debido a que sus padres eran maestros (Massana 2004: 3).

⁴⁵⁶ “Finalment vam decidir-nos per la clamor d’Almacelles i la Noguera Ribagorçana per la banda oest, per l’Ebre al seu pas pel terme municipal d’Almatret per la banda sud, la Panadella per la banda est i pel nord vam tallar pel Montsec d’Ares i de Rúbies” (Masip *et al.* 2010: 14).

⁴⁵⁷ Por ejemplo, hay léxicos sobre el *pallarès*, como el de Coll (1991), el de Sistac (1998) o el de Ramoneda (coord. 2007). Sobre las hablas extremo-occidentales de *la Franja*, puede consultarse Galan y Moret (1995), Navarro (2005) o Sistac (1993). Agradecemos desde aquí la información proporcionada amablemente al respecto por el profesor Ramon Sistac, de la Universidad de Lleida.

comunidades autónomas. Si observamos con atención el mapa anterior (*supra*, mapa 4.3), se ve claramente cómo el catalán *nord-occidental* rebasa, con creces, el límite provincial. La situación sociolingüística en estos municipios, que pertenecen administrativamente a Aragón pero son de habla catalana, es peculiar ya que una mayoría amplia de la población habla catalán, no obstante, la conciencia lingüística al respecto es débil, aunque este hecho parece estar cambiando en la actualidad (Badia Capdevila 2002: 83). Por otro lado, los datos más recientes señalan que es precisamente en esta área donde, de entre todos los territorios de habla catalana —esto es, Cataluña, Andorra, las Islas Baleares, Alguer, Valencia y la denominada Cataluña Norte—, el porcentaje de población que es capaz de entender y hablar el catalán es más alto⁴⁵⁸. A pesar de la vitalidad que tiene la lengua catalana en esta zona, los datos también indican que, en la transmisión intergeneracional de la lengua, es decir, en la transmisión de padres a hijos, el catalán no avanza (Querol 2007).

Así como la frontera que separa el catalán *nord-occidental* del aragonés en el oeste de la provincia no es nítida, tampoco lo es la que separa la variedad occidental del catalán oriental. Concretamente, la provincia limita al este con el catalán central, dialecto del catalán oriental dotado de un gran prestigio al ser la lengua hablada en la capital de la comunidad autónoma, cuya principal característica es la reducción del vocalismo átono en: [ə] (>a/e), [i] y [u] (>o/u).

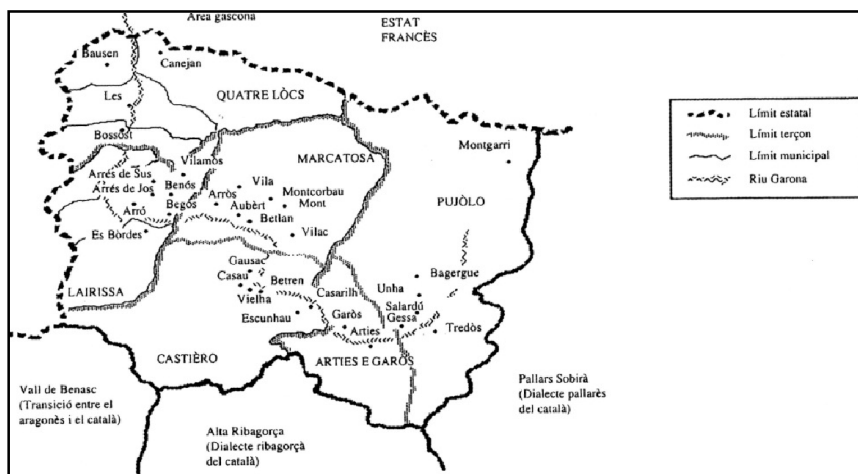
Dado que los límites territoriales no coinciden con las isoglosas lingüísticas, algunas zonas de Lleida —como *la Cerdanya*— se hallan de pleno dentro del bloque del catalán central, y en otras —como *l'Alt Urgell* o *el Solsonès*— se habla esta variedad en parte de su territorio, aunque, como ya hemos dicho, la separación entre los dos zonas lingüísticas no es tajante. Además, estas comarcas, y algunas otras como la *Segarra*, *l'Urgell* o *les Garrigues*, limitan con la variedad de transición denominada *xipella*. Se trata de una isoglosa discontinua —correspondiente a la [i] resultado del cierre de la /e/ final átona—, que baja desde *l'Alt Urgell* hasta *la Conca de Barberà*, ya en la provincia de Tarragona (Veny 1984: 49).

⁴⁵⁸ El primer índice se sitúa en el 98.5%, mientras que la expresión oral está en el 88.8%. Cataluña y Andorra, por este orden, son las otras dos demarcaciones que registran los mejores índices a continuación de *la Franja*. Sin embargo, la situación cambia con respecto a las capacidades escritas: en estas, Cataluña, seguida de Andorra, superan esta área (Querol 2007).

En el noroeste de la provincia, hallamos el aranés, variedad de la lengua occitana propia de *la Vall d'Aran*. En este pequeño territorio, de 620 km², se habla esta modalidad del gascón —que, a su vez, es un dialecto principal del occitano— desde el siglo XI. Si se observa el mapa 4.3. (*supra*), se aprecia cómo la frontera entre esta lengua y la variedad *nord-occidental* del catalán está mucho más definida que las fronteras anteriormente comentadas entre el catalán y el aragonés, por un lado, y el catalán *nord-occidental* y el central, por otro. Este hecho se debe a las peculiares condiciones orográficas de *la Vall d'Aran*, que han hecho que históricamente esta región oriente sus relaciones hacia tierras ultrapirenaicas. Cabe recordar que, hasta la construcción del Túnel de Viella, inaugurado en 1958, este valle quedaba aislado por la nieve durante muchos meses y la comunicación con Cataluña solo era posible en verano (Veny 1984: 121).

El aranés presenta como rasgos peculiares dentro de la morfología nominal, la desinencia plural en *-i* —en determinados nombres, adjetivos y pronombres—, las formas del artículo definido *eth/era/es*, o el morfema *-es* para las palabras acabadas en *-a* en singular. Es posible hallar, en aranés, interferencias del catalán, y aún muchas más del castellano, como muestra de la subordinación que sufre esta variedad del occitano con respecto al español, y también al catalán, aunque en menor medida (Badia Capdevila 2002: 46-47)⁴⁵⁹.

Mapa 4.4. Mapa de la Vall d'Aran.



Fuente: *Suïls* (2004: 53).

⁴⁵⁹ Sobre la lengua aranés y su gramática, véase Carrera (2003 y 2007).

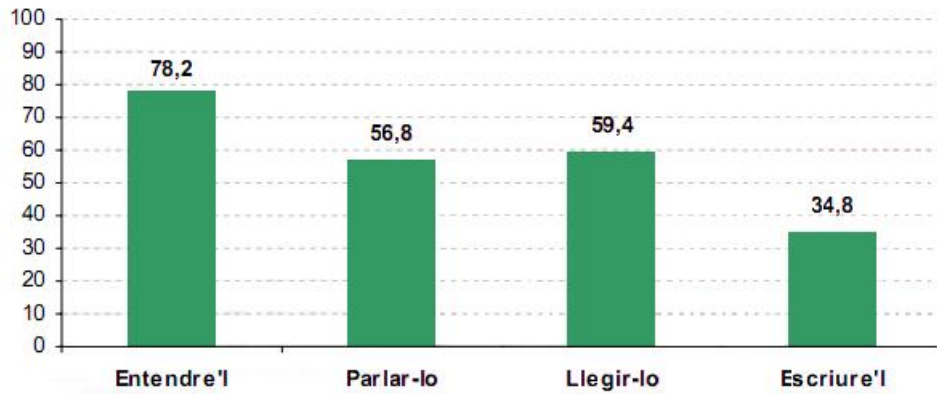
El aranés es la lengua propia de *la Vall d'Aran* donde comparte cooficialidad con el castellano y el catalán⁴⁶⁰. De hecho, el último *Estatut d'Autonomia de Catalunya* reconoce el aranés como lengua oficial en todo el territorio catalán⁴⁶¹. Actualmente, esta variedad del occitano se emplea en *la Vall d'Aran* en los medios de comunicación, en la enseñanza y en la administración. El aranés cuenta con apoyo institucional —de la *Generalitat* y del *Conselh Generau d'Aran*—, por lo que son muchas las medidas que se adoptan destinadas a promover la creación y la difusión cultural en esta lengua. Con respecto al conocimiento del aranés por parte de los residentes en la comarca, Vila i Moreno (2000: 136-137), tras analizar la evolución de esta variable, concluye que es la población nacida en el resto del Estado español la que presenta un conocimiento más deficitario de esta lengua. Por el contrario, la población autóctona es la que presenta un dominio superior, seguida de los originarios de otras zonas catalanohablantes (Cataluña, Valencia y las Islas Baleares) y de los nacidos en un país extranjero. También es interesante notar cómo el autor detecta un conocimiento muy desigual de la lengua en función del municipio del valle analizado (Vila i Moreno 2000: 130).

En el panorama actual, las últimas encuestas señalan que, aunque las competencias pasivas son bastante altas —el porcentaje con respecto a la comprensión oral es de casi un 80%— poco más de la mitad de los habitantes de la región es capaz de hablar su lengua propia, y muchos menos son capaces de escribirla. Nótese cómo los cuatro índices se sitúan bastante por debajo de los mismos parámetros con respecto al catalán en la provincia (gráfico 4.16).

⁴⁶⁰ A partir de 1990 se reconoce la oficialidad del aranés en el territorio, junto con el catalán y el castellano (Badia Capdevila 2002: 46).

⁴⁶¹ Así lo indica en el punto 5 del artículo 6 del *Estatut d'Autonomia de Catalunya* (2006). Sin embargo, según los datos de la última *Enquesta d'Usos Lingüístics de la Població* (2008), más de la mitad de los catalanes ignora este hecho, concretamente el 56.9%.

Gráfico 4.27. Conocimientos de aranés de la población de 15 años y más en la Vall d'Aran, en porcentajes. Año 2008.



Fuente: EULP 2008⁴⁶².

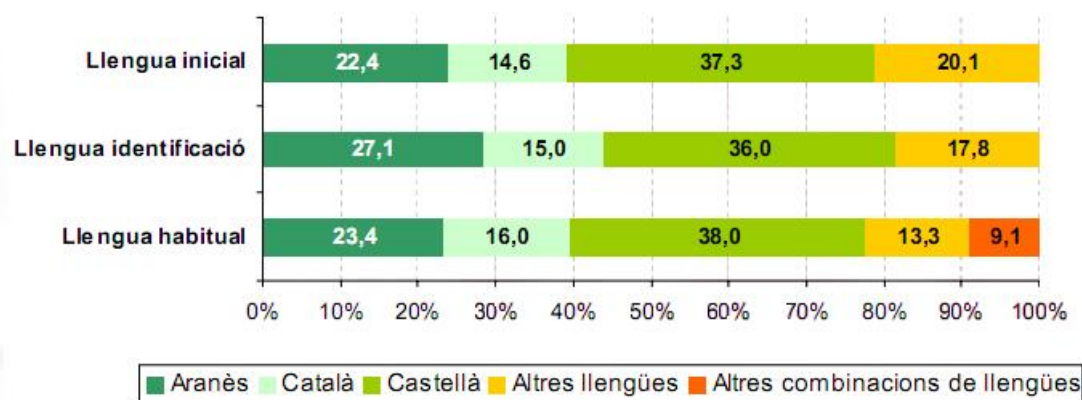
Con respecto a la lengua inicial o materna, la de identificación y la de uso habitual, cabe señalar cómo la situación se polariza en *la Vall d'Aran* en torno al aranés y al castellano, mientras que el catalán se sitúa por debajo en las tres variables analizadas. Es de esperar que haya una correlación entre el uso del aranés y el origen autóctono del individuo, del mismo modo que la población nacida en otros lugares del Estado se decanta mayormente por el uso del castellano⁴⁶³. Llama la atención que la variable en que el aranés alcanza un porcentaje superior sea en la de lengua de identificación, lo que nos da una idea de las actitudes favorables de la población hacia la lengua y de la vitalidad de la misma en la región⁴⁶⁴. Por último, también es importante destacar la presencia de otras lenguas en el panorama lingüístico del valle, entre las que muy posiblemente destaque el francés por la proximidad de la región con el país galó.

⁴⁶² Utilizamos esta fuente de datos, puesto que la encuesta sobre usos lingüísticos del año 2003 (*Institut d'Estadística de Catalunya*, 2004) solo recopila datos referentes a la lengua catalana.

⁴⁶³ Como ya hemos comentado, Vila i Moreno (2000) aprecia esta relación. Por otro lado, la distribución de la población en la comarca según su origen difiere de la que hemos visto en general para la provincia. Concretamente, el porcentaje de población nacida en el resto de la Península asciende y alcanza el 26.6 % (año 2002, dato de *Idescat*).

⁴⁶⁴ Sobre este tema y sobre el aranés en general y las circunstancias que lo rodean, puede consultarse Suïls (2004).

Gráfico 4.28. Lengua inicial, de identificación y habitual de la población de 15 años y más en la Vall d'Aran, en porcentajes⁴⁶⁵. Año 2008.



Fuente: EULP 2008.

A lo largo de este capítulo, hemos trazado el panorama general del contacto entre el castellano y el catalán en Cataluña, atendiendo a las peculiaridades que el mismo presenta en la provincia de Lleida. Como ya se ha constatado en la bibliografía al respecto (*vid.* 4.2), faltan estudios sobre nuestro ámbito geográfico, que se basen en corpus y que analicen cuantitativamente el peso que alcanzan determinados fenómenos de contacto lingüístico, tarea que esperamos realizar en este trabajo. Pero antes, veamos, en el próximo capítulo, las condiciones metodológicas que rodean el mismo.

⁴⁶⁵ Adviértase la gran diferencia que hay en las tres variables entre este gráfico y el 4.20. Además de la diferencia con respecto al año de la encuesta (2003 en la 4.20 y 2008 en la presente), la principal razón de las divergencias se halla en que en el gráfico 4.20 se engloba las comarcas del *Alt Pirineu i Aran*, y en cambio, este gráfico se refiere únicamente a la comarca de *la Vall d'Aran*.

5. METODOLOGÍA

En el presente capítulo, exponemos detalladamente las características de la investigación que hemos llevado a cabo. Como comentamos (*vid.* 2.1.2), nuestro trabajo se adscribe a un macroproyecto de estudio de la disponibilidad léxica en todo el territorio hispanohablante. Por esta razón, las decisiones metodológicas adoptadas se han visto condicionadas, en gran parte, por las directrices establecidas por consenso entre los miembros del equipo que lleva a cabo este Proyecto, y siguen, *grosso modo*, las pautas de las investigaciones españolas en curso o ya finalizadas. La uniformidad metodológica es una herramienta imprescindible que posibilita la comparación de los resultados *a posteriori*.

En los dos primeros apartados, ofrecemos una visión, lo más completa posible, de las características de la muestra utilizada para crear nuestros corpus, especificando las variables sociales tomadas en consideración. A continuación, en el tercer apartado, detallamos tanto el instrumento que hemos usado en la investigación como la técnica de obtención de los datos. El cuarto y último apartado, el más extenso, está dedicado al conjunto de decisiones adoptadas con respecto a la homogeneización de los materiales, uno de los aspectos más complicados de los estudios de léxico disponible.

5.1. LA MUESTRA

La sociolingüística variacionista, apoyándose en la estadística inferencial, considera que es posible deducir el comportamiento lingüístico de una población mediante el análisis de una muestra representativa de ella⁴⁶⁶. En nuestro caso, el universo objeto de este estudio queda conformado por el alumnado matriculado en 2º de bachillerato⁴⁶⁷ en la provincia de Lleida durante el año 2002. Tanto la edad de los informantes (17-18 años) como el nivel educativo en cuestión se ajustan al criterio acordado por los componentes del Proyecto Panhispánico. Esta elección se basa en el convencimiento de que los estudiantes preuniversitarios ya están formados léxicamente pero no han incorporado aún terminología específica de un área profesional

⁴⁶⁶ Para ello se parte del “principio de homogeneidad lingüística”, que establece que el comportamiento lingüístico es lo suficientemente homogéneo como para poder ser representado por un número reducido de datos (Labov 1966 *apud* Moreno Fernández 1990: 70 y 89-90).

⁴⁶⁷ Este nivel educativo corresponde al antiguo C.O.U., totalmente extinguido en Cataluña en el año de aplicación de nuestro estudio.

determinada. Por esta razón, se cree que pueden ser un buen ejemplo de la norma léxica adulta general (Benítez Pérez 1992b: 73, López Morales 1999a: 28)⁴⁶⁸. Hay otra razón implícita: la facilidad metodológica que conlleva este procedimiento puesto que permite recabar datos de muchos informantes al mismo tiempo⁴⁶⁹.

La base de los dos corpus que analizamos en la presente monografía son los datos obtenidos a partir de 480 encuestas que completaron 240 estudiantes de 2º de bachillerato de distintos puntos de la provincia de Lleida. Como explicamos más adelante (*vid.* 5.3), los mismos informantes completaron dos veces la prueba: la primera, en castellano y, la segunda, en catalán. La mayoría de las encuestas se cumplimentó en invierno del año 2002, concretamente, durante el primer trimestre del curso 2002/2003. Teniendo en cuenta que el alumnado matriculado, en este nivel educativo, ascendió a 2471 en dicho curso académico⁴⁷⁰, el volumen de la muestra constituye el 9.71% del universo estudiado, cifra que supera con creces la ratio de 0.025 que estableció Labov (1972) como suficientemente representativa para los estudios de corte variacionista⁴⁷¹. A su vez, los 240 informantes suponen el 6.23% de la franja de edad analizada (17 años), y un 0.065% del total de población de la provincia⁴⁷².

Esta cantidad también supera las 200 unidades determinadas por Mackey (1971b, *cfr.* 3.1), aunque es inferior a los 400 encuestados por área de estudio acordados en la primera reunión del Proyecto de Estudio de la Disponibilidad Léxica, llevada a cabo en Bilbao en 1999 (Samper Padilla y Samper Hernández 2006: 45). Si bien, buena parte de las investigaciones respeta esta cifra, e incluso la rebasa (*v.gr.* Gómez Molina y Gómez Devís 2004, Martínez Olmos 2008 o Ávila 2006⁴⁷³) también se han realizado

⁴⁶⁸ Borrego y Fernández Juncal (2003) llevan a cabo una investigación con estudiantes del último curso de universidad y analizan las divergencias halladas con respecto a los resultados de los preuniversitarios. En general, las diferencias observadas son mínimas, aunque sí queda patente, en los universitarios, una mayor maduración en el proceso de lexificación.

⁴⁶⁹ Como también habían ya notado Gougenheim *et al.* (1964: 153): “Nous avons choisi les écoliers parce qu’eux seuls fournissaient des témoins en nombre suffisant”.

⁴⁷⁰ Este y los siguientes datos sobre la población de 2º de bachillerato del curso académico 2002/2003 nos fueron proporcionados por la Delegación Territorial de Lleida del *Departament d’Ensenyament*, actual *Departament d’Educació*.

⁴⁷¹ Moreno Fernández (1990: 69-71) resume las críticas expuestas acerca del límite estipulado por Labov.

⁴⁷² Datos de *Idescat*, año 2002. Todos los datos de este capítulo que no son propiamente sobre la población de 2º de bachillerato proceden de esta misma fuente (www.idescat.cat).

⁴⁷³ Los primeros utilizan 465 informantes para analizar el léxico disponible de la provincia de Valencia, Martínez Olmos (2008), 460 para la provincia de Alicante, y Ávila (2006), 458 en su trabajo sobre Málaga.

muchas otras investigaciones manejando muestras numéricamente inferiores⁴⁷⁴. A nuestro modo de ver, esta cantidad ha de ponerse en relación, necesariamente, con el universo estudiado, esto es, el conjunto de alumnado matriculado en el curso preuniversitario, que, a su vez, depende del número de jóvenes que se hallan en esa franja de edad, reflejo final de la densidad de población del área de estudio. Por esta razón, no puede necesitarse el mismo número de informantes para reflejar la norma léxica disponible de una ciudad (como es el caso de Ceuta o Melilla), de una provincia (como es nuestro caso —o los muchos realizados con anterioridad sobre otras provincias españolas—), o de toda una comunidad autónoma (Galicia, Aragón o Castilla-La Mancha). Sin embargo, también reconocemos que un número similar de encuestados sería más que deseable si el propósito principal de una investigación es la comparación dialectal.

La provincia de Lleida, como ya observamos, no destaca por su densidad de población, tanto en relación con el resto de provincias catalanas (*vid.* 4.3.1, gráficos 4.4 y 4.5) o españolas, sino que se sitúa, más bien, en una posición intermedia. La cantidad de informantes que hemos elegido es cercana a la utilizada en estudios realizados en provincias con una densidad similar, como por ejemplo, Huelva, donde se utilizaron 220 informantes (Prado y Galloso 2005), o Burgos, donde se tuvieron en consideración los datos de 175 individuos (Fernández Juncal 2008)⁴⁷⁵. También se ha de tener en cuenta que, aunque la muestra reúna 240 informantes, los datos recabados se duplican, ya que, como hemos indicado, todos los encuestados rellenaron la prueba en dos ocasiones, una por cada lengua que aquí se analiza y que conviven en el territorio.

Los 240 informantes seleccionados proceden de 12 centros educativos, la mitad de ellos ubicados en distintos puntos de la provincia, y la otra mitad situados en la ciudad de Lleida, tal como se aprecia en la siguiente tabla:

⁴⁷⁴ Para conocer la cantidad de informantes utilizados en la mayor parte de las investigaciones españolas, puede consultarse el cuadro de Samper Padilla y Samper Hernández (2006: 46) donde los informantes se distribuyen según la variable *sexo*.

⁴⁷⁵ De hecho, la relación con la densidad de población del marco de estudio se tiene, de alguna forma, siempre en cuenta, toda vez que, por ejemplo, Bartol (2004) utiliza solo 100 informantes para reflejar el léxico disponible de Soria, mientras que Gómez Molina y Gómez Devís (2004) se valen de 465 informantes al estudiar la disponibilidad de la provincia de Valencia. Por poner solo otro ejemplo, en la vecina comunidad autónoma de Aragón, Arnal *et al.* (2004) emplean 417 individuos para ilustrar el léxico disponible de las tres provincias, las cuales en conjunto, obviamente, superan demográficamente en mucho la población de nuestro estudio.

Tabla 5.1. Centros educativos seleccionados con su titularidad y ubicación territorial.

Nombre del centro	Titularidad	Municipio	Habitantes del municipio	Comarca
<i>Col·legi Lestonnac</i>	Privada	Lleida	115000	<i>El Segrià</i>
<i>Col·legi Maristes Montserrat</i>	Privada	Lleida	115000	<i>El Segrià</i>
<i>Col·legi Episcopal</i>	Privada	Lleida	115000	<i>El Segrià</i>
<i>IES Ronda</i>	Pública	Lleida	115000	<i>El Segrià</i>
<i>IES Joan Oró</i>	Pública	Lleida	115000	<i>El Segrià</i>
<i>IES Escola del Treball</i>	Pública	Lleida	115000	<i>El Segrià</i>
<i>IES Canigó</i>	Pública	Almacelles	5753	<i>El Segrià</i>
<i>IES Josep Vallverdú</i>	Pública	Les Borges Blanques	5313	<i>Les Garrigues</i>
<i>IES Alfons Costafreda</i>	Pública	Tàrrega	13210	<i>L' Urgell</i>
<i>IES Francesc Ribalta</i>	Pública	Solsona	7689	<i>El Solsonès</i>
<i>IES Joan Brudieu</i>	Pública	La Seu d'Urgell	11434	<i>L'Alt Urgell</i>
<i>IES d'Aran</i>	Pública	Vielha e Mijaran	4352	<i>La Vall d'Aran</i>

Siguiendo el criterio que también han tomado otros investigadores⁴⁷⁶, hemos eliminado los datos que nos facilitaron los encuestados que llevaban menos de 10 años viviendo en el mismo municipio —o en la comarca respectiva— donde se situaba el centro educativo seleccionado, por considerar que era probable que no hubieran adoptado las características léxicas de la variedad autóctona. De modo similar, hemos suprimido también las encuestas de los informantes cuyos progenitores u ellos mismos hubieran nacido en el extranjero, ya que, a pesar de haber residido la mayor parte de su vida en nuestra provincia, consideramos que, con toda probabilidad, otras lenguas maternas distintas al castellano o al catalán influirían en su caudal léxico. De este modo, reflejamos la propia realidad de la provincia de Lleida cuya población se caracteriza, como ya expusimos con todo detalle (*vid.* 4.3.2.), por un fuerte carácter autóctono y en la que los originarios —tanto de otros lugares de España como de otros países—

⁴⁷⁶ Por ejemplo, López Morales (1999a: 28) o González Martínez (2002: 14) eliminan los datos de los individuos que no han residido la mayor parte de su vida en el lugar donde se lleva a cabo el estudio.

suponen solo el 20.99% del total de la población (*vid.* gráfico 4.11)⁴⁷⁷. Adoptando estas decisiones, además, tenemos la certeza de que solo hemos recabado los datos de informantes bilingües o que han estado en contacto con el catalán desde la infancia (con el castellano tal contacto existe a través de los medios de comunicación, del cine, de la escuela, incluso en los lugares donde no hay inmigración castellanohablante, *vid.* 4.1.), y seguimos fielmente nuestra intención de estudiar las características del léxico español en un territorio bilingüe.

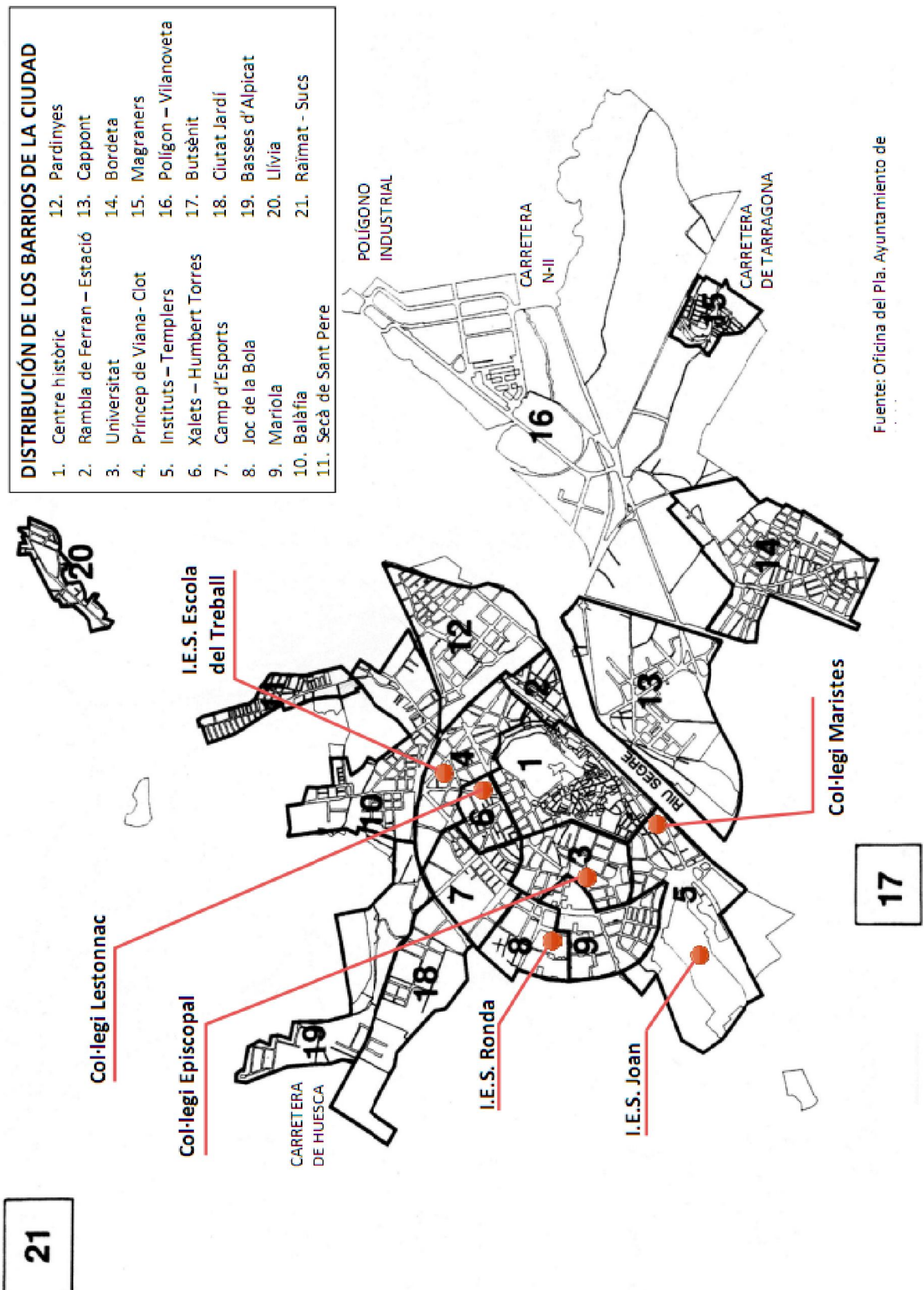
Se procuró seleccionar 20 informantes de cada centro educativo⁴⁷⁸. En la capital, escogimos tres centros públicos (*IES Ronda*, *IES Joan Oró* e *IES Escola del Treball*) y tres privados (*Col·legi Lestonnac*, *Col·legi Episcopal* y *Col·legi Maristes Montserrat*) distribuidos estratégicamente en la ciudad de modo que quedaran representados los distintos barrios de Lleida; tarea que no fue sencilla, ya que, de un lado, no existen institutos de educación secundaria en todos los barrios⁴⁷⁹ y, de otro, la mayoría se concentra en los barrios más céntricos, especialmente los de titularidad privada. A continuación, describimos las características más relevantes de los centros escolares urbanos.

⁴⁷⁷ Además, este porcentaje desciende hasta el 10.22 % en la franja de edad analizada (15 a 19 años, datos de *Idescat*, año 2002).

⁴⁷⁸ Si bien este criterio se cumplió a rajatabla en las instituciones educativas de la provincia, en los centros ubicados en la capital no fue siempre posible, bien porque algunos alumnos no completaron alguna de las dos encuestas o porque no cumplían con los requisitos de selección de la muestra que acabamos de explicar, aunque las cifras manejadas son siempre muy cercanas a esta cantidad: *Maristes* (18), *Episcopal* (19), *Lestonnac* (23), *IES Ronda* (20), *IES Joan Oró* (16) e *IES Escola del Treball* (24).

⁴⁷⁹ Por ejemplo, *Cappont*, uno de los barrios que ha registrado un mayor crecimiento demográfico en los últimos años, carece de un centro educativo de educación secundaria de titularidad pública o privada. En la misma situación se halla el barrio de *Magraners*, a las afueras de la ciudad.

Mapa. 5.1. Mapa de la ciudad de Lleida dividida en barrios con los centros educativos seleccionados.



Fuente: Oficina del Pla. Ayuntamiento de

El *IES Joan Oró* está situado al sur de la urbe, tal como se observa en el mapa anterior, cerca del río Segre y de varias de las carreteras que permiten comunicar la ciudad con otras poblaciones. Es uno de los centros públicos de bachillerato más antiguos de la ciudad y se ubica, junto con otros dos institutos, en un mismo recinto que da nombre al barrio en el que se halla, *Instituts-Templers*. Además de la educación secundaria obligatoria, es posible cursar en él distintas ramas de bachillerato y algunos ciclos formativos. En cuanto a los otros dos centros educativos públicos seleccionados, el *IES Escola del Treball* está situado en una de las partes céntricas de la ciudad y cuenta con una larga tradición en los estudios de formación profesional. Este centro ofrecía a su alumnado, en el curso 2002/2003, tanto la posibilidad de cursar una modalidad de bachillerato como de aprender un oficio mediante alguno de los numerosos ciclos formativos de grado medio y superior que se ofrecen, tras acabar la educación secundaria⁴⁸⁰.

Por su lado, el *IES Ronda* es un centro escolar cuya ubicación es singular puesto que se sitúa justo en el límite entre dos barrios de índole completamente distinta, la *Mariola* y el *Joc de la Bola*. Este hecho ocasiona que su alumnado proceda de estratos sociales muy diferentes, ya que la *Mariola* es un barrio de tradición obrera, donde abundan los pisos de protección oficial, caracterizado —entre otras particularidades— por el bajo nivel de formación de la población que habita en él (Mir 1995: 30-46, Vilagrasa 2003: 75-78)⁴⁸¹. Los residentes del barrio *Joc de la Bola*, en cambio, ostentan un nivel socioeconómico alto, ya que este es uno de los lugares de la ciudad donde el precio de la vivienda alcanza cotas más elevadas y donde han ido apareciendo urbanizaciones exclusivas, antes de que estas fueran comunes en los pueblos más próximos a la ciudad y se popularizaran por sus mejores precios. El *IES Ronda* ofrece estudios de bachillerato, dos ciclos formativos y la educación secundaria obligatoria.

⁴⁸⁰ El curso académico 2012/2013 es el primero en que el instituto se dedica de forma exclusiva a la formación profesional y, por tanto, ya no es posible cursar otros estudios en él. Por otro lado, en la actualidad, las siglas *IES* (*Institut d'Educació Secundària*) de todos los institutos públicos se han sustituido por *INS* (*Institut*).

⁴⁸¹ En la *Mariola*, se ubican los bloques Juan Carlos, un complejo de edificios de viviendas en los que se intentó concentrar a la población de etnia gitana, situada anteriormente en casetas próximas al barrio y en otros barrios periféricos de la ciudad. De otra parte, en Mir (1995: 30-46) es posible hallar una descripción completa de la estructura de la población de Lleida según los distintos barrios, aunque, lógicamente, la información que se ofrece está un tanto desfasada.

Con respecto a los colegios privados, todos ellos son confesionales católicos y en ellos se imparten todos los niveles educativos desde la educación infantil hasta el bachillerato, lo que supone que los alumnos del nivel preuniversitario han cursado la mayor parte de sus estudios en él, e, incluso, algunos acuden allí desde el parvulario. Mientras que las etapas de escolarización obligatoria —la educación primaria y la secundaria— tienen carácter concertado, la educación infantil y el bachillerato, no⁴⁸². El *Col·legi Lestonnac “L’ensenyança”* está ubicado en el barrio *Xalets-Humbert Torres*, en una zona céntrica de la ciudad. Está gestionado en la actualidad por una cooperativa de padres, ya que las religiosas que se encargaban de su dirección, pertenecientes a la Compañía de María, abandonaron el ejercicio docente en 1989. No muy lejos de allí, se encuentran los otros dos centros privados que hemos seleccionado, el *Col·legi Maristes Montserrat* y el *Col·legi Episcopal*. El primero de ellos se ubica en el barrio *Instituts Templers* y es una de las diez escuelas que existen en Cataluña dirigidas por la Congregación de los Hermanos Maristas. En Lleida, es uno de los colegios más antiguos, ya que la presencia de esta comunidad en la ciudad data de 1895. El último colegio privado, el *Col·legi Episcopal-Mare de Déu de l’Acadèmia*, cuya titularidad ostenta el obispo de la ciudad, es el único de los tres que hemos seleccionado que incorpora en su oferta educativa ciclos formativos, además de todas las restantes etapas. Se ubica en el barrio *Universitat*, otra zona concurrida de la ciudad.

Tal como puede observarse en el mapa anterior (5.1), algunos de los centros educativos se sitúan geográficamente muy cerca unos de otros. Esto ocurre, por ejemplo, con *Maristes* e *IES Joan Oró*, ubicados en el mismo barrio. También se da la misma situación con el *Col·legi Lestonnac*, que dista solo unas decenas de metros del *IES Escola del Treball*, a pesar de que pertenecen a dos barrios distintos. Este hecho se debe, como ya hemos comentado, a la concentración de centros educativos, sobre todo de índole privada, en el corazón de la ciudad. Por otro lado, consideramos que esta circunstancia nos permite analizar mejor la posible incidencia de la variable *tipo de centro* (público/privado) en los datos, ya que demuestra, con claridad, que la familia opta por un tipo de escolaridad u otra.

⁴⁸² No existe ningún colegio privado en Lleida que no sea concertado. Cabe señalar que desde el curso académico 1999/2000 tanto el bachillerato como los ciclos formativos del *Col·legi Episcopal-Mare de Déu de l’Acadèmia* tienen concierto económico con el *Departament d’Educació*.

Por otro lado, los centros rurales se escogieron principalmente en función de su ubicación territorial para que pudieran ser representativos de la provincia. Para realizar tal elección tuvimos en cuenta las diversas fronteras territoriales y, sobre todo, lingüísticas del área de estudio. Como ya explicamos anteriormente (*vid.* 4.3.2., pp. 215-223), en la provincia predomina una variedad del catalán occidental denominada *nord-occidental*, aunque también es posible hallar otras subvariedades como el *lleidatà*, el *pallarès* o el *ribargorçà*. Además, limita, en toda su vertiente este, con el dialecto central del catalán oriental, y concretamente en algunos puntos, con una variedad de transición denominada *xipella* (Veny 1984: 49, *supra* mapa 4.3). Al oeste hallamos la franja lingüística extendida verticalmente entre las comunidades autónomas de Cataluña y Aragón, denominada *la Franja de Ponent*, donde se habla catalán, castellano e incluso algunas hablas de transición que actúan de puente entre el catalán y el aragonés. Con respecto al norte, Lleida colinda con dos países, Andorra, donde la única lengua oficial es el catalán —aunque esta convive con el castellano, el francés y el portugués— y Francia. De otra parte, es muy interesante, y ha sido poco estudiada, la situación lingüística que se da en el noroeste de la provincia, en *la Vall d’Aran*, donde como ya explicamos (*vid.* 4.3.2.), existe una lengua propia, el aranés, modalidad del gascón —un dialecto del occitano— que comparte cooficialidad con el catalán y el castellano.

Mapa 5.2. Mapa de la provincia de Lleida con los municipios seleccionados.



Teniendo en cuenta las peculiaridades lingüísticas que acabamos de señalar, los centros educativos finalmente elegidos son los situados en las siguientes poblaciones: *les Borges Blanques* (I.E.S. *Josep Vallverdú*), *Tàrrega* (I.E.S. *Alfons Costafreda*), *Solsona* (I.E.S. *Francesc Ribalta*), *la Seu d'Urgell* (I.E.S. *Joan Budrieu*), *Vielha e Mijaran* (I.E.S. *d'Aran*) y *Almacelles* (I.E.S. *Canigó*). Las tres primeras poblaciones —*les Borges Blanques*, *Tàrrega* y *Solsona*— se corresponden con las capitales de comarca que más se aproximan al límite geográfico este de la provincia y limitan, respectivamente, con las provincias de Tarragona y Barcelona. *Les Borges Blanques*, capital de *les Garrigues* —mundialmente conocida por la excelencia de su aceite— se encuentra en la ruta de camino hacia Tarragona, y los otros dos municipios, *Tàrrega* y *Solsona* son los puntos más poblados que delimitan lingüística y geográficamente con la provincia de Barcelona, el primero de ellos (*Tàrrega*) con una variedad dialectal autóctona similar a la de la capital de la provincia y, el segundo (*Solsona*), con una variedad más próxima al dialecto oriental del catalán.

Por otro lado, *Viella* y *la Seu d'Urgell* son los núcleos más importantes situados en el noroeste y nordeste —respectivamente— de la provincia, en plena cordillera pirenaica. *Viella*, *Vielha e Mijaran* en aranés, es la capital de *la Vall d'Aran*, cuya situación lingüística tan singular ya hemos comentado. *La Seu d'Urgell*, por su lado, debe gran parte de su desarrollo socioeconómico a la cercana Andorra, de la que dista tan solo diez kilómetros. La última población escogida, *Almacelles*, supone el límite oeste provincial. Delimita, administrativamente, con la comunidad autónoma de Aragón y, lingüísticamente, con los municipios que conforman *la Franja de Ponent*. Con la elección de estas poblaciones, esperamos que se reflejen las diversas influencias lingüísticas que pueden observarse en nuestra provincia.

5.2. LAS VARIABLES SOCIALES

En cuanto a las variables sociológicas utilizadas para la estratificación de la muestra, estas coinciden, —como seguidamente se detallará—, con las acordadas por los miembros del Proyecto Panhispánico, y ya se han contemplado en otros lugares del mundo hispánico (vid. 2.1.2). De este modo hemos tenido en cuenta el *sexo* del informante, el *tipo de centro*, la *zona geográfica*, el *nivel sociocultural* de la familia, y a estas variables, hemos añadido una más —la *lengua materna*— que, como es lógico, solo se incluye en los estudios realizados en territorios bilingües (vid. 3.4.3).

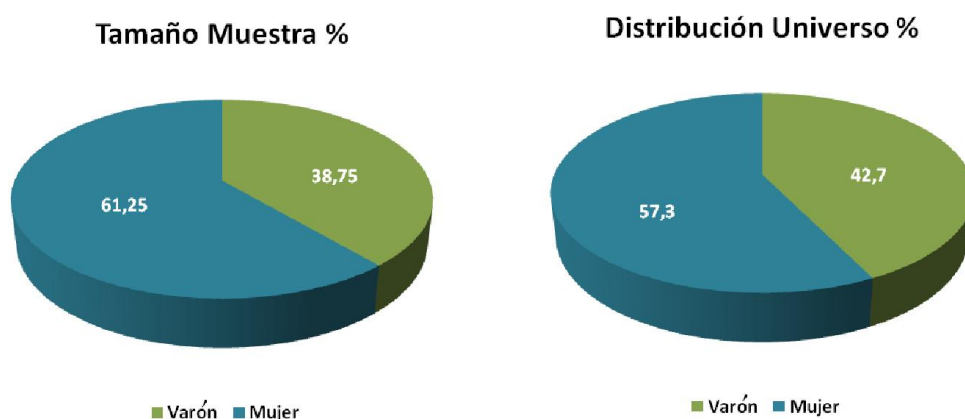
Con respecto al método de muestreo, hemos llevado a cabo —en la medida de lo posible— un muestreo estratificado con asignación proporcional al universo estudiado. Por esta razón, iremos comprobando, en las líneas sucesivas, la correspondencia observada entre la estratificación de la muestra y la respectiva en la población estudiada. Como Fernández Smith *et al.* (2008: 27) indican acertadamente, debido a la necesidad de respetar las pautas metodológicas exigidas por el proyecto al que nos sumamos así como los criterios de cualquier método estadístico, no siempre es fácil lograr las cuotas de proporcionalidad requeridas; ya que, por ejemplo, es muy difícil conseguir algunos datos sobre la población en la que nos basamos, como los relativos al *nivel sociocultural*. A causa de estas limitaciones, algunas variables —el *sexo*, el *nivel sociocultural* o la *lengua materna*— han sido post-estratificadas, mientras que en otras —como el *tipo de centro* o la *zona geográfica*— ha podido preverse con antelación los informantes que escogeríamos dentro de cada una de las opciones.

Con respecto a la variable *sexo*, tal como ya se ha señalado que ocurre en la mayoría de las investigaciones (Samper Padilla y Samper Hernández 2006: 46), se constata una mayor presencia femenina en este sector poblacional, a pesar de que, en el conjunto general de la población y en la franja de edad analizada, la tendencia se invierte (*cf.* 4.3.2., gráficos 4.9 y 4.10). Este hecho se debe a que hay más mujeres que optan por continuar con sus estudios, y concretamente, se decantan por los de bachillerato.

Tabla 5.2. Distribución por sexo según distintos segmentos de población.

Sexo	Varones	%	Mujeres	%	Total
Población general	186 026	50.13	185 029	49.87	371 055
Población de 17 años	2015	52.36	1833	47.64	3848
Población en 2º de bachillerato	1055	42.7	1416	57.3	2471
Muestra	93	38.75	147	61.25	240

Gráfico 5.3. Distribución según la variable sexo.



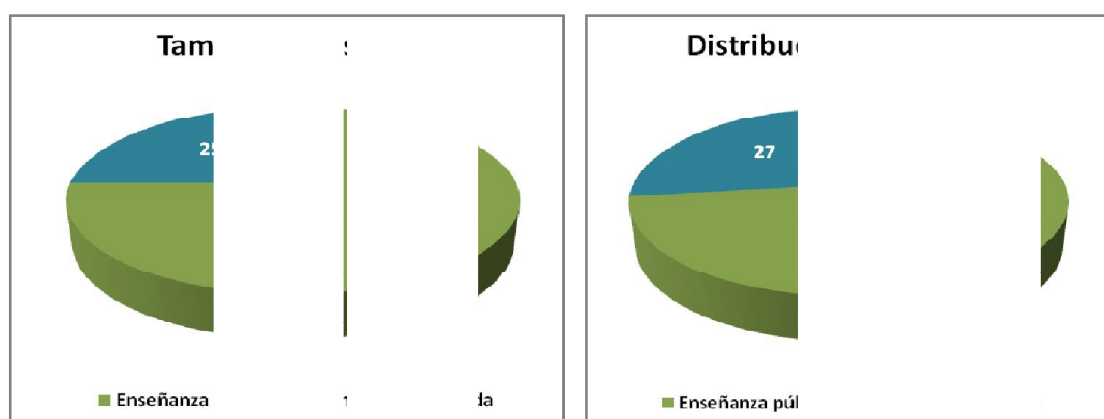
La segunda variable considerada es la del *tipo de centro* (público/privado). En nuestra provincia, hay muchos más centros públicos que privados, y concretamente, no existen centros privados en todos los municipios de la provincia que, además, impartan bachillerato. Por esta razón, en un primer momento decidimos tener en cuenta este parámetro únicamente en los informantes de Lleida ciudad (*cf.* Serrano Zapata 2003a: 933), habida cuenta de que, en algunos de los núcleos rurales analizados, la enseñanza pública era la única opción posible en el nivel estudiado. No obstante, tras realizar un análisis más pormenorizado de cada uno de los municipios donde habíamos llevado a cabo el estudio, nos decantamos por incluirla como variable en todos los informantes. Si bien es cierto que en dos de los municipios elegidos, *la Seu d’Urgell* y *Vielha e Mijaran*, no existe ninguna institución privada donde se pueda cursar 2º de bachillerato, e incluso sería necesario recorrer diariamente bastantes kilómetros para que fuera posible esta opción, en otros dos (*Tàrrrega* y *Solsona*) sí es posible y existen sendos colegios donde es posible estudiar el curso preuniversitario. Por otro lado, ni *Almacelles* ni *les Borges Blanques* cuentan con un centro de titularidad privada pero la poca distancia que separa estos municipios de la capital hace posible que las familias puedan optar por este tipo de enseñanza, si lo desean. Además, existe una serie de colegios concertados situados en las afueras de Lleida, —como *Mater Salvatoris*, *Arabell*, *Terraferma* o *Claver*—, a los que sí podrían acudir, sobre todo los residentes en Almacelles, ya que incluso les quedan más cerca que la propia capital.

A continuación, ofrecemos los datos relativos a la proporción entre informantes de instituciones privadas y públicas en la muestra y en el universo que analizamos, teniendo en cuenta tanto el número de alumnos como de centros educativos:

Tabla 5.3. Distribución por tipo de centro según distintos parámetros.

Tipo de centro	Enseñanza pública	%	Enseñanza privada	%	Total
Número de centros educativos	36	70.59	15	29.41	51
Población de 2º de bachillerato	1804	73	667	27	2471
Muestra	180	75	60	25	240

Gráfico 5.4. Distribución según la variable tipo de centro.



De otra parte, sobra decir que pretendemos reflejar en esta variable el tipo de enseñanza que el alumno ha recibido durante sus años de escolaridad. Por esta razón, incluimos en la encuesta una pregunta para determinar si el informante había cursado la mayoría de las etapas educativas en el centro en el que lo encontrábamos. Muy pronto nos percatamos de que las características del actual sistema educativo en el que el bachillerato constituye un ciclo educativo de dos años no nos permitían establecer esta concreción: mientras que la mayoría de los alumnos de los centros educativos privados sí había realizado allí la educación secundaria y, por tanto, llevaba seis años como mínimo de escolaridad privada, en los centros públicos no ocurría lo mismo y contábamos con informantes que habían realizado la secundaria obligatoria en un centro privado⁴⁸³. Esta circunstancia nos obligó a tomar como determinación asegurar, por lo menos, la realización de todo el bachillerato de nuestros informantes en el centro educativo en el que los encontrábamos.

⁴⁸³ La proporción de estos informantes en la muestra nos impedía eliminarlos. Esta circunstancia fue frecuente en los informantes de *la Seu d'Urgell*, municipio que cuenta con un centro privado donde es posible cursar la enseñanza primaria y secundaria, pero no el bachillerato.

De otra parte, se ha discutido mucho sobre la relación entre esta variable y la del *nivel sociocultural*. Parece evidente que los contenidos que se imparten, tanto en centros públicos como privados, no pueden diferir, por cuanto todos ellos siguen las directrices marcadas, en nuestro caso, por el *Departament d'Educació*, y, en última instancia, por el Ministerio de Educación y Ciencia. Lo que sí los diferencia, como afirman Fernández Smith *et al.* (2008: 27, nota 4) son las características económicas y jurídicas que rigen su funcionamiento. Los mismos autores opinan que el interés por incluir esta variable se basa en la mayor probabilidad de hallar informantes de estratos socioculturales altos en los centros educativos privados. No obstante, hasta la fecha se ha comprobado no solo que es un factor importante en la estratificación del léxico disponible, sino que los resultados que arroja esta variable no son siempre equiparables a los obtenidos mediante la diferenciación social de los informantes (Samper Padilla y Samper Hernández 2006: 72-73). En el caso concreto de nuestra área de estudio, generalmente, las familias con menos recursos económicos acostumbran a llevar a sus hijos a centros públicos, para evitar así un gasto añadido, pero también es cierto que los hijos de ciertos profesionales son sistemáticamente enviados a centros públicos de reconocido prestigio. En la ciudad de Lleida, por ejemplo, esto sucede tanto en la enseñanza primaria (*CEIP Joc de la Bola, CEIP Escola Alba, CEIP Escola Espiga*) como en la secundaria (*IES Ronda, IES Samuel Gili Gaya*); en el resto de la provincia, como ya hemos señalado, no siempre es posible esta opción.

En cuanto a la variable *zona geográfica*, los informantes de nuestra investigación se distribuyen en urbanos o rurales, según la ubicación del centro educativo en el que estudian⁴⁸⁴. Cabe señalar que, en otros estudios —como Arnal *et al.* (2004) o Pastor y Sánchez (2008)—, se considera de forma independiente la ubicación del centro educativo y la de la residencia de los padres, toda vez que, según el tipo de comunidad de habla, puede haber divergencias importantes entre estas dos circunstancias. Al respecto, consideramos que las características de nuestra provincia hacen innecesaria tal distinción. Como ya comentamos (*vid.* 4.3.1), en nuestro ámbito de estudio es muy importante el peso demográfico y socioeconómico ejercido tradicionalmente por la capital, incluso no existe ningún otro foco de población superior

⁴⁸⁴ No faltan tampoco investigaciones donde se distinguen varias zonas según las características demográficas y orográficas del área estudiada (*v.gr.* Mateo 1998, Carcedo 2001, Gómez Molina y Gómez Devís 2004, Martínez Olmos 2008).

a los 20 000 habitantes. El reparto de la población en la misma tampoco es equitativo, sino que se caracteriza por su gran dispersión y, al mismo tiempo, por la concentración en la capital y su comarca, donde vive casi la mitad de los habitantes de la provincia (*vid.* 4.3.1., gráficos 4.6 y 4.7).

Además, siguiendo a Samper *et al.* (2000: 44) consideramos que, para que un núcleo de población sea urbano, se tiene que tener en cuenta una serie de parámetros como los porcentajes de población activa en los diferentes sectores, los equipamientos municipales y los bienes de consumo cultural, así como una serie de valores simbólicos y de identidad con los que se construya un determinado estilo de vida, distinto al de la vida rural. Los mismos autores, en su estudio sobre la estructura social en Lleida, son contundentes al afirmar que “més de la meitat dels municipis de Lleida tenen menys de 500 habitants, cosa que accentua les diferències entre la ciutat i el territori” (Samper *et al.* 2000: 44). Hay que destacar, también, el notable desarrollo reciente de algunas capitales de comarca —como *Tàrrrega*, *Mollerussa* o *Balaguer*—, aunque su nivel de influencia y actividad económica es incomparable a todas luces con el que ejerce la capital (*vid.* 4.3.1).

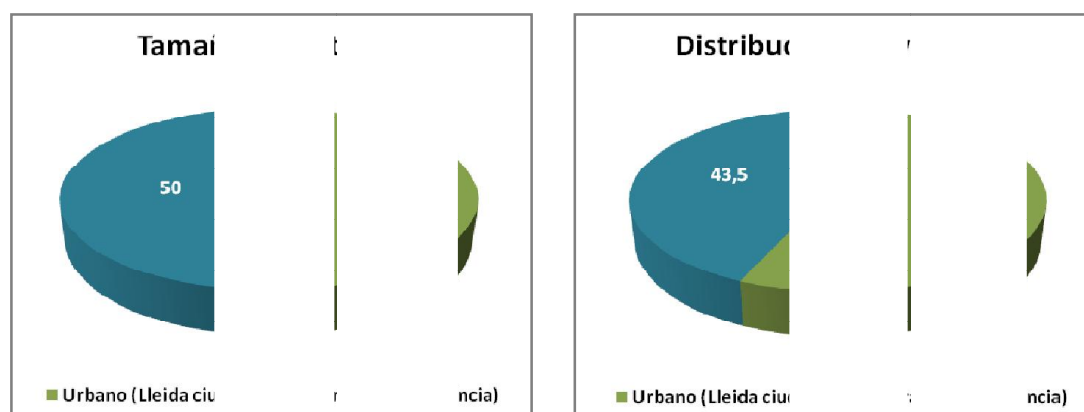
Por todas las razones argüidas, y tal como han hecho también otros autores (Carcedo 2001, Bellón 2011, Arnal *et al.* 2004, Ávila 2006), hemos considerado como urbanos los informantes procedentes de los seis centros educativos de la ciudad de Lleida y, como rurales, los procedentes de los restantes institutos de la provincia. En la capital, para ser coherentes, hemos eliminado los datos de aquellos informantes que no residían en la misma ciudad sino en poblaciones cercanas, tanto de Lleida como de Huesca⁴⁸⁵. Creemos que los centros urbanos deben representar a la colectividad urbana y que la inclusión de estos informantes podía distorsionar el peso de esta variable. A continuación podemos observar la estratificación de la muestra según este parámetro:

⁴⁸⁵ No han sido pocos los encuestados en esta situación ya que no existe en todos los pueblos de ambas provincias un centro público o privado donde pueda estudiarse el curso preuniversitario.

5.4. Distribución por zona geográfica según distintos parámetros.

Zona geográfica	Urbano (Lleida ciudad)	%	Rural (resto provincia)	%	Total
Población general (año 2002, Idescat)	115 000	30.99	256 055	69.01	371 055
Número de centros educativos	20	39.22	31	60.78	51
Población en 2º de bachillerato	1396	56.50	1075	43.50	2471
Muestra	120	50	120	50	240

Gráfico 5.5. Distribución según la variable zona geográfica.



Si observamos estrictamente la población estudiada —el conjunto de alumnado matriculado en 2º de bachillerato en el curso académico 2002/2003— se observa que predominan, aunque no con mucha diferencia, los informantes residentes en la ciudad de Lleida. No obstante, la situación cambia ostensiblemente si se tiene en cuenta la ubicación de los centros educativos, ya que, en este caso, son más los situados en el resto de la provincia que cuentan, sin embargo, con menos alumnado que los de la ciudad. Como indican Bartol y Hernández Muñoz (2006: 728), si el diseño de la muestra se basa en los centros educativos más que en el alumnado, se favorecen las zonas con menos habitantes, y, en nuestro caso, creemos que se reflejan mejor las distintas variedades lingüísticas que se aprecian en la zona. Ante esta disyuntiva, hemos optado por equiparar el número de alumnos de los centros urbanos y rurales, opción que facilita, por otro lado, el análisis cuantitativo de los datos.

Respecto al *nivel sociocultural* del alumnado, este fue determinado teniendo en cuenta la profesión y los estudios del padre y de la madre, y el nivel cultural de la familia. A pesar de que los estudios de disponibilidad léxica han advertido de forma categórica sobre la importancia de este factor (Samper Padilla y Samper Hernández 2006: 54-63), no hay un acuerdo unánime sobre la forma de establecerlo⁴⁸⁶. Algunas de las investigaciones llevadas a cabo (López Morales 1999a, Alba 1995a, Carcedo 2001 o Ayora 2004) han tenido en cuenta tanto los estudios y la profesión de los padres como los ingresos familiares. En general, la mayoría de las investigaciones españolas sigue muy de cerca la clasificación establecida por Samper Padilla y Hernández Cabrera (1997: 230) y contempla solo dos parámetros: la profesión y el nivel educativo alcanzado por los progenitores. En otras ocasiones (Benítez Pérez 1992a, Arnal *et al.* 2004, López Meirama 2008 o Pastor y Sánchez 2008), el nivel sociocultural se ha establecido basándose únicamente en un factor, a saber, el nivel educativo de los padres.

Lo que resulta evidente es que resulta muy difícil establecer los parámetros que identifican fielmente este factor social. Si bien existen distintos niveles socioculturales o socioeconómicos en cualquier comunidad de habla —tal y como la mayoría de sus integrantes percibe—, es muy complicado saber qué es realmente lo que divide en estratos sociales distintos a los diversos individuos de una misma sociedad⁴⁸⁷. Nos pareció que, por lo menos en nuestra área de estudio, no siempre el nivel cultural de una familia tiene un paralelismo exacto con el grado de instrucción de los padres ni, mucho menos, con el estatus social de las profesiones que desempeñan⁴⁸⁸. Es decir, podemos hallar una familia en la que los padres desempeñen profesiones de poco prestigio social pero que tengan, por sus inquietudes personales, un gran bagaje cultural; de la misma manera que una profesión prestigiosa, como puede ser cualquier ingeniería u otras como farmacéutico, químico, etc. —y no pretendemos hacer ninguna valoración—, no necesariamente va acompañada de una preocupación por mantener, acrecentar o gozar

⁴⁸⁶ Hernández Muñoz (2006: 346) recalca que se deberían establecer unas normas comunes a todas las investigaciones con respecto a la clasificación de los niveles socioculturales.

⁴⁸⁷ Moreno Fernández (1990: 85) afirma acertadamente: “[a] priori no puede especificarse cuáles son los factores que funcionan en todas las sociedades para distinguir clases sociales, porque ni siquiera se puede afirmar que todas las sociedades estén estratificadas de esta manera”.

⁴⁸⁸ Mateo (1998: 77) también expresa sus reservas respecto al establecimiento del factor sociocultural de un informante y constata, siguiendo a García Marcos, que otros autores, en otro tipo de estudios, distinguen con mayor nitidez entre clase social y cultura. En la ciudad de Lleida ya se ha realizado una tesis doctoral en la que se plantea la influencia del estatus sociocultural de la familia en las construcciones sintácticas del lenguaje infantil (Mayoral 1994).

de algunos aspectos de la cultura que contribuyen al cultivo intelectual y espiritual de las personas. De modo similar, tampoco un nivel de estudios alto equivale a una profesión prestigiosa, tal como puede apreciarse en el panorama actual de crisis económica en el que nos hallamos inmersos.

Con el fin de establecer este parámetro, que denominamos “nivel cultural de la familia”, incorporamos en la encuesta una serie de preguntas que versaban sobre distintos aspectos: la cantidad de libros que leía el informante y que se compraban en su casa, la asiduidad con que se compraba el periódico en su hogar y si el informante lo leía o no, la pertenencia del alumno a alguna asociación de carácter cultural y la frecuencia con la que asistían sus progenitores a distintos acontecimientos culturales como espectáculos teatrales, de danza o conciertos musicales⁴⁸⁹. Este conjunto de preguntas —un total de siete cuya formulación exacta puede verse en el apéndice 2— fue valorado con tres puntos por respuesta, con lo que todos los informantes se clasificaron en una escala que iba de 0 a 21. Hemos de señalar que algunas preguntas permitían establecer intervalos de 0 a 3 en la puntuación, mientras que otras solo permitían una respuesta cerrada y, por ende, solo podía obtenerse la máxima puntuación o ninguna⁴⁹⁰. Con respecto al gran número de asociaciones recabadas, se descartaron las de tipo político, deportivo o religioso y solo se consideraron las estrictamente culturales. Entre estas abundan los grupos de teatro, las *colles sardanistes*, los conjuntos de canto coral o las asociaciones tradicionales como la de moros y cristianos. La relación de todas las que fueron tenidas en cuenta puede consultarse al final del apéndice 1.

Para la clasificación del grado de instrucción y de la actividad profesional seguimos los índices que detallan Samper Padilla y Hernández Cabrera (1997: 230 nota 6) según los cuales se establecen seis niveles educativos (1: sin estudios, 2: enseñanza primaria, 3: bachillerato elemental, 4: bachillerato superior, 5: estudios universitarios de grado medio, 6: estudios universitarios de grado superior) y seis categorías profesionales (1: obrero sin cualificar, 2: obrero con especialización, 3: empleados medios, 4: pequeños empresarios, autónomos, personal técnico de tipo medio, 5:

⁴⁸⁹ Estas preguntas están inspiradas en la noción de “capital cultural” de Mayoral (1994: 109-110 y 122-125).

⁴⁹⁰ Por ejemplo, a la primera pregunta sobre la cantidad de libros que se compraban al año en casa del informante se podía responder con 0 (0 puntos), entre 1 y 6 (1 punto), entre 6 y 12 (2 puntos) o entre 12 y 24 o más de 24 (3 puntos). En cambio, si el alumno leía cada día el periódico o pertenecía a alguna asociación cultural se valoraba con 3 puntos, y en caso negativo, con ninguno.

medianos empresarios, personal técnico de titulación superior, 6: profesionales liberales, altos directivos y grandes empresarios). Al valorar estos factores, el alumnado recibió, por tanto, una puntuación de entre 2 y 24 puntos en cada uno de los dos parámetros. En los casos de familias monoparentales o en los que uno de los dos progenitores había fallecido o no se aportaba información sobre él, multiplicamos por dos la información que poseíamos, normalmente la de la madre. Por otro lado, si bien el nivel educativo alcanzado no presentó grandes problemas, la profesión sí nos reportó algunos. Como ya señalan Fernández Smith *et al.* (2008: 49), uno de los inconvenientes fue decidir qué estatus tiene la población inactiva, o en qué categoría se incluyen determinadas profesiones denominadas con etiquetas amplias como “funcionario”⁴⁹¹.

En último lugar, consideramos los tres factores de forma conjunta (las profesiones y los estudios de los progenitores y el nivel cultural de la familia) mediante una sumatoria no ponderada en la que cada parámetro tenía el mismo valor y establecimos un *continuum* que dividimos en tres grupos correspondientes al nivel sociocultural alto, medio y bajo⁴⁹². La muestra se distribuye, pues, en estos estratos. La mayor parte de informantes se concentra en el estrato medio, mayoritario también en la población general, y, por ende, en el universo estudiado⁴⁹³. Concretamente, en el nivel bajo hallamos 69 informantes, en el medio, 135 y, en el alto, 36. Ya que esta forma de calcular el nivel sociocultural del informante supone una novedad metodológica dentro de nuestra disciplina, hemos optado por estratificar también la muestra considerando únicamente el nivel de estudios y las profesiones de los progenitores, tal como realizan las investigaciones que siguen a Samper Padilla y Hernández Cabrera (1997):

⁴⁹¹ Al no hallar una solución mejor, incluimos al funcionariado en la categoría 3 de empleado medio (cuando el alumno no brindaba más información al respecto) y otorgamos un 1 a los casos de progenitores en paro o inválidos.

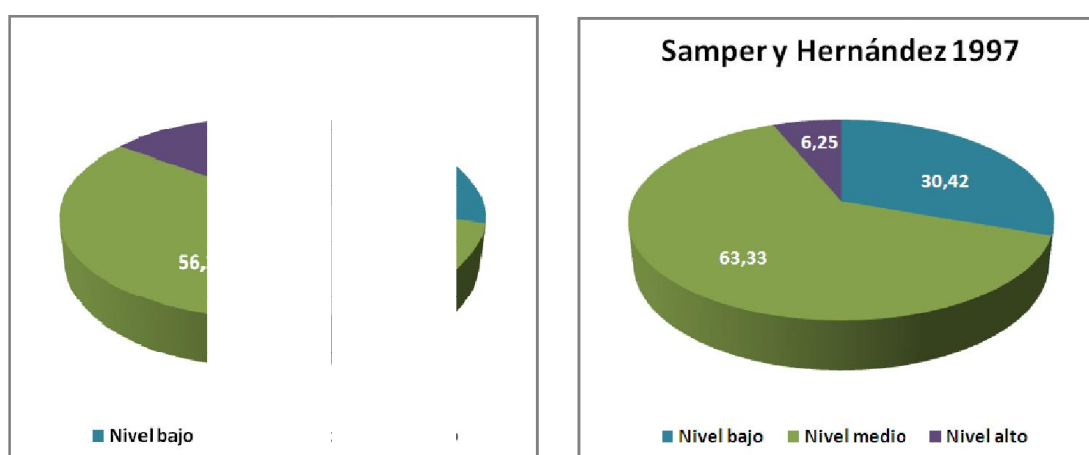
⁴⁹² Otras investigaciones dividen la variable del nivel sociocultural en cuatro grupos (*v.gr.* Hernández Muñoz 2006 o Pastor y Sánchez 2008).

⁴⁹³ A pesar de que, como ya hemos dicho, no poseemos datos sobre este parámetro en relación con los dos segmentos de población citados. Véase al respecto Fernández Smith *et al.* (2008: 45, nota 19).

Tabla 5.5. Distribución de los informantes según distintos criterios de estratificación social.

Nivel sociocultural	Nivel bajo	Nivel medio	Nivel alto
Lleida (n° informantes)	69	135	36
Lleida (%)	28.75	56.25	15
Samper y Hernández 1997 (n° informantes)	73	152	15
Samper y Hernández 1997 (%)	30.42	63.33	6.25

Gráfico 5.6. Distribución según la variable nivel sociocultural (%).



Según se observa en la tabla 5.5 y el gráfico 5.6, al añadir el parámetro del nivel cultural de la familia hay una cantidad de informantes, sobre todo del nivel medio, que se incorporan al nivel alto. Nuestro cálculo del nivel sociocultural supone una estratificación más proporcionada de los distintos informantes que, si bien otorga un peso considerable a la clase media —tal como también ocurre en la realidad—, aporta un mayor número de informantes al estrato alto. En el caso concreto de nuestra muestra, consideramos más adecuada nuestra distribución, toda vez que la cantidad de informantes que obtendríamos con la clasificación de Samper Padilla y Hernández Cabrera (1997) en el nivel más elevado sería muy baja, sin llegar a alcanzar siquiera el 7% del total. De cualquier modo, será el análisis de los datos el que determine, realmente, si la estratificación mediante el índice que hemos ideado resulta o no satisfactoria, esto es, si covaría o no significativamente con la producción lingüística.

La última variable contemplada se refiere a la adscripción lingüística del informante. Como ya vimos ampliamente (*vid.* 4.3.2), nuestros informantes pertenecen

al conjunto, ampliamente mayoritario en nuestra provincia, de personas cuya lengua materna es el castellano o el catalán pero que son capaces de utilizar con fluidez ambas lenguas⁴⁹⁴. Con la finalidad de conocer cuáles eran las lenguas generalmente empleadas por los informantes, incorporamos en la prueba una serie de cuestiones. Primero, preguntamos directamente cuál era la lengua materna y, en segundo lugar, para conocer cuál era la lengua de uso habitual del informante, detallamos los principales ámbitos en los que se mueve un estudiante de 2º de bachillerato y solicitamos qué lengua empleaba habitualmente en ellos. Concretamente, preguntamos qué lengua utilizaba en su casa, con sus amigos y en la escuela. La lengua imperante en el conjunto de estos tres ámbitos fue la considerada habitual del alumno⁴⁹⁵.

Cabe precisar que, al aplicar la encuesta, observamos que algunos informantes dudaban acerca de la etiqueta de “lengua materna”, con lo que, además de explicar en voz alta el concepto cuando así lo solicitaban, adoptamos como criterio comprobar la correlación entre los datos sociológicos de las dos encuestas realizadas⁴⁹⁶ y también entre la lengua materna y la lengua que daba el informante como habitualmente empleada en el hogar, correspondencia que se dio en casi todas las ocasiones. Cuando no fue así, el alumnado incluyó las dos lenguas (catalán y castellano) como maternas o de uso habitual en el hogar. Finalmente, las variantes que tuvimos en cuenta fueron cuatro: catalán, castellano, ambas (castellano/catalán)⁴⁹⁷ y aranés. Esta última fue añadida por unos pocos informantes ubicados en *la Vall d’Aran*. A pesar de que la proporción de estos dentro del conjunto de la muestra es muy baja, consideramos imprescindible mantenerlos como variante, ante la imposibilidad de incorporarlos a ninguna de las otras opciones.

⁴⁹⁴ Si recordamos los datos que ofrecía el censo lingüístico de 2001 sobre el catalán, el 98.49% de los jóvenes de entre 15 y 19 años lo entienden, y un 95.32% se declara capaz de hablarlo (*cf.* 4.3.2. gráfico 4.17).

⁴⁹⁵ Como la pregunta versaba sobre tres ámbitos, fue fácil, de entrada, establecer qué lengua predominaba. Los informantes que se decantaron por el uso de ambas (castellano y catalán) en al menos dos ámbitos, fueron catalogados dentro de la opción “ambas (lenguas)”.

⁴⁹⁶ Como a continuación detallamos (*vid.* 5.3), la encuesta fue aplicada en dos ocasiones, una en castellano y otra en catalán, de modo que el apartado inicial con las respuestas a todos los datos sociológicos fue completado dos veces por los mismos estudiantes.

⁴⁹⁷ Tal como puede verse en el apéndice 2, en la encuesta no se ofrecía ninguna alternativa a la pregunta sobre la lengua materna, en la lengua habitualmente empleada, en cambio, sí se nombraban las dos posibilidades: castellano o catalán. Sin embargo, algunos informantes se decantaron por incluir ambas lenguas (castellano y catalán) en las dos opciones.

Por otro lado, en un principio, pensamos contemplar las dos variables, es decir, tanto la lengua materna como la de uso habitual (*cf.* Serrano Zapata 2003a: 934). Sin embargo, finalmente, optamos por considerar solo una, la lengua materna, cuya influencia sobre los resultados pensamos, de forma intuitiva, que sería más contundente⁴⁹⁸. No obstante, puesto que estratificamos la muestra según las dos variables, detallamos a continuación la distribución de las mismas, sin descartar el análisis futuro de la variable *lengua habitual*:

Tabla 5.6. Distribución de los informantes según las variables de adscripción lingüística.

Adscripción lingüística	Castellano	Catalán	Ambas	Aranés
Lengua materna (n° informantes)	66	157	10	7
Lengua materna (%)	27.50 %	65.42 %	4.17 %	2.92 %
Lengua habitual (n° informantes)	56	160	23	1
Lengua habitual (%)	23.33 %	66.67 %	9.58 %	0.42 %

Tal como se puede apreciar, la lengua que predomina de forma amplia en las dos variables es el catalán. Esto es lógico, puesto que, como explicamos con anterioridad, la provincia de Lleida se caracteriza por la fuerte presencia de la lengua catalana tanto en los ámbitos interpersonales como en los públicos. De hecho, si recordamos los datos de la encuesta de usos lingüísticos de 2003 que comentamos en el capítulo anterior (*vid.* 4.3.2.), los porcentajes de uso que se recaban para la provincia son muy similares a los que se detallan en nuestra muestra, tal como puede observarse en los siguientes gráficos.

⁴⁹⁸ La decisión de considerar solo una de las dos variables de adscripción lingüística se debió, en parte, a las limitaciones del programa informático, que, como a continuación veremos, permite únicamente considerar cinco variables.

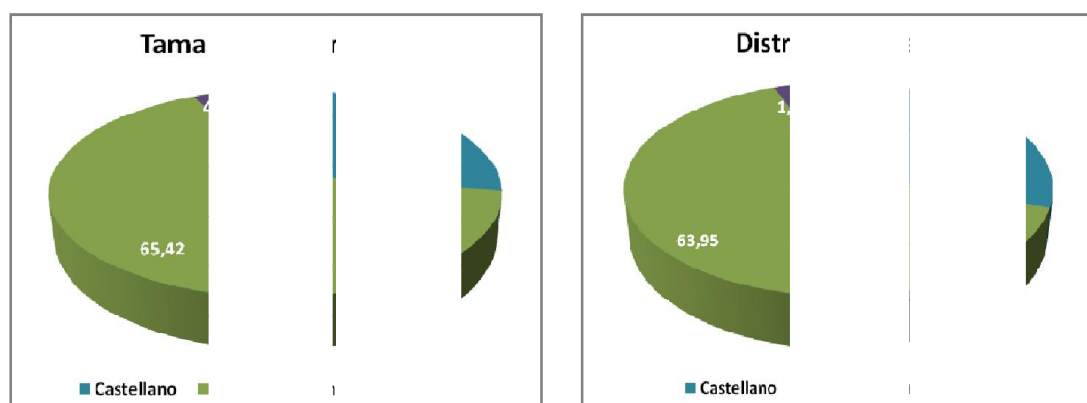
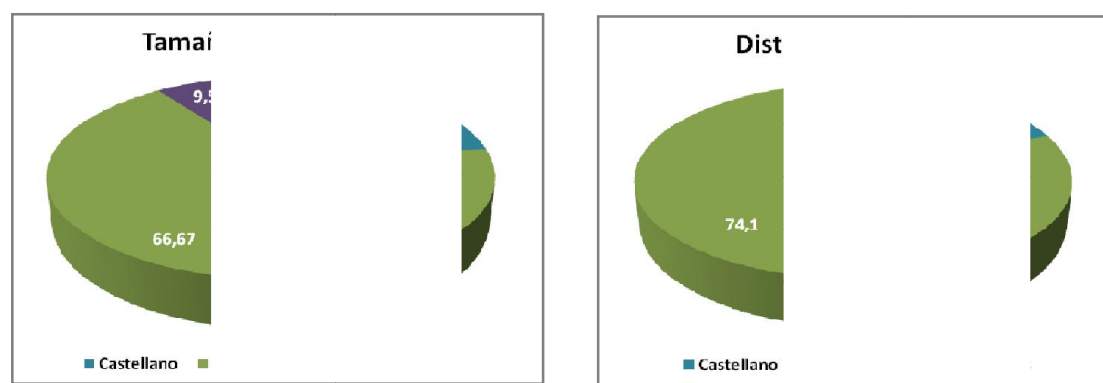
Gráfico 5.7. Distribución según la variable lengua materna⁴⁹⁹.

Gráfico 5.8. Distribución según la variable lengua habitual.



Es curioso constatar que, de modo similar a lo que ocurre en el conjunto de la población, son pocos los informantes que se decantan por el uso de las dos lenguas ya que la mayoría opta por una de las dos. También son más, en los dos segmentos de población, los que eligen esta opción como *lengua habitual* antes que como *lengua materna*. Otra coincidencia con el comportamiento real es el mayor índice de personas que utilizan el catalán como lengua habitual en comparación con las que lo tienen como lengua materna o inicial, hecho que corrobora que parte de la población de origen castellanohablante, como ya vimos con anterioridad (*vid.* 4.1 y 4.3.2.), no solo adopta el catalán en sus interacciones cotidianas sino que se identifica con esta lengua a lo largo de su vida.

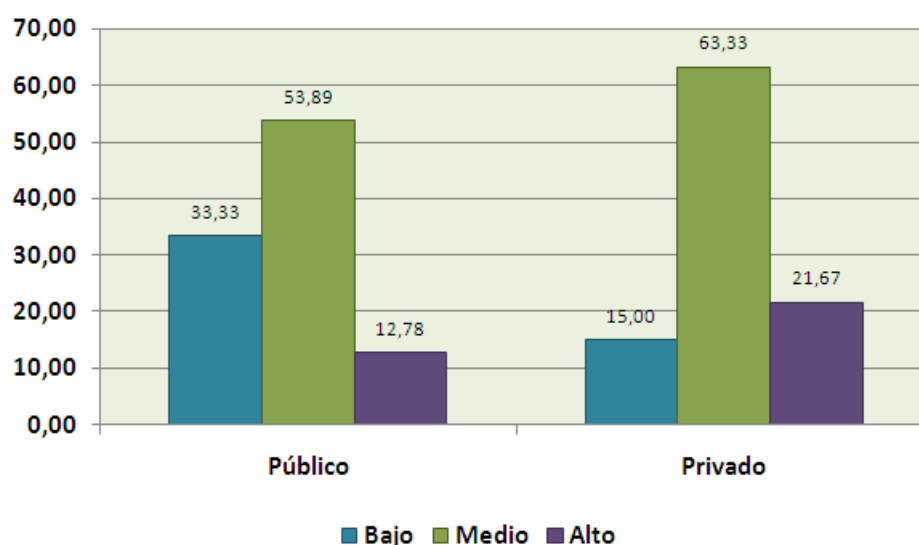
De otra parte, en algunos territorios bilingües se ha estudiado la incidencia de la variable *modelo de enseñanza* o *lengua utilizada en la enseñanza*⁵⁰⁰ para observar hasta

⁴⁹⁹ En este gráfico y el siguiente, para obtener la distribución real hemos calculado la media de los porcentajes de los gráficos 4.19 y 4.20 (*vid.* 4.3.2.) correspondientes a los dos ámbitos de la provincia (*Àmbit de Ponent* y *Alt Pirineu i Aran*).

qué punto la lengua en que se reciben los conocimientos influye en el léxico disponible del alumno en ambas lenguas. Sin embargo, la situación sociológica de nuestra comunidad impide este análisis, puesto que todos los centros educativos usan el catalán como lengua vehicular⁵⁰¹.

Después de establecer todas las variables, consideramos interesante observar si había algún tipo de relación entre ellas, con lo que analizamos todas las combinaciones posibles mediante una tabla dinámica de Excel. Tras realizar este análisis nos percatamos de que, como ya se ha detectado en ocasiones, hay una cierta relación entre el *nivel sociocultural* del alumno y el *tipo de centro* al que acude, de modo que hay una mayor presencia de alumnado del estrato bajo en los centros públicos y, de forma inversa, del nivel alto en los centros privados:

Gráfico 5.9. Distribución según el tipo de centro y el nivel sociocultural (%).



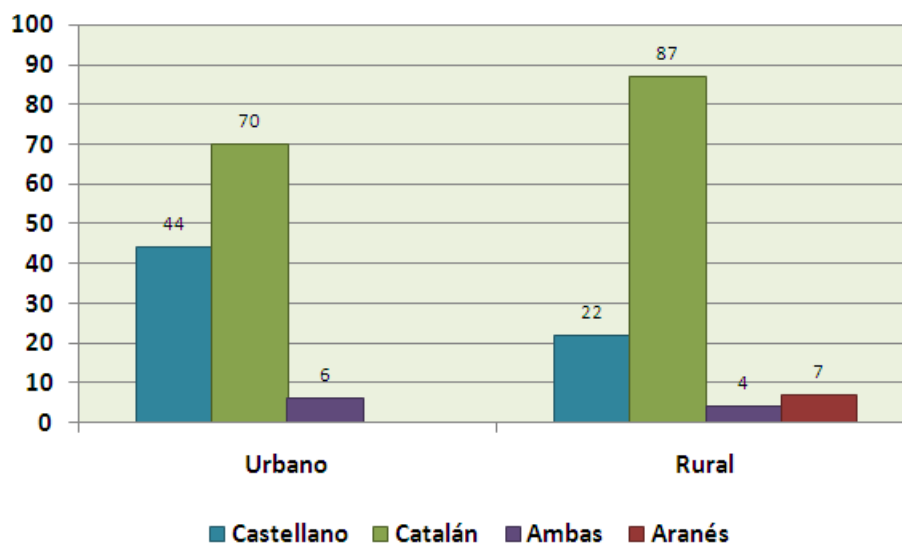
Otras variables que se hayan relacionadas son la *lengua materna* y la *zona geográfica*. En la capital, aunque predomina el catalán, las dos lenguas se hallan más igualadas con respecto al número de informantes que las poseen como lenguas maternas. En cambio, en el resto de la provincia, la presencia del catalán es mayoritaria.

⁵⁰⁰ Sobre la incidencia de esta variable en otros territorios bilingües, véanse algunas de las conclusiones del capítulo 3 (vid. 3.4.3).

⁵⁰¹ La *Llei de Normalització Lingüística* del año 83 y el actual *Estatut d'Autonomia* determinan que el catalán es la lengua propia de la enseñanza en todos los niveles educativos (vid. 4.1.). A pesar de ello, recopilamos información sobre el número de horas semanales que recibía el alumnado en las dos lenguas cooficiales, por si se pudiera constatar alguna repercusión en los resultados; no obstante, finalmente desechamos esta información, ya que la proporción era muy similar en todas las instituciones educativas analizadas.

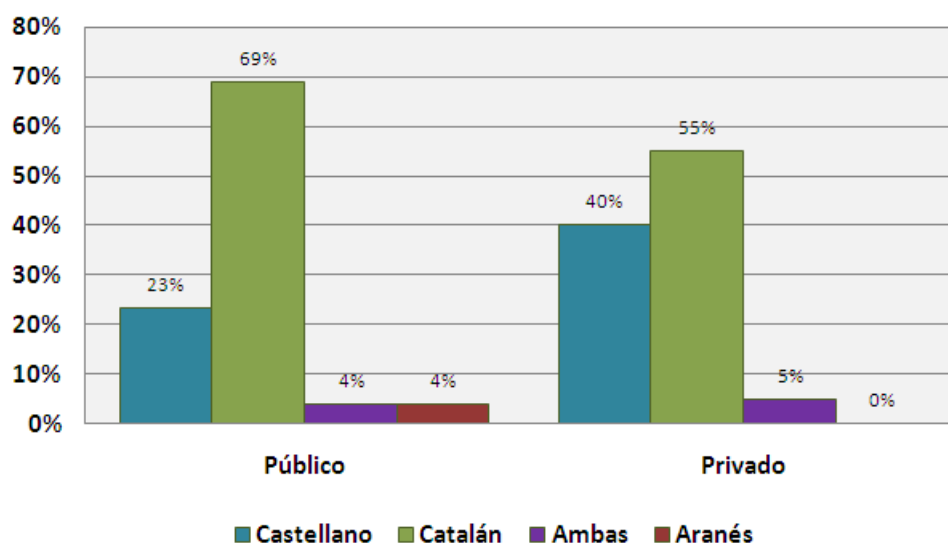
Esto se debe a que, como vimos en el capítulo anterior (*cf.* 4.1. y 4.3.2., especialmente los gráficos 4.11 y 4.12), el contingente obrero venido de tierras castellanohablantes durante la segunda mitad del siglo XX tendió a concentrarse en las ciudades.

Gráfico 5.10. Distribución según la zona geográfica y la lengua materna (*nº de informantes*).



Como consecuencia de la relación entre las variables anteriores, y dado que todos los centros privados se ubican en la capital, se observa la misma relación entre el *tipo de centro* al que acude el alumnado y la *lengua materna* del mismo. En las instituciones concertadas las dos lenguas tienen una presencia más equilibrada que en las públicas, donde la situación se decanta, claramente, a favor del catalán como lengua materna de la mayoría del alumnado.

Gráfico 5.11. Distribución según el tipo de centro y la lengua materna (%).



Finalmente, la muestra, según las variables sociales analizadas, se distribuye de la siguiente forma:

Tabla 5.7. Distribución de las variables sociales analizadas en la muestra.

Variable	Variantes			
Sexo	1-Varones: 93		2-Mujeres: 147	
Tipo de centro	1-Público: 120		2-Privado: 60	
Zona geográfica	1-Urbana: 120		2-Rural: 120	
Nivel sociocultural	1-Bajo: 69	2-Medio: 135	3-Alto: 36	
Lengua materna	1-Castellano: 66	2-Catalán: 157	3-Ambas: 10	4-Aranés: 7

5.3. OBTENCIÓN Y PROCESAMIENTO DE LOS DATOS

El mecanismo de obtención de los datos fue el siguiente. Como ya dijimos, enviamos una carta a todos los centros que impartían 2º de bachillerato en la capital y a los situados en las poblaciones que nos interesaban del resto de la provincia. En esta carta explicamos las directrices básicas de la investigación y solicitamos ciertos datos estadísticos sobre el alumnado que nos resultaban de interés. Concretamente, la información que requerimos hizo referencia a la cantidad, la procedencia y el porcentaje

de alumnos y alumnas que cursaban el nivel preuniversitario en dicho centro⁵⁰². Tras recibir las respuestas de cada uno de ellos y ver, además, cuáles mostraban más interés en participar en nuestro estudio, procedimos a seleccionar los seis que nos parecieron más adecuados de la capital. Después de elegir todos los centros, nos pusimos de nuevo en contacto con sus respectivos directores con el fin de establecer un día y una hora en la que pudiéramos pasar la encuesta. El grupo de bachillerato nos fue asignado completamente al azar y la hora lectiva destinada a tal fin acostumbró a ser la de tutoría o la de lengua castellana. Tanto el contacto con el centro como el acto de pasar todas las encuestas en el aula fueron realizados por la autora de este trabajo. Nos pareció que era imprescindible controlar de forma estricta las condiciones de reacción de los informantes y estábamos convencidos de que estas serían muy similares si aplicaba las encuestas la misma persona⁵⁰³. En cuanto a las características de la investigadora, el hecho de pertenecer a la misma comunidad de habla que los encuestados fue una ventaja; por el contrario, pudimos constatar que el ser una mujer y además joven provocó que parte del alumnado no tomara con seriedad la prueba. Debido a estas circunstancias, solicitamos a partir de la primera experiencia, que el tutor del grupo o un profesor estuviera presente en el aula para garantizar, de este modo, el correcto desarrollo de la prueba.

Una vez allí, se repartió un cuestionario a cada alumno y se explicó brevemente la finalidad del estudio. Las instrucciones con respecto al mecanismo de cumplimentación de la encuesta fueron claras y sencillas. Subrayamos que contarían con un tiempo limitado de dos minutos para escribir todos aquellos términos que se les ocurrieran en relación con la palabra que se hallaba como epígrafe al inicio de cada hoja, pudiendo sobrepasar las líneas numeradas que había debajo de cada campo léxico. También realizamos oralmente un ejemplo, para nada relacionado con los que luego tenían que hacer⁵⁰⁴, e hicimos especial hincapié en que la prueba no puntuaba ni tendría

⁵⁰² Con esta medida intentamos asegurarnos de no pasar las encuestas en centros donde el alumnado de este nivel académico predominara en uno de los dos sexos, procediera mayoritariamente de otros lugares o no alcanzara la cantidad que requería nuestro estudio.

⁵⁰³ En Hernández Muñoz (2006: 346-354) pueden verse algunos de los inconvenientes importantes que puede tener el hecho de que diferentes personas controlen la aplicación de la prueba en una misma investigación.

⁵⁰⁴ Concretamente, propusimos “astronomía” como estímulo, ante el cual se podría escribir palabras como “luna” o “cohete”.

repercusión alguna en sus calificaciones⁵⁰⁵, y que, por tanto, debían escribir todas aquellas palabras que les vinieran a la mente, aunque no supieran a ciencia cierta cómo se escribían, ya que la ortografía no se tendría en cuenta. Cabe destacar que, a diferencia de las instrucciones dictadas en las investigaciones valencianas (Gómez Molina y Gómez Devís 2004: 49, nota 54 y Llopis 2009: 73), en ningún momento se dijo que podían escribir palabras en otra lengua que no fuera en la que se realizaba la encuesta, esto es, incluir palabras en catalán si la encuesta era en castellano, y a la inversa. Tanto la explicación inicial como las instrucciones se llevaron a cabo en la lengua en la que los alumnos debían completar en aquella ocasión la encuesta. Es decir, la primera vez en castellano, y la segunda, en catalán.

La encuesta, que puede consultarse en el apéndice 2 de este trabajo, consta de 19 páginas⁵⁰⁶. En las dos primeras se halla una serie de preguntas encaminadas a determinar las variables sociológicas del informante —como se ha dicho más arriba— y, a partir de la tercera página, se distribuyen los centros de interés, a razón de uno por página, en torno a los cuales el alumnado tenía que producir, en el tiempo máximo de dos minutos, las listas de unidades léxicas. Esta es una novedad metodológica de nuestra investigación puesto que en la mayoría de estudios hispánicos se distribuyen tres centros de interés por página. Consideramos que, de esta manera, es muy difícil controlar que el alumno dedique realmente dos minutos a cada campo puesto que, al terminar el primero, ya tiene el segundo al lado para ir escribiendo las palabras o simplemente para ir pensando en él. En cambio, como nuestro sistema exige pasar la hoja para cambiar de centro, de esta forma puede controlarse mejor la tarea del alumno. Otra solución que nos parece adecuada es la de Hernández Muñoz (2004: 28, 2006: 262-263), que optó, en sus estudios sobre Cuenca y sobre Castilla-La Mancha, por no dar los nombres de los centros de interés, sino por ir dictándolos a medida que los informantes los iban completando.

⁵⁰⁵ Este punto nos pareció interesante subrayarlo con la finalidad de que el alumnado no cayera en la tentación de compartir sus respuestas con los compañeros más cercanos.

⁵⁰⁶ Por razones de espacio no se incluye toda la encuesta en el apéndice 2. De cualquier modo, de la tercera página hasta la última solo varía el título de cada centro de interés.

La duración total de la encuesta osciló entre 45 y 50 minutos durante los cuales los alumnos contestaron en silencio a la misma⁵⁰⁷. El mecanismo de la prueba se repitió al cabo de un mínimo de dos semanas de forma idéntica con el cuestionario sobre disponibilidad léxica catalana⁵⁰⁸. El hecho de dejar dos semanas de tiempo entre una prueba y otra respondía a nuestro deseo de que el alumnado se viera obligado a realizar de nuevo el mecanismo asociativo sin traducir las palabras que, de otro modo, pudiera recordar de la primera prueba hecha en castellano, toda vez que los informantes eran exactamente los mismos⁵⁰⁹. Optamos, de forma intuitiva, por este intervalo, al no conocer referencias bibliográficas sobre cuál debía ser el margen necesario entre las dos pruebas. No obstante, los pocos estudios existentes realizados también con los mismos informantes en distintas lenguas eligen también casi todos esta opción (*vid.* 3.4.2.). Por otro lado, dado que la encuesta era anónima, dedicamos algún tiempo a encontrar la “pareja” correspondiente a la prueba en la otra lengua. Este problema lo solucionamos gracias a la abundante información sociológica que los informantes aportaban en la primera parte de la prueba⁵¹⁰ y, en muchos casos, ni siquiera fue necesario recurrir a esta información porque era evidente, por la caligrafía, que se trataba de un mismo informante. Como ya hemos comentado (*vid.* 5.2), esta metodología facilitó el establecimiento de algunas variantes, por ejemplo, la lengua materna o de uso del informante⁵¹¹.

⁵⁰⁷ Las investigaciones valencianas (Gómez Molina y Gómez Devís 2004: 49, nota 54, Llopis 2009: 73), por el contrario, dieron algunas instrucciones antes de cada centro de interés, del tipo “La ciudad-imagina que paseas y todo lo que ves”. En Gómez Devís (2007) se demuestra que esta forma de contextualizar las áreas nocionales no afecta a los resultados. De otra parte, Hernández Muñoz (2006: 120-121) lleva a cabo un interesante experimento donde compara el léxico disponible recogido de forma escrita, tal como nosotros llevamos a cabo, y de forma oral. Comprueba cómo el método de recogida de datos no influye en el nivel cuantitativo, en cambio, sí se aprecian diferencias en el nivel cualitativo, ya que los procesos de lexificación son distintos.

⁵⁰⁸ Algún alumno preguntaba con sorna, al despedirnos, si volveríamos al cabo de 15 días con la misma prueba en inglés.

⁵⁰⁹ En un principio (*cfr.* Serrano Zapata 2003a: 931) pensamos manejar dos muestras paralelas en las que la mayor parte de informantes coincidía. Sin embargo, decidimos, finalmente, eliminar las encuestas de los informantes que no habían realizado alguna de las dos pruebas y mantener una sola muestra con la finalidad de poder explotar mejor el hecho, en parte novedoso en los estudios actuales de disponibilidad (*vid.* 3.4.2.), de manejar los datos de los mismos informantes en dos lenguas.

⁵¹⁰ Por ejemplo, en un mismo grupo escolar era muy difícil que hubiera coincidencia entre el lugar de nacimiento de ambos progenitores, el nivel de estudios alcanzado o su profesión. Otra solución hubiera sido que se perdiera el anonimato, pero en contra de esta opción estaba el hecho de que pensábamos que los informantes podrían verse de algún modo coaccionados y dedicaran más atención a los aspectos ortográficos de la prueba.

⁵¹¹ También en otras ocasiones nos ayudó al establecimiento del nivel sociocultural ya que el alumno concretó, en una de las dos pruebas, la información referente a la profesión de sus padres.

Los centros de interés que utilizamos fueron los siguientes⁵¹²:

01. <i>El cuerpo humano</i>	01. <i>El cos humà</i>
02. <i>Ropa y complementos</i>	02. <i>Roba i complementos</i>
03. <i>Partes de la casa (sin los muebles)</i>	03. <i>Parts de la casa (sense els mobles)</i>
04. <i>Muebles de la casa</i>	04. <i>Mobles de la casa</i>
05. <i>Alimentos y bebidas</i>	05. <i>Aliments i begudes</i>
06. <i>Objetos colocados en la mesa para la comida</i>	06. <i>Objectes col·locats a la taula per menjar</i>
07. <i>La cocina y sus utensilios</i>	07. <i>La cuina i els seus utensilis</i>
08. <i>La escuela: muebles y materiales</i>	08. <i>L'escola: mobles i materials</i>
09. <i>Iluminación, calefacción y medios para airear un recinto</i>	09. <i>Il·luminació, calefacció i mitjans per airejar un recinte</i>
10. <i>La ciudad</i>	10. <i>La ciutat</i>
11. <i>El campo</i>	11. <i>El camp</i>
12. <i>Medios de transporte</i>	12. <i>Mitjans de transport</i>
13. <i>Trabajos del campo y del jardín</i>	13. <i>Treballs del camp i del jardí</i>
14. <i>Animales</i>	14. <i>Animals</i>
15. <i>Juegos y distracciones</i>	15. <i>Jocs i distraccions</i>
16. <i>Profesiones y oficios</i>	16. <i>Professions i oficis</i>
17. <i>Colores</i>	17. <i>Colors</i>

Como se verá, estos centros son, salvando algunas pequeñas diferencias, los mismos que utilizaron en su estudio pionero Gougenheim *et al.* (1964, *vid.* 2.1.1.). A pesar de que algunas veces se ha comentado la conveniencia de añadir otros centros e incluso de cambiar la denominación de los mismos⁵¹³, todos los estudios que se incorporan al Proyecto Panhispánico mantienen los dieciséis centros clásicos en aras de la uniformidad metodológica⁵¹⁴. Las únicas modificaciones que hemos realizado han sido la inclusión —al igual que han hecho otras investigaciones— del centro

⁵¹² Hernández Muñoz (2006: 153) clasifica estos campos léxicos según su estructura interna (natural, bien definida, *ad hoc* y radial) y sus niveles de inclusividad (inclusivo o relacional), desde una novedosa perspectiva que analiza los fundamentos cognitivos de la disponibilidad.

⁵¹³ Véanse las interesantes reflexiones de Gómez Molina y Gómez Devís (2004: 273) al respecto.

⁵¹⁴ Algunas investigaciones incorporan otros, como Hernández Muñoz (2004, 2006), que añade *La inteligencia* con la finalidad de analizar las diferencias entre los centros de índole concreta o abstracta. Véase un recorrido exhaustivo por los distintos centros de interés empleados en Samper, Bellón y Samper Hernández (2003: 46 y ss.) y en Samper Padilla y Samper Hernández (2006: 22 y ss.).

17. *Colores*, considerado un aspecto básico de la percepción del entorno⁵¹⁵, y la ampliación del segundo estímulo verbal, *La ropa a Ropa y complementos*, puesto que algunos autores (Carcedo 2000a: 76, Samper Padilla 1997: 318) ya advierten de que los alumnos enumeran en este campo tanto prendas de vestir como accesorios y complementos⁵¹⁶.

Los materiales obtenidos mediante los cuestionarios se procesaron, en primer lugar, mediante *Lexidisp* (versión 1.02). Este programa informático, creado *ex professo* para el estudio de la disponibilidad léxica por J.E. Moreno Fernández y A. J. García de las Heras⁵¹⁷, ha sido, durante muchos años, el único instrumento utilizado para el cálculo de la disponibilidad por numerosos investigadores (v.gr. González Martínez 2001, Bartol 2004, Martínez Olmos 2008...). En nuestro caso, su uso se limitó a la tarea de homogeneización de los listados, puesto que dicho programa facilita, de forma muy cómoda, los vocablos ordenados alfabéticamente por cada centro de interés. Este hecho permite poder tomar decisiones observando el conjunto de los datos y, por tanto, viendo los posibles problemas que se plantean en cada área léxica. Además, también se puede constatar rápidamente cómo afectan las pequeñas modificaciones que se realizan al conjunto global de palabras recabadas, con lo que también el investigador puede corregir numerosos detalles que se le hayan pasado por alto con anterioridad. Para poder introducir los datos y realizar los análisis pertinentes se ha de disponer previamente de unos archivos con todas las respuestas de los informantes. Existen unas normas precisas sobre el formato de estos ficheros: deben ser .txt y, antes de las respuestas de cada informante, se ha de incorporar una cadena de dígitos que indique las variables sociológicas, el número del informante y el centro de interés al que se refieren las respuestas. En el siguiente ejemplo se puede comprobar que el informante número 9, varón de un centro urbano privado de nivel sociocultural bajo y lengua materna

⁵¹⁵ Concretamente López Morales y García Marcos (1995: 70) afirman: “«los colores» nos introducen en un aspecto básico de la verbalización de la percepción del entorno, de constante referencia desde los trabajos de los antropólogos lingüistas para el examen de las relaciones lengua/pensamiento/realidad”.

⁵¹⁶ Con respecto al primer centro de interés, modificamos levemente el enunciado de *Partes del cuerpo*, que algunos autores incluyen como *Partes del cuerpo humano*, y lo sustituimos por *El cuerpo humano*, como también realizan Gómez Molina y Gómez Devís (2004) y Bartol (2004). Consideramos, como Samper, Bellón y Samper Hernández (2003: 47), que esta modificación no supone ningún cambio significativo.

⁵¹⁷ Consúltese, para mayor información sobre el programa, F. Moreno Fernández, E. Moreno Fernández y García de las Heras (1995).

castellana, ha escrito la retahíla de palabras que se observan a continuación en el centro de interés 5, correspondiente a *Alimentos y bebidas*⁵¹⁸.

12111 009 05 pollo, pescado, gamba, refresco, (coca) cola, Fanta, atún, queso, jamón, chorizo, sangría, vino, cava, verdura, fruta, naranja, manzana, melocotón, sandía, melón, lechuga, tomate, carne, zumo, plátano, pera, pepino

Cabe destacar que el programa *Lexidisp* permite hacer los cálculos matemáticos de los índices cuantitativos tradicionalmente usados en los trabajos de disponibilidad (palabras, vocablos, índice de cohesión...). No obstante, a excepción del número de vocablos, el resto de índices deben extraerse manualmente a partir de otros datos. Para evitarnos ese tedioso trabajo, una vez realizada la larga tarea de revisión de los materiales, volcamos de nuevo los archivos en la página web www.dispoplex.com. Esta página, instrumento de comunicación y coordinación entre los distintos investigadores del Proyecto Panhispánico, cuenta con un banco de datos donde es posible introducir la información para su posterior cuantificación matemática. Al igual que *Lexidisp*, permite calcular la disponibilidad léxica de los vocablos según cada centro de interés. Los dos programas utilizan la misma fórmula matemática desarrollada por López Chávez y Strassburguer (1991: 92) que tiene en cuenta no solo el número de apariciones de un determinado vocablo, sino también la posición alcanzada en los listados por el mismo⁵¹⁹. Una vez introducidos los archivos, el programa ordena las palabras en función del *grado de disponibilidad*. Además, también nos ofrece la *frecuencia relativa*, o el número de veces que aparece esa palabra con respecto al total; la *aparición*, es decir, el porcentaje de alumnos que han escrito ese término; y, por último, la *frecuencia acumulada*, magnitud referida a la suma de las frecuencias relativas. Los resultados, en forma de diccionarios con los cuatro índices matemáticos que acabamos de mencionar, constituyen los capítulos 10 y 11 de este trabajo. Para facilitar la consulta, los diccionarios están ordenados según el índice de disponibilidad (10.1. y 11.1.) y de forma alfabética (10.2 y 11.2.).

⁵¹⁸ Los cinco primeros dígitos indican las variables sociales y siguen la numeración que indicamos en la estratificación de la muestra de la tabla 5.7., es decir, por ejemplo el primer número indica el sexo (1-varón/2-mujer).

⁵¹⁹ Los primeros estudios de disponibilidad léxica solo tuvieron en consideración la frecuencia conseguida por las unidades. Hoy en día, no se considera que es más disponible el término que más veces aparece en las encuestas, sino aquel que lo hace más veces en los primeros lugares.

Además del cálculo de la disponibilidad general, en www.dispoplex.com se obtienen de forma inmediata todos los índices cuantitativos (número de palabras, número de vocablos, promedio de palabras e índice de cohesión). De otra parte, el banco de datos también permite realizar otros análisis interesantes como la comparación entre palabras e índices de distintos proyectos. Al respecto, hemos de aclarar que los archivos que utilizamos para *Lexidisp* tuvieron que modificarse antes de ser introducidos en *Dispoplex*, ya que en la página web se procesa la información según las instituciones educativas y no según los centros de interés, tal como los dispusimos para la tarea de edición de los listados. Por dicha razón, reordenamos los datos según el centro de estudio al que correspondiera el informante, con la única salvedad de que agrupamos los centros educativos públicos y privados ubicados en la ciudad de Lleida, por considerar que diferenciar entre uno u otro no aportaba nada relevante a la investigación⁵²⁰. En último lugar, subdividimos la información para facilitar el volcado de los datos, con lo que los 34 archivos que manejamos para *Lexidisp* —17 en cada lengua según los centros de interés— se convirtieron en 40, correspondientes a la subdivisión en 8 centros educativos⁵²¹. Otro circunstancia importante que queremos advertir desde aquí a futuros investigadores es que *Dispoplex* no está preparado para editar otra lengua que no sea la española y, por tanto, no procesa correctamente caracteres ajenos a ella, como los acentos abiertos o el apóstrofe entre distintas palabras, formas frecuentísimas en catalán y, por ende, también en nuestro corpus. Como nuestra investigación es la primera que vuelca datos de otras lenguas peninsulares, no tuvimos conocimiento de estos problemas hasta después de haber introducido todos los datos. En relación con las tildes abiertas o los grafemas inexistentes en castellano —como la *ce trencada* (ç) o la *ele geminada* (l·l)—, el programa los admite pero no los ordena correctamente en la edición alfabética de los datos, lo que supuso que realizáramos a mano este trabajo. Más peliagudo es el caso del apóstrofe entre palabras (*v.gr. taulell d'anuncis*), toda vez que estas lexías se computan de forma adecuada en todos los índices cuantitativos pero son eliminadas drásticamente de los listados de palabras. La

⁵²⁰ Además, dimos nuestra palabra al director del colegio de los Maristas de Lleida de que nunca ofreceríamos los resultados discriminados de su colegio.

⁵²¹ A los 32 archivos resultantes de la subdivisión de los 8 centros educativos (8x2x2 lenguas), hay que sumar 8 archivos más que corresponden al centro 17 que tuvimos que volcar aparte en un nuevo proyecto, ya que la página web no admite más centros que los 16 clásicos en los proyectos panhispánicos, que son los únicos que permiten la opción de importar archivos. *Dispoplex* también ofrece la opción de introducir a mano los datos de cada informante.

solución que hallamos, al respecto, fue suprimir este carácter en todos los archivos y volver a introducirlo *a posteriori* en la edición final de los diccionarios.

Tras utilizar estas dos herramientas informáticas, la información se procesó mediante el paquete estadístico SPSS (*Statistical Package for the Social Sciences*, versión 21.0.0). Con este tercer análisis pretendimos conocer la aplicación inferencial de nuestros datos, es decir, si las diferencias observadas en la producción de palabras según las variables sociales y los distintos centros de interés, se darían, o no, si hubiéramos elegido otro informante al azar. Dicho de otro modo, con el SPSS comprobamos si los datos recogidos podían extrapolarse al conjunto de la población estudiada. Para poder procesar la información mediante este programa, creamos *ad hoc* una hoja de *Excel* con 25 columnas, donde dispusimos mediante dígitos los datos referentes a las cinco variables sociales contempladas junto con el total de palabras de cada informante por cada campo léxico⁵²². Finalmente, añadimos, además de una columna para numerar la muestra, otra con el promedio de palabras de cada informante y, en último lugar, otra donde se especificaba la lengua en la que se había completado la encuesta (catalán/castellano).

5.4. CRITERIOS DE EDICIÓN

La homogeneización de los materiales es una de las tareas que conlleva mayores dificultades en los estudios de disponibilidad léxica, sin duda alguna. Una vez recabadas las listas de palabras, estas se someten a un riguroso proceso de edición para el cual, como ya hemos indicado, utilizamos el programa *Lexidisp*. Como acabamos de ver, los informantes aportan una serie de palabras ante los estímulos verbales de la prueba, que son la base sobre la que se elaboran los diccionarios de disponibilidad léxica. En el caso anterior del informante 9 ante el centro de interés 5, lo que realmente escribió el informante, tal como se mantiene en la primera versión de la transcripción de las encuestas, fue:

12111 009 05 pollo, pescado, gambas, refresco, coca cola, Fanta, atún, queso, jamón, chorizo, sangría, vino, cava, verdura, fruta, naranja, manzana, melocotón, sandía, melón, lechuga, tomate, carne, zumo, plátano, pera, pepino

⁵²² Para poder conocer este dato construimos una tabla dinámica de Excel e importamos los archivos de *Lexidisp* distribuidos por centro de interés. La misma información puede obtenerse a través de *Dispolex*, aunque de una forma mucho más lenta.

Como se puede observar, hay algunas ligeras diferencias entre los dos listados. A este proceso de depuración —por llamarlo de algún modo— y estandarización de los resultados, dedicamos el presente apartado, en el que exponemos los criterios generales adoptados y los específicos en relación con los fenómenos de contacto hallados y, también exponemos con detalle las decisiones que hemos tomado en cada una de las distintas áreas léxicas.

5.4.1. Pautas generales⁵²³

Existen ya algunos acuerdos generales sobre los criterios que deben seguir las investigaciones que se enmarcan dentro del Proyecto Panhispánico de disponibilidad. A pesar de que la gran mayoría de estas pautas son ampliamente aceptadas por todos los investigadores, hay algunos puntos concretos en los que todavía existen discrepancias. Contar con unos criterios de edición comunes es imprescindible, como ya hemos señalado en diversas ocasiones, para poder realizar comparaciones dialectales entre distintas investigaciones. En realidad, en nuestra modesta opinión, y tal como otros autores ya han señalado anteriormente (Galoso 2003a: 32-33, Hernández Muñoz 2006⁵²⁴), el problema radica en la disparidad de objetivos a los que se pretende dar respuesta mediante estos estudios. Por un lado, se intenta confeccionar un diccionario de lengua y, por otro, se trata de realizar estudios de distintos ámbitos, principalmente de carácter sociolingüístico⁵²⁵.

El hecho de uniformar los materiales implica que no puedan realizarse algunos análisis interesantísimos —como el estudio de la ortografía o el del género gramatical—

⁵²³ A lo largo de este apartado transcribimos los ejemplos referidos al castellano entre comillas y los del catalán, en cursiva, independientemente del corpus (castellano o catalán) en el que hayan sido encontrados. Cuando una palabra esté en cursiva y entrecomillada significa que es homógrafa en las dos lenguas. Los extranjerismos, en general, se consignan en cursiva, salvo que ya formen parte del acervo léxico de cada lengua y se aluda a esta circunstancia. En tal caso, se transcriben entre comillas o en cursiva, según la lengua (castellano/catalán) a la que pertenezcan. Las marcas comerciales también se registran en cursiva, mientras que las formas híbridas entre las dos lenguas, se señalan con un asterisco al final del término.

⁵²⁴ En concreto, Hernández Muñoz (2006: 266-273) propone una serie de cambios en los criterios de edición con la finalidad de estudiar las implicaciones psicolingüísticas del léxico disponible. La autora no lleva a cabo dichas propuestas en su trabajo para poder comparar sus resultados con los del resto del Proyecto.

⁵²⁵ Véase en Hernández Muñoz (2006: 127-141) una interesante reflexión sobre el porqué de muchos de los problemas que se plantean al investigador en la edición de los materiales. Según la autora, la causa reside en la falta de coherencia del nivel categorial de las respuestas de los informantes.

basados en las listas “en bruto”⁵²⁶. De este modo, una opción que han seguido muchas investigaciones es la de realizar distintas versiones y de esta forma conservar siempre las respuestas tal cual las han dado los informantes. En nuestro caso, hemos optado por esta solución y, por ello, en un primer momento, hemos reproducido con total fidelidad el contenido de los cuestionarios. Más tarde, hemos editado los materiales —tal y como se encuentran en los capítulos 10 y 11—, procurando adecuarnos a las pautas de edición preestablecidas. Aunque no descartamos utilizar los datos de los archivos sin pulir en posteriores trabajos —para analizar, por ejemplo, las interferencias ortográficas o las variantes dialectales del catalán—, en esta ocasión nos basamos únicamente en los materiales ya convenientemente modificados.

También hay que señalar que, en nuestro caso, a las dificultades generales de editar los materiales se añade la de tener que homogeneizar no una, sino dos lenguas; por esta razón, tanto las decisiones que deben ser tomadas como los problemas que presentan los materiales y las justificaciones consiguientes que deben darse en este apartado, se han duplicado. A nuestro modo de ver, los principales inconvenientes que hemos debido solventar en esta fase de la investigación han sido, dejando de lado la complejidad y pesadez de la tarea en sí, la ausencia de pautas relativas a la edición del corpus catalán⁵²⁷ y la abundante presencia de elementos a caballo entre las dos lenguas tanto en las respuestas al cuestionario sobre el castellano como en las obtenidas en el dedicado al catalán. Si tenemos en cuenta, además, la gran similitud entre estos dos sistemas lingüísticos, emparentados diacrónicamente, podremos entender que decidir dónde estaba la frontera entre los dos, o, dicho de otro modo, qué palabras eran del castellano y cuáles del catalán, no ha sido una tarea tan fácil como pudiera parecer de antemano⁵²⁸.

Por otra parte, vale la pena destacar que la mayoría de las decisiones que detallamos en este apartado se ha basado, en primera instancia, en la fidelidad a la

⁵²⁶ En el léxico disponible de Galicia (López Meirama 2008) se utiliza un programa informático creado *ad hoc* por Guillermo Rojo que permite conservar las distintas variantes actualizadas por los informantes debajo de cada lema.

⁵²⁷ Durante la elaboración de esta investigación vio la luz la tesis de Llopis (2009), trabajo pionero realizado sobre disponibilidad léxica en lengua catalana, gracias al cual solventamos algunas de nuestras dudas. Sin embargo, en muchos otros casos no nos sirvió de ayuda, ya que este estudio se realiza sobre el valenciano, variedad del catalán que presenta notables diferencias en algunos niveles de la lengua.

⁵²⁸ Más si atendemos a que algunos informantes parecen desconocer por completo qué normas ortográficas y qué grafemas corresponden a cada lengua.

norma de uso de nuestra comunidad, y, en segundo lugar, en los criterios comunes ya adoptados por los miembros del Proyecto Panhispánico con el fin de que nuestros materiales puedan compararse, en el futuro, con los elaborados en otras sintopías. Otra pauta que hemos secundado de forma firme ha sido la de adoptar posturas paralelas en los dos corpus, puesto que nos parecía evidente que, si adoptábamos criterios dispares con respecto al castellano y al catalán, podía no ser efectiva la comparación cuantitativa de los resultados, en la que se basa nuestro análisis.

Los protocolos de edición que se adoptan habitualmente y que nosotros hemos seguido son los siguientes⁵²⁹:

a) Depuración de las listas. Hemos transcrito las respuestas de los informantes siguiendo fielmente el orden de las entradas consignado en las encuestas. Los casos en los que se ha repetido un mismo término han sido más bien escasos. La solución que hemos tomado en estas ocasiones, como en el resto de estudios, es la de eliminar la segunda alusión. Del mismo modo, hemos suprimido todos aquellos signos gráficos que los informantes habían escrito junto a los lexemas. Hemos hallado numerosos casos en los que las palabras se actualizaban entrecomilladas, señalando la posible obscenidad del término (“coño”) o que la palabra estaba escrita en un idioma distinto al de la encuesta (“*rebot*”). Hemos desechado también, en algunos casos aislados, los paréntesis o las barras con las que los informantes se referían a más de un término o precisaban el significado al que querían aludir. Tampoco hemos mantenido los pocos artículos hallados ante algunas lexías⁵³⁰. Cuando el informante daba dos ítems en una misma línea, hemos optado por colocar la segunda palabra debajo. Por último, hemos desarrollado las abreviaturas que hemos encontrado (*tlf.*> “teléfono”, *cat. express*> “cataluña exprés”).

b) Corrección y unificación ortográfica. Las distintas entradas se han transcrito en minúscula y de forma normativa. Las mayúsculas solo se han conservado en las siglas y en los nombres propios (ADN, TV, *arc de Sant Martí*, *trompa de*

⁵²⁹ Las pautas generales de edición pueden hallarse en Samper Padilla (1998) y en Samper, Bellón y Samper Hernández (2003). Los protocolos de edición han sido ampliamente debatidos en las reuniones de integrantes del proyecto realizadas hasta ahora, sobre todo en la realizada en San Millán de la Cogolla en el año 2003.

⁵³⁰ En cambio, los pioneros franceses solicitaban explícitamente un artículo delante de cada sustantivo (Gougenheim *et al.* 1964: 154).

Fal·lopi...). La corrección ortográfica nos ha exigido un gran esfuerzo puesto que, a pesar del grado de instrucción de los informantes, los errores constatados han sido innumerables, no solo en las palabras castellanas, sino también en las catalanas⁵³¹. Los equívocos más comunes han sido la confusión entre grafemas homófonos del tipo “x”/“ch”⁵³², “ny”/“ñ” y los aspectos referentes a la acentuación, sobre todo la relativa a los diptongos e hiatos, punto en el que las dos lenguas llegan a soluciones dispares y, en muchos casos, contrapuestas (por ejemplo, *ciència*/“ciencia”, *dial*/“día”, etc)⁵³³.

Con respecto a la unificación ortográfica la fuente que hemos utilizado para la corrección ha sido, con relación al castellano, la grafía aceptada por el DRAE⁵³⁴ y, para el catalán, hemos usado como criterio de autoridad la información que proporciona el DIEC⁵³⁵. En las ocasiones en las que los ítems léxicos no aparecían en estos diccionarios, hemos recurrido a los diccionarios de uso de cada lengua⁵³⁶. Hay que destacar que ha sido especialmente útil la posibilidad de realizar búsquedas a través del servicio TERMCAT, base terminológica electrónica especializada en la introducción de neologismos en la lengua catalana⁵³⁷. En cuanto a los dialectalismos de la lengua catalana presentes en los listados, nos han sido de gran utilidad los recuentos léxicos del catalán de la zona como el *Diccionari de Lleidatans* y *Lo nou diccionari català-*

⁵³¹ La gran cantidad de interferencias ortográficas halladas demuestra que el alumnado preuniversitario de Lleida no distingue con suficiente claridad los límites entre la norma ortográfica del español y la del catalán. Sería muy interesante abordar este aspecto en trabajos futuros a partir de los archivos conservados en bruto.

⁵³² En la variedad del catalán *nord-occidental*, los sonidos prepalatales sordos se pronuncian, generalmente, de forma africada, y no de modo fricativo como en otras variedades del catalán (Julià 2004: 41).

⁵³³ También es notable el número de lexías con acentos abiertos que se han detectado en las encuestas en castellano. En favor del alumnado, hay que recordar que en las instrucciones de la encuesta se dijo explícitamente que no se preocuparan por la ortografía.

⁵³⁴ Con esta siglas aludimos, a partir de ahora, a la vigésima segunda edición del *Diccionario General de la Lengua Española*, editado por la RAE, la de 2001, cuya referencia completa se halla en el apartado bibliográfico así como la de todos los diccionarios y bases de datos que mencionamos seguidamente. Cabe señalar que hemos utilizado la edición electrónica del diccionario, con lo que, cuando la información sobre la vigésima tercera edición ya estaba disponible, hemos utilizado esta última versión.

⁵³⁵ *Diccionari de l'Institut d'Estudis Catalans*, 2ª edición (2007). Véase la referencia completa en el capítulo bibliográfico.

⁵³⁶ Para el castellano hemos usado el *Diccionario del Español Actual* (DEA, de ahora en adelante), el *Diccionario de uso del Español* (DUE, a partir de ahora) y la base de datos CREA (*Corpus de Referencia del Español Actual*). Por lo que refiere a la lengua catalana, nos ha sido de gran utilidad el *Gran Diccionari de la Llengua Catalana de l'Enciclopèdia Catalana* (GDLC, a partir de este momento).

⁵³⁷ La *Generalitat de Catalunya*, el *Institut d'Estudis Catalans* y el *Consorci per a la Normalització Lingüística* son los creadores de esta base de datos (<http://www.termcat.net>) cuya finalidad es promover, elaborar y difundir recursos terminológicos, normalizar los neologismos catalanes y prestar servicio de asesoramiento.

lleidatà, cuyas características ya reseñamos en el capítulo anterior (*vid.* 4.3.2.)⁵³⁸. Con respecto a la edición del corpus catalán, nos ha sido de gran ayuda la obra de Julià (2004), donde se detallan las características de la pronunciación noroccidental del catalán, ampliamente plasmada en nuestros materiales en bruto.

Por otro lado, hay que decir que en ocasiones —aunque contadas— es posible escribir correctamente una misma palabra de varias formas. En estos casos nos hemos inclinado por la solución que el DRAE o el DIEC daban como preferente. Sin embargo, en los ejemplos en que la solución no preferida por la autoridad era la mayoritaria o la única presente en los listados, hemos optado por mantenerla considerándola reflejo de la norma de nuestra comunidad. También supone una excepción a esta pauta los ejemplos en los que el DRAE o el DIEC admiten dos variantes como igualmente válidas, en esta situación, hemos escogido de nuevo la opción más frecuente⁵³⁹. En cuanto a los numerosos dialectalismos de nuestra zona, que reflejan la variedad del catalán de Lleida, hemos optado por mantener la solución normativa, generalmente la más frecuente en los listados, a no ser que, como en las ocasiones anteriores, la forma dialectal fuera la única presente o la que predominara en nuestro corpus⁵⁴⁰.

Resulta especialmente complejo el caso de los extranjerismos, registrados en las encuestas con infinidad de grafías. En general, hemos seguido las pautas de Samper Padilla (1998) y adoptado la forma recogida en el DRAE o en el DPD⁵⁴¹ y en el TERMCAT⁵⁴², a pesar de que, en muchas ocasiones, no era la predominante en las encuestas. De este modo, en vez de *bikini* o *whisky* nos hemos decantado por incluir *biquini* o *güisqui*, que aparecen en pocas ocasiones en nuestros materiales en bruto. Sin

⁵³⁸ A partir de este momento usamos DLL para referirnos al *Diccionari de Lleidatanismes* de Massana (2004) y NDLLC para *Lo nou diccionari lleidatà-català* de Masip *et al.* (2010). Otro diccionario que hemos usado para localizar dialectalismos del catalán ha sido el *Diccionari Català-Valencià-Balear* (a partir de ahora DCVB) de Alcover y Moll.

⁵³⁹ Esto sucede, por ejemplo, con “vermut”, que el DRAE recoge como “vermú” o “vermut” sin establecer ninguna preferencia. Otros autores, como Gómez Molina y Gómez Devís (2004) deciden mantener estas variantes por la posible información que pudieran aportar al estudio de las lenguas en contacto.

⁵⁴⁰ Nos referimos a pares como *aubergínia/albergínia* o *pernil/pernill*. Hay que decir que, para elegir la forma predominante, consideramos cada centro de interés como un corpus independiente. Por esta razón, hay pares que se resuelven de un modo en un centro y de otro, en otro. Es el caso del par *ametlla/ametlla*, resuelto a favor del término normativo *ametlla* en el centro 5, por ser más frecuente que *ametlla*, que, no obstante, aparece en el 17 así, puesto que en ese centro ni tan siquiera aparece la otra alternativa.

⁵⁴¹ *Diccionario Panhispánico de Dudas*, publicado en 2005 por la RAE. Véase la referencia completa en el capítulo bibliográfico.

⁵⁴² Muchos de los extranjerismos que se han incorporado a la lengua catalana no aparecen todavía en el DIEC.

embargo, cuando ningún informante ha consignado el extranjerismo con la grafía aceptada por los censores de la lengua, para advertir de esta circunstancia —la cual puede estar indicando que no está tan adaptada en el uso— lo hemos respetado con la grafía original⁵⁴³, al igual que hemos transcrito según su lengua de origen las palabras foráneas que aún no están recogidas en los diccionarios. Los extranjerismos se consignan en cursiva —en la edición final de los datos— cuando así se recogen en el DRAE o cuando no aparecen en él, y, con respecto al catalán, cuando no aparecen en los diccionarios ni en las fuentes terminológicas normativas, es decir en el DIEC o en el TERMCAT⁵⁴⁴. Cabe destacar que, en general, hay un número importante de voces ya incluidas en las fuentes catalanas que, sin embargo, aún no se recogen en el DRAE.

c) Neutralización de las distintas variantes. Todas las formas recabadas se han unificado en la forma no marcada del paradigma, siguiendo los criterios comunes de edición —en este caso, de origen lexicográfico—. De este modo, los sustantivos y los adjetivos aparecen en masculino singular, y los verbos, en infinitivo. Constituyen excepciones a esta regla los sustantivos que tienen la entrada o la acepción correspondiente únicamente en plural (“bermudas”) o aquellas voces que presentan un plural lexicalizado equivalente al singular (“lavavajillas”), en especial cuando el plural es más usual que el singular, a tenor de las informaciones lexicográficas⁵⁴⁵. En estas ocasiones transcribimos la palabra en plural cuando todos los informantes la transcriban así y dejamos la forma singular cuando haya vacilación entre las dos opciones o únicamente se presente en singular⁵⁴⁶. Asimismo hemos dejado en plural aquellas lexías cuya singularización suponía una pérdida del significado original, como ocurre en “lazos”, *spirals* o *extensions*⁵⁴⁷. Como veremos más adelante (*vid.*5.4.2.), el plural también se mantiene en algunas lexías donde se constatan distintos fenómenos de contacto lingüístico.

⁵⁴³ La misma solución adoptan otros investigadores como Carcedo (2001), Arnal *et al.* (2004) o Hernández Muñoz (2006).

⁵⁴⁴ Incluimos también en cursiva las lexías marcadas con un asterisco en el GDLC, lo cual indica que la voz analizada es un préstamo.

⁵⁴⁵ Con respecto al catalán, este tipo de información la proporciona más a menudo el GDLC que el DIEC.

⁵⁴⁶ Un ejemplo de este tipo de casos es la palabra “gafas”, que transcribimos en plural por ser esta la única opción que han propuesto nuestros informantes. Seguimos así el mismo criterio que Hernández Muñoz (2006: 275-276) quien además utiliza la barra para mostrar la variación de género y número en la edición final de los datos. Bartol (2004) ya utiliza la barra y el paréntesis para la variación de género y número, respectivamente.

⁵⁴⁷ Las dos primeras aparecen en *Alimentos y bebidas* —y se refieren a un tipo de pasta—, y la tercera en *Ropa y complementos* aludiendo a un tipo de arreglo del cabello.

Con respecto a la variación de género, hemos mantenido los nombres cuya distinción de género conlleva una alteración del significado (“bolsa/a”). Cuando no existe tal oposición semántica, nos decantamos por la opción más frecuente de las dos. Si solo aparece la variante femenina, la conservamos, aunque no sea la normativa. En el centro de las profesiones y oficios (16), optamos por la forma masculina solo cuando tenemos respuestas en ambos géneros, y mantenemos la femenina cuando aparece como única. Lo hemos hecho así siguiendo las mismas pautas de los académicos quienes presentan en el DRAE algunos lemas de oficios solo en masculino o en femenino basándose precisamente, en la documentación existente, en concreto la que ofrecen sus propias bases de datos (el CORDE y el CREA)⁵⁴⁸.

En cuanto a los derivados, se han eliminado aquellos aumentativos o diminutivos que no implican lexicalización (“barquita” se une a “barca” o “pajarito” o “pájaro”) y se han mantenido cuando suponen una alteración del significado léxico y, por tanto, se recogen como entradas distintas en los diccionarios (“mesilla”, “tortilla”, “camiseta” o “botín”). Hemos actuado de idéntica forma en aquellos casos en los que su alto índice de aparición recomienda conservar los diminutivos puesto que puede ser que estén en vías de lexicalización o que alternen con el sufijo lexicalizado en el resto de territorios hispánicos (“cucharita/cucharilla”). Con respecto a los compuestos sintagmáticos en los que se da variación en el tipo de nexos que une los dos miembros, hemos optado, en nuestro intento por reflejar el uso lingüístico de nuestra comunidad, por la opción más frecuente; de este modo, las distintas variantes “goma de/del/para el pelo” se reducen a una sola “goma del pelo”, el sintagma más usual en el corpus español⁵⁴⁹.

⁵⁴⁸ En el léxico disponible de Huelva (Prado y Galloso 2005), las autoras optan por marcar, únicamente en este centro de interés, la vacilación de género con una barra. De un modo similar procede Ávila (2006) en su trabajo sobre Málaga donde, además de la barra, incorpora el uso del corchete para destacar las profesiones que solo aparecen en femenino en las encuestas y que los diccionarios documentan en masculino. De hecho, Galloso (2003a) es la primera en introducir el uso de la barra para mostrar la variación de género en este centro de interés.

⁵⁴⁹ El mismo criterio sigue Hernández Muñoz (2004) en su investigación sobre la provincia de Cuenca y Martínez Olmos (2008) sobre el léxico de los preuniversitarios alicantinos. Otros autores (*v.gr.* Gómez Devís y Gómez Molina 2004, Llopis 2005, Prado y Galloso 2005 o Hernández Muñoz 2006) optan por incluir el conjunto sintagmático más disponible.

d) Uso del paréntesis. Hemos utilizado el paréntesis en distintas ocasiones para indicar la alternancia de algún elemento. En primer lugar, para uniformar las formas plenas y los acortamientos del tipo “bici(cleta)”, “tele(visión)” o “(auto)bús”. En segundo lugar, para indicar que algunas lexías aparecen unas veces con todos sus componentes y otras con solo algunos: *palmell (de la mà)*, “columna (vertebral)”, “(horno) microondas”... También hemos usado el paréntesis para dar fe de la existencia de ciertas variantes del catalán como *(es)tovalles* o *(es)golfa*, que considerábamos que debían conservarse de algún modo por su alto índice de aparición. Por otra parte, hemos utilizado el corchete, en un uso novedoso, para indicar la ausencia de un elemento que ningún informante aporta pero que, sin embargo, nos parece necesario para comprender la lexía evocada y, por ende, su relación con el estímulo verbal. Esto sucede en escasas ocasiones como en “[policía] secreta” (centro 16) o “[producto] congelado” (centro 5).

e) Tratamiento de marcas comerciales. La inclusión o exclusión de marcas comerciales en los listados es uno de los aspectos más controvertidos. A pesar de que el criterio propuesto, en principio, era incluir solo aquellas que estuvieran lexicalizadas⁵⁵⁰; muchos autores se han decantado por admitir todas las marcas recabadas por la imposibilidad de decidir cuándo una marca podía considerarse lexicalizada o no⁵⁵¹. En nuestro trabajo, hemos tomado como pauta la inclusión solo de aquellas marcas que puedan considerarse en vías de lexicalización o completamente lexicalizadas —siguiendo así el criterio de otras investigaciones precedentes⁵⁵²—. Sin embargo, creemos que sería necesario en este punto alcanzar un acuerdo unánime que pudieran secundar todos los miembros del Proyecto, lo que facilitaría el cotejo futuro de los datos.

En nuestro caso hemos tomado como criterios para decidir la conservación o no de las marcas comerciales, la comparación del índice que alcanza la marca y su correspondiente genérico en las listas de disponibilidad y su aparición en los listados de

⁵⁵⁰ Tal como indica el fundamental artículo de Samper Padilla (1998: 317) donde el profesor canario sienta las bases de los criterios de edición comúnmente aceptados por los integrantes del Proyecto Panhispánico.

⁵⁵¹ Se sitúan en esta corriente de opinión los estudios sobre léxico disponible de Mateo (1998), Galloso (2003a), Gómez Molina y Gómez Devís (2004), Bartol (2004), Hernández (2004 y 2006), Prado y Galloso (2005), Ávila (2006) y Llopis (2009).

⁵⁵² Por ejemplo, Samper Padilla (1998), González Martínez (2002), López Morales (1999a), Carcedo (2001), Bellón (2011), Arnal *et al.* (2004), Ayora (2006), Martínez Olmos (2008), Fernández Smith *et al.* (2008), López Meirama (2008) y Samper Hernández (2009).

ambas lenguas. Nos parecía más que improbable que una marca comercial estuviera lexicalizada solo en una de las dos lenguas de la comunidad, tal y como hemos podido corroborar, en la mayor parte de ejemplos, al contrastar los datos obtenidos. En algunos casos dudosos también hemos recurrido a nuestra propia experiencia como miembros integrantes de la comunidad de habla estudiada. En el siguiente apartado (*vid.* 5.4.3), ofrecemos más información acerca de las decisiones tomadas en este punto en relación con los distintos centros de interés.

f) Asociaciones secundarias. Otro problema es el concerniente a las palabras que se relacionan solo indirectamente con cada centro de interés. Según hemos podido observar en nuestros corpus, suelen aparecer en los listados algunos términos que no tienen cabida en el enunciado estricto del centro. Algunos investigadores efectúan algunas restricciones y eliminan ciertos vocablos, mientras que otros opinan que deben incluirse todas las palabras actualizadas por el estudiante por la importante información que pudieran aportar de tipo sociolingüístico o psicolingüístico y porque —como afirman acertadamente—, la supresión de términos puede desvirtuar los resultados finales, que se someten en todos los casos a análisis cuantitativos, además de que dicha supresión parece contradecirse con la propia definición de “léxico disponible”⁵⁵³.

En nuestra investigación, la manipulación de los datos se ha limitado, en primer lugar, a la exclusión de términos incluidos con una evidente finalidad jocosa y carentes, en alguna ocasión aislada, de una mínima ética⁵⁵⁴ y, en segundo lugar, a algunas restricciones semánticas muy concretas de algunos centros de interés, como en *Partes de la casa (sin los muebles)*, donde ya se especifica en el enunciado que la información relativa al mobiliario no debe incluirse, puesto que se solicita en el siguiente estímulo⁵⁵⁵, o los nombres de lugares concretos de la capital del *Segrià* en *La ciudad*. Hemos tenido especial cuidado en excluir las mismas asociaciones en las dos lenguas, con la finalidad de no interferir en los resultados cuantitativos.

⁵⁵³ Esto último afirma Galloso (2002: 1145). Otros autores de la misma opinión son Mateo (1998), Hernández Muñoz (2004 y 2006), Bartol (2004), Prado y Galloso (2005) y Ávila (2006).

⁵⁵⁴ Nos referimos, por ejemplo, a algún nombre propio de profesor que aparecía en el centro *14. Animales* o a entradas del tipo “atropellar a ancianos”, mencionada en *Juegos y distracciones*.

⁵⁵⁵ Otros autores que nos preceden como Samper Padilla (1998), Carcedo (2001), Samper Hernández (2009), Bellón (2011), Arnal *et al.* (2004) o Llopis (2009) ya habían realizado restricciones semejantes en este centro de interés. Creemos, una vez más, que sería necesario acordar una pauta idéntica para todos los estudios de disponibilidad léxica.

g) Creaciones de los informantes. Finalmente, constatamos en nuestros corpus la existencia de algunas voces creadas por los mismos informantes para aludir a referentes cuyo nombre desconocen en castellano o en catalán (como “guardazapatos”, “reposacubiertos”, *biquinera*...). Hemos creído conveniente conservar todos esos términos, algunos de los cuales ya se documentan en otros léxicos disponibles, dado que podrían llegar a convertirse en innovaciones lingüísticas⁵⁵⁶. La frecuencia de aparición de cada palabra diferenciará la invención esporádica, por un lado, de un posible neologismo, por el otro⁵⁵⁷.

5.4.2. Pautas relativas a los fenómenos de contacto lingüístico

Con respecto a la edición de la amplia tipología de fenómenos de contacto lingüístico hallados, hemos de aclarar, para empezar, que todas las palabras escritas en catalán que aparecían en las encuestas en castellano, y a la inversa, se han incorporado a la edición final de los datos en cursiva, así como los extranjerismos que aparecen de este modo en el DRAE o no tienen cabida en los diccionarios normativos, tal y como explicamos en el punto anterior. La cursiva también se mantiene en los casos de sintagmas compuestos donde alguno de los miembros está en la otra lengua (*zapatilla esportiva*, *samarreta* corta) ante la imposibilidad de decidir a cuál de las dos lenguas pertenece la lexía en global. En relación con los términos agramaticales —híbridos entre las dos lenguas o adaptaciones lingüísticas en distintos grados—, estos se señalan con un asterisco para mostrar que no existen en ninguno de los dos idiomas estudiados. Los calcos sintácticos o semánticos, al entender que, en realidad, los términos que los componen son correctos —y que lo excepcional es, si acaso, la combinación sintáctica de los mismos o el significado que se añade— se incorporan a las encuestas sin ninguna marca específica.

Por otro lado, no es tan fácil como en principio podría parecer, detectar y decidir qué palabras del catalán o del castellano se incluyen en los diccionarios y bajo qué forma. En primer lugar, tenemos el problema de los homógrafos (“copa”/*copa*), en los

⁵⁵⁶ Por ejemplo, “pelapatatas”, ausente en el DRAE, también está en la nómina léxica del español de Asturias y de Valencia (Carcedo 2001, Gómez Molina y Gómez Devís 2004). Bartol (2003) analiza este tipo de creaciones y las compara en distintos léxicos disponibles.

⁵⁵⁷ Aprovechamos para explicar aquí que hemos suprimido paráfrasis del tipo “lo del aceite” o *el que es posa davall de l’olla* por entender, en nuestra modesta opinión, que no se acogen a la definición de “caudal léxico disponible en una situación determinada”.

que es imposible saber en qué lengua está pensando el estudiante⁵⁵⁸. En estos casos, hemos deducido siempre que la palabra se ha escrito en la lengua en que se ha pasado la encuesta. Más difícil es secundar este criterio cuando nos hallamos con términos en los que puede haber una interferencia ortográfica o léxica (“chimpancé”/ximpanzé). En estos casos, como la encuesta es escrita y no oral, no podemos recurrir a la pronunciación, mediante la cual deduciríamos la lengua en la que se actualiza el término. Teniendo en cuenta la escasez de casos hallados en esta situación y que, a pesar del nivel educativo, no estamos seguros de que la distinción ortográfica conlleve realmente una diferencia léxica, hemos llegado a la conclusión, de nuevo, de que el hablante incorpora la palabra en la lengua en que está completando la encuesta, y, por ende, la hemos mantenido o asimilado a la grafía de la lengua correspondiente. Por último, las interferencias fonéticas (del tipo sumo* por “zumo” o *salut* por “salud”) se han corregido. De un lado, se ha tenido en cuenta que no abordamos el estudio del nivel fonético en este trabajo; de otro, algunas de estas grafías simplemente retratan la pronunciación del castellano en boca de catalanohablantes (*cf.* 4.2.).

Además, siguiendo las pautas generales de neutralización de las distintas variantes, nos hemos encontrado con casos en los que, al singularizar el término en cuestión, se pierde la adscripción lingüística de la palabra; es decir, al pasar al singular la palabra, esta coincide con una forma de la otra lengua (*v.gr.* *olles* > *olla*/ “olla”). En estas ocasiones, consideramos que la baja frecuencia de este tipo de fenómenos hace irrelevante que se pierda esta información. Más complicadas han sido las decisiones relativas a los innumerables casos de híbridos detectados, tanto en los que concierne a la raíz (tiburó*) como en los que únicamente se presenta en la otra lengua el morfema flexivo (pomas*, pajarites*). En todos estos casos, nos debatíamos entre dos voluntades contrapuestas: el deseo de reflejar todas las variedades de fenómenos de contacto lingüístico presentes en los listados, y la homogeneización necesaria para poder editar el diccionario y hacer nuestros datos comparables con los de las demás investigaciones. Si bien es cierto que plasmar todas las soluciones sería deseable para un análisis exhaustivo sobre los trasvases entre las dos lenguas, también es cierto que mantener

⁵⁵⁸ Moreno Fernández (2007: 50) detalla el mismo problema en sus materiales sobre el español de estudiantes hispanos de Chicago. En su caso, como la encuesta es únicamente en castellano, opta por deducir a qué lengua pertenece la palabra según las palabras que preceden o siguen al vocablo en cuestión en la lista del estudiante. Otros investigadores de disponibilidad léxica en zonas bilingües, como Fernández Smith *et al.* (2012), López Meirama (2008: 42) o Taberner (2008: 554-555), también aportan numerosos ejemplos de la influencia de la segunda lengua en las grafías del castellano.

todas las variantes provocaría el descenso en el índice de disponibilidad de muchas palabras⁵⁵⁹. Tras todas estas dudas, hemos optado por asimilar estos casos a la lengua en la que está la raíz de la palabra (esto es “tiburón”, *poma*, “pajarita”, con respecto a los ejemplos anteriores). Sin embargo, cabe destacar que constituyen excepciones a esta norma los casos en los que no se halla la palabra correspondiente en la otra lengua en el corpus —por no haber sido mencionada por ningún informante— o cuando, existiendo las dos opciones, el híbrido alcanza un índice de disponibilidad mayor, lo que da cuenta de la importancia del fenómeno⁵⁶⁰. Por otro lado, se mantienen tal cual todos los préstamos asimilados, que muestran distintos grados de adaptación fonética a la otra lengua, del tipo calajera* (del cat. *calaixera*).

5.4.3. Pautas específicas

A continuación, exponemos las decisiones tomadas con respecto a los anteriores puntos en cada uno de los centros de interés.

5.4.3.1. *El cuerpo humano/El cos humà*

Tal y como ocurre en el resto de sintopías, el léxico recogido en torno a este centro de interés ha resultado ser bastante homogéneo, en el sentido de que no se incluyen apenas extranjerismos o dialectalismos ni ninguna marca comercial en los listados, únicamente hemos recabado dos extranjerismos —*piercing* y *sexy*—, dos términos dialectales del catalán —*manoquilla*⁵⁶¹ y *xut*⁵⁶²— y una sigla —ADN—.

⁵⁵⁹ Véanse las reflexiones de Carcedo (2001: 38-39) sobre la misma cuestión al editar sus materiales de la provincia de Asturias.

⁵⁶⁰ Por ejemplo, si *pomas** alcanzara un índice mayor de disponibilidad que *pomes* o fuera la única opción presente en los listados, porque no apareciera siquiera la opción de *pomes* en las encuestas en castellano, se mantendría tal cual en los diccionarios finales. Creemos que es importante mantener, en estos casos, esta información, que puede reflejar así una innovación léxica en nuestro ámbito de estudio. Seguimos en este punto el mismo criterio que Llopis (2009: 75-76) quien también asimila todos los derivados de este tipo a favor del término correspondiente de la raíz del lexema con las mismas excepciones ya mencionadas.

⁵⁶¹ Incluimos este dialectalismo que únicamente se halla recogido en NDLLC y que, por nuestra propia experiencia como miembros de la comunidad de habla, constatamos que se halla bastante extendido con el significado de “muñeca” (*canell*, en catalán estándar). También se incluía en alguna iniciativa lexicográfica menos seria como el *Dizionario Belverino*, donde se consignaban algunas palabras del habla de Belver de Cinca, un pueblo de la franja catalanoaragonesa. Lamentablemente, la página web en la que estaba el diccionario (www.belver.es/dizionario/index.html) ya no está disponible en la actualidad.

⁵⁶² Insertamos de este modo este dialectalismo, cuyo significado recoge también NDLLC bajo la tercera acepción de *txut*. Cabe señalar que, no obstante, en el DCVB y en el diccionario erótico de Vinyoles (1989) se incluye como *xot*.

En cuanto a las variantes de género y número, hemos efectuado, al igual que en otros recuentos de léxico disponible, una gran cantidad de singularizaciones de términos que designan realidades duales y que, por tanto, aparecían generalmente en plural (v.gr. “ojos”/*ulls*, “brazos”/*braços*, “fosas nasales”/*fosses nasals*). Un caso curioso es el adjetivo catalán *gluti*, que todos los informantes han transcrito en plural como si se tratara de un sustantivo (muy probablemente por la influencia de “glúteo”)⁵⁶³. Respecto a los derivados incluidos, hemos mantenido los pares *nina/niñeta* —puesto que ambos se recogen en el diccionario como “pupila”—, *falange/falangea* —con distintos significados— y “campanita”, donde se observa la influencia del catalán. En los compuestos sintagmáticos hemos mantenido la oposición correcta *intestí gros/prim* junto con la popular *intestí gran/petit*, para reflejar la vacilación de los informantes en este punto.

Hemos utilizado los paréntesis para indicar la alternancia de uno de los dos elementos de un sintagma en las siguientes ocasiones: “*columna (vertebral)*”⁵⁶⁴, “palma (de la mano)”, “tronco (del cuerpo)” y *palmell (de la mà)*. En cuanto al tipo de asociaciones que se han efectuado en torno a este primer estímulo, hemos de señalar que los estudiantes han actualizado términos, además de los relacionados directamente con el cuerpo humano, de alcance semántico más amplio —como “alma”, “belleza”, “medicina”, “tatuaje”, *pentinat*, *procreació* o *suavitat*— y algunos adjetivos referidos a las cualidades del cuerpo humano (“bronceado”, “escultural”, “vivo”, *alt*, *esvelt* o *moreno*). También se han incorporado algunas entradas propias del vocabulario científico médico (“duodeno”, “isquiotibial”, “mitocondria”, “ribosoma”, *diàstasi*)⁵⁶⁵, así como algún coloquialismo (“michelin”), y una buena muestra de vulgarismos que designan partes del cuerpo humano tabuizadas (“chocho”, “coño”, “huevo”, “*polla*”, “rabo”, “teta”, *cony*, *ou*, *popa*, *tita*, *titola*, *tranca* o *xut*)⁵⁶⁶.

⁵⁶³ Hay que señalar aquí también que hemos consignado en singular incluso formas como “pechos” cuya diferencia con respecto al singular pudiera conllevar una distinción semántica, por entender que —al menos en este caso—, el referente al que se alude es el mismo. En las ocasiones en las que la oposición semántica es clara, hemos mantenido la diferencia de número, tal y como veremos más adelante.

⁵⁶⁴ Tanto en castellano como en catalán, las dos formas son homógrafas y homónimas.

⁵⁶⁵ Lo que confirmaría un aspecto algunas veces comentado: la importancia de las clases que tengan lugar el mismo día, antes y después de la encuesta.

⁵⁶⁶ Llama la atención que algunos vulgarismos castellanos como “chocho” o “rabo” aparezcan en el corpus catalán, cosa que no sucede a la inversa, esto es, no aparecen vulgarismos catalanes en la encuesta en castellano. Algunos términos que designan las partes sexuales no se incluyen en los diccionarios normativos del catalán, con lo que hemos tenido que buscarlos en diccionarios terminológicos especializados o de argot (v.gr. *polla*, *tranca* o *xut*).

5.4.3.2. Ropa y complementos/*Roba i complements*

El segundo de los centros de interés aglutina un tipo de léxico muy rico y variado. Hemos encontrado e incluido solamente una palabra —“camellera”— que no recogen los diccionarios normativos ni de uso y con la que se alude a un bolso pequeño donde se guarda algún tipo de droga blanda. Asimismo, hemos mantenido el término catalán *xancla*, a pesar de que ningún informante lo transcribe así (optan por *xancla* o, mayoritariamente, usan el plural *xanclas*, que no es de gran ayuda). Sin embargo, por nuestra propia experiencia como miembros de la comunidad de habla, podemos afirmar categóricamente que la forma normativa terminada en *-e* no se utiliza⁵⁶⁷.

Las palabras que provienen de otras lenguas halladas en este campo asociativo no son pocas⁵⁶⁸. Entre ellas hemos podido observar, en algunas ocasiones, diferencias ortográficas en su proceso de adaptación en cada una de las dos lenguas estudiadas (v.gr. “body”/*bodi*, “parka”/*parca*). En otros casos, la solución es idéntica en ambas lenguas —“biquini”—, o solo está adaptada en una de ellas —“piercing”/*pírcing*—, o ningún diccionario normativo recoge el significado pertinente —“boxer”/*bòxers*—⁵⁶⁹.

Sobre los criterios de unificación de las distintas variantes, hemos de decir que mantenemos en plural —ya que así se documenta la respectiva entrada en los distintos diccionarios— “bermudas”, en castellano, y *abarqués, bermudes, calces, calcetes, calçotets, leotards, malles, texans* y *ulleres*, en catalán⁵⁷⁰; así como los compuestos “pasamontañas”/*passamuntanyes*, “pintalabios”/*pintallavis* y *pintaungles*. Cuando la entrada está en singular pero el diccionario advierte de un uso mayoritario o equivalente en plural, mantenemos la forma plural cuando esta es la única opción aportada por los informantes (v.gr. “gafas”, “calentadores”/*caletantadors, elàstics*) y, en cambio, la

⁵⁶⁷ En ello pueden influir distintos factores. Para empezar, generalmente se usa la forma plural *xanclas*, ya que alude a un referente dual, de lo que el hablante podría deducir que la forma singular es en *-a*, puesto que la solución “*-a/-es*” para el singular y el plural, respectivamente, es muy frecuente en catalán. Por otro lado, también podría deberse a la influencia de la forma homónima del castellano “chancla” o a la influencia del catalán oriental, sobre el que se ha construido la norma estándar catalana, donde la *-e* final se pronuncia de un modo más abierto que en el catalán occidental.

⁵⁶⁸ Esto resulta lógico ya que en la indumentaria se refleja la influencia de otras culturas y, como es sabido, algunos países tienen más importancia que otros en el mundo de la moda occidental.

⁵⁶⁹ A pesar de que los diccionarios normativos, el DRAE y el DIEC respectivamente, recogen este término, ninguno de ellos aporta el significado más habitual de prenda masculina de ropa interior. En el DEA o en el TERMCAT, en cambio, sí consta.

⁵⁷⁰ *Malles* se recoge con este significado solo en el GDLC; *abarqués* aparece tal cual en el DLL mientras que el DCVB, por su lado, recoge *albarca*.

dejamos en singular en caso contrario (v.gr. “braga”⁵⁷¹, “pantalón”/pantaló). De igual modo mantenemos en plural, a pesar de que no se recoja así en ningún diccionario, las palabras “cuadros”, “rayas” y *extensions*, puesto que consideramos que el significado que justifica su inclusión en este centro de interés se perdería de otro modo.

Con respecto a la distinción de género, hemos conservado los pares “bolso/a” y “gorro/a”. También hemos optado por mantener en femenino las entradas *menorquina*⁵⁷² y “náutica”/nàutica, ya que así es como las han consignado los informantes en las encuestas. En cuanto a las formas derivadas, hemos respetado los siguientes aumentativos y diminutivos ya lexicalizados: “batín”, “blusón”, “botín”, “camiseta”, “camisola”, “camisón”, “chancleta”, “chaquetón”, “corbatín”, “fajín”, “manguito”, “plumón” y, *barretina*, *botí*, *calcetes*, *collaret*, *corbatí*, y *xancleta*. El diminutivo “collarete”, que refleja el diminutivo catalán *-et/-eta*⁵⁷³ ha sido incluido porque no se daba la forma correlativa en castellano. Hemos admitido asimismo el diminutivo *mitgeta* por su alto índice de aparición, a pesar de que no se recoge en ningún diccionario. Por su parte, en los sintagmas compuestos —“cinta de pelo”, “goma del pelo”— hemos optado por la forma más frecuente entre las posibles variantes halladas (“cinta de/del/para el pelo”), como ya especificamos en las pautas generales.

La formulación parentética ha afectado a los siguientes sintagmas: “(forro)(polar)”⁵⁷⁴, *lliga(cama)*, “(pantalón) pirata”/(pantaló) pirata, (pantalons) texans y (sabatilla) esportiva/“(zapatilla) deportiva”⁵⁷⁵. Por otro lado, este es uno de los

⁵⁷¹ Este es un caso curioso puesto que la diferencia de número conlleva una diferencia de significado en determinados informantes, puesto que alude en su forma singular a una prenda exterior (tercera acepción del DRAE) y, en plural, a la ropa interior femenina (primera acepción del DRAE). Por esta razón, en un principio nos planteamos mantener esta diferencia numérica; sin embargo, al revisar las encuestas y las asociaciones anteriores y posteriores, nos percatamos de que el singular no siempre equivale a la tercera acepción. Así pues, decidimos finalmente unificar en singular todas las entradas pero no podemos olvidar que, seguramente, la palabra alcanzará un índice de disponibilidad alto, al referirse a dos realidades distintas.

⁵⁷² Con este nombre se alude a un tipo de calzado muy popular que proviene, como su nombre indica, de Menorca, y que también se suele denominar “mallorquinas”, lexía ausente en nuestros corpus.

⁵⁷³ En catalán sí existe como diminutivo lexicalizado *collaret*.

⁵⁷⁴ Hemos unificado igualmente de este modo el sintagma en catalán, ya que todos los informantes han transcrito esta forma y no la correcta *folre polar*. Incorporamos el paréntesis en ambos vocablos para mostrar que puede vacilar la presencia de los dos elementos del compuesto sin que cambie el referente al que se alude. Hemos llegado a esta conclusión tras comprobar las palabras que aparecían antes y después del primer miembro del sintagma.

⁵⁷⁵ En castellano hemos detectado, además, el caso mixto de “zapatilla esportiva”. En el corpus catalán mantenemos *esportiva*, porque la presencia de *sabata esportiva* y *sabatilla esportiva* nos impide saber a cuál de las dos se refieren los informantes.

estímulos que más propicia la aparición de marcas comerciales. De entre todas las recabadas, hemos decidido conservar las que se encuentran lexicalizadas en nuestra comunidad de habla, a tenor de la información lexicográfica hallada y del porcentaje de aparición en los listados de ambas lenguas⁵⁷⁶, esto es, *bamba*, *buff*, *chiruca*, *quet* y *wonderbra*. La primera de estas marcas, que alude a un tipo concreto de zapatilla de deporte, alcanza una posición alta en ambas listas y, además, está ya incluida en los diccionarios normativos de las dos lenguas (“*bamba*”/*vamba*). La segunda marca comercial equivale a la forma genérica catalana *tapaboques*, completamente ausente en los dos corpus, o a “braga (del cuello)” en castellano. La tercera marca lexicalizada, en cambio, solo está aceptada en catalán (como *xiruca*), pero no en castellano, aunque se halla en el DEA. Y, la última marca la incluimos —aunque no está presente en ningún diccionario— por no registrarse el correspondiente genérico “sujetador con relleno” o, más bien, “sujetador que realza el pecho”. Por lo que se refiere a la marca *quet* —proveniente de la marca comercial *keds*—, nuestros datos demuestran que se halla plenamente lexicalizada en nuestra provincia: tanto su índice de disponibilidad como de aparición superan ampliamente las denominaciones genéricas “zapatilla deportiva”/*sabata esportiva*, *sabatilla esportiva*. De hecho, esta marca, aunque no se incluye tampoco en ningún diccionario normativo del catalán, se recoge como *Ked(s)* en NDLLC y como *quet/quets* en el DLL, forma bajo la cual la incluimos en los dos corpus⁵⁷⁷.

Por lo que respecta a las asociaciones realizadas, hemos hallado además de todas las palabras relacionadas directamente con la indumentaria, todo tipo de complementos y accesorios (“chapela”, “móvil”, “turbante”...), así como algunas otras evocaciones un tanto más secundarias (“armario”, “estilo”, “perfume”, *purpurina*...).

⁵⁷⁶ Concretamente, las marcas que incluimos aparecen en los dos corpus y superan las cinco menciones en al menos uno de los dos listados, con la salvedad de “chiruca”/*xiruca*, que admitimos por registrarse ya en el DIEC. Un caso realmente sorprendente es el de “vamba”/*bamba*, que mencionan 58 informantes en catalán, y en cambio, solo uno en castellano. En el resto de ocasiones, como veremos, el porcentaje de aparición de las marcas comerciales alcanza posiciones muy igualadas en los dos diccionarios.

⁵⁷⁷ Entre las encuestas en castellano y en catalán, hemos sumado un total de once grafías distintas para la marca *keds*, siendo mayoritarias *quets* y *kets*. La variación de número también está presente (*quet/s*), lo que confirma la pérdida de conciencia de que se trata de una denominación comercial. Sobre esta marca, NDLLC dice literalmente que es una “paraula d’adopció lleidatana per excel·lència” y la califica como “passaport lleidatà exclusiu i intransferible” (Masip *et al.* 2010: 119), hecho que corroboramos por nuestra propia experiencia.

5.4.3.3. Partes de la casa (sin los muebles)/Parts de la casa (sense els mobles)

Con respecto a este centro, los extranjerismos recabados no han sido muchos (*cuina-office, hall, pista de pàdel, suite*) y, entre ellos, el más importante ha sido “parking” que cuenta con un alto índice de disponibilidad. También hemos hallado una sigla en este campo asociativo: WC.

Como venimos haciendo en los centros anteriores, hemos mantenido en plural las palabras catalanas *baixos* y *fonaments*, puesto que en el diccionario ya se advierte de su uso general en plural y porque esta es la única forma que aportan los informantes⁵⁷⁸. En cambio, en el caso de *(es)golfa* hemos conservado la forma en singular, si bien hay que señalar que la opción mayoritaria era el plural *(es)golfes*. También hemos mantenido en plural los núcleos de los complementos que formaban parte de sintagmas nominales del tipo “habitación de invitados”/*habitació dels convidats*. Hay que señalar que las unidades compuestas de este tipo son numerosas en este centro de interés. Los núcleos léxicos más utilizados han sido los mismos para las dos lenguas “cuarto”/*quarto/quartet*, “dormitorio”/*dormitori*, “habitación”/*habitació* y “sala”— y han dado lugar a una retahíla de formas analíticas: “cuarto de la caldera, de la lavadora, de la limpieza, de la plancha, de la ropa...”. En los casos en los que se utilizaban distintas preposiciones para un mismo sintagma hemos optado por la más frecuente (“habitación de/para invitados” > “habitación de invitados”).

En cuanto a los diminutivos, hemos respetado “barandilla”, “caseta/casita del perro”, “cuartillo” y “descansillo”. También hemos conservado el binomio “viga/vigueta” y *biga/bigueta*, que tienen sus respectivas entradas en los diccionarios, así como *saleta* y *quartet*⁵⁷⁹, cuya inclusión resultaba obligada por el alto índice de aparición en las encuestas. Del mismo modo hemos mantenido los aumentativos que ya están plenamente lexicalizados tanto en castellano (“salón”, “ventanal”), como en catalán (*balconada, finestró*).

Hemos utilizado el paréntesis para indicar la parte del sintagma que suele omitirse en los casos “salita (de estar)”/*saleta (d’estar)* y “(cuarto de) baño”/*(quarto de)*

⁵⁷⁸ Por el contrario, en castellano se registra en singular “bajo”.

⁵⁷⁹ Esta forma no aparece, con este significado, en los diccionarios. La forma de la que deriva, *quarto*, se incluye únicamente en el DLL y en el GDLC.

bany. En este último ejemplo, no asimilamos *bany* a *quart de bany* ni a *quartet de bany* puesto que estas opciones son mucho menos frecuentes⁵⁸⁰. También hemos usado el paréntesis para dejar constancia de las variantes dialectales de *golfa*, esto es *engolfa* y *esgolfa*, que hemos homogeneizado en *(es)golfa*⁵⁸¹, la opción más frecuente. Hemos añadido los corchetes a la lexía “taza”, que ha quedado registrada como “taza [del váter]”, y a *campana*, que se ha consignado como *campana [extractora]* en el corpus catalán. Nos parecía necesario especificar estos significados para justificar así su inclusión en el centro. A pesar de que ningún informante había aportado estas explicaciones, eran fácilmente deducibles a partir de las palabras anteriores y posteriores a su enumeración.

En cuanto a las asociaciones que ha suscitado este estímulo verbal, hay que señalar que numerosas lexías fueron rechazadas por no pertenecer al ámbito de este centro de interés, sino al del siguiente, tal y como explicamos en la introducción a este apartado. Por tanto, a pesar de la claridad del enunciado, hemos tenido que eliminar todas las palabras que aludían a muebles (“cama”, “sillón”, “librería”, “zapatero”...), a electrodomésticos (“congelador”, “DVD”, “horno”, “nevera”, “lavadora”...) y, en general, a una amplia tipología de objetos variados (“aceitero”, “cortina”, “figura”, “maleta”, “sábana”, “tumbona”...)⁵⁸². El criterio que hemos seguido ha sido el de incluir aquellas realidades de carácter estático⁵⁸³ o que se podrían hallar en una casa vacía (“chimenea”, “escalera”, “interruptor”), y todos aquellos elementos que se podían relacionar tanto con las partes de la vivienda (“cocina”, “comedor”, “vestíbulo”, *dormitori, garatge, rebedor*...) como con la construcción de la misma (“cemento”, “estructura”, “ladrillo”, *formigó, llosa, pintura, plànol*).

⁵⁸⁰ Concretamente, cada una había sido aportada por un solo informante frente a los 16 que incluyeron el sintagma *quarto de bany*. Extraña también que la forma *campra de bany* —la normativa en lengua catalana— se mencione en una única ocasión.

⁵⁸¹ Esta unificación también afecta al híbrido *(es)golfas**, del corpus castellano, que también reúne las tres opciones: *golfas/esgolfas/engolfas*.

⁵⁸² Creemos que la presencia de estas voces puede deberse a la dificultad —de algunos informantes— de mencionar las partes de la casa en lengua castellana, puesto que en las encuestas sobre el catalán no se han detectado tantas asociaciones secundarias.

⁵⁸³ Igual que hace Carcedo (2001: 41). Hay que señalar que, en la edición de algunos materiales, sí se han incluido bajo este epígrafe los electrodomésticos (*v.gr.* Bellón 2011 y Samper Hernández 2003).

5.4.3.4. Muebles de la casa/Mobles de la casa

Las tareas de unificación de este centro de interés nos han llevado a adoptar algunas decisiones contrarias a las propuestas por el DRAE. Estos han sido los casos de “fregaplatos” —que consignamos así puesto que la forma normativa diptongada no aparecía en ninguna encuesta—, y el de “bufet”, registrado bajo la forma no preferente del DRAE —que remite a “bufé”⁵⁸⁴—. De esta forma hemos pretendido reflejar cuáles son las preferencias de la norma lingüística de nuestra comunidad. Otro ejemplo es el de “sinfonier”, transcrito por nuestros informantes de este modo o como “simfonier”. Aunque el galicismo ya está admitido por el DRAE, la forma por la que ha optado el diccionario académico —“chifonier”— no parece coincidir con la que utilizan los hablantes, tal y como muestra además su inclusión como “sinfonier” en otros léxicos disponibles regionales. Algunos extranjerismos detectados en nuestras encuestas se encuentran ya plenamente asimilados a la lengua castellana (“póster”, “puf”, “váter”...) o catalana (*bufet*, *buró*, *escàner*...). Por el contrario, otros los hemos transcrito en cursiva, tal y como constan en el DRAE (*jacuzzi*), o para indicar que no aparecen en los diccionarios normativos de la lengua correspondiente (*webcam*). También hemos hallado una palabra del aranés, *estatgèra*, y un dialectalismo, *empostada*, que sí se recoge en los vocabularios propios de la región⁵⁸⁵.

En cuanto a la homogeneización de las variantes, hemos dado cabida a los pares “arca/arco” y “secador/secadora”, toda vez que se oponen semánticamente. También hemos encontrado, en este centro, algunas lexías compuestas que se caracterizan por que el segundo miembro se halla en plural: “guardazapatos”, “lavamanos”, “lavaplatos”, “lavavajillas”, “reposapiés”, *eixugaplats*, *guardasabates*, *penja-robres*, *rentamans* y *rentaplats*. Hemos respetado los derivados lexicalizados, tanto en catalán (*armariet*, *cadireta*, *tauleta/taulell/tauló*) como en castellano (“jarrón”, “mesita”, “mesilla”, “sillín”, “sillón”, “taquillón”...). Con las palabras “sillita” y “carrito”, hemos secundado un criterio especial ya que, a pesar de no registrarse en el DRAE, las incluimos puesto que designan realidades distintas a “silla” y “carro”. En el primer caso, resulta evidente puesto que el informante escribe el sintagma “sillita del bebé” y, en el segundo, el

⁵⁸⁴ En cambio, en el caso similar de “bidé”, al no incluir el DRAE *bidet*, hemos considerado que la palabra está escrita en catalán (cfr. 7.1.).

⁵⁸⁵ El DLL incluye *empostada* y NDLLC, *ampostada*. *Estatgèra*, por su parte, significa “estante”/*prestatge*.

encuestado parece referirse a “carrito de las bebidas”, aunque no se incluya esta especificación.

Al igual que ocurría en el centro anterior, este campo asociativo provoca la aparición de bastantes compuestos sintagmáticos, cuyos núcleos suelen ser *armari* / “armario” (corredero, de cocina, empotrado...), “mesa” / *taula* (*de cafè, de la cuina, de l’ordinador...*) y *moble* / “mueble” (de cocina, del recibidor, del televisor...). Respecto a las marcas comerciales, hemos prescindido de *playstation* y de *ikea*, aportadas solo en una ocasión, y, en cambio, hemos conservado *lot*, puesto que esta denominación se suele utilizar de forma habitual en el catalán de nuestra comunidad para designar cualquier linterna en general⁵⁸⁶. Las siglas que hemos recogido bajo este estímulo verbal son CD, DVD, TV y WC; a veces, se han presentado en sintagmas como en “estante de los CD”.

Se ha hecho uso del paréntesis para la unificación de “ropero” y “armario ropero” > “(armario) ropero”, “zapatero” y “armario zapatero” > “(armario) zapatero”⁵⁸⁷, “nido” y “cama nido” > “(cama) nido”, “bar” y “mueble bar” > “(mueble) bar” y “orejero”, “sillón orejero” > “(sillón) orejero” y “tele” y “televisión” > “tele(visión)”. En catalán este criterio solo ha sido posible en los casos de *capçal* y *capçal del llit* > *capçal (del llit)*, *bar* y *moble bar* > *(moble) bar* y *tele* y *televisió* > *tele(visió)*. En cuanto a los casos de “mesita” / “mesita de noche”, “mesilla” / “mesilla de noche”, *tauleta* / *tauleta de nit* no se ha efectuado la combinación, puesto que en todos ellos aparecían otros sintagmas que tenían por núcleo los respectivos diminutivos, de modo que resultaba difícil saber a qué tipo de mesa se referían. Con respecto a las variantes dialectales, hemos unificado mediante paréntesis las variantes *empostada/postada* > *(em)postada*, que aparecían en el corpus castellano. También hemos creído conveniente incorporar entre corchetes [*llit*] a la lexía *niu*, incluida así en el corpus catalán.

⁵⁸⁶ A pesar de que en este estímulo solo la consigna un informante, esta marca está totalmente lexicalizada, tal y como veremos en el centro 09. Solo el GDLC incorpora esta lexía y, además, nos informa sobre su origen. Al parecer, el fundador de la empresa que las fabricaba, “Ràdio-Lot” de Barcelona, se inspiró en el personaje bíblico de Lot a la hora de elegir el nombre, según la información que ofrece el GDLC.

⁵⁸⁷ En el corpus catalán no es posible esta unión dado que solo aparece “zapatero”.

También cabe destacar que hemos hallado algunas creaciones con las que los hablantes trataban de denominar realidades cuyo nombre correcto desconocían, como “guardazapatos”/*guardasabates* o la curiosa presencia de “salpicadero”, que hasta tres informantes de distintos puntos de la provincia aportan como equivalente a “fregadero”, tras realizar una interesante asociación por influencia, sin duda, de la palabra *pica* (“pila”) y del verbo “salpicar”.

Las asociaciones que han realizado los informantes en esta área léxica han sido mayoritariamente referidas al mobiliario (*calaixera, llit, prestatgeria, sofà...*); aunque tampoco son nada desdeñables las alusiones a electrodomésticos variados (“frigorífico”, “lavavajillas”, “microondas”, “nevera”...) o a objetos de toda índole (*ganivet, mirall, ordinador, piano, rellotge...*). Al igual que en otros léxicos disponibles, aparecen también los nombres de las partes de la casa, ya incluidos en el centro anterior, puesto que designan metonímicamente los muebles que se encuentran en ellas⁵⁸⁸.

5.4.3.5. Alimentos y bebidas/*Aliments i begudes*

Este es uno de los centros de interés más productivos donde, por tanto, los problemas que se plantean en la edición son más numerosos. Cuando el DRAE nos ofrecía dos posibilidades para un mismo término, hemos optado siempre por la que fuera más habitual; por esta razón incluimos “butifarra”, “mayonesa” y “vermut”, tras comprobar que “botifarra”, “mahonesa” y “vermú” eran mucho menos frecuentes. Lo mismo sucede en el corpus catalán con *soja*, que siempre aparece así y no bajo la forma preferida por el DIEC, *soia*. También hemos observado que, algunas veces, el DRAE no recoge el significado de algunos términos muy habituales en nuestra comunidad como “*napolitana*” o “chucho”⁵⁸⁹. En otras ocasiones, hemos constatado la ausencia en el diccionario académico de coloquialismos (como “birra”⁵⁹⁰, “cubalitra”) y de algunos tipos de bebidas alcohólicas (“orgasmo”, “pitufo”, “pomada” o “champú”), aunque en este caso, resulta lógico puesto que algunas de estas palabras son verdaderos localismos⁵⁹¹. En este centro de interés también hemos recogido un gran número de

⁵⁸⁸ De acuerdo con el criterio de Samper Padilla (1998: 319-329).

⁵⁸⁹ Las dos lexías aluden a distintas formas de bollería industrial; el DUE recoge este significado de “napolitana” en la segunda acepción del término. En otros casos semejantes, por el contrario, el DRAE sí especifica este significado (v.gr. “trenza” o “palmera”).

⁵⁹⁰ El DEA y el DUE, en castellano, y el DCVB, en catalán, por el contrario, sí la incluyen.

⁵⁹¹ Carcedo (2001) también incluye “pitufo” que, al igual que “orgasmo”, son nombres de distintos chupitos. “Champú”/*xampú* es un localismo muy habitual —no solo en nuestra comunidad (aparece en

dialectalismos del catalán (*albergínia/aubergínia*, *api/àpit*, *caneló/canaló*, *carbassa/carabassa*, *cargol/caragol*, *nou/anou*, *pernil/pernill*, etc.). Siguiendo el criterio que expusimos en las pautas generales, nos hemos decantado siempre por la opción normativa, excepto en los pocos casos en los que la forma dialectal era la única opción que aparecía (*llagosta/llangosta*, *llagostí/llangosti*⁵⁹²). A esta lista hay que añadir dos palabras más que aparecen en los diccionarios dialectales y que, en cambio, no se registran en los obras lexicográficas generales: *llomillo* y *safranòria*⁵⁹³.

Otro aspecto interesante, en este campo asociativo, es el tratamiento de los extranjerismos, que han sido más frecuentes que en otros centros debido a la influencia de otras culturas en la gastronomía de nuestra tierra⁵⁹⁴. La unificación ortográfica de algunos de ellos ha sido bastante laboriosa por la infinidad de grafías bajo las que los estudiantes incluían el término⁵⁹⁵. A pesar de que muchos de ellos ya se admiten en el DRAE (“bíter”, “champán”, “güisqui”, “roquefort”...) —aunque algunos se incorporen en cursiva (*brandy*, *gin-tonic*, *mousse*)⁵⁹⁶—, otros aún no aparecen (*frankfurt*, *tortellini*), aunque son ya de uso corriente en la comunidad hispanohablante⁵⁹⁷. En cuanto al catalán, hemos de señalar que la inmensa mayoría de extranjerismos se halla en el TERMCAT, completamente adaptados a la lengua catalana (*caipirinha*, *patxaran*, *tortel·lini*...). La única excepción son algunas bebidas combinadas —“güisqui *peach*” y “gin *lemon*”—. Resulta curioso que, a veces, los informantes aportan el extranjerismo bajo la forma en que este se acepta en castellano y no bajo la que es normativa en lengua catalana. Esto ocurre con *bacó*, *quetxup* y *sandvitx*, incluidas por los informantes como “beicon”, “kétchup” y “sándwich”. En estos casos, por fidelidad a los hablantes,

NDLLC como *txampú*) sino también en la provincia de Tarragona—, con el que se designa la bebida compuesta por cerveza y limonada. En cuanto a “pomada”, al parecer, es una bebida típica menorquina compuesta de ginebra y limonada.

⁵⁹² En el DLL aparece *llangosta* y en NDLLC, *llangost*. En esta secuencia y en la anterior, la primera palabra es la variante normativa y la segunda, la dialectal.

⁵⁹³ Se incluyen en el DLL, en NDLLC y en el DCVB.

⁵⁹⁴ A pesar de ello también hemos encontrado las comidas más típicas de la región como *pa amb tomata*, *escalivada*, *escudella*, *caragols* o *cassola de tros*.

⁵⁹⁵ Por poner solo un ejemplo, para los pares “cruasán”/croissant se han recabado 12 grafías distintas en castellano y 11 en catalán.

⁵⁹⁶ De esta forma se advierte de que su representación gráfica o su pronunciación son ajenas a las convenciones de nuestra lengua (DRAE 2001: XXXV).

⁵⁹⁷ El DEA, en cambio, sí registra “frankfurt”. En cuanto al italianismo *gnochi*, lo dejamos en su lengua original puesto que ningún informante lo transcribe como “ñoqui”, la forma que admite el DRAE.

hemos respetado la forma que ellos consignaban, por ende, la del castellano, aunque en algunas ocasiones las dos sean completamente homófonas⁵⁹⁸.

En cuanto a la unificación de las variantes de género, solo contamos con el ejemplo de “pimiento/a”, que se oponen semánticamente; o el caso de “*napolitana*” que incorporamos en femenino, tal como la incluyen todos los informantes. Llamativo es el caso de *fruit sec* que, a pesar de ser incorrecto, incorporamos en masculino por ser la única opción hallada, hecho que refleja, sin duda, el uso habitual de este compuesto sintagmático en el catalán de nuestra provincia. Respecto a la variación de número, hay que destacar que hemos mantenido los pares “pasta/pastas”, *pasta/pastes*, al entender que designan referentes distintos identificables según el singular o el plural⁵⁹⁹. También hemos mantenido en plural las formas que así se consignaban en los diccionarios (*espinacs, herbes*) o aquellas cuya singularización supondría, a nuestro modo de ver, perder el verdadero significado de la palabra como ocurre en *espirals* y *llaços*/“lazos”⁶⁰⁰. Cabe destacar el error extendidísimo del plural “especies”, que ningún informante ha incluido como “especias”⁶⁰¹. Sobre las formas derivadas halladas, hemos mantenido los siguientes diminutivos lexicalizados: “empanadilla”, “ensaladilla”, “manzanilla”, “palomita”, “pepinillo”, “pescadilla”, “pescadito”, “quesito”, “tortilla”, “tortita”...; y también, a pesar de no figurar en el DRAE, “montadito” y “rollito” (en el sintagma “rollito de primavera”). En el corpus catalán hemos conservado los derivados *orelleta, panadó, panellet* y *pastisset*, que cuentan con su propia entrada en el diccionario y se refieren, en su mayoría, a repostería típica de la zona.

Las marcas comerciales son, sin duda, otra de las características básicas de este centro de interés. Como ya dijimos al empezar este apartado, hemos dado cabida en los listados finales únicamente a aquellas marcas que pudieran considerarse lexicalizadas o en vías de lexicalización. Para tal efecto, hemos tenido en cuenta que dicha marca apareciera tanto en los listados del castellano como en los del catalán, que fuera más frecuente que su correspondiente nombre genérico y que un mínimo de cinco

⁵⁹⁸ Creemos que esto puede suceder por la influencia del castellano en los mercados de consumo, ya que el etiquetaje es mayoritario en esa lengua y no en la propia de la comunidad. Nótese que en todos los casos el informante prefiere la forma castellana que es también la más cercana a la lengua original, en estos ejemplos, el inglés.

⁵⁹⁹ Nos parece que este caso es distinto al de “pecho/pechos” donde se alude al mismo referente.

⁶⁰⁰ Los dos términos se refieren a distintos tipos de pasta que siempre se presentan en bolsas.

⁶⁰¹ Quizá se deba a la influencia del plural de la lengua catalana. De hecho, “especias” no se oye habitualmente en el castellano de los habitantes de Lleida.

informantes la mencionara en cada una de las dos lenguas⁶⁰². Se hallan en dicha situación las siguientes marcas comerciales: *aquarius*, *cacaolat*, *(coca) cola*, *cola cao*, *donut*, *fanta*, *J&B*, *martini*, *malibú*, *nestea*, *pepsi cola*, *red bull* y *vichy*⁶⁰³. Todos los términos anteriores superan con creces la frecuencia del producto genérico al que nombran⁶⁰⁴, e incluso, en algún caso como *donut*, no existe siquiera un nombre común para designar los referentes con exactitud.

Al listado anteriormente mencionado hemos añadido algunos casos más que, a pesar de tener escasas menciones, consideramos que debieran mantenerse. En esta situación están las marcas *nocilla* y *bimbo* (en “pan bimbo”/*pa bimbo*), mencionadas por tres y dos informantes en cada una de las lenguas respectivamente. En el primer caso, hemos decidido incluirla ya que el correspondiente nombre común “crema de cacao” no aparece. En cuanto al segundo ejemplo, hemos de reconocer que la lexía “pan bimbo” ha tenido solo dos menciones pero que únicamente un estudiante ha mencionado “pan de molde” en castellano y ninguno *pa de motlle* en catalán, con lo que las dos menciones podrían ser significativas.

Solo hemos hallado una sigla en este centro, TGV —que aparece tanto en castellano como en catalán—, con la que se alude, en el lenguaje juvenil, a la combinación de tres bebidas alcohólicas: tequila, ginebra y vodka. Con relación a la unificación parentética, se han visto afectadas las siguientes lexías del castellano: “ala (de pollo)”, “(carne de) caballo”, “(carne de) cerdo”, “(carne de) cordero”, “(carne de) pollo”, “(carne de) ternera”, “(carne de) tocino”, “(carne de) vaca”, “(coca) cola⁶⁰⁵” y “chuche(ría)”. En catalán, la unificación se ha usado en casos muy similares: *(carn de) bou*, *(carn de) corder*, *(carn de) pollastre*, *(carn de) porc*, *(carn de) vaca* y *(coca) cola*.

⁶⁰² La misma cantidad fija Bellón (2011: 97); en nuestro caso esto supone el 2% del total de la muestra. En realidad, decidir qué marcas están lexicalizadas o en vías, y cuáles no, es una tarea bastante arriesgada. Por esta razón, no podemos tener la completa seguridad de que, a pesar de cumplir todos los requisitos mencionados, una marca equivalga al producto para todos los miembros de nuestra comunidad, o que, al contrario, lo haga otra que no figura en nuestros listados.

⁶⁰³ Hay que señalar que contamos tanto el número de apariciones de la marca como las especificaciones (*fanta naranja/taronja*, *fanta limón/llimona*) o combinaciones de la misma con otras bebidas (*malibú* con piña). La hipótesis de la que partimos —que las marcas estarían lexicalizadas en las dos lenguas— se ve ampliamente corroborada puesto que su aparición es bastante similar en los dos corpus.

⁶⁰⁴ Solo tres informantes en castellano, y seis en catalán, mencionan “vermút” —frente a las numerosísimas menciones de *martini*— o más llamativo es aún el caso de *cacaolat*, ya que solo un informante en catalán aporta la lexía equivalente *batut de xocolata*.

⁶⁰⁵ Llegamos a esta conclusión por nuestra propia experiencia como integrantes de la comunidad de habla y tras comprobar que todos los informantes que incluyen “cola” lo hacen antes o después de alguna otra bebida, con lo que la posible alusión a otro referente parece más que improbable.

Por otro lado, hemos usado los corchetes en “pescadito [frito]” y en “[producto] congelado”.

Para finalizar, hemos de destacar que hemos hallado, junto con el amplio abanico de bebidas y alimentos, algunos nombres generales que designan tipos de alimentos (“alimento precocinado”, “comida basura”, “comida sana”...), bebidas (“bebida alcohólica”, “bebida isotónica”...) o comidas (“desayuno”, “merienda”, “régimen”...), así como el nombre de algún recipiente como “litrona” que pudiera designar metonímicamente su contenido.

5.4.3.6. *Objetos colocados en la mesa para la comida/Objectes col·locats a taula pel dinar*

Para empezar, hemos corregido, en este centro de interés, la palabra “panadera”, que nombraban muchos informantes por ultracorrección al pensar que “panera” no podía ser una palabra del castellano⁶⁰⁶. De otra parte, han sido muy escasos los extranjerismos hallados, únicamente *fondue*. Las variantes de género se han mantenido cuando había una distinción semántica como en *culler/a*⁶⁰⁷ o “cesto/a del pan” y su equivalente catalán *cistell/a*. Este último caso resulta curioso puesto que los dos nombres masculinos se refieren a objetos de mayor tamaño y que, raramente, pueden situarse encima de una mesa. Sin embargo, *cistella*, sinónimo de *cistell*, equivale en su segunda acepción a *panera* y, a nuestro modo de ver, de ahí surge la confusión y la consiguiente extensión a la forma masculina catalana y castellana, tal y como se plasma en los sintagmas *cistell del pa* y “cesto del pan”. En los casos en que no hay oposición semántica entre los dos géneros hemos optado por la forma más frecuente, siguiendo el criterio general. De este modo, hemos admitido “aceitera”, “frutero” y “vinagrera” en castellano, y *fruitera*, *oliera*, *setrill* y *sucrera*, en catalán.

En cuanto a la variación de número, hemos dado cabida a los compuestos recabados en este centro, cuyo segundo miembro se consigna siempre en plural, como sucede en los siguientes casos: “abrebotellas”, “abrelatas”, “posavasos”, “salvamanteles”, “sacacorchos”, *escuradents*, *llevataps*, *obreampolles*, *obrellaunes*,

⁶⁰⁶ El sufijo “-era” es mucho más frecuente en catalán que en castellano, además, también en catalán existe *panera* con el mismo significado.

⁶⁰⁷ Equivaldría a la diferencia en castellano entre “cucharón” y “cuchara”.

trecanous y *trespeus*. Cuando los diccionarios incluyen la forma en plural, así se ha incorporado en el corpus (“vinagreras”, *estalvis*, (*es*)*tovalles*, *postres*⁶⁰⁸ y *vinagreres*). De la misma forma, hemos incluido *pinces* en plural por ser esta la única opción aportada —y la habitual según el diccionario—, y “tijera”, en singular, puesto que la variación numérica estaba presente en nuestras encuestas, y, como ya dijimos, en caso de aparecer las dos opciones nos inclinamos siempre por la forma singular. Excepcionalmente, hemos incluido la forma en plural “toallas”, puesto que así la habían incluido todos los informantes, seguramente por influencia de *estovalles*⁶⁰⁹. Por otro lado, a pesar de que el DRAE solo admite el plural “aceiteras”, hemos respetado la forma “aceitera” ya que, además de ser aportada por más informantes, en nuestra opinión cada forma designa un objeto distinto, del mismo modo que se diferencian “vinagrera” y “vinagreras”. En la parte relativa al catalán, hemos conservado el plural *vinagreres* —a pesar de que no existe como tal en esa lengua— puesto que nos parece interesante su inclusión como reflejo de la influencia, una vez más, del castellano sobre el catalán.

En cuanto a los derivados nominales, hemos admitido los diminutivos lexicalizados “botellín”, “cucharilla”, “platillo”, *cullereta*, *tassó*⁶¹⁰... y otros, que, a pesar de no tener entrada propia en el diccionario, caracterizan las preferencias de nuestra comunidad de habla (v.gr. “cesteta”, “cucharita”, “pajita”, “platito”, *forquilleta*, *platet*). De otra parte, los aumentativos recabados están ya plenamente lexicalizados (“cucharón”, “jarrón”, “tazón”, *cullerot*...).

El criterio que hemos seguido respecto a los compuestos sintagmáticos (formados habitualmente por núcleos como “botella”, “cuchara”, “cuchillo”, “plato”, “tenedor”...) ha sido el de elegir la forma más frecuente entre los posibles nexos que

⁶⁰⁸ El “postre” castellano se denomina en catalán con un sustantivo femenino plural. Sin embargo, en algunos sintagmas de nuestro diccionario se incluye en singular o en plural (*cullereta del postre* vs. *plata de postres*) según el criterio mencionado de frecuencia. De esta forma, queremos reflejar el uso real por parte de nuestra comunidad, donde muchas veces se singulariza y masculiniza el término autóctono catalán, por influencia —sin lugar a dudas— del castellano. Téngase en cuenta, pues, que la forma correcta, por ejemplo para *plata de postres*, sería *plata de les postres*, completamente ausente en nuestros listados.

⁶⁰⁹ Así justificamos también la inclusión de este término en el centro, donde no tendría en principio relación alguna con el enunciado. Al parecer, nuestros informantes ignoran el significado castellano e incluyen la palabra como equivalente a *estovalles*, es decir, a “mantel”.

⁶¹⁰ A pesar de que el significado de este término es “taza pequeña” creemos con rotundidad que los hablantes de nuestra provincia lo utilizan como aumentativo, por influencia de “tazón”.

aportaban los informantes. Así, las distintas variantes como “cuchillo de pan/del pan/de cortar pan/para el pan” o *ganivet del pa/de pa/per tallar el pa* quedan reducidas a: “cuchillo del pan” y *ganivet del pa*. Hemos incorporado del modo que sigue las siguientes lexías complejas que aparecen algunas veces con todos sus elementos y otras con parte de ellos: “mando de la tele(visión)”, “papel (de cocina)”, “(paquete de) tabaco” y “(tetra)brik”—en castellano— y *anella (del tovalló), botella de (coca) cola, fusta per a (tallar) l'embotit*, “mando de la” *tele(visió)* y *(tetra)brik*. No hemos unificado “mando” y “mando de la televisión” porque hemos considerado la posibilidad de que se refiera a otra clase de “mando” (del aire acondicionado, de la cadena de música, del DVD, del toldo...). También hemos unido mediante el paréntesis las variantes dialectales *estovalles* y *tovalles*, aunque es de justicia señalar que la primera opción era la preferida por nuestros informantes⁶¹¹. Por otro lado, hemos utilizado los corchetes solo una vez en el sintagma [*estovalles*] *individuals* puesto que nos parecía que era necesario aclarar el sentido del adjetivo. La únicas marcas incluidas han sido *coca cola*, que también aparece en los sintagmas “botella de coca cola” y *copa de coca cola*, y *(tetra)brik*.

Por lo que respecta al tipo de asociaciones que despierta este estímulo verbal, estas no se han alejado, en general, de lo que se solicitaba en el enunciado. Además de algunos nombres genéricos del tipo “bebida”, “comida”, “postre”, “primer plato”..., se han evocado aquellos alimentos que acostumbran a estar presentes en todas las comidas, como el agua o el pan. Del mismo modo, como se viene haciendo tradicionalmente en los trabajos de léxico disponible, también se han mantenido los sustantivos que pueden designar metonímicamente el recipiente en el que se hallan (“aceite”, “azúcar”, “mayonesa”, “sal”, “salsa”...) ⁶¹². Sin embargo, las asociaciones más abundantes han sido las referidas a objetos de toda índole que pueden hallarse sobre la mesa, tal y como se especifica en el enunciado (*v.gr. cendrer, comandament, diari, mòbil, mocador...*).

Para finalizar, cabría comentar que hemos conservado también, tal y como explicamos en la introducción, algunas creaciones inventadas por los hablantes como

⁶¹¹ Concretamente, en la encuesta en catalán, un 45% incluyó *estovalles* frente a un 2.5% que prefirió la forma no prefijada. En la encuesta en castellano solo aparece la primera opción, que menciona el 5% de los informantes. La misma vacilación se ha observado en las formas híbridas *estovallas** y *tovallas**, asimiladas a la forma normativa catalana, puesto que eran menos frecuentes que esta.

⁶¹² Nos hemos basado en la comparación con otros léxicos disponibles para dilucidar qué alimentos debíamos incluir.

“guardaservilletas”, “reposacubiertos” y “reposaollas” porque cumplen perfectamente con las reglas de composición y remedan la estructura “verbo en tercera persona del plural + sustantivo en plural” en compuestos fijados en la lengua como son “lavavajillas” o “reposacabezas”.

5.4.3.7. *La cocina y sus utensilios/La cuina i els seus utensilis*

Dentro de las tareas de unificación de este centro de interés, hemos optado por mantener la forma no diptongada “fregaplatos”, por ser la mayoritaria entre nuestros informantes. También hemos detectado, en este campo semántico, algunos extranjerismos muy usuales como *grill* o *sandwichera*, y otros que nos advierten de la influencia de otras culturas en nuestra cocina (*raclette*, *wok*)⁶¹³. Por otro lado, y como ya ocurría en el centro anterior, hemos corregido los numerosos casos en los que se confundía *ganivet* con *gavinet*. Con respecto a los dialectalismos recogidos en este centro, hemos hallado *escarsadora* y *tenasses*, palabras que, al parecer, se limitan a la geografía de nuestra provincia⁶¹⁴, y una palabra en aranés, *padena* (“sartén”).

La variación de género ha presentado algunos problemas que hemos resuelto de la forma que detallamos a continuación. En principio, hemos mantenido las dos formas genéricas en los casos en los que pudiera existir una diferencia semántica entre los pares; esto ocurre en “caldero/a”, *culler/a*, *escorredor/a*, *fregador/a*, *gerro/a*, *pal/a*, *tap/a* y *ratllador/a*. En los ejemplos en que tal distinción no se daba, hemos optado, de nuevo, por incluir la opción más frecuente. Por esta razón hemos mantenido “aceitera”, “azucarero”, “batidora”, “cocinero”, “escurridor”, “escurridero⁶¹⁵”, “fregadero”, “frutero”, “picadora”, “pimentera”, “pelador”, “tostadora”, “tritadora”, y; *batedora*, *espremedora*, *fruiter*⁶¹⁶ y *tritadora*. En las escasas ocasiones en que la frecuencia era la misma para los dos términos, nos hemos decantado por el criterio normativo; esto ha

⁶¹³ Únicamente el primero de los extranjerismos recabados, *grill*, ya se admite en el DRAE. Todos los demás no aparecen, aunque sí se incluya “sándwich”, étimo de *sandwichera*. Hemos conservado *raclette*, que aparece en el corpus catalán, con su grafía original, puesto que así la escribe el único informante que la incluye. Sin embargo, este vocablo ya está adaptado al catalán bajo la forma *raclet* (TERMCAT). Mientras este instrumento se utiliza para confeccionar el plato culinario que lleva el mismo nombre y que es típico del cantón suizo de Valais, el *wok*, que también admite el TERMCAT, es una cazuela metálica originaria de China.

⁶¹⁴ El DLL y NDLLC incluyen *tenasses*; *escarsadora* aparece en el DCBV y, en NDLLC, como *ascarsadora*.

⁶¹⁵ Nótese que, en algunos casos como en este, la otra opción genérica ni siquiera tiene cabida en los diccionarios normativos.

⁶¹⁶ A pesar de que esta forma no existe en catalán, la incluimos dado que supera con creces el índice de aparición de la correcta *fruitera*, suponemos que por influencia directa de la forma castellana.

ocurrido con “aspirador/a”, *refrigerador/a* y *sucrer/a*, que han quedado reducidos a “aspirador”, *refrigerador* y *sucrera*.

En cuanto a la variación de número, hemos mantenido la oposición “vinagrera/vinagreras”, que ya aparecía en el campo anterior, y *tenalla/(es)tenalles*, puesto que, como en los casos anteriores, se refieren a objetos distintos, y, por otro lado, hemos respetado el plural en las formas catalanas *passapurés* y *rentavaixelles*, a pesar de que las formas normativas son las singulares (*passapuré* y *rentavaixel·la*), y, por tanto, se evidencia una influencia morfológica del castellano. También hemos decidido mantener en plural la lexía “hierbas”, ya que nos parecía que su singularización podía acarrear una distorsión del significado⁶¹⁷. Por último, y al igual que ocurría con el estímulo anterior, hemos detectado numerosas lexías compuestas: “abrebotellas”, “abrelatas”, “cortapatatas”, “lavaplatos”, “lavavajillas”, “pelapatatas”, “rompenueces”, “sacacorchos”, “salvamanteles”, y *eixugaplats*, *escuradents*, *llevataps*, *obrellaunes*, *rentaplats*, *setrilleres*, *trencanous* y *trespeus*.

Los derivados nominales que hemos admitido han sido aquellos que podíamos dar por lexicalizados, a saber, “cacillo”, “cucharilla”, “cucharón”, “molinete”, “paleta”, “platillo”, *cassoleta*, *cullereta*, *cullerot*... Además de estos casos, hemos incluido otros como “pajita” y *armariet*, que ya aparecían en centros anteriores. Hubiéramos podido unir “carrito” a “carro de la compra”, pero hemos pensado que podría tratarse de otra asociación del tipo “carro de las bebidas”. En los compuestos sintagmáticos hemos seguido, igual que en el centro anterior, la opción mayoritaria al hallar nexos distintos para un mismo sintagma: “cuchillo del/para el pan” > “cuchillo del pan”.

En cuanto a la formulación parentética, esta ha sido utilizada en los siguientes casos: “campana (extractora)”, “guante (de cocina)”, “(horno) microondas”, “madera para (cortar) el pan”, *extractor (de fum)*, *tele(visió)*, *tupper(ware)* y *vitro(cerámica)*. Como en otros campos léxicos, hemos usado también los paréntesis para las variantes *(es)tenalles* y *(es)tisores* del catalán. Por otro lado, las marcas comerciales no han sido muy abundantes en esta área. Hemos recabado únicamente *minipimer* —que aparece en el DEA—, *túrmix* —que ya se registra en el DRAE—, “papel albal”/*paper albal*

⁶¹⁷ En el DEA sí se reconoce que con esta acepción la palabra frecuentemente se pluraliza. Sin embargo, el DRAE no proporciona información al respecto.

—que compite con el genérico “papel de plata” en nuestra comunidad de habla⁶¹⁸—, *termomix*, que cuenta con el mismo número de apariciones que “robot”, y *tupperware*.

En torno a este centro de interés se han realizado todo tipo de asociaciones. Las más frecuentes han sido las relacionadas con utensilios (“abrebotellas”, “coctelera”, “mortero”...), electrodomésticos (“batidora”, “licuadora”, “nevera”...), instrumentos culinarios (“bayeta”, “pala”, “rodillo”...), mobiliario de cocina (“encimera”, “estante”, “mueble”...) o incluso algunos nombres de alimentos, concretamente muchos de los que ya aparecían en el centro anterior (“aceite”, “alimento”, “pan”...). También hemos hallado numerosas creaciones léxicas, aunque algunas de ellas ya aparecen en otros léxicos disponibles (“cortapatatas”, “pelapatatas”) o en los diccionarios de uso (“secaplatos”), otras, sin embargo, son simples derivaciones creativas (como *biquinera* por *sandvitchera*).

5.4.3.8. La escuela: muebles y materiales/L'escola: mobles i materials

El léxico recabado bajo este estímulo verbal ha sido bastante homogéneo y estándar. Cabe destacar, no obstante, la sorprendente ausencia en los diccionarios de términos tan comunes en nuestra comunidad como “carpesano”/*carpessà* o “subrayador/subratllador”⁶¹⁹. Es destacable también la confusión entre los adjetivos “florescente”/*florescent* y “fluorescente”/*fluorescent*, que hemos corregido en ambas lenguas⁶²⁰. Otra confusión frecuentísima, en ambas lenguas, ha sido la de *retolador* / “rotulador”, motivada por la pronunciación del catalán en Lleida donde es frecuente la asimilación de la /e/ en /o/⁶²¹. Por dicha razón, hemos corregido estas variantes adoptando la solución normativa apropiada en cada idioma⁶²². De otra parte,

⁶¹⁸ Los informantes que los han nombrado no son tantos como para saber cuál de los dos es más habitual. Solomencionan la marca dos informantes, uno en cada lengua de la encuesta, y el correspondiente genérico “papel de plata”, uno en castellano y dos en catalán.

⁶¹⁹ NDLLC sí incluye *carpessà* como “carpeta separadora per anar a estudi”.

⁶²⁰ En catalán, concretamente, el porcentaje del alumnado que ha escrito mal el término es superior al que se ha decantado por la opción correcta. Por otro lado, consideramos más que probable la confusión entre los referentes a los que se alude con los términos “fluorescente”/*fluorescent* y “fosforescente”/*fosforescent*, que explicaría el alto índice de disponibilidad que alcanza el primer par de lexías.

⁶²¹ Julià (2004: 33) cita este mismo ejemplo en su libro sobre la descripción del catalán noroccidental con la vacilación, además, interna de la o/u (*roto/ulador*).

⁶²² Es decir, en castellano “rotulador” y en catalán *retolador*. Mientras que, en castellano, una aplastante mayoría ha elegido la opción correcta, en catalán, en cambio, los dos términos presentaban porcentajes de aparición más ajustados. Sin duda, sería interesante para el catalán de la zona conocer el peso concreto de las dos variantes. De otra parte, es posible que, además de la variante fonética dialectal influya también en la dicción la forma homófona del castellano.

solo han tenido cabida, bajo este epígrafe, algunos extranjerismos como *disquete*, *casete*⁶²³, *dossier*, *fax*, *guixeta*⁶²⁴, *minitramp*, *webcam* o *zoom*. Se incluye en nuestros corpus *estatgèra*, palabra del aranés, y algunos dialectalismos, como *(em)postada* y *tenasses* que aparecen, sobre todo, en el corpus catalán y a los que ya nos hemos referido en centros anteriores.

En relación con la variación de género, hemos mantenido los pares “bolsa/a”, *anell/a* y *plom/a*, donde se alude a distintos referentes⁶²⁵. En cambio, hemos conservado en femenino términos que solo aparecen así en el corpus (v.gr. *nena* o “calculadora”⁶²⁶). Respetamos en plural algunos términos, como “apuntes”, “portafolios”, “portaminas”, “portarrollos”, *apunts*, *deures*, *escombraries*, *es(tisores)*, *penja-robes*, *pinces* o *portamines*, cuya entrada en los diccionarios ofrece también esta información. E igualmente mantenemos esta variante cuando esta ha sido la única opción registrada en nuestros listados en términos como “gafas” o “tenazas” que admiten las dos posibilidades. Hemos hallado, además, muchos derivados ya lexicalizados y con su propia entrada en los recuentos lexicográficos como “alfombrilla”, “barandilla”, “bombilla”, “cuadernillo”, *armariet*, *bombeta*, *escaleta* o *gomet*. La única excepción ha sido el aumentativo *calaixó*, cuyo referente concreto ignoramos, que hemos incluido en el diccionario porque su índice de aparición parece indicar que no se trata de un fenómeno individual.

La unificación mediante paréntesis ha afectado a las siguientes palabras y combinaciones: “bolí(grafo)"/*bolí(graf)*, “goma (de borrar)"/*goma (d'esborrar)*, *maquineta (de fer punta)* —cuya especificación solo se recoge en el corpus catalán— “pega(mento)”⁶²⁷, “(rotulador) permanente”/(*retolador*) *permanent*,

⁶²³ Hemos incluido en catalán también la forma *casete*, en lugar de la normativa *casset*, debido a que, como ya ocurría en el centro número 5, ningún informante optaba por la forma adecuada. En cambio, en el caso similar de *disquet*, los informantes sí incluyen así la forma, es decir, tal como aparece en el DIEC.

⁶²⁴ Este galicismo, ausente en los diccionarios normativos catalanes, aparece en el corpus catalán como sinónimo de *taquilla*.

⁶²⁵ En el corpus catalán, es considerable el índice de error entre *fulla* y *full*, bajo la influencia, sin duda, del término castellano “hoja”. Aunque en un principio pensamos incluir tal distinción, por si la variante femenina pudiera referirse a un objeto del aula de tecnología, finalmente hemos asimilamos las dos opciones a la solución masculina, normativa y mayoritaria, porque los sintagmas hallados (*fulla quadriculada*, *fulla blanca*, *fulla reciclada*) nos confirmaban el significado equivalente a *full* del término.

⁶²⁶ Es extraño que en el avance de la vigésima tercera edición, el DRAE no señale en femenino esta entrada, tal y como aparece en la cuarta acepción de la edición anterior.

⁶²⁷ Por el contrario, no unificamos en catalán *pega* y *pegament**, por tratarse esta última de una lexía inexistente, interferencia clara del castellano, cuya presencia en nuestro corpus analizaremos en el

“profe(sor)”/profe(ssor), “tele(visión)”/tele(visió) y *taulell (d’anuncis)*⁶²⁸. También en los dialectalismos *(es)tisores* y *(em)postada* hemos usado la misma solución⁶²⁹. Con respecto a las marcas comerciales, hemos conservado las ya lexicalizadas (como *tippex*) o en vías (*dacs*, *pilot*, *plastidecor* y *rotring*). Como en los casos anteriores, el número de informantes que las mencionan y el hecho de que aparezcan en las dos lenguas sugerían su inclusión⁶³⁰. No han sido abundantes, de otro lado, las siglas recabadas: CD (y CD-ROM), DIN A3, DVD, TV y WC. Con respecto a las asociaciones que ha despertado este centro de interés cabe destacar el predominio, en primer lugar, de los objetos presentes en cualquier aula escolar (“diccionario”, “libro”, “mesa”, “mochila”...), seguidos de las personas que trabajan en un centro educativo (“director”, “maestro”, “profesor”...) y de las distintas partes del mismo (“aula”, “biblioteca”, “laboratorio”...).

5.4.3.9. Iluminación, calefacción y medios para airear un recinto/Il·luminació, calefacció i mitjans per airejar un recinte

Este centro de interés no plantea demasiados problemas de edición. Al igual que en el centro anterior, también aquí hemos hallado la confusión entre “fluorescente”/“florescente” y *fluorescent/florescent*. Los extranjerismos han sido escasísimos, únicamente recogemos *flash* en las encuestas en castellano. También vale la pena constatar que la aparición de “pila”, en ambas lenguas, se justifica por su equivalencia a “linterna”, hecho que corroboramos por nuestra propia experiencia como miembros de la comunidad de habla estudiada, a pesar de que este significado añadido no se recoge en ningún diccionario del catalán de la zona, ni siquiera en los de ámbito dialectal.

Con respecto a las distintas variantes, hemos mantenido los pares *calder-caldera*⁶³¹ y consignado en plural el segundo miembro de los sintagmas “abrir puertas”,

capítulo correspondiente (cfr. 7.4.). A pesar de mantener las dos variantes, por nuestra propia experiencia como miembros de la comunidad de habla, constatamos que son completamente equivalentes.

⁶²⁸ La única opción apropiada y normativa, *tauler (d’anuncis)*, la aportaba un solo informante con lo que la hemos asimilado a la más frecuente *taulell* (“mostrador” en castellano). La confusión sobre este referente alcanza también otros sintagmas como *tauló d’anuncis* o el híbrido *tabló* d’anuncis*.

⁶²⁹ En estos dos casos, la opciones estándares *postada* y *tisores* eran mayoritarias.

⁶³⁰ También aparecen en este centro algunas marcas ya plenamente lexicalizadas e incluidas así en los diccionarios como “plastilina” o “celo”. Con respecto a esta última el TERM CAT recoge su denominación comercial “Cello” y propone el término *cinta adhesiva*, en su lugar.

⁶³¹ Además, en el corpus catalán, hemos corregido la asignación de género incorrecta que se daba a *corrent*, dando lugar a los sintagmas *corrent alterna*, *corrent elèctrica*; por influencia sin duda, del castellano. Cabe decir que el índice de aparición de cada palabra es el que nos ayuda a dilucidar entre los

“abrir ventanas”, “correr cortinas”/*obrir finestres, obrir portes*, toda vez que creíamos que así se respetaba el referente al que aludía el informante. También hemos conservado el plural, tal como reconocen los diccionarios de cada lengua, en “plomos”, “portalámparas”⁶³² y *portabombetes*. En relación con los aumentativos y diminutivos incluimos los siguientes, la mayor parte de los cuales ya están lexicalizados: “bombilla”, “cajetín”, “lamparilla”, “lamparita”, “rejilla”, “ventanal”, “ventanilla”, *bombeta*, *lampareta** y *reixeta*.

Hemos utilizado el paréntesis en las siguientes lexías compuestas: “(aparato de) aire acondicionado”/*(aparell d’)aire condicionat/(aparell d’)aire condicionat**, “extractor (de humo)”/*extractor (de fums)*, “(gas) butano”/*(gas) butà* y *(raig) làser*. Las marcas comerciales enumeradas no han sido muchas. Según su porcentaje de aparición, hemos mantenido *tropicano* como posible marca en vías de lexicalización, y por otro lado, *lot*, nombre comercial que equivale a “linterna”, significado que ya se incluye en el GDLC⁶³³. De otra parte, hemos recogido la sigla PVC (en *tub de PVC*), en alusión a un tipo de tubo que se hace con Cloruro de Polivinilo, y la unidad de potencia *Kw*, que insertamos en el corpus tal como la aporta el único informante que la cita.

En general, se han actualizado, en torno a esta área temática, un gran abanico de medios de calefacción (*carbó, electricitat, gas natural, llenya...*) con diferentes combustibles, la mayor parte de ellos derivados del petróleo como la gasolina, el gasóleo, el gasoil y diferentes gases (“butano”, “propano” o “queroseno”). Dentro de los medios de iluminación y de las formas de airear un recinto, se han enumerado tanto procedimientos naturales (“abrir puertas”, “aire”, “calor humano”, “sol”) como artificiales de todo tipo (“bombilla”, “climatizador”, “lámpara”...). Nos hemos visto obligados a descartar algunas asociaciones que se alejaban de los parámetros del centro de interés o eran de tipo jocoso como, por ejemplo, “colonia”, “desodorante”, “ ducharse”, “foco de Batman”, “vodka”; no obstante, hemos procurado que estas fueran las mínimas posibles⁶³⁴.

posibles errores, como este, de las verdaderas tendencias en las lenguas de nuestra provincia. *Cfr.*, por ejemplo, la opción contraria que tomamos al mantener *fruit sec*, en el centro 5.

⁶³² En este caso el DRAE admite las dos opciones y todos nuestros informantes la transcribieron con el segundo miembro del compuesto en plural.

⁶³³ También aparecía en el centro 3, recuérdese lo dicho en la nota 586.

⁶³⁴ Cabe señalar que la inclusión de algunos vocablos se justifica por ser asociaciones secundarias que realizan los informantes tras una retahíla de términos (*v.gr.* “cable”, “pegamento”, *filferro*). Por otro lado,

5.4.3.10. La ciudad/La ciutat

Este centro de interés ha sido muy rico y variado en el tipo de léxico actualizado. Los extranjerismos han sido abundantes; se transcriben —siguiendo los criterios generales— tal como marcan los respectivos diccionarios (“campus”, “club”, “stop”, “tráiler”, “vodka”, “yonqui”, *bulevard, càmping, campus, cibercafé...*) o en su lengua original (*container, parking, pizza, pub, school, scooter, sex-shop, skatepark*)⁶³⁵. Cabe destacar que nos hemos decantado por “chalet” y “metrópolis”, frente a “chalé”⁶³⁶ y “metrópoli”, claramente minoritarias. Asimismo, hemos detectado algunos términos propios del argot urbano como “chulo”⁶³⁷ o “puticlub”, relacionados con el mundo de la prostitución, y “munipa” o *condoneria*, palabra que ya incluye el TERMCAT. Hemos hallado dos dialectalismos, *cemintiri*⁶³⁸ y *colègit*, reflejo de la dicción del catalán de Lleida, que hemos asimilado a las formas normativas *cementiri* y *col·legi*, más frecuentes en el corpus.

Con respecto al género, hemos unificado, como ya viene siendo habitual, en masculino singular la mayoría de lexías, a excepción de algunos adjetivos (como “compleja”, “intranquila”, “sosa”, *petita, trista...*) que se actualizaban únicamente en femenino por referirse a “la ciudad”. El plural es de uso obligado cuando el sustantivo se encuentra en un sintagma preposicional que actúa como complemento del nombre, por ejemplo en “máquina de bebidas”, “parque de atracciones/de bomberos”, “tienda de alimentos/de electrodomésticos”, *centre de comunicacions/de jubilats, llar de jubilats/d’infants, sala d’actes/de jocs/de màquines/d’exposicions...* También aparece en algunas expresiones como “más oportunidades”; y, por supuesto, hemos mantenido en plural aquellos términos cuyo significado ya se documenta así en los diccionarios (*v.gr.* “afueras”, “alrededores”, “correos”, “grandes almacenes”, “grandes superficies”,

“pingüino” se admite, al igual que Arnal *et al.* (2004: 35), ya que es una forma de denominar un tipo de aparato refrigerador.

⁶³⁵ El DRAE incluye en cursiva “pizza”, “scooter” y “stop”. El DEA, por su parte, ya registra “container”, “parking” y “sex-shop”. El TERMCAT incluye *pub*, marcado con asterisco en el GDLC, y *sex-shop*. En el corpus catalán ningún informante escribe la adaptación homófona correcta “escúter”, por esta razón, lo incorporamos tal cual está en su lengua original.

⁶³⁶ En la preferencia por esta adaptación del galicismo influye, sin duda, la solución homófona catalana *xalet*; además, la terminación del plural *-ts* es habitual en la fonología catalana, y por extensión, en el castellano de la zona (*vid.* 4.2.)

⁶³⁷ Aunque podría tener también otros significados, hemos llegado a esta conclusión tras observar las palabras que aportaba el informante antes y después del término en cuestión.

⁶³⁸ Julià (2004: 32) cita precisamente este caso como ejemplo de cierre de e> i, seguida de [i] tónica o algún sonido palatal, en el catalán noroccidental.

“nervios”, *afores*, *correus*, *diners*, *estudis*, *grans magatzems* y *nervis*⁶³⁹). Dejamos constancia aquí, que —sin duda por influencia de “rascacielos”— el compuesto *gratacel* aparecía más frecuentemente en plural, aunque en la edición final de los datos hemos incluido la opción normativa en singular, que también aparecía.

Los derivados que hemos mantenido han sido “callejón”, “callejuela” —y el híbrido *carrejuela**— “carrito”, “cohecito” y “placeta”, donde puede verse la influencia morfológica del catalán. En cambio, hemos desechado “cohecillo” y “chavalita”, voces incluidas en “coche” y “chaval”, respectivamente, por no referirse a realidades distintas a las que tiene el sustantivo del que derivan. En el corpus catalán, hemos conservado *carreró*, *carretó*, *caseta*, *cotxet* y *paradeta*⁶⁴⁰.

Mediante el paréntesis hemos unido las variantes con apócope como “(auto)bús”, “estación de (auto)buses”/*estació d’(auto)busos/parada d’(auto)bús*, “bici(cleta)”, *cine(ma)*⁶⁴¹, “disco(teca)”, “insti(tut)”, “zoo(lógico)”/*zoo(lògic)*. Sin embargo, al igual que han hecho otros autores con anterioridad (Carcedo 2001, Bartol 2004 o Hernández Muñoz 2006), mantenemos separadamente las lexías “moto” y “motocicleta”, por considerar que tienen dos referentes claramente distintos según la cilindrada del vehículo. También hemos usado el paréntesis en los siguientes grupos sintagmáticos: “bloque (de pisos)”⁶⁴², *caixa (d’estalvis)*, *forn (de pa)* y *mosso (d’esquadra)*. Con respecto a las siglas, hemos hallado AVE, CAP, CD, ONCE y el acrónimo “Renfe”.

En este estímulo léxico se han evocado numerosas realidades de distinto tipo. Todas han sido admitidas con las únicas excepciones de los nombres propios que aparecían —nombres de ciudades como Barcelona, Madrid, Lleida, Lérida o Solsona, así como el apellido de quien fue el primer alcalde de Lleida de la transición y permaneció durante veintitrés años, Siurana— y también las denominaciones concretas

⁶³⁹ En realidad, el significado en plural se recoge en el DIEC bajo la construcción *tenir nervis*, que equivale a *estar nerviós*.

⁶⁴⁰ Esta palabra, con el significado de “puesto” (de mercado o de mercadillo) sustituye, frecuentemente, el término castellano “puesto”, que no ha aparecido en nuestro corpus.

⁶⁴¹ En castellano hemos separado las dos lexías “cine”/*cinema*, por considerar que la segunda es un cambio de código y, por tanto, pertenece al acervo léxico del catalán. La razón que nos lleva a pensar así, a pesar de que este acortamiento se incluye en el DRAE, es su total ausencia en los restantes léxicos disponibles del español de España. En el capítulo 7 ofrecemos más información al respecto (*vid.* 7.1.).

⁶⁴² También así en Bartol (2004: 28).

de calles, plazas y partes de la capital han sido excluidas (*La Seu Vella, La Suda, La Mitjana, Lluís Companys, Pavelló 11 de Setembre, Plaça Europa y Gardeny*). También hemos eliminado los nombres de comercios, referidos principalmente a grandes almacenes y hamburgueserías (*Burguer King, El Corte Inglés, La Caixa, Mac Donalds, Mercadona, Pans&Company, Plus, Pryca, Carrefour, Pull&Bear, Pinky...*)⁶⁴³. Cabe decir, al respecto, que la mayor parte de estas asociaciones habían sido citadas en solo una ocasión. Los únicos términos que tenían más de una mención han sido: el nombre de la capital —Lleida—, su monumento más emblemático —*La Seu Vella*⁶⁴⁴— y los grandes almacenes *El Corte Inglés*, de los que, paradójicamente, carece la ciudad. No obstante, hemos conservado los nombres de aquellos organismos e instituciones que podían ser genéricos y, por ende, hallarse en cualquier ciudad como “ayuntamiento”, “diputación” “consejo comarcal”/*consell comarcal, hisenda...* también hemos incluido “eje comercial”/*eix comercial* y “calle mayor”/*carrer major*, denominaciones referidas a Lleida, pero también extrapolables a otros núcleos urbanos.

5.4.3.11. *El campo/El camp*

Este centro de interés es uno de los que aglutina un mayor número de voces y, por tanto, son bastantes los dilemas que se plantean a la hora de editar los datos. En primer lugar, hemos incluido en cursiva, como en los casos anteriores, los extranjerismos no registrados en los diccionarios normativos, esto es, *camping, golden* (en el sintagma “manzana *golden*”), *picnic* y *rafting*, en el corpus castellano, y *container*, en el catalán⁶⁴⁵. En relación con las palabras que pueden consignarse bajo distintas opciones, las hemos unificado, una vez más, bajo la forma mayoritaria entre nuestros informantes. De este modo, hemos incluido “chalet”, igual que sucedía en el centro anterior, y *llogarret*, dado que no aparecían en los respectivos corpus las formas

⁶⁴³ Carcedo (2001), Arnal *et al.* (2004) o López Meirama (2008) también optan por suprimir todas estas asociaciones (nombres propios, lugares de la ciudad y marcas comerciales referidas a comercios). López Meirama (2008: 68), no obstante, incluye *Carrefour* y *El Corte Inglés* como testimonio del peso que han tenido estos elementos en las encuestas. También hay que decir que muchos otros investigadores conservan todas estas denominaciones en los listados definitivos. Por otro lado, hemos mantenido “día” en nuestros listados porque, a partir de las asociaciones evocadas antes y después, nos parecía que el informante en cuestión no se refería al centro comercial que lleva el mismo nombre.

⁶⁴⁴ Otra forma de denominar este conjunto histórico es “el castillo”, porque, tras la Guerra de Sucesión, la catedral y todo su entorno arquitectónico y urbanístico fue transformado en cuartel militar fortificado como uno de los castigos que recibió la ciudad del entronizado rey Felipe V, y así permaneció hasta poco después de la Guerra Civil al construirse nuevos cuarteles para las tropas, razón que explica el alto índice de aparición de esta palabra.

⁶⁴⁵ *Camping* ya aparece en el DRAE aunque en cursiva; *picnic* y *rafting*, por su lado, ya se incluyen en el DEA. En el corpus catalán hemos incorporado *càmping* y *pícnic*, bajo las formas ya sancionadas por el TERMCAT.

preferentes según los diccionarios normativos (“chalé” y *llogaret*). Ante la extensa nómina de variantes halladas en catalán, hemos optado por el primer miembro de las siguientes parejas: *alfals/aufals*, *àguila/àliga*, *ametlla/atmetlla*, *ametller/atmetller*, *cabana/cabanya*, *caragol/cargol*, *mixó/moixó*, *presseguer/pressiguer*, *taup/talp*⁶⁴⁶. También hemos dado cabida a los dialectalismos *anouer*, *mangrana*, *palot*, *safranòria* y *tocino*⁶⁴⁷. En cuanto a la unificación ortográfica, cabe decir que hemos dado cabida a los pares diacríticos *sol/sòl* en el corpus catalán, toda vez que los dos informantes que los incluían lo hacían seguidamente, con lo que se constata que evocan dos realidades distintas.

La unificación en el género masculino no afecta, en este estímulo verbal, a las parejas compuestas por el árbol y su fruto (“almendro/a”, “cerezo/a”, “ciruelo/a”, “manzano/a”, “naranja/a”, “olivo/a”), por la persona encargada de alguna labor y la máquina que la lleva a cabo (“sembrador/a”, *llaurador/a*, *segador/a*) o cuando puede existir alguna diferencia semántica entre las dos palabras (“fruto/a”, “huerto/a”, *fruit/a*, *hort/a*, *roc/a*)⁶⁴⁸. Hemos respetado el femenino en las siguientes voces —que solo se han consignado de este modo— que denominan distinta maquinaria agrícola o bien reflejan el papel tradicional de las mujeres en las labores del campo: “abonadora”, “cosechadora”, “embaladora”, “trilladora”, (*en*)*sulfatadora*, *menjadora*, *munyidora*, *polidora*, *sembradora* y *trinxadora*. Por el contrario, hemos reducido bajo la opción más frecuente en los listados, ya fuera la masculina (“granjero”, “perro”, “zorro”, *oliver*, *pomer*, *pruner*) o la femenina (*aixada*, *perera*), las voces que no presentaban alteración en el significado en función del género⁶⁴⁹.

⁶⁴⁶ Las voces preferidas por los informantes coinciden con la voz estándar y preferente en el DIEC, excepto en el caso de *caragol*, *mixó* y *taup*. El DLL incluye las voces dialectales *aufals*, *mixó* y *taup*, NDLLC, por su parte, recoge *aufals*, *mixó*, *taup* y *pressiguer*; *atmetlla* y *atmetller* se consignan en el DCVB.

⁶⁴⁷ Todas ellas se consignan en el DLL y en NDLLC, excepto *anouer*, aunque sí aparece en ambos diccionarios *anou*. *Palot*, cuyo uso está bastante extendido en nuestra provincia tal como muestra su inclusión en las listas del castellano, solo se registra en NDLLC donde se define como “construcció de fusta de forma cúbica amb base de palet, amb la cara superior descoberta i d’un metre d’alçada, usat per abocar-hi fruita al tros, per després transportar-la en tractors a la comparativa” [sic].

⁶⁴⁸ En el caso de “arado”/ *arada* mantenemos la distinción de género para mostrar la variante castellana y la catalana del mismo concepto. Nos parece más que improbable que los hablantes se refieran en femenino a la “acción de arar” o a la “tierra labrada con el arado”, definiciones que otorga el DRAE al respecto.

⁶⁴⁹ Igual que sucedía en el centro 5, hemos incorporado *fruit sec* en género masculino, a pesar de que la forma normativa catalana es la femenina.

En cuanto a la variación de número, hemos conservado el plural en aquellas lexías cuyo uso plural, o generalmente plural, ya se especifica en los diccionarios: “tijeras”, *diners* y “purines”/*purins*⁶⁵⁰. También se ha respetado en las palabras compuestas (“espantapájaros”, “saltamontes”, *centpeus*, *espantaocells*) o cuando se exige en determinados sintagmas (“pocos vecinos”, “sin coches”, “sin obligaciones”). Los derivados lexicalizados que aparecen en este centro, junto con la voz de la que proceden, no son pocos: “cabra/cabrito/cabrón”, “carreta/carretilla”, “casa/caseta/casita”, “fresco/frescor/frescura”, “fruta/frutal/frutero”, “ganado/ganadería/ganadero”, “hierba/hierbajo”, “manzana/manzanar/manzano”, “pollo/pollito”, “río/riachuelo”, *carreta/carretó/carreter*, *fruit/frutal/fruiter*, *gall/gallina/galliner*, *mosca/mosquit*, *poll/polli/pollet*, *ria/riera/rierol*, etc. Los diminutivos lexicalizados son frecuentes en la designación de algunos insectos o anfibios como “mariquita”, “tijerilla”, *cullereta*, *estisoreta* o *marieta*. En cambio, hemos unido “pajarito” a “pájaro”, *animalet* y *animaló* a *animal*, *ocellet* a *ocell*, *pedreta* a *pedra* y *riuuet* a *riu*, siguiendo el mismo criterio que en anteriores ocasiones.

En esta área léxica hemos utilizado el paréntesis únicamente en “*bici(cleta)*” y en *natura(lesa)* para mostrar la vacilación entre la forma completa o apocopada⁶⁵¹. También lo hemos usado en la forma *(en)sulfatadora*, con la vacilación dialectal del prefijo que ya hemos visto en otras ocasiones. Por otro lado, hemos incluido *retro[excavadora]*, aunque usamos el corchete, en esta ocasión, para constatar que ningún informante ha actualizado la lexía completa.

Al igual que otros investigadores con anterioridad (Gómez Molina y Gómez Devís 2004 o Hernández Muñoz 2004), constatamos la sorprendente aparición, en este campo léxico, de palabras que se adecúan más al título del centro anterior. Algunos ejemplos son “autobús”, “calle”, “coche”, “hotel”, “ladrillo”, “restaurante”, *autocar*, *col·legi*, *escola*, *plaça*... Hemos mantenido todas estas asociaciones e incluso otras más alejadas como “elfo” o *fada* con las únicas excepciones de algunos nombres de una conocida serie que se desarrollaba en las montañas suizas (“Pedro”, “Heidi”), un

⁶⁵⁰ Aunque, tal como indican las fuentes lexicográficas, el uso general de *fem* es en plural, lo hemos consignado en singular por ser la opción más frecuente en nuestros datos.

⁶⁵¹ En castellano, en cambio, mantenemos separadamente “naturaleza” y *natura*, por considerar esta última una transferencia lingüística del catalán, conclusión a la que llegamos tras comprobar la ausencia de esta palabra en los léxicos disponibles del castellano publicados hasta el momento (*cfr.* 7.1.).

fragmento del himno nacional de Cataluña (*bon cop de falç*) y algunos sintagmas u oraciones que sobrepasaban los límites de la palabra (“poca disponibilidad para viajar” o “por la noche se pueden ver las estrellas”).

5.4.3.12. Medios de transporte/Mitjans de transport

Este centro de interés ha suscitado menos problemas de edición que el centro anterior, principalmente por la escasez de fenómenos de contacto lingüístico y de variantes dialectales. Es notable, no obstante, el número de extranjerismos registrados en él: *dumper*, *ferry*, *hovercraft*, *kart*, *quad*, *mountain bike*, *rafting*, *roulotte*, *skate*, *snowboard*..., a pesar de que muchos de ellos ya se recogen en el DRAE (“catamarán”, “kayac”, “limusina”, “sidecar”, “surf”, “tráiler”...) o en el TERMCAT (*bugui*, *caiac*, *escúter*, *hovercraft*, *quad*...) ⁶⁵². En otras ocasiones, no aparecen o se consignan en cursiva en los diccionarios y, por ende, los mantenemos así en la edición final de los datos: *ferry*, *jet*, *motocross*, *roulotte*, *scooter*, *windsurf*, en castellano, y *bobsleigh*, *skate*, *skater*, *snow* (en “tabla de snow”/*taula de snow*), *snowboard* y *windsurf*, en el corpus catalán.

En relación con las variantes, hemos conservado el género femenino en la denominación de distinta maquinaria agrícola —susceptible de ser empleada como medio de transporte según nuestros informantes— como “apisonadora”, “cosechadora”, “excavadora” y “segadora”, o cuando era necesario mantener los dos géneros porque aludían a referentes distintos (“barco/a”). El plural se ha respetado en las palabras compuestas (“paracaídas”, “portaaviones”, *parabrises* ⁶⁵³, *paracaigudes*, *portaavions* e, incluso, la creación “*portacontainers*”), en locuciones adverbiales (“a gatas”, “[a] hombros”, *a cunetes*, *de quatre grapes*) y en el segundo miembro de algunas formas analíticas (“coche de caballos/bomberos”, “camión de mercancías”, *canvi de marxés*...). Siguiendo las pautas generales, hemos conservado los siguientes derivados

⁶⁵² Como en otras ocasiones, el DEA señala algunas voces ausentes en el diccionario académico, como *dumper*, *kart* o *skate*. Al igual que sucedía en otros estímulos verbales, no parece haber coincidencia entre las voces que registran las autoridades de cada lengua; así, por ejemplo, el TERMCAT incluye *kart* y *quad*, que no se recogen en el DRAE, o a la inversa, el diccionario académico recoge “windsurf” o “zodiaco”, que no aparecen en las fuentes normativas catalanas. Parece un contrasentido, ya que, al igual que sucede con las marcas comerciales, es extraño que el uso de un extranjerismo se difunda solo en una de las dos lenguas en contacto de un mismo territorio. Con respecto al corpus catalán, extraña la acentuación de *dúmp*er en el DIEC, y, sin embargo, la ausencia de la misma en el TERMCAT como sinónimo complementario de *mototràfic*.

⁶⁵³ Incorporamos esta palabra en plural, a pesar de que la forma normativa es en singular, para reflejar el uso en nuestra comunidad. Una vez más, la influencia del castellano se percibe en el catalán.

lexicalizados: “*avioneta*”, “*camioneta*”, “*carretilla*”/*carretó*/*carrossa*/*carruatge*, “*cochecito*”/*cotxet*⁶⁵⁴, “*furgoneta*”, “*patinete*”/*patinet*, etc.

En cuanto a los sintagmas que aparecían con bastante frecuencia del tipo “*pie/a pie/tus pies/a propio pie*”, los hemos unificado siempre en la variante más frecuente en el corpus (en el caso anterior “*a pie*”). De este modo registramos: “*a pata*”, “*en brazos*”, *a peu*... También hemos usado el criterio de frecuencia en la variación entre infinitivo y gerundio en las formas verbales: “*andar/andando*”> “*andando*”, “*caminar/caminando*”> “*caminar*”, “*correr/corriendo*”> “*corriendo*”, *nedar/nedant*> *nedant*, *volar/volant*> *volant*⁶⁵⁵... Se ha recurrido al uso del paréntesis en “*(auto)bús*” y en *bici(cleta)*⁶⁵⁶, aunque hemos separado “*moto*” de “*motocicleta*”, al igual que hicimos en el centro 10 y por la misma razón. Con paréntesis se señala, además, la vacilación en la adjetivación del sustantivo en “*cohete (espacial)*”/*coet (espacial)*, “*globo (aerostático)*”/*globus (aerostàtic)* aunque no se produce en los casos en los que se podría tratar de diferentes realidades (“*nave/nave espacial*”). Cabe comentar en este punto que no han sido pocos los adjetivos enumerados en torno a este campo semántico sin que aparezca el núcleo al que complementan (“*deportivo*”, “*descapotable*”, “*dirigible*”...). Hemos utilizado el corchete en la locución adverbial “[*a*] *hombros*”, donde nos parecía imprescindible la presencia de la preposición *e*, igual que en el centro anterior, en *retro[excavadora]*. Con respecto a las siglas y acrónimos, aparecen en los corpus AVE, F-16, Renfe, TAV y TGV.

Las marcas comerciales, profusas en esta área léxica, se han mantenido en el caso de que estuvieran lexicalizadas o en vías de lexicalización. Igual que hicimos en el centro 5, en principio, hemos decidido mantener aquellas que contaran con un mínimo de cinco menciones en ambas lenguas. Sin embargo, en esta circunstancia hallamos solo *jeep*, cuya entrada ya consta en el DIEC. Por esta razón, tras un análisis más

⁶⁵⁴ “*Cochecito*” no aparece en el DRAE. Lo hemos mantenido porque alude a un referente distinto para el que también se emplea “*carrito*”, como se observa al incluirse las formas “*carrito de los niños*” y “*carrito de bebé*” en el corpus. Por otro lado, hemos unificado en “*barco/a*” los diminutivos “*barquito/a*” y “*trenecillo*” en “*tren*”.

⁶⁵⁵ En el corpus catalán, *volant* parece referirse, al menos en una ocasión con bastante seguridad, a la pieza con la que un conductor dirige un automóvil, puesto que el informante lo añade entre *roda* y *fre*. Esto puede explicar el índice de disponibilidad que alcanza en catalán, puesto que en castellano ni siquiera aparece el gerundio “*volando*”, que es el significado normativo de *volant* en catalán.

⁶⁵⁶ Mantenemos el paréntesis en los sintagmas según haya o no vacilación en los informantes que los incluyan, eso explica que incorporemos *(auto)bús local* y, en cambio, *bus turística*.

pormenorizado de los corpus, hemos optado por incluir las marcas *audi*, *mercedes* y *volkswagen*, toda vez que, al menos en catalán, sí alcanzaban las cinco menciones e, incluso las dos primeras, las rebasaban. Hemos conservado también las marcas comerciales que ya se incluyen en alguno de los diccionarios normativos de las dos lenguas como síntoma de su lexicalización. En esta situación hallamos *bamba*, que ya incluimos en el segundo centro de interés, o *vespa*, *vespino* y *zodiak*, todas ellas presentes en el DRAE. El resto de denominaciones comerciales, principalmente referidas a marcas y modelos de coches y motos (*aprilia*, *boeing*, *BMW*, *chrysler*, *citröen* (*zx*, *xsara*), *daewo*, *fiat*, *ford*, *funicamp*, *golf*, *harley davidson*, *honda*, *hyundai*, *iberia*, *jaguar*, *kawasaki*, *landrover*, *laguna*, *mobylette*, *nissan*, *opel* (*corsa*), *patrol*, *renault* (*laguna*), *saab*, *scenic*, *scoopy*, *seat* (*ibiza*, *león*, *panda*), *skoda*, *smart*, *snowblade*, *suzuki*, *transmediterránea* y *toyota*) han sido desechadas; cabe aclarar que una gran mayoría se mencionaba en solo una ocasión. Todas las restantes asociaciones se han conservado, incluyendo los distintos tipos de aviones (*concorde*, F-16, *jumbo*) o trenes (*euromed*, *talgo*, *tren delta*, *tren regional*, etc.), y a pesar de que algunos se refirieran a medios de transporte fantásticos (“alfombra mágica/voladora”, “escoba”, *casquet volador*), simbólicos (“droga”, “teletransporte”) o incluso bíblicos (*arca*).

5.4.3.13. Trabajos del campo y del jardín/Treballs del camp i del jardí

Destaca, en este centro de interés, la gran cantidad de verbos recabados en comparación con las restantes áreas léxicas y también, aunque en menor número, la abundante presencia de adjetivos derivados de dichos verbos con los que se alude a las personas que realizan los distintos trabajos del campo y que funcionan como nombres de oficios (“arador”, “cosechador”, “cuidador”, “fumigador”, “plantador”, “podador”, “recolector”, “regador”, “talador”, *adobador*, *collidor*, *criador d’animals*, *cultivador*, *empeltador*, *fumigador*, *munyidor*, *sembrador*, *tallador*...). También, como veremos en el análisis cualitativo, es abundante la presencia de voces catalanas en los listados castellanos; por el contrario, los extranjerismos brillan por su ausencia. Con respecto a las pautas concretas de edición, hemos incluido, en el corpus castellano, “invernar”, que agrupa las variantes “invernar/ívernar”, y también “anivellar”, “cequias” (en el sintagma “abrir cequias”) y “encepar”, puesto que eran las únicas opciones que aparecían en nuestros datos, aunque son palabras que remiten a otras entradas en el DRAE⁶⁵⁷. En

⁶⁵⁷ Se percibe, por tanto, la influencia del catalán donde existen *anivellar*, *séquia* y *encepar*. De otra parte, extraña la ausencia de “recolecta” en el DRAE, que sí recoge el DEA —aunque con otro significado— y

ambas lenguas, hemos corregido el error, bastante extendido, de “transplantar” por “trasplantar” y, en catalán también hemos subsanado algunas variantes fonéticas —como la elisión de la [e] pretónica en *veremar*, frecuentemente escrito como *bremar*— que ya se detallan, en casos semejantes, en la bibliografía sobre el catalán noroccidental (Julià 2004: 34)⁶⁵⁸. Algunos dialectalismos se han mantenido con un elemento epentético inicial, y hemos mostrado esta vacilación mediante paréntesis cuando aparecían las dos opciones: *enfemar*, *(en)sulfatar*, *(en)sulfatador*... También se han detectado otros dialectalismos —ausentes en las obras sobre el catalán estándar— como *currar*, *corronar*, *espampar*, *gradejar*, *llimpiar*, *rampilladora* y *trisellar*⁶⁵⁹.

Con respecto a las variantes, hemos mantenido los dos géneros en “fruto/a” (en los sintagmas “coger/recoger fruta/frutos”, *collir/recollir fruita/els fruits*) o en *hort/a* y también en los casos de *jardiner/a*, *recol·lector/a*, *regador/a* y “*segador/a*”, donde puede que se distinga el sexo de la persona que desempeña el oficio o entre la persona y la máquina que sirve para realizar la actividad. También hemos conservado la variante femenina cuando solo está presente esta forma por referirse a tareas tradicionalmente realizadas por mujeres (“costurera”, *minyona*). Asimismo, hemos respetado en la variación de número la opción indicada por nuestros informantes cuando las dos constaban en los diccionarios, de tal modo, incorporamos “guardabosques” y “tijera”, y lógicamente, cuando la opción en plural es la única posible (“quitanieves”, *espantaocells*). Son escasísimos los derivados hallados en torno a esta área (“carretón”, “paleta”) y, en cambio, muy abundantes los sintagmas con un núcleo verbal y un complemento, muchos de ellos consignados en plural, que hacen referencia a distintas tareas agrícolas (“contratar jornaleros”, “limpiar (malas) hierbas”, “matar insectos”, “recoger aceitunas”, *arreglar les flors*, *cremar fulles*, *desinfectar animals*, *retallar les branques*, etc.).

tampoco consta como sustantivo en las fuentes normativas catalanas que sí consignan *recol·lecte/a*, como adjetivo, y el verbo *recol·lectar*.

⁶⁵⁸ También hemos corregido *desllucar>esllucar*, y en el corpus castellano, *cardar>escardar*, donde creemos que se omite el primer elemento por ultracorrección.

⁶⁵⁹ La mayoría de ellos son verbos derivados de instrumentos que, unidos a un tractor, sirven para trabajar la tierra, como por ejemplo, *currar* (esto es, *passar la curra* —que aparece en el corpus castellano—) o *corronar* (*passar el corró*, una pieza cilíndrica que se utiliza en su rotación), *gradejar* (*passarla grada*, instrumento que consta de dos filas de rejas) y *trisellar* (*passar la trisella*, utensilio cóncavo que sirve para allanar el terreno). *Rampilladora* es la versión dialectal de la (máquina) *rampinadora* y *espampar* consiste en aclarar los pámpanos presentes en una cepa o parra. NDLLC recoge la mayoría de estos instrumentos agrícolas (*curra*, *curro*, *grada*, *trisella*...) y el verbo *curronar*.

La simplificación de lexías mediante el paréntesis ha afectado en este centro a “adobar (la tierra)”, *adobar (la terra)*, “limpiar/quitar (malas) hierbas”, “remover (la tierra)”, “talar (árboles)”, *airejar (la terra)*, *conrear (els camps)*, *(enginyer) agrònom*, *remoure (la terra)*, *sembrar (la terra)*, *talar (arbres)*, *tallar (les males) herbes* y *treure (les males) herbes*⁶⁶⁰.

Aunque otros autores suprimen algunas asociaciones en este centro de interés, como los nombres de los productos (Carcedo 2001: 49), en nuestro caso hemos mantenido todas las palabras recogidas; no obstante, al igual que hemos hecho en otras áreas, hemos suprimido las paráfrasis “el que arregla las flores”, “el que se cuida de los animales” y *persona que s’encarrega dels horts*. Asimismo, hemos simplificado, de la forma que sigue, las siguientes expresiones: “comer lo del campo” > “comer” y *retallar les branques dels arbres* > *retallar les branques*. Hemos admitido todas las palabras que habían evocado los informantes, aunque algunas se alejasen bastante del estímulo en cuestión como “artesano”, “barrendero”, “bombero”, “cocinero”, “constructor”, “costurera”, “decorador”, “panadero”, “policía”... Consideramos que algunas de estas (“caballero”, “domador”, “enterrador”...), más que asociaciones secundarias son reflejo del desconocimiento, por parte del informante, del léxico que sería adecuado enunciar en torno a este estímulo.

5.4.3.14. *Animales/Animals*

En este centro, como en los anteriores, hemos optado por la forma más frecuente entre las dos posibilidades que ofrece el diccionario en los siguientes casos del catalán: *euga/egua* > *euga*, *àliga/àguila* > *àguila*, *cargol/caragol* > *caragol*. Este proceso de lematización según la frecuencia de mención también ha afectado a las siguientes parejas compuestas por un término normativo y uno dialectal: *llagosta/llangosta* > *llagosta*, *llangardaix/llargandaix* > *llangardaix*, *mixó/moixó* > *mixó* y *talp/taup* > *talp*⁶⁶¹.

⁶⁶⁰ Nos habría gustado extender esta reducción a otros compuestos como “abonar (la tierra)” o “cultivar (la tierra)”, *cultivar (la terra)* puesto que la presencia de las dos formas hace descender el índice de disponibilidad de las dos. Sin embargo, el hecho de que los informantes aportasen otras asociaciones, en los ejemplos anteriormente citados, como “abonar las plantas” o “abonar el campo”, y la escasez de menciones que incluyen el complemento, nos han hecho decantarnos finalmente por incluir todas las opciones.

⁶⁶¹ Las variantes dialectales, esto es *llangosta*, *llargandaix*, *mixó* y *taup*, aparecen tanto en el DLL como en NDLLC. Otros dialectalismos presentes en este centro son *esquitxa-robès*, *gat fer* y *puó*, que también aparecen en los dos vocabularios regionales mencionados (*esquitxa-robès* aparece en NDLLC como *asquixarrobes* y en el DLL como *esqueixa-robès*).

Los extranjerismos actualizados en torno a este estímulo son escasos y, mayoritariamente, ya se incluyen en el DRAE (“hámster”, “poni”) o en el TERMCAT (*bulldog*, *rottweiler*), salvo algunas excepciones como *black-bass*, que aparece en los dos corpus, o *pit bull*, en el corpus del castellano. También hemos documentado una palabra del aranés, *taure*⁶⁶², y algún caso como “helicóptero”/*helicòpter*, cuyo significado coloquial equivalente a un insecto no se recoge en los diccionarios normativos⁶⁶³.

Con respecto a la unificación de variantes, al igual que los autores que nos preceden, en general, hemos igualado los dos géneros a favor del masculino salvo en las ocasiones en que el femenino es la única opción presente (“mula”) o se designa mediante un heterónimo (*cavall/euga*, *porc/truja*) o un sufijo específico (“gallo”/“gallina”, *abella/abellot*) o cuando existe una oposición semántica de distinto tipo entre ambos (“caballo/a”, *cuc/a*, *poll/a*). En relación con los derivados, se han conservado, como de costumbre, los lexicalizados (“abejorro”, “aguilucho”, “caballito de mar”, “cabrito”, “cabrón”, “caracolina”, “cochinillo”, “copito de nieve”, “lagartija”, “mariquita”, “moscardón”, “mosquito”, “ratón”, “vaquilla”, *abellot*, *cabrit*, *cabró*, *floquet de neu*, *marieta*, *mosquit*, *peixet* y *tisoreta*)⁶⁶⁴. El resto de diminutivos, que no han sido pocos —sobre todo en el corpus catalán—, se han asimilado a la forma primitiva: *cavallet de mar*> *cavall de mar*, *foqueta*>*foca*, *gosset*>*gos*, *mixonet*> *mixó*, *ocellet*>*ocell*, “pececito”> “pez”, “pollito”> “pollo”, “sardinilla”> “sardina”... De otra parte, el plural se ha mantenido, al igual que en los centros anteriores, en algunas formas compuestas como “ciempiés”/*centpeus*, “milpiés”/*milpeus*, “quebrantahuesos” o “saltamontes”. La formulación parentética ha afectado a las siguientes lexías: “(oso) panda”/(*ós*) *panda*, “mantis (religiosa)”, “(pájaro) carpintero”, “(paloma) torcaz”, (*ésser*) *humà* y (*porc*) *senglar*.

⁶⁶² Es curioso que las interferencias del aranés se realicen siempre en el corpus catalán, y no cuando se está utilizando la lengua castellana, lo que nos lleva a pensar que los informantes de esa zona distinguen mejor entre su lengua materna y el castellano, que entre el aranés y el catalán. Sobre el uso de las tres lenguas en *la Vall d’Aran*, véase el final del capítulo cuarto.

⁶⁶³ En NDLLC sí aparece *helicòptero* y es definido en la segunda acepción como equivalente a “libélula” (*vid.* 7.2.).

⁶⁶⁴ En los casos de “caballito de mar” y “caracolina”, ni siquiera aparecían “caballo de mar” o “caracola”. El diminutivo *peixet* se conserva porque puede referirse a un subtipo de pez: el *peixet de coure* o el *de plata*. Con el término “copito de nieve”/*floquet de neu* se alude a un determinado primate, el único ejemplar mundial de gorila albino, que vivió la mayor parte de su vida en el zoo de Barcelona. Adquirió una gran fama al ser portada, en el año 67, de la revista *National Geographic Magazine* y en Cataluña ha sido un icono para los niños y niñas de diversas generaciones.

En esta área léxica se han recogido abundantes especies de animales, aunque cabe destacar que, muchas de ellas, se mencionan sin indicar el nombre genérico al que se refieren, sobre todo en el caso de perros y gatos (“dálmata”, “siamés”, *persa...*) pero también en muchos otros ejemplos (“cascabel”, “diamante”, “mandarín”, “manta”, “raya”, *collverd...*). Tras debatirnos entre diferentes posturas, finalmente hemos optado por considerar innecesario establecer el genérico al que se refiere cada término, al igual que López Meirama (2008: 77), puesto que el uso parece ya establecerlo así: prueba de ello es que algunas de estas especificaciones ya se registran en el diccionario (*v.gr.* “dálmata”, *manta*). Además, en algunos ejemplos, es también imposible averiguar a qué nombre genérico se refiere el estudiante (“mandarín” > “pato mandarín” o “diamante mandarín”).

En cuanto a las asociaciones halladas en este estímulo verbal, la mayor parte, como hemos indicado, han sido nombres de animales y de distintos subtipos y especies. Los informantes han entendido el concepto “animal” en sentido amplio, y por tanto, han incluido referencias al ser humano (“hombre”, “humano”, *dona, (ésser) humà...*) y también al medio, natural o artificial, donde los animales se desarrollan o viven (“mar”, “naturaleza”, “selva”, *colomer, galliner, teranyina...*). No faltan las designaciones de todo tipo de seres microscópicos (*bacteri, microorganisme, virus...*) —incluso en sus estadios más tempranos (*caviar, “liendre”*)—, los adjetivos que los clasifican o califican (*argollat, herbívor, invertebrat, ovípar, racional, salvatge...*), los nombres de especies ya desaparecidas (“diplodocus”, “mamut”, “tiranosaurio”, *dinosaure...*) o las alusiones a seres mitológicos (*drac, “tritón”/tritó ...*). Al igual que realiza Arnal *et al.* (2004: 43), las únicas asociaciones que hemos desechado han sido las referidas a nombres propios con intención claramente jocosa (“Héctor”, “Miriam”, “van Gaal”...) y las que designan animales que aparecen en series de televisión o películas (“bambi”, “chiwaka”, “pájaro loco”, “pantera rosa”, “rana gustavo” y “pikachu”), denominaciones todas ellas escasísimas en los dos corpus.

5.4.3.15. Juegos y distracciones/Jocs i distraccions

Si algo caracteriza este centro de interés, además de la profusión de formas verbales y marcas comerciales, es, sin duda, la abundancia de extranjerismos, que proceden, sobre todo, del inglés. Muchos de ellos ya forman parte del acervo léxico de la lengua española (“aikido”, “aerobic”, “bádminton”, “básquet”, “béisbol”, “cómic”, “club”, “diábolo”, “golf”, “karaoke”, “póquer”, “taichí”, “taekwondo”, “waterpolo”...) y catalana (*càmping, criquet, handbol, karate, kart, pub, ràfting, ral·li, ràpel, rugbi, surf, tennis, xat...*). En cambio, son menos los que aún no aparecen en el DRAE (*fitness, funky, kick-boxing, picnic, pub, rafting, rappel, skate, snowboard, squash, steps, strip poker...*) ni en el TERMCAT (*joker, messenger, pinball, skate, snowboard, strip poker...*), aunque sí lo hagan en algunos diccionarios de uso⁶⁶⁵, o los que se incluyen en el diccionario académico en cursiva, a pesar de que son vocablos muy extendidos en el uso (*ballet, camping, hobby, hockey, motocross, rally, rugby y stop*). Al igual que realizamos en otros centros de interés, cuando ningún informante consigna la forma del préstamo sancionada por los diccionarios, lo transcribimos según su forma de origen, por dicha razón, incluimos en el corpus catalán *footing, motocroos* y *squash*, y *no fúting, motocrós* y *esquaix*⁶⁶⁶. También hemos recabado algunas adaptaciones de anglicismos como “edredoning”, “grafitear”, “puenting” o “surfear” y algunas siglas (CD, BTT, DVD, PC), la mayoría de las cuales también son anglicismos.

Hemos homogeneizado las variantes siguiendo el criterio de frecuencia en las formas verbales *xarrar/xerrar*>*xerrar*, *nadar/nedar*> *nedar*, y también con respecto a los distintos nombres que recibía un juego tradicional infantil, cuando teníamos la seguridad de que se trataba del mismo referente: *cucut i amagar, cuit i amagar, cuc i amagar*> *cucut i amagar*. El mismo criterio, basado en el número de menciones, hemos usado para transcribir algunas marcas comerciales, asunto del que hablaremos a continuación, toda vez que algunas de ellas —en el corpus catalán— eran traducidas por

⁶⁶⁵ El DEA incluye la mayor parte de estos extranjerismos, ausentes en el diccionario académico. Es significativo que el DRAE, en el avance de su 23ª edición, admita ya algunos de estos vocablos —antes ausentes o incluidos con su grafía original— como *chat, chatear, pádel* o *pimpón*. En catalán, el GDLC admite *snowboard*, aunque lo marca con un asterisco.

⁶⁶⁶ Nótese cómo, una vez más, estas formas coinciden con las que se utilizan en lengua castellana para estos mismos préstamos, aunque la diferencia sea solo gráfica. El mismo fenómeno se detectaba en el centro 5.

los informantes: “buscaminas”/buscamines> *buscamines*, “hundir”/enfonsar (la flota)> “hundir la flota”, “quién es quién”/qui és qui> *qui és qui*⁶⁶⁷.

De otra parte, hemos conservado los dos géneros en los pares “anillo/a” y “bolos/as”, donde se alude a juegos distintos; el resto de lexías se han unificado en la forma masculina, a no ser que la femenina fuera la única opción hallada (“novia”, *noies*). Un aspecto especialmente problemático, en este campo léxico, es la lematización en singular: son muy abundantes las voces incluidas en plural, cuyo significado parece perderse o tergiversarse al pasarlas al singular (v.gr. “barcos”, “cartas”, “damas”, “mamás y papás”, “sacos”...), aunque esta diferencia de apreciación no sea tan clara en algunas ocasiones (como en “amigos”, “apuestas” o “caramelos”). Por otro lado, los mismos diccionarios ofrecen, algunas veces, esta información acerca del uso frecuente en plural (así sucede en “damas”, “maracas”, “naipes”, *escacs* o *malabars*), a pesar de que no se dé esta circunstancia en muchos de los ejemplos recabados. Ante esta disyuntiva, hemos optado por incorporar en plural aquellas lexías que han sido nombradas así por todos los informantes y, por mantener el singular en aquellas en las que hubiera vacilación (“película/s”> “película”). De este modo, finalmente se consignan en plural una gran retahíla de voces tanto en castellano (“atracciones”, “barcos”, “canicas”, “cocinas”, “construcciones”, “dados”, “moros y cristianos”, etc.) como en catalán (*amics, bitlles, cadires, carreres, castells, cromos, cuinetes, disbarats, dones, endevinalles, firetes, focs, imitacions, pares i mares, parelles*, etc.).

También son frecuentes, en este centro, los aumentativos o diminutivos lexicalizados que aluden a juegos y distracciones diversas: “botellón”, “cinquillo”, “cohecito”, “maquinita”, “palitos”, “platillo”, “soldadito inglés”, *cuinetes, ganxet, manetes, ninot, patinet, siset*... En cambio, hemos reducido a la forma no marcada del paradigma en caso de hallar más de una opción: “gallina/gallinita ciega”> “gallina ciega”, “conejo/conejito de la suerte”> “conejo de la suerte”. Hemos mantenido, en cambio, separadamente “oso” y “osito de peluche” o *cadires/cadireta* y *fira/firetes*, por entender que nombran realidades distintas.

⁶⁶⁷ Sin embargo, no adoptamos este criterio cuando son denominaciones de juegos tradicionales, donde hemos respetado las dos opciones (v.gr. “conejo de la suerte”/conill de la sort, “soldadito inglés”/soldadet anglès) que aparecen únicamente en el corpus catalán.

Como la mayoría de investigadores dedicados a este campo — Gómez Molina y Gómez Devís (2004), Arnal *et al.* (2004), Fernández Smith *et al.* (2008), etc., quienes a su vez siguen el criterio de Samper Padilla (1998: 330-331)—, unificamos bajo el sustantivo las formas compuestas por un verbo y un nombre que aparecen en los listados junto con el sustantivo a secas. Son construcciones del tipo “ver la televisión/televisión> televisión”, “ver películas/película> película”, “escuchar la radio/radio> radio”, *anar al cinema/cinema> cinema*, *anar a concerts/concert> concert*, *llegir un llibre/llibre> llibre*, *saltar la corda/corda> corda*, etc.; sin embargo, respetamos el verbo cuando se alude a una acción distinta a la habitual (“actuar en televisión”). Tampoco hemos unido las entradas cuando se trata de locuciones (“ir de compras”/ “compras”, *beure/fer un beure...*) o, como en el resto de centros de interés, cuando son dos verbos los que alternan, aunque nombren acciones sinónimas (“ver básquet”/ “mirar básquet”, “ir de copas”/“salir de copas”).

Se ha recurrido al uso del paréntesis en los siguientes casos: “*bici(cleta)*”, “(juego de) rol”/(*joc de rol*), “(máquina) tragaperras”, “pelí(cula)”, “*play(station)*”, “snow(board)”, “tele(visión)”, “(video)consola”, “volei(bol)”, *biblio(teca)* —en el sintagma *anar a la biblio(teca)*— *cine(ma)*, *disco(teca)*, *poli(cie)s i lladres* y *tele(visió)*⁶⁶⁸.

Junto con el gran número de extranjerismos, la otra característica de este estímulo verbal es la profusión de marcas comerciales, con las que se alude, principalmente a juegos de mesa, pero también a videoconsolas y videojuegos. En este caso, decidir cuál está lexicalizada o no es difícilísimo, y el criterio de ver si sustituye o no a un nombre genérico no parece de utilidad, por cuanto no hay un equivalente existente para cada juego o muñeco particular⁶⁶⁹. Por otro lado, la importancia de las denominaciones comerciales en este centro está fuera de toda duda, ya que, por ejemplo, dos de ellas, —*monopoly* y *playstation*— se sitúan entre los veinte vocablos más disponibles en ambas lenguas. Ante tal confrontación, finalmente, hemos decidido

⁶⁶⁸ Hemos unificado, en los dos idiomas, las lexías “consola” y “videoconsola”, ya que —por nuestra propia experiencia— constatamos que se emplean con el mismo significado, hecho que se corrobora porque ningún informante incluye las dos voces en su listado. Contrariamente a lo que hicimos en el centro 10, en castellano, separamos “disco” y “discoteca” por entender que aquí el acortamiento puede referirse a un objeto, y no únicamente al lugar, como sucedía en aquella área léxica. Hemos conservado tal cual los acortamientos “ciber” y “chelo” por ser estas las únicas opciones halladas.

⁶⁶⁹ Véanse los acertados comentarios al respecto de López Meirama (2008: 78-80). La autora gallega toma como criterio la inclusión de las marcas comerciales mencionadas como mínimo en diez ocasiones.

incluir, como hicimos en el centro 12, aquellas marcas que han alcanzado un mínimo de cinco menciones en alguna de las dos lenguas (esto es, *barbie*, *buscaminas*, *cluedo*, *dreamcast*, *game boy*, *hotel*, *hundir la flota*, *master system*, *monopoly*, *nintendo*, *pictionary*, *playstation*, *quién es quién/qui és qui*, *scalextric*, *scattergories*, *scrabble*, *tabú*, *tetris* y *trivial*), a pesar de que reconocemos cierta arbitrariedad en esta decisión. Del mismo modo, hemos obrado con respecto a los nombres propios que denominaban programas de televisión o parques temáticos⁶⁷⁰.

También hay que señalar que hemos respetado el nombre de algunos juegos que parecen similares pero cuya coincidencia exacta no podemos corroborar, como “bote”/“bote bote”, “cruz y cara”/“cruz y raya”, *agafar/agafar i parar*. Asimismo, hemos conservado algunas denominaciones cuyo referente desconocemos (“abracadabra”, “cachumba”, “sopapo”, “trompa”, “tumbalibre”, *canari*...). Finalmente, constatamos la presencia de algunos adjetivos (“estéreo”, “latino”, “olímpico”) que se han actualizado tal cual, sin el sustantivo al que acostumbran a acompañar.

5.4.3.16. Profesiones y oficios/*Professions i oficis*

La edición de este campo semántico ha resultado más sencilla que la del centro anterior. Los extranjerismos recogidos son muchos menos, aunque siguen apareciendo algunos que no constan en el DRAE (*boy*, *drag queen*, *gigoló*, *showman*, *stripper*) o que se incluyen en cursiva (*striptease* —en “bailarín de *striptease*”—, *hockey* —en “jugador de *hockey*”—, *maître*, *paparazzi*...) indicando así que su representación gráfica o su pronunciación son ajenas a las convenciones de nuestra lengua. El catalán va por delante, una vez más, puesto que la mayoría de extranjerismos ya se registra en el DIEC o en el TERMCAT, y son solo unos pocos los que no figuran (v.gr. *cracker*, *hacker*, *paparazzi* o *showman*). Cabe destacar también la presencia en nuestros listados de voces muy habituales que, sin embargo, no aparecen en el diccionario académico, como “esteticien”, “planchista”, “reponedor” o “teleoperador(a)”. Además, hemos hallado términos jergales (“chirlero”, “segurata”) y algunos neologismos, habituales entre los jóvenes, como “pizzero” o “internauta”. En catalán, hemos unificado las variantes *algutzir/agutzil* en *algutzir*. Asimismo, hemos recogido algunas siglas, principalmente

⁶⁷⁰ Con la excepción de *PortAventura* y *pasapalabra*, estas entradas han sido eliminadas al contar con escasísimas menciones. Cabe destacar, en este punto, que *trivial* alcanza un elevado índice de disponibilidad al coincidir con el nombre de un bar juvenil de Lleida. Resulta imposible, en la práctica, diferenciar cuándo el informante se refiere al juego de mesa y cuándo al lugar de ocio nocturno.

referidas a distintos tipos de estudios: ADE, ATS, DJ, INEF, y un acrónimo del castellano en el corpus catalán, “geo”.

Es propio de este centro de interés el número de entradas con variación con respecto al género, reflejo final de la inserción laboral de la mujer en la sociedad. Por ser esta una información privilegiada de tipo sociolingüístico, algunos investigadores deciden lematizar de forma distinta a los restantes centros de interés, haciendo constar de algún modo la opción por la que se decantan los informantes⁶⁷¹. Nosotros, como ya dijimos en la introducción a este apartado, hemos homogeneizado todas las variantes en masculino singular con independencia de cómo figure la misma entrada en el diccionario académico⁶⁷². La excepción en este punto la conforman aquellas entradas, que no son pocas, transcritas por la totalidad de nuestros informantes bajo la forma femenina, lo cual nos puede dar una idea de las labores que se reservan para la mujer en la sociedad: “ama de casa”, “criada”, “escaladora”, “folclórica”, “niñera”, “teleoperadora”, “vividora”, *costurera, famosa, filóloga inglesa, llevadora, mainadera, mecanògrafa, pediatra o tendera*⁶⁷³. También mantenemos la opción del femenino cuando hay diferencia entre el significado de los dos términos, circunstancia que se da únicamente en nuestros corpus en “asistente/a”, “monje/a” y en las parejas que distinguen entre una profesión y unos estudios (“informático/a”, “mecánico/a”, “músico/a”...). Del mismo modo, los dos géneros se respetan también cuando se utilizan distintos sufijos en su formación (“actor/actriz”, *actor/actriu*) o en el caso de los heterónimos (*dona/home del temps, dona/home de la neteja*).

⁶⁷¹ Por ejemplo, Galloso (2003a, 2003b) utiliza la barra para señalar la variación con respecto al género en este centro; López Meirama (2008) lematiza en la opción más frecuente y también utiliza la barra para señalar los vocablos que se consignan con variación de género. Otros autores (como Hernández Muñoz 2006 o Bartol 2004), como ya señalamos al inicio de este apartado, utilizan el mismo método para señalar la variación según este morfema en todos los centros de interés.

⁶⁷² Por tanto, nuestra solución sigue la línea de Carcedo (2001), Gomez Molina y Gómez Devís (2004) o Arnal *et al.* (2004).

⁶⁷³ Secundamos el criterio de la variación en cada corpus de forma independiente. Esto explica que en castellano se registre “ama de casa” o *mainadera*, por ser estas las únicas opciones que aparecen, y en cambio, en catalán se incluya *amo de casa* y *mainader*, toda vez que aparecen las dos formas. De modo contrario, en castellano unificamos “costurero/a” bajo la forma masculina, por entender que no hay diferencia semántica entre las profesiones, pero en catalán mantenemos *costurera*, por ser la única opción presente. Un caso interesante es el del catalán *pediatra* que transcribimos en femenino porque todos los informantes que lo mencionan lo incorporan así, no sabemos —y resultaría imposible averiguarlo— si, en este caso concreto, se da un error de grafía, ya que es frecuente en Lleida la pronunciación [ε] de toda —a final, o todos los informantes actualizan el término en femenino. Aunque nos parece más probable la primera opción, también es cierto que en el caso similar de *psiquiatre/a*, las dos opciones aparecen en los listados. La preferencia de la voz femenina también puede deberse a la influencia de la forma castellana. De otra parte, la presencia de “escaladora” se debe a la fama de Araceli Segarra, alpinista leridana de gran renombre.

De otra parte, hemos conservado el morfema de plural, como es lógico, en las formas compuestas (“guardabosques” / *guardaboscós*, (“guardaespaldas” / *guardaespatlles*, “trotamundos”), y también en otras voces cuyo significado se distorsionaría al incluirlas en singular (“botones”, “ciencias ambientales”/*ciències ambientals*, “empresariales”, “fuerzas armadas”, “obras públicas”, “relaciones públicas”/*relacions públiques*). Por otro lado, hemos recurrido al uso del paréntesis en: “(agente) comercial”, “dirección (y administración) de empresas”, “(guardia) urbano”/*(guardia) urbà*, *mosso (d’esquadra)* e “ingeniero de teleco(municacione)s”/*enginyer de teleco(municacion)s*. Hemos utilizado el corchete en “[agente] inmobiliario” y “[policía] secreta”, dos casos en los que nos parecía imprescindible aportar el núcleo para la correcta comprensión del sintagma⁶⁷⁴.

Con respecto a las asociaciones recabadas, hemos eliminado únicamente algunos nombres propios (“Boris”, “Aznar”), y mantenemos —como la mayoría de autores— tanto las denominaciones de carreras y estudios como las de lugares donde se llevan a cabo determinadas profesiones (“cocina”, “correos”, “hotel”, “juzgado”...). También hemos recabado algunos verbos (“limpiar”, *netejar* y *robar*) y el sintagma “sus labores”, de cuya presencia también dan cuenta Carcedo (2001) y Arnal *et al.* (2004) en sus respectivos corpus. Al igual que nuestros informantes, hemos entendido el campo de las profesiones en sentido amplio, lo que justifica la inclusión de numerosas actividades, aunque algunas de ellas no gocen de prestigio social (“asesino”, “bohemio”, “camello”, “chapero”, “curandero”, “famoso”, “fantasma”, “héroe”, “hombre bala”, “liante”, “moroso”, “rey”, “terrorista”, “vividora”...).

5.4.3.17. Colores/Colors

El último de los centros de interés es, sin duda, el que menos problemas de edición plantea, tal como se señala en algunas de las pocas investigaciones que lo incluyen (Arnal *et al.* 2004: 47, López Meirama 2008: 86). En primer lugar, cabe

⁶⁷⁴ Por el contrario, hemos respetado *comercial*, que aparece en catalán sin el sustantivo *agent*, toda vez que en el uso es frecuente oír frases del tipo “trabaja como comercial” pero que no serían posibles en el caso de “inmobiliario”, transcrito como “[agente] inmobiliario”, como ya hemos indicado. A pesar de que contamos en el corpus con la lexía “promotor inmobiliario”, no creemos que los informantes se refirieran a esta combinación al consignar solo el adjetivo, prueba de ello es la aparición en catalán de *agent immobiliari*. Hemos mantenido, asimismo, “seguridad” tal cual, porque en este caso contamos con los sintagmas “guardia de seguridad” y “vigilante de seguridad”, lo que nos impide unificarlo.

señalar que —al igual que hacen Gómez Molina y Gómez Devís (2004: 60) y Llopis (2010: 31-32)— hemos corregido la frecuente confusión entre “fluorescente”/“fosforescente” y *florescent/fluorescent/fosforescent*, resuelta a favor del último término en ambas lenguas. El léxico recabado en torno a este estímulo es muy homogéneo, con lo que no se detectan apenas extranjerismos, dialectalismos o marcas comerciales. Las únicas excepciones son, en el corpus catalán, *atmetlla* y *carbassa/carabassa*, pareja que unificamos —siguiendo el mismo criterio de frecuencia utilizado de forma reiterada en el presente apartado— en *carbassa* y la aparición de la marca “camel”, habitualmente usada para denominar un tipo de color concreto.

Al igual que Arnal *et al.* (2004: 47), hemos editado como entradas independientes algunas combinaciones de colores como *blanc verd* o *verd groc* caracterizadas por su presencia escasísima en el corpus. Al contrario, otras como *blau verd* o “rosa fucsia” se han conservado tal cual, porque así lo aconsejaba su porcentaje de aparición. Con respecto a las variantes, son numerosas las formas derivadas que aparecen (“amarillento”, “anaranjado”, “azulado”, “azulón”, “grisáceo”, *ataronjat*, *blavós*, *grogueuc*, *rosat*, *vermellós...*), e incluso no son pocas, en este sentido, las creaciones de los informantes que, por ende, no aparecen en los diccionarios generales (“amarinado”, “avainillado”, “verdi azul”, *agranatat*, *granatós*, *lilós*, *taronjós...*). De otra parte, son muy abundantes las ocasiones en que hemos recurrido al uso del paréntesis en este centro de interés, siguiendo el criterio que hemos utilizado a lo largo de este capítulo: “(azul) aguamarina”, “(azul) celeste”, “(azul) cian”, “(gris) perla”, “(gris) piedra”, “(rojo) vino”, “(verde) aceituna”, “(verde) caqui”, “(verde) esmeralda”, “(verde) manzana”, “(verde) militar”, “(verde) oliva”, “(verde) pistacho”, *(blau) cel*, *(blau) celeste*, *(blau) cian*, *(blau) marí*, *(gris) perla*, *(groc) llimó*, *(marró) caca d’oca*, *(marró) xocolata*, *(verd) caqui*, *(verd) oliva*, *(verd) pistatxo*, *(verd) poma*. Algunas veces, la unificación mediante este recurso no ha sido posible, toda vez que se hallaba más de un núcleo con el mismo adyacente (“amarillo limón”/“limón”/“verde limón”, “azul turquesa”/ “turquesa”/ “verde turquesa”, *blau turquesa/turquesa/verd turquesa*)⁶⁷⁵.

⁶⁷⁵En alguna otra ocasión aislada tampoco hemos utilizado este recurso por entender que no aparecía junto con el color básico al que normalmente acompaña (*negre violí/violí*).

La uniformidad de este centro de interés también se aprecia en el tipo de léxico evocado. La inmensa mayoría de informantes actualiza colores primarios y distintas subespecificaciones. También se utilizan sustantivos, cuyo significado en principio no contiene ninguna alusión al color, para designar distintos matices cromáticos (“calabaza”, “ceniza”, “ciruela”, *arena, borgonya, mostassa, vi, xampany...*). Por último, cabe mencionar que se recogen adjetivos, que nombran cualidades del color en general (“alegre”, “brillante”, “cálido”, “relajante”...), así como algunos sustantivos, donde se evidencian algunas relaciones secundarias, algo más alejadas ya, del campo semántico estricto de los colores (*aquarel·la, arc de Sant Martí, dibuix, pintor o pintura*).

Tras esta extensa exposición sobre las características de la investigación que hemos llevado a cabo y sobre las decisiones que hemos tomado con respecto a la edición de los dos corpus, abordamos —en el siguiente capítulo— el análisis de los datos en cada una de las dos lenguas. Estos se referirán a la covariación de los índices cuantitativos tanto con las variables lingüísticas analizadas, esto es, con los distintos centros de interés (variables dependientes), como con las variables sociales escogidas (variables independientes).

6. ANÁLISIS DE LOS DATOS

En el presente capítulo acometemos el análisis de los resultados obtenidos en nuestra investigación sobre la disponibilidad léxica en castellano y en catalán en la provincia de Lleida. En la actualidad, son muchas ya las investigaciones realizadas dentro de este ámbito de la léxico-estadística y, gracias a ellas, hoy en día no solo se conocen los índices que mejor describen la productividad léxica de un conjunto de informantes, sino que además sabemos cuáles de ellos influyen a su vez sobre otros, cuáles son más estables —y resultan, por tanto, idóneos en las comparaciones interdialectales—, y cuáles —por el contrario— se ven influidos en mayor medida por las variables propias de cada investigación, como el número de informantes que la componen.

Sin embargo, a nuestro modo de ver, el principal avance realizado hasta la fecha en el análisis de la disponibilidad es el uso de programas estadísticos que corroboran el alcance de los resultados obtenidos. Desde Galloso (2003a) —pionera en el uso del SPSS en los estudios de léxico disponible— son muchas las investigaciones que han sometido sus datos a paquetes estadísticos (*v.gr.* Gómez Molina y Gómez Devís 2004, Llopis 2009, López Meirama 2008, Hernández Muñoz 2006, Fernández Smith *et al.* 2008, Areta 2009, etc.) y la nómina de trabajos tiende a aumentar en la actualidad⁶⁷⁶. A estas alturas, parece evidente que el análisis del léxico disponible no puede limitarse, por ejemplo, a la descripción de las diferencias halladas entre los distintos subgrupos sociales, sino que debe comprobarse, de alguna manera, que dichas diferencias no se deben al más puro azar, tal y como se ha venido repitiendo acertadamente (Hernández Muñoz 2004: 48). Dicho de otro modo, antes de abordar la causa que puede llevar a que los informantes de un estrato social, por ejemplo, ostenten un promedio de palabras más alto que los de otro estrato, debe primero hacerse efectiva la comprobación de que dichas diferencias se darían también en el caso de haber tomado otros informantes, y que, por ende, los resultados son, efectivamente, extrapolables al conjunto del universo estudiado.

⁶⁷⁶ Se ha de tener en cuenta que no todas los trabajos hacen el mismo tipo de análisis ni lo extienden a todos los datos. Galloso (2003a) y López Meirama (2008) basan su análisis estadístico en los promedios generales obtenidos en cada variable social. Las restantes investigaciones citadas, además de los promedios, estudian también las diferencias, con respecto a cada variable social, en cada uno de los centros de interés.

En el presente trabajo, hemos utilizado también el programa SPSS (*Statistical Package for the Social Sciences*), en su versión 21.0.0, para determinar la influencia que tienen las distintas variables lingüísticas —los centros de interés— y las cinco variables sociales elegidas sobre la productividad léxica de los informantes en lengua castellana y catalana. Partiendo del análisis de las palabras que enumera cada informante bajo cada estímulo verbal, se han utilizado distintas técnicas estadísticas para observar si las diferencias en la producción en cada una de las dos lenguas son significativas, así como para constatar si lo son también las halladas entre las medias de cada uno de los subgrupos sociales en todas las áreas semánticas y en cada una de las dos lenguas (*vid.* 6.1.3. y 6.2.).

Como ya indicamos (*vid.* 5.1.), la muestra es representativa de la población analizada —el alumnado de 2º de bachillerato de la provincia de Lleida— y el nivel de confianza establecido para los cálculos estadísticos es del 95%, como acostumbra a ser habitual en este tipo de trabajos. La regla general de decisión se basa en el valor obtenido (*p-value*) que permite al 5% de significación refutar o aceptar la hipótesis nula tratada en cada prueba estadística. El procedimiento seguido ha sido el siguiente: con el objetivo de comparar muestras independientes —las submuestras sociales para un conjunto de variables— se ha realizado la prueba de homogeneidad de varianzas, y, en caso de que dicha igualdad se diera, hemos sometido los datos al cálculo con ANOVA (análisis de la varianza). En caso contrario, hemos analizado los resultados mediante una prueba no paramétrica denominada KRUSKAL-WALLIS (Kruskal y Wallis 1952). Con la finalidad de no entorpecer la lectura, todos los análisis realizados se hallan debidamente clasificados y ordenados en el tercer apéndice, salvo algunas excepciones que, por su importancia, se muestran en el texto principal⁶⁷⁷.

Por otro lado, vale la pena subrayar que las características tan peculiares de nuestra investigación —donde los mismos informantes han completado la encuesta en las dos lenguas— ofrecen muchas ventajas en el análisis que realizamos a continuación. En primer lugar, el número de informantes, parámetro que tantas veces distorsiona los

⁶⁷⁷ Para comprobar la homogeneidad de varianzas hemos solicitado al programa que calculara el estadístico de Levene (*vid.* apéndice 3). Agradecemos desde aquí, de nuevo, la ayuda inestimable de la profesora M^a Àngels Cabasés en la programación y el análisis estadístico de los datos.

resultados e impide que algunos índices sean comparables entre sí, es el mismo en los dos inventarios léxicos y, además, todos los subgrupos que se derivan de la muestra general (el número de mujeres catalanohablantes, por poner solo un ejemplo) coinciden numéricamente en los dos corpus lingüísticos, lo que amplía las posibilidades de comparación y análisis⁶⁷⁸.

Otro de los inconvenientes que influye generalmente en los cotejos entre distintas investigaciones, aunque desconocemos en qué grado, es el relativo a los criterios de edición o, dicho de otro modo, a la cantidad de vocablos que el investigador elimina tras realizar el proceso de “depuración” de los datos en bruto⁶⁷⁹. Dejando de lado, pues, la intervención que haya podido realizar en cada uno de sus corpus cada investigador y cómo esta pueda influir en las comparaciones interdialectales que se realicen, en nuestro caso resulta claro que este factor no cobra importancia, por cuanto los criterios de edición —explicados con todo detalle en el capítulo anterior (*vid.* 5.4.)— se han aplicado exactamente de la misma forma en las dos lenguas estudiadas. Esta es, por tanto, y a nuestro modo de ver, otra de las principales ventajas del análisis comparado de nuestro estudio.

Hechas ya las salvedades oportunas, a continuación (6.1.) cotejamos los índices cuantitativos que alcanzan, en distintos parámetros, las dos lenguas estudiadas y comparamos estos resultados con los obtenidos en otros trabajos realizados con anterioridad. En un segundo apartado (6.2.), estudiamos la incidencia de las variables sociales elegidas sobre la producción léxica en cada una de las dos lenguas. Sobra decir que, dada la ingente cantidad de datos obtenida en esta investigación, sería posible realizar muchos análisis distintos. En este capítulo, y en el siguiente, nos centramos en los resultados más importantes adoptando una perspectiva eminentemente sociolingüística. Relegamos pues, para el futuro otro tipo de enfoques.

⁶⁷⁸ Los índices que se ven influidos por el número de informantes son los totales de palabras y de vocablos, los vocablos concretos de un grupo social e, incluso, el índice de cohesión de todos los subgrupos. Hernández Muñoz (2004: 63) afirma, al respecto: “Éste último [el número de sujetos] condiciona todos los valores numéricos, sobre todo el índice de cohesión”.

⁶⁷⁹ Sería deseable valorar en algún momento esa información, puesto que “el panorama léxico a evaluar sería mucho más fidedigno” (Hernández Muñoz 2004: 53).

6.1. RESULTADOS GLOBALES DE LAS VARIABLES LINGÜÍSTICAS

Antes de analizar los resultados globales en las dos lenguas, hemos de aclarar la distinción que realizamos a lo largo de todo este trabajo, entre *palabra*, de un lado, y *vocablo*, del otro. A pesar de que esta terminología es clásica y habitual en los estudios de disponibilidad —y en los de léxico-estadística, en general— puede que presente problemas a un lector no especializado en el tema, motivo por el cual algunos investigadores, como Martínez Olmos (2008: 152), prefieren sustituirla por la dicotomía entre *tipos léxicos diferentes* (*types*, esto es “vocablos”) y *total de ocurrencias* (*tokens*, “palabras”).

Así pues, es necesario explicitar que por *palabra* entendemos “toda actualización repetida o no de unidades léxicas que se registran en las respuestas de los informantes”, mientras que el término *vocablo* designa “cada una de las entradas léxicas correspondientes pero únicas que compendian dichas actualizaciones y que forman parte de un diccionario de disponibilidad, donde se constatan” (Fernández Smith *et al.* 2008: 65). Dicho de otro modo, ante un mismo estímulo verbal —por ejemplo, *El cuerpo humano*— cada informante puede escribir una retahíla de términos (“ojo”, “mano”, “dedo”, etc.) que conformarán la cantidad de *palabras* que dicho hablante actualice. Con el término *vocablo*, en cambio, se alude a las palabras *diferentes* que los informantes actualizan en conjunto bajo ese mismo estímulo. Sobra decir que si evaluamos un solo informante, el número de palabras y de vocablos coincide.

Como podrá observarse en las tablas que aparecen más abajo (tablas 6.1 y 6.2), el total de vocablos en ambas lenguas es siempre inferior al total de palabras recabadas y, concretamente, supone, aproximadamente, un 9% de esa magnitud, tal como se aprecia en la tabla 6.2. A continuación, en el examen de cada centro de interés (*vid.* 6.1.1.), veremos cómo el número de palabras recogidas en cada área temática será, asimismo, superior a la cifra de vocablos, toda vez que la actuación lingüística de los informantes tiende a ser similar y las coincidencias en las asociaciones que se realizan a partir del enunciado son abundantes, como puede verse en los diccionarios finales (*vid.*

capítulos 10 y 11)⁶⁸⁰. Otro parámetro que estudiaremos, el índice de cohesión, analiza la relación entre las palabras y los vocablos en un mismo centro de interés (*vid.* 6.1.5.).

Es necesario y conveniente distinguir con claridad no solo entre *palabra* y *vocablo*, sino también entre lo que miden estas dos magnitudes. Con respecto a las palabras, el valor de este índice es de suma importancia puesto que evalúa la *productividad léxica* del hablante, obtenida, claro está, a partir de la metodología “artificial” de la enumeración controlada de unidades léxicas⁶⁸¹. En concreto, el promedio de palabras por sujeto (resultado de dividir el total de palabras de un centro de interés entre el número de informantes que componen la muestra) se ha reconocido, en reiteradas ocasiones, como el índice más estable y adecuado para realizar comparaciones dentro de un mismo estudio y entre distintas investigaciones.

De otra parte, el número de vocablos nos informa sobre la *riqueza léxica* de un determinado centro de interés. A veces se han utilizado estos dos parámetros —*riqueza* y *productividad*— como sinónimos, por lo que creemos necesario distinguirlos con claridad en este trabajo⁶⁸². La *productividad* se relaciona, como acabamos de ver, con el número de palabras que emite cada informante en cada centro de interés, o a lo largo de toda la encuesta. En cambio, la *riqueza* léxica depende del número de vocablos, es decir, de las palabras distintas que se recogen bajo cada estímulo verbal. Por su propia naturaleza, este último índice no puede utilizarse para medir aspectos individuales, ya que ofrece información acerca del comportamiento global de un grupo de informantes ante una misma área léxica. Un inconveniente importante que presenta es que puede alterarse fácilmente por el comportamiento de un solo sujeto atípico que, ante un estímulo verbal y por distintas causas, dé una retahíla de voces muy alejadas de las formuladas por sus compañeros. Este comportamiento, que apenas modificaría el promedio de palabras por sujeto —o lo haría levemente según las características concretas de la actuación lingüística del informante—, provocaría, en seguida, el aumento del total de vocablos recabados en torno a esa área temática.

⁶⁸⁰ A pesar de ello, también se observan diferencias, claro está. No en vano numerosos estudios de la disponibilidad dedican páginas al cotejo de los vocablos más disponibles, estableciendo distintos cortes para ello (las 10, 20 ó 50 primeras palabras) o analizan la compatibilidad entre distintos subgrupos.

⁶⁸¹ Hernández Muñoz (2006: 320) afirma: “que un estudiante aporte más palabras que otro no significa necesariamente que conozca más palabras que el que ha aportado menos términos, sino que ha desarrollado unas estrategias más efectivas para producir mayor número de palabras”.

⁶⁸² Es muy interesante, al respecto, la distinción que realiza Hernández Muñoz (2004: 52-53) entre los términos “disponibilidad” y “riqueza” léxicas, y su empleo inadecuado en los análisis.

Como dijimos en la introducción a este apartado, el hecho de contar con una sola muestra posibilita la comparación entre las cifras totales, a las que dedicamos las siguientes tablas y gráficos.

Tabla 6.1. Índices cuantitativos generales en las dos lenguas con respecto al número de palabras.

Total de palabras		Promedio de palabras por informante		Promedio de palabras por centro de interés		Promedio de palabras por informante y centro de interés	
Castellano	Catalán	Castellano	Catalán	Castellano	Catalán	Castellano	Catalán
73038	76061	304.32	316.92	4296.35	4474.18	17.90	18.64

Gráfico 6.1. Total de palabras en castellano y en catalán.

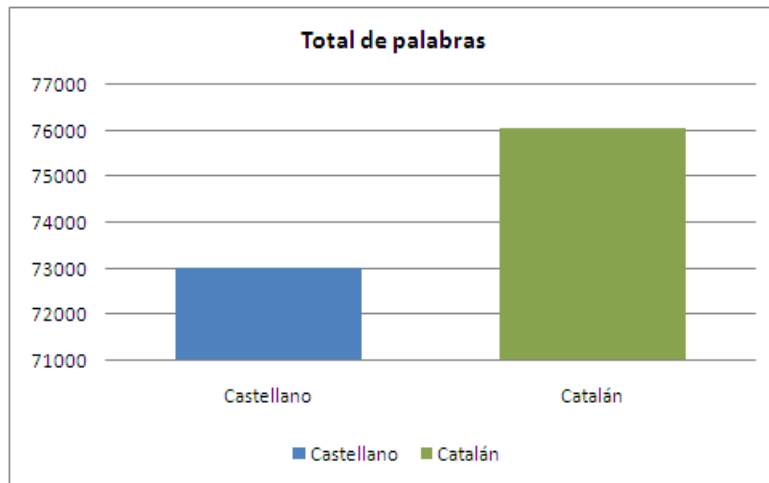
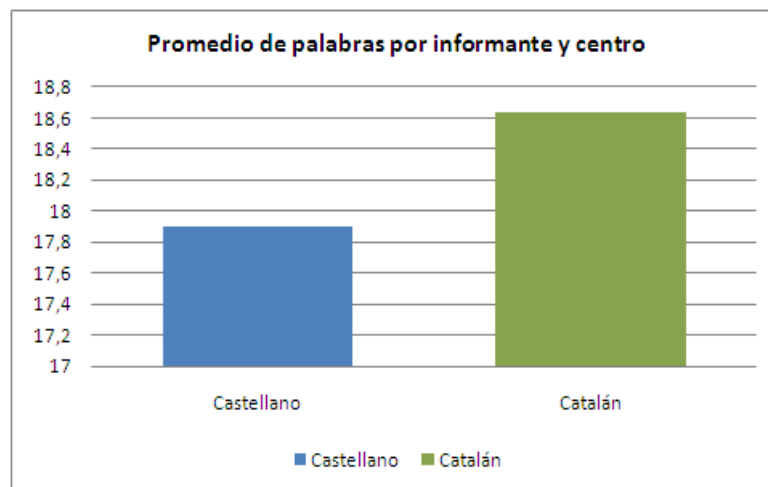


Gráfico 6.2. Promedio de palabras por informante y centro de interés en castellano y en catalán.



En la tabla y en los gráficos anteriores, se recoge el total de palabras recabadas en las dos lenguas. Se aprecia, con claridad, cómo el catalán sobrepasa al castellano y, concretamente, se obtienen en lengua autóctona 3023 palabras más (73038 en castellano frente a 76061, en catalán). En las restantes magnitudes, los promedios de palabras por informante o por centro de interés y por informante y centro de interés, se aprecia la misma tendencia a favor de la lengua catalana. Esto resulta lógico, puesto que estos parámetros derivan siempre del total de palabras⁶⁸³.

Además de la abrumadora diferencia en los totales, el catalán sobrepasa en casi 178 palabras (177.83) en el promedio de palabras por centro de interés. Las diferencias, no obstante, se reducen si observamos la media de palabras por informante a lo largo de toda la encuesta (304.32 en castellano, frente a 316.92 en catalán) y constatamos cómo en lengua catalana se recogen solo 12.6 palabras más. La distancia es aún menor en los promedios por informante y centro de interés y se sitúa en 0.74, esto es, no alcanza una respuesta de diferencia (17.90 en castellano y 18.64 en catalán).

Sin embargo, hay que reconocer el absoluto predominio del catalán en los totales de palabras y en los parámetros que derivan de esta magnitud. Como ya hemos explicado, este índice sirve para medir la productividad léxica de los informantes. Por tanto, ante los datos globales obtenidos, no podemos más que afirmar que, en general, la muestra presenta mejores resultados, en cuanto a la *producción léxica*, en lengua catalana que en castellana. Este hecho no deja de ser lógico toda vez que, como vimos ampliamente en el capítulo cuarto (*vid.* 4.3.2), la provincia de Lleida presenta unos índices de uso del catalán muy elevados respecto a otras provincias catalanas. Además, el porcentaje de población autóctona (nacida en la misma provincia e incluso comarca) es también alto, circunstancia que influye en el mantenimiento de la lengua vernácula y en su uso en el hogar. En última instancia, la muestra es un reflejo de esta realidad, ya que el 65.42% de los informantes que la componen tienen como lengua materna el catalán (*vid.* 5.2).

⁶⁸³ Son resultado de dividir esta cifra entre los informantes de la investigación, 240, entre los centros de interés, 17 y esta última cifra, a su vez, de nuevo entre los informantes.

Tabla 6.2. Índices cuantitativos generales en las dos lenguas con respecto al número de vocablos.

Total de vocablos		Promedio de vocablos por centro de interés		Porcentaje de vocablos sobre el total de palabras	
Castellano	Catalán	Castellano	Catalán	Castellano	Catalán
6824	7064	401.41	415.53	9.34%	9.29%

Gráfico 6.3. Total de vocablos en castellano y en catalán.

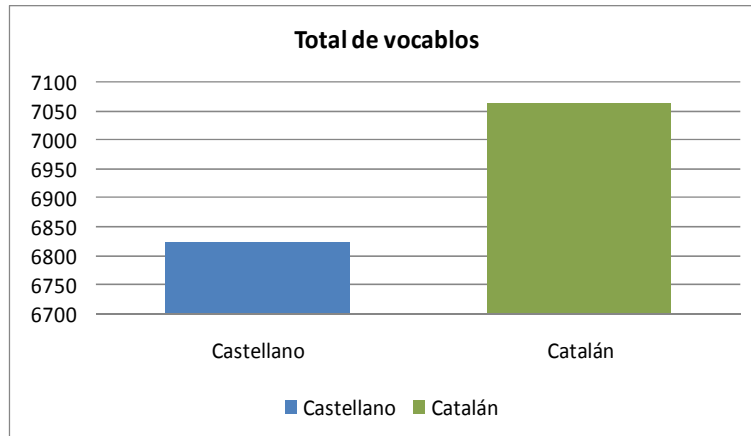
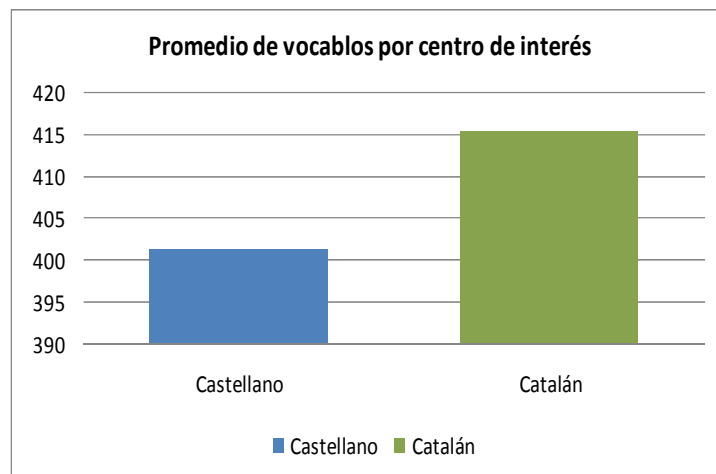


Gráfico 6.4. Promedio de vocablos por centro de interés en castellano y en catalán



Con respecto al número total de vocablos, o de palabras distintas emitidas, el catalán también descuella, aunque las diferencias cuantitativas que se observan esta vez son menores. Como ya explicamos, este parámetro puede alterarse por individuos cuyo comportamiento lingüístico escape a los patrones habituales. Podría argumentarse, no obstante, que, al tratarse de la misma muestra, un mismo comportamiento atípico por parte de un individuo tenderá a darse en las dos lenguas; sin embargo, ni siquiera

podemos estar seguros de un hecho así: todo lo contrario, podríamos hallar montones de razones por las cuales un informante actualizara vocablos dispares o insospechados en solo una de las dos lenguas, la que le fuera más familiar o la que manejara mejor en ese ámbito⁶⁸⁴. Sin profundizar en esta cuestión, está claro que los centros de interés contienen más vocablos en catalán que en castellano (7064 frente a 6824, respectivamente; esto es, 240 vocablos más), al menos en el análisis global de las cifras, lo que da de media poco más de 14 vocablos por centro de interés. Constatamos, pues, que, de igual modo que sucedía con la producción de palabras, el número de vocablos recabados es también superior en lengua catalana, lo que determina una mayor *riqueza léxica* de los categorías semánticas en esta lengua⁶⁸⁵. Es decir, no solo la producción (el número de palabras actualizadas bajo cada enunciado) es superior en catalán, sino que también lo es el número de asociaciones, o vocablos evocados, a partir de un mismo estímulo verbal. El castellano solo aventaja al catalán en la proporción que suponen los vocablos sobre el total palabras, aunque la cifra es realmente cercana en las dos lenguas (9.34% en castellano y 9.29% en catalán).

En la introducción a este trabajo, partimos de la hipótesis de que la producción y la riqueza léxicas de los informantes serían mejores en lengua catalana, por ser esta la lengua materna predominante en el universo estudiado —y por tanto, en la muestra concreta analizada— así como la lengua en que han sido escolarizados todos nuestros informantes. Es el momento de confirmar que esta hipótesis, como acabamos de ver, se cumple con respecto a los totales de palabras y de vocablos.

6.1.1. Total de palabras por centro de interés

En los estudios de disponibilidad léxica, los centros de interés se consideran un tipo especial de variables lingüísticas que contienen distintas variantes, que son las palabras que se aportan al respecto⁶⁸⁶. De hecho, la producción del informante varía más según el área nocional ante la que se halle, que en función de algunas de las variables sociales analizadas, como veremos a lo largo de todo este capítulo. La diferencia entre

⁶⁸⁴ Por ejemplo, un alumno cuya lengua materna fuera el castellano y acostumbrara a pasar las vacaciones con sus abuelos, también castellanohablantes, en un pueblo de La Mancha, podría realizar un sinfín de asociaciones bajo el centro *El campo*, distintas a las de sus compañeros. Dichas asociaciones, no tendrían correspondencia en la misma encuesta completada por el alumno en lengua catalana.

⁶⁸⁵ Recordemos en este punto al lector la total convergencia entre los criterios de edición de ambas lenguas. Se han desechado o aceptado las mismas asociaciones, y se ha adoptado el mismo criterio respecto a las marcas comerciales o a los nombres propios (*vid.* 5.4).

⁶⁸⁶ Martínez Olmos (2008) y Gómez Molina y Gómez Devís (2004) también las denominan así.

estos dos tipos de factores (las variables lingüísticas, esto es, los centros de interés, y las sociales como el *sexo* o el *nivel sociocultural*) radica en que, mientras las segundas muestran un comportamiento dispar en función de la comunidad que se analice —por ejemplo, el *sexo* influye en algunos lugares y en otros no, e incluso cuando lo hace es de distinto modo—, las diferencias que se observan según las variables lingüísticas tienen una gran estabilidad por encima de la sintopía que estemos analizando y muestran grandes coincidencias en todas las provincias españolas en las que se han desarrollado estos estudios, como ya han notado diversos autores con anterioridad (Samper, Bellón y Samper Hernández 2003, Samper Padilla y Samper Hernández 2006)⁶⁸⁷.

En nuestra investigación, también se aprecian diferencias notables en la producción en función de las variables lingüísticas, es decir, según el ámbito de la realidad que se suscite, tal como se observa en la siguiente tabla⁶⁸⁸:

Tabla 6.3. Total de palabras por centro de interés en castellano y en catalán⁶⁸⁹.

Nº	Centro de Interés	Castellano	Catalán
1	El cuerpo humano	6129	6924
2	Ropa y complementos	5189	5566
3	Partes de la casa (sin los muebles)	3400	3503
4	Muebles de la casa	2975	3072
5	Comidas y bebidas	5811	6357
6	Objetos colocados en la mesa para la comida	3169	3325
7	La cocina y sus utensilios	3691	3954
8	La escuela (muebles y materiales)	5080	5310
9	Calefacción, iluminación y formas de airear un recinto	2660	2652
10	La ciudad	5124	5100
11	El campo	4369	4414
12	Medios de transporte	4043	4102
13	Trabajos del campo y del jardín	2125	2332
14	Animales	6245	6317
15	Juegos y distracciones	3914	3771
16	Profesiones y oficios	4709	4888
17	Colores	4405	4474

⁶⁸⁷ Por dicha razón, Hernández Muñoz (2004: 63) apunta a que el promedio de palabras por sujeto depende antes de la composición interna de cada centro de interés que de las características sociales de cada informante. Volveremos más adelante sobre esta cuestión.

⁶⁸⁸ Relegamos para el siguiente apartado el análisis de las diferencias que se observan entre la producción en las dos lenguas, para el cual utilizaremos tanto los totales como los promedios de palabras según cada centro de interés.

⁶⁸⁹ Como resulta lógico, en la encuesta en catalán los nombres de los centros de interés estaban en dicha lengua (*vid.* 5.3.). Para facilitar la lectura, y evitar redundancias, citamos en el presente capítulo todos los centros de interés en castellano.

En castellano, el centro más productivo es *Animales*, que aporta 6245 palabras, mientras que, en catalán, la mayor posición la ostenta el campo *El cuerpo humano* con 6924. Tras esta posición, siguen, en lengua castellana, *El cuerpo humano* (6129) y *Comidas y bebidas* (5811). En catalán, tras el primero de los centros de interés, destacan *Comidas y bebidas* (6357) y *Animales* (6317). Es digno de mención, en definitiva, el predominio de estas tres áreas léxicas (1, 14 y 5) que, aunque en distinto orden, ostentan los rangos de mayor productividad en ambas lenguas (*vid.* tabla 6.4, *infra*). En las últimas posiciones, las coincidencias entre los centros más improductivos son aún más destacables. Los campos *Muebles de la casa*, con 2975 palabras en castellano y 3072 en catalán, *Calefacción e iluminación y formas de airear un recinto* con 2660 y 2652, respectivamente, y, finalmente, *Trabajos del campo y del jardín*, con 2125 y 2332, ocupan los últimos lugares en cuanto a los totales de palabras recogidas. Si echamos un vistazo a las cifras globales de palabras, los tres centros que alcanzan un mayor número en ambas lenguas (1, 14 y 5), sobrepasan las 6000 unidades en lengua catalana —de hecho, *El cuerpo humano* casi alcanza las 7000—, mientras que en castellano solo los dos primeros, *Animales* y *El cuerpo humano*, alcanzan esta cifra. Por la parte inferior, solo las dos últimas áreas (*Calefacción, iluminación y formas de airear un recinto* y *Trabajos del campo y del jardín*) no superan las 3000 palabras en catalán; en castellano, en cambio, los tres últimos centros (4, 9 y 13) ni siquiera llegan a esta cifra.

De otra parte, la media respectiva en cada lengua se sitúa en algo más de 4000 palabras (4296.35 palabras en castellano y en 4474.18 en catalán, *vid.* tabla 6.1). Por encima de este valor medio, se posicionan en ambas lenguas: *El cuerpo humano*, *Ropa y complementos*, *Comidas y bebidas*, *La escuela*, *La ciudad*, *Animales* y *Profesiones y oficios*; y quedan por debajo de este valor: *Partes de la casa*, *Muebles de la casa*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios*, *Calefacción e iluminación y formas de airear un recinto*, *Medios de transporte*, *Trabajos del campo y del jardín* y *Juegos y distracciones*⁶⁹⁰. *El campo* y *Colores* quedan por encima de la media en lengua castellana, pero por debajo, en la lengua autóctona. Este hecho se debe

⁶⁹⁰ Con el fin de agilizar la lectura, omitimos la especificación “muebles y materiales” del centro *La escuela*, y la de “sin los muebles”, de *Partes de la casa*. Recordemos que estas dos aclaraciones se hallaban en paréntesis en las encuestas originales; únicamente las obviamos en el texto, no así en las numerosas tablas que contiene el presente capítulo.

a que los centros que ocupan las primeras posiciones en el ranquin⁶⁹¹ catalán son muy productivos, mientras que en castellano se observa un reparto más proporcionado del volumen de respuestas por centro de interés (*vid.* tabla 6.4, *infra*). Cabe señalar que *Colores*, con 4474 respuestas, coincide prácticamente con la media en lengua catalana.

Por otro lado, si reparamos en la distancia mayor en la productividad de cada lengua, vemos que en castellano dicha diferencia es de 4120 palabras (las que hay entre el centro 14, el más fecundo, y el centro 13, el menos) y en catalán, de 4592 (las que hay entre el centro 1 y el 13). Por tanto, se constata una mayor diferencia global en la producción de palabras en torno a cada área léxica en lengua catalana, analizando cada lengua por separado. El mismo hecho se observa en los gráficos que incluimos a continuación (gráficos 6.5 y 6.6), si se compara el porcentaje que suponen sobre el total los centros más y menos improductivos en cada lengua⁶⁹². En ellos se plasma, de forma muy ilustrativa, el volumen que supone cada área temática en relación con la producción total. De este modo, es fácil percibir cómo *El cuerpo humano*, *Comidas y bebidas* y *Animales* alcanzan, en las dos lenguas, los mayores porcentajes (8.39 en castellano y 9.10 en catalán en el primero de los centros de interés; 7.96 y 8.36, respectivamente, el centro 5, y 8.55 y 8.31 el 14). También vale la pena mencionar la diferencia tan llamativa que hay entre la aportación que hacen las distintas áreas, como se observa en el gráfico 6.6 donde *El cuerpo humano* supone casi el 10% del total de palabras, mientras que *Trabajos del campo y del jardín* apenas aporta un 3.07%.

⁶⁹¹ Utilizamos, en este trabajo, la adaptación gráfica propuesta por el *Diccionario Panhispánico de Dudas* para este extranjerismo.

⁶⁹² Una posible causa de la mayor divergencia que se observa en el corpus catalán puede ser la diferencia entre el perfil de los dos grupos mayoritarios de la muestra según la adscripción lingüística (esto es, entre los que tienen como lengua materna el catalán y los que tienen como lengua materna el castellano). La relación que estos dos colectivos mantienen con su segunda lengua puede ocasionar que se observen mayores diferencias en la producción del catalán que en la del castellano. Confirmaremos más adelante esta hipótesis (*vid.* 6.2.5. y 6.2.6.).

Gráfico 6.5. Frecuencia relativa del total de palabras por centro de interés en castellano.

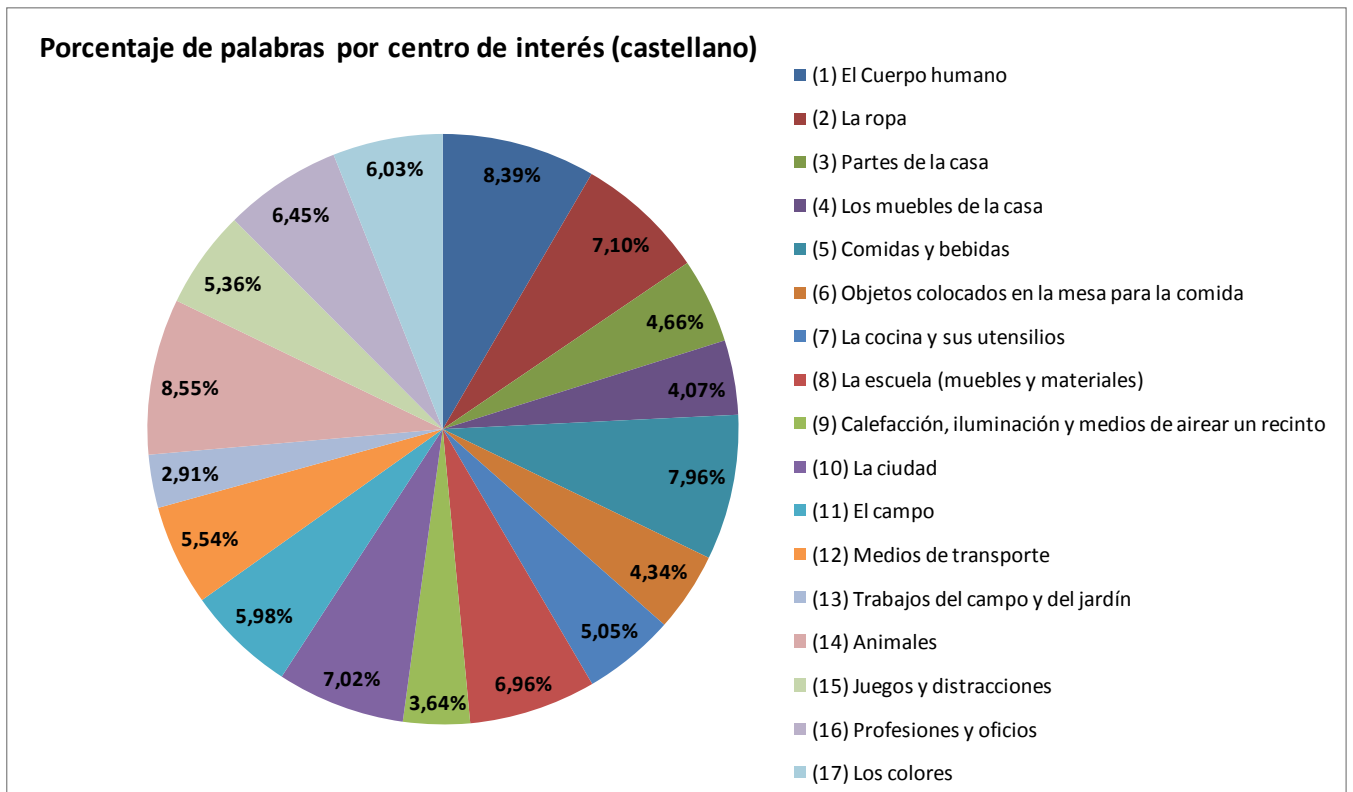
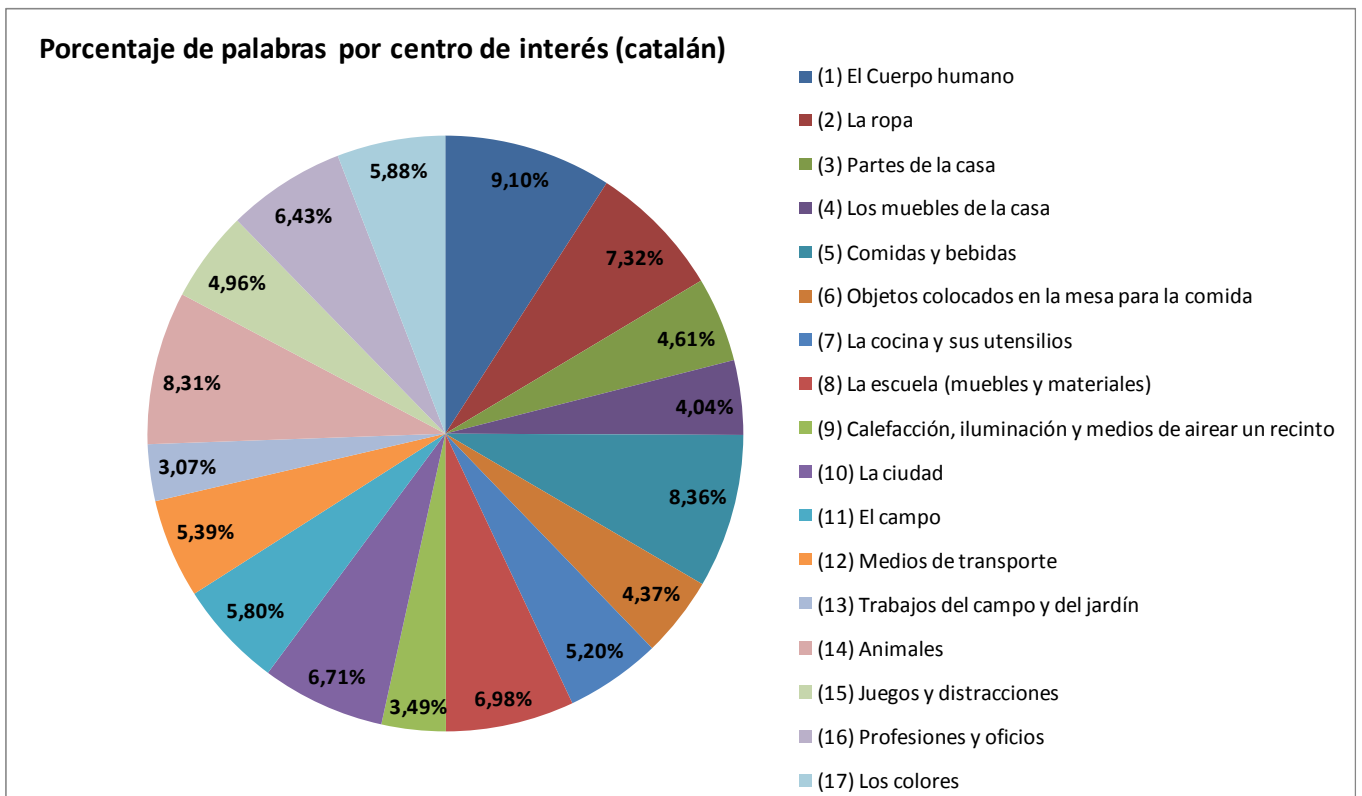


Gráfico 6.6. Frecuencia relativa del total de palabras por centro de interés en catalán.



A pesar de que, como hemos visto, los resultados globales en las dos lenguas difieren, si nos fijamos en el orden de los centros de interés según el número de palabras, las posiciones que alcanzan ambos idiomas son muy similares, tal como se aprecia en la siguiente tabla:

Tabla 6.4. Rango de los centros de interés en cada una de las lenguas según la producción de palabras⁶⁹³.

Rango	Castellano	Palabras	Catalán	Palabras
1	14. Animales	6245	1. El cuerpo humano	6924
2	1. El cuerpo humano	6129	5. Comidas y bebidas	6357
3	5. Comidas y bebidas	5811	14. Animales	6317
4	2. Ropa y complementos	5189	2. Ropa y complementos	5566
5	10. La ciudad	5124	8. La escuela	5310
6	8. La escuela	5080	10. La ciudad	5100
7	16. Profesiones y oficios	4709	16. Profesiones y oficios	4888
8	17. Colores	4405	17. Colores	4474
9	11. El campo	4369	11. El campo	4414
10	12. Medios de transporte	4043	12. Medios de transporte	4102
11	15. Juegos y distracciones	3914	7. La cocina y sus utensilios	3954
12	7. La cocina y sus utensilios	3691	15. Juegos y distracciones	3771
13	3. Partes de la casa	3400	3. Partes de la casa	3503
14	6. Objetos colocados en la mesa para la comida	3169	6. Objetos colocados en la mesa para la comida	3325
15	4. Muebles de la casa	2975	4. Muebles de la casa	3072
16	9. Calefacción, iluminación y formas de airear un recinto	2660	9. Calefacción, iluminación y formas de airear un recinto	2652
17	13. Trabajos del campo y del jardín	2125	13. Trabajos del campo y del jardín	2332

De hecho, hay casi una total coincidencia en el rango que alcanza cada centro de interés según la producción de palabras. Como ya hemos indicado, los tres centros más productivos se ordenan de un modo distinto, sin embargo, en cuarta posición coinciden castellano y catalán con *Ropa y complementos*, y también en séptima, octava, novena y décima con *Profesiones y oficios*, *Colores*, *El campo* y *Medios de transporte*. En la parte final de la tabla (de la decimotercera a la decimoséptima posición), también la coincidencia es absoluta en los centros menos productivos, como ya hemos comentado.

⁶⁹³ Las áreas que coinciden en el mismo rango en las dos lenguas se han coloreado en la tabla. Con una línea más gruesa se indica donde se sitúa la media respectiva.

Las diferencias, cuando las hay, son mínimas: en castellano *La ciudad* alcanza la quinta posición, mientras que en catalán se sitúa en la sexta, y, de modo contrario, el centro *La escuela* es el quinto en lengua autóctona y sexto, en lengua española. De modo similar, las áreas *La cocina y sus utensilios* y *Juegos y distracciones*, ambas situadas por debajo de la media en las dos lenguas, alternan las posiciones undécima y duodécima. Quizá la mayor disparidad hallada sea, pues, el ranquin de las tres áreas léxicas más productivas, a la que intentaremos encontrar una explicación más adelante.

Como acabamos de comprobar, las diferencias en la productividad de cada centro de interés son constantes en las dos lenguas, lo que indica que, por encima de la lengua en la que se completa la encuesta, los enunciados de cada área léxica influyen en la producción del informante. Además, todos los estudios anteriores, tanto los realizados sobre la lengua española como sobre la catalana, señalan las mismas tendencias (*vid.* 6.1.7.).

Se ha intentado hallar una explicación a la fuerte divergencia que se observa en la producción en función de las variables lingüísticas. Gómez Molina y Gómez Devís (2004: 74) afirman que estas divergencias son “reflejo de realidades peculiares tanto psicosociales como mediambientales”. Nosotros mismos, en un trabajo anterior (Serrano Zapata 2003b: 101), argumentamos que cabía esperar que los encuestados produjeran un mayor número de palabras alrededor de aquellas áreas temáticas que sintieran como más cercanas, es decir, que guardaran relación con sus actividades más cotidianas y habituales. Del mismo modo, también apuntamos, en aquel momento, que la mayor o menor concreción del centro de interés podía influir en el número de palabras obtenido. De forma similar, López Meirama (2008: 104) sostiene que en la productividad influye tanto el hecho de que el área léxica sea más o menos accesible para la población que conforma la muestra, como la organización interna del centro de interés, que puede ser jerarquizado o conformarse de forma más extensa, dando pie a una mayor variedad en las evocaciones.

Hernández Muñoz (2006: 127-189) aporta, en nuestra opinión, una gran luz alrededor de estas cuestiones. La investigadora manchega constata cómo los centros de interés son categorías semánticas dispares y los divide, desde una perspectiva cognitiva novedosa, según su estructura de organización interna (natural, bien definida, *ad hoc* y

radial) y sus niveles de “inclusividad” (inclusiva o relacional)⁶⁹⁴. La autora afirma, al respecto, que “quizá sean las discrepancias observadas en la naturaleza interna de los campos nocionales las causantes de esta falta de homogeneidad en la producción cuantitativa de las respuestas” (Hernández Muñoz 2006: 154)⁶⁹⁵. Además, comprueba cómo la disponibilidad se relaciona con tres variables cognitivas: la tipicidad, la familiaridad y la edad de adquisición, de forma que las palabras más típicas de una categoría, las más familiares y las más tempranamente aprendidas son más disponibles. La primera de estas características, la tipicidad, que además es la variable que mejor predice la disponibilidad, sería la responsable de que los índices cuantitativos entre las distintas investigaciones coincidieran (Hernández Muñoz 2006: 188 y 321, nota 216).

6.1.2. Promedio de palabras por centro de interés

A continuación, analizamos los promedios de palabras por informante y por centro de interés, magnitud que deriva del total de palabras⁶⁹⁶, y que se considera el índice más adecuado para realizar comparaciones interdialectales entre distintas investigaciones.

Tabla 6.5. Promedio de palabras por informante y centro de interés en castellano y en catalán.

Nº	Centro de Interés	Castellano	Catalán
1	El cuerpo humano	25.54	28.85
2	Ropa y complementos	21.62	23.19
3	Partes de la casa (sin los muebles)	14.17	14.60
4	Muebles de la casa	12.40	12.80
5	Comidas y bebidas	24.21	26.49
6	Objetos colocados en la mesa para la comida	13.20	13.85
7	La cocina y sus utensilios	15.38	16.48
8	La escuela (muebles y materiales)	21.17	22.13
9	Calefacción, iluminación y formas de airear un recinto	11.08	11.05
10	La ciudad	21.35	21.25

⁶⁹⁴ Las categorías naturales son las que responden a las características clásicas de una categoría semántica; las bien definidas, por su parte, tienen una estructura de lista cerrada. Un caso muy distinto es el de las categorías *ad hoc*, donde sus miembros se definen por ser elementos para la consecución de un objetivo concreto o las radiales, donde los conceptos se agrupan en torno a una subcategoría central que cuenta con extensiones o ramificaciones. En cuanto a los niveles de “inclusividad”, en las categorías inclusivas los elementos se relacionan con la etiqueta categorial según una relación de pertenencia estricta, mientras que esto no sucede en las categorías relacionales (Hernández Muñoz 2006: 127-156).

⁶⁹⁵ También señala que la caracterización propia de cada centro de interés es el primer condicionante de la obtención de los índices de disponibilidad de las palabras y que, además, la facilidad para evocar un concepto está motivada por elementos distintos según el campo nocional en que se encuentre (Hernández Muñoz 2006: 130 y 185).

⁶⁹⁶ Es resultado de dividir el total de palabras entre los informantes que forman parte de la investigación, esto es, 240.

11	El campo	18.20	18.39
12	Medios de transporte	16.85	17.09
13	Trabajos del campo y del jardín	8.85	9.72
14	Animales	26.02	26.32
15	Juegos y distracciones	16.31	15.71
16	Profesiones y oficios	19.62	20.37
17	Colores	18.35	18.64

Igual que ocurre con los totales de palabras, los centros más productivos en castellano son *Animales* (26.02), *El cuerpo humano* (25.54) y *Comidas y bebidas* (24.21). En torno a estas tres áreas léxicas, los informantes producen, como mínimo, 24 palabras de promedio. En lengua catalana, las categorías semánticas con mejores promedios son *El cuerpo humano* (28.85), *Comidas y bebidas* (26.49) y *Animales* (26.32), todas ellas con valores por encima de las 26 palabras. Obsérvese que incluso la tercera área con mejor promedio en lengua catalana (*Animales*, 26.32) tiene un resultado superior al que alcanza el centro más productivo en lengua castellana (también *Animales*, 26.02). De forma paralela a lo observado en los totales, los centros de interés más infecundos se repiten ahora y son los mismos para las dos lenguas: *Muebles de la casa* (12.40 de media por informante en castellano y 12.80 en catalán), *Calefacción e iluminación y formas de airear un recinto* (11.08 y 11.05, respectivamente) y *Trabajos del campo y del jardín* (8.85 y 9.72). Tal como sucede al examinar las cifras globales de palabras, la diferencia en la producción de cada centro de interés analizando cada lengua *per se* es mayor en lengua catalana, toda vez que en castellano la mayor diferencia es de 17 palabras (17.7) —las que separan de media la producción ante el estímulo 14. *Animales* (26.02) y el 13. *Trabajos del campo y del jardín* (8.85)— mientras que en catalán es de 19 palabras (19.13), las que hay de media entre el centro 1. *El cuerpo humano* (28.85) y el 13. *Trabajos del campo y del jardín* (9.72).

Pasemos ahora a analizar las interesantes diferencias que se observan al comparar los resultados según la productividad léxica en las dos lenguas, uno de los objetivos primordiales de nuestro estudio. Como ya indicamos en el apartado anterior, la productividad parece verse influida por la denominación concreta de cada centro de interés —con organizaciones internas distintas— y por la cercanía del alumnado con respecto al área en cuestión. Habrá que buscar más allá de estos dos presupuestos las causas de que la producción difiera según la lengua en que el informante completa la encuesta, dado que las variables lingüísticas son exactamente las mismas para los dos

idiomas. Para analizar las diferencias que se registran en la producción utilizaremos los datos de las tablas anteriores (tablas 6.3, 6.4 y 6.5), referidos tanto a los promedios como a los totales de palabras y a la clasificación de los centros según este parámetro. En el siguiente gráfico, basado en los datos de la tabla 6.5, puede observarse la diferencia en la producción de cada ámbito según la lengua de la encuesta y, en la siguiente tabla (6.6), incluimos los centros de interés ordenados de mayor a menor según la diferencia hallada entre la productividad de las dos lenguas. Los datos se refieren al catalán, la lengua que muestra una mayor preponderancia en los resultados cuantitativos.

Gráfico 6.7. Promedio de palabras por centro de interés en castellano y en catalán.

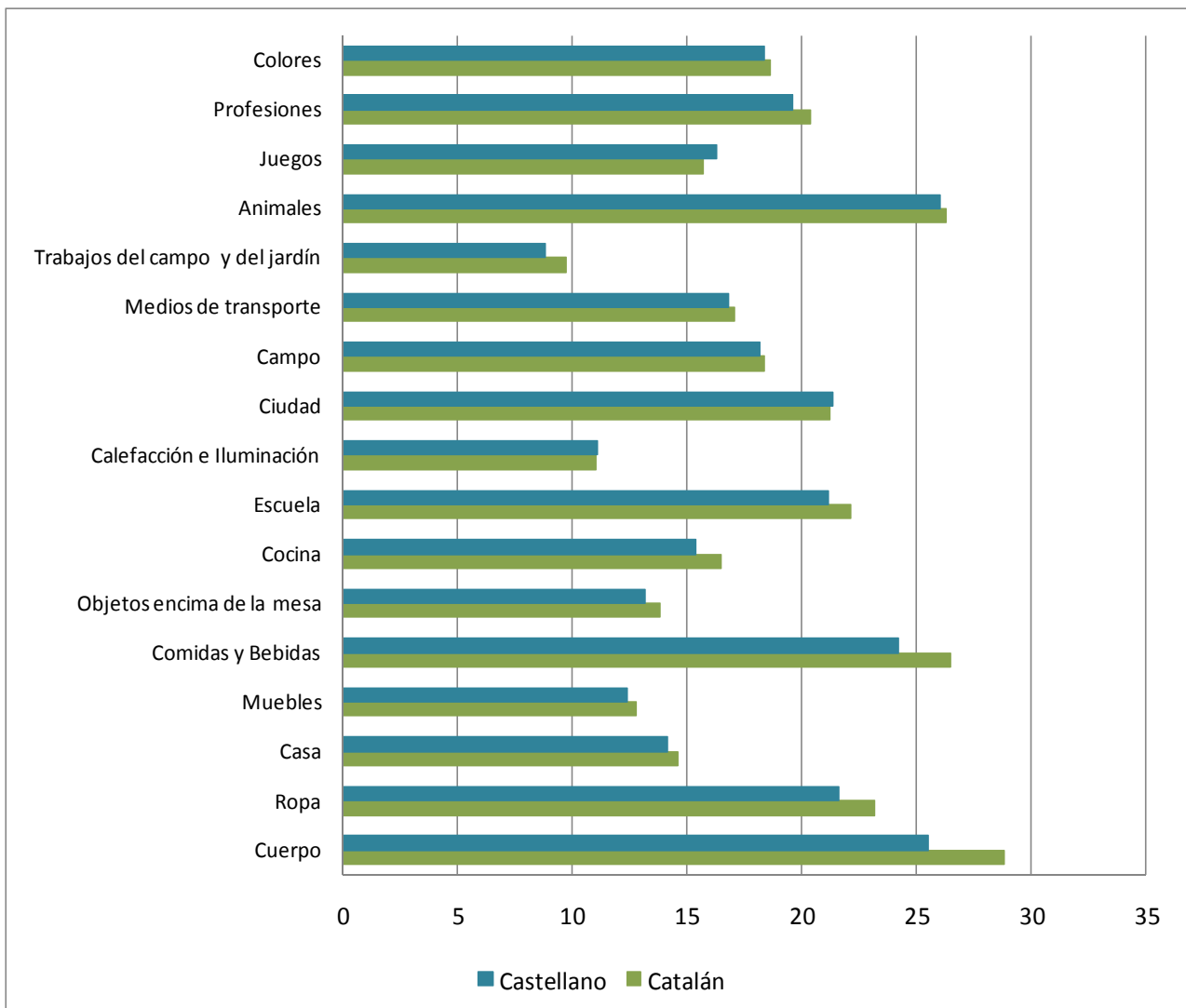


Tabla 6.6. Diferencias entre el promedio y el total de producción de las dos lenguas con respecto al catalán (de más a menos).

Nº	Centro de Interés	Diferencia Promedio	Diferencia Total
1	1. El cuerpo humano	3.31	795
2	5. Comidas y bebidas	2.28	546
3	2. Ropa y complementos	1.57	377
4	7. La cocina y sus utensilios	1.1	263
5	8. La escuela (muebles y materiales)	0.96	230
6	13. Trabajos del campo y del jardín	0.87	207
7	16. Profesiones y oficios	0.75	179
8	6. Objetos colocados en la mesa para la comida	0.65	156
9	3. Partes de la casa (sin los muebles)	0.43	103
10	4. Muebles	0.4	97
11	14. Animales	0.3	72
12	17. Colores	0.29	69
13	12. Medios de transporte	0.24	59
14	11. El campo	0.19	45
15	9. Calefacción, iluminación y formas de airear un recinto	-0.03	-8
16	10. La ciudad	-0.1	-24
17	15. Juegos y distracciones	-0.6	-143

Como se observa en el gráfico 6.7, el catalán supera al castellano en la mayoría de las áreas, concretamente en catorce de las diecisiete existentes. Solo en tres —*Calefacción, iluminación y formas de airear un recinto, La ciudad y Juegos y distracciones*— los promedios son mejores en castellano. Sin embargo, la diferencia entre las dos lenguas en dos de estos centros de interés es prácticamente mínima; en *Calefacción, iluminación y formas de airear un recinto* es de 0.03 en el promedio (11.08, en castellano y 11.05, en catalán) y en *La ciudad*, de 0.1 (21.35 y 21.25, respectivamente). Si nos fijamos en los datos relativos a la diferencia en los totales de palabras (tabla 6.6), estas diferencias se traducen únicamente en 8 palabras más en *Calefacción, iluminación y formas de airear un recinto* y 24 en *La ciudad*⁶⁹⁷.

Consideramos que un caso distinto es el que atañe a *Juegos y distracciones*, donde la distancia entre las dos lenguas en el promedio por informante, a favor del castellano, es de algo más de media palabra (0.6) y representa una diferencia de 143 voces en el total. A pesar de no ser una divergencia abismal, puede explicarse, a nuestro parecer, porque este campo léxico está relacionado con las formas de ocio adolescente

⁶⁹⁷ Recordamos, para poder valorar adecuadamente estas diferencias, que estos últimos datos dan fe de la actuación global de los 240 informantes.

donde la presencia del castellano es notable⁶⁹⁸. Si recordamos la información que expusimos en el capítulo 4 —donde analizamos el uso de las dos lenguas en distintos contextos como las relaciones interpersonales, el ámbito del consumo y el comercio y el de los medios de comunicación—, en general, el catalán tiene un predominio claro sobre el castellano en las dos demarcaciones provinciales. Sin embargo, concretamente en los medios de comunicación, la situación favorece claramente el uso del castellano en internet y muestra un cierto equilibrio entre las dos lenguas en la televisión. Sobra decir que la televisión y las nuevas tecnologías tienen una mayor vitalidad entre los jóvenes, población de nuestro estudio, que otros medios de comunicación como la radio, más favorables al uso de la lengua catalana (*vid.* 4.3.2.)⁶⁹⁹. Otras formas de ocio culturales, como el teatro o el cine, muestran un uso desigual de la lengua catalana: mientras esta tiene gran vitalidad en el teatro, su uso no está equilibrado con respecto al castellano en el cine, ya que el doblaje de películas en catalán es escaso (*vid.* 4.1). Igual que sucede con la televisión e internet, entre la franja de población de nuestro estudio, tiene más influencia el cine que el teatro.

De otra parte, los informantes también asocian este estímulo a los juegos infantiles típicos de la edad escolar. Como ya vimos en el capítulo cuarto, Galindo (2006: 397), en una amplia investigación sobre los usos lingüísticos de los escolares de primaria a la hora del recreo, es tajante al concluir que en los patios de las escuelas catalanas se habla más castellano que catalán⁷⁰⁰. Todas estas razones nos llevan a afirmar que, entre las formas de ocio adolescente —y las específicamente propias del entorno escolar—, tiene una gran presencia el castellano, o, como poco, es mayor su importancia en este ámbito que en otros donde predomina mucho más el uso de la lengua materna de la mayoría de la población.

⁶⁹⁸ Llama la atención que en el estudio de Bastardas (1986) sea precisamente este centro de interés, *Juegos, deportes y distracciones*, el que muestre unos resultados más favorables al uso del catalán entre los castellanohablantes. Este hecho puede señalar las diferencias en el uso de las dos lenguas en los distintos territorios catalanes o el cambio tan importante experimentado en la situación sociolingüística catalana en los últimos 20 años.

⁶⁹⁹ También el *Institut de Sociolingüística (s.d.)* reconoce la poca vitalidad del catalán en el terreno de las nuevas tecnologías. De otra parte, como también comentamos en el capítulo 4, estudios recientes sobre oferta lingüística otorgan una posición baja al uso del catalán en los medios de comunicación (Fabà 2008: 42).

⁷⁰⁰ Aunque el catalán se prefiere en el uso con los maestros. Volveremos sobre esta última afirmación en el siguiente apartado (6.1.3.). También en los datos que acabamos de comentar del capítulo cuarto, el ámbito más bajo con respecto al catalán, en los usos interpersonales, es el referente a los compañeros de estudio.

De hecho, nuestros propios datos hablan por sí solos. Si observamos la naturaleza del léxico evocado bajo este estímulo, vemos que los juegos tradicionales están presentes (*botifarra, cucut i amagar, pica paret, tocar i parar*), al igual que otras manifestaciones culturales propias de la vida catalana (*castells, esplai, sardana...*), donde es esperable que el uso del catalán predomine. No obstante, alcanzan rangos superiores en los índices de disponibilidad, en ambas lenguas, los nombres en castellano de juegos escolares (“pilla pilla”, “escondite”...) ⁷⁰¹, y, en general, los de juegos y distracciones relacionados tanto con la televisión como con las nuevas tecnologías, ámbitos donde, como hemos dicho, impera el uso del castellano ⁷⁰². También nuestros datos corroboran que el cine (18-14) parece estar más de moda que el teatro (33-39) o la radio (46-48), ámbitos donde tiene una mayor presencia el catalán ⁷⁰³. Finalmente, no podemos pasar por alto la importancia que cobran los deportes y la música en este centro de interés, sin duda, también influidos por la presencia mayor del castellano en los medios de comunicación ⁷⁰⁴.

De otra parte, *Juegos y distracciones* se ha clasificado ya anteriormente como un área léxica donde tienen un gran peso las marcas comerciales (en nuestros datos, *monopoly* ocupa la undécima posición en castellano y la novena en catalán), lo que nos lleva a los mercados de consumo, donde la presencia del castellano también es mayor. Asimismo, la naturaleza del léxico evocado se ha mostrado favorable al uso de otras lenguas bajo este estímulo. En las investigaciones anteriores a la nuestra, se han

⁷⁰¹ A continuación, detallamos mediante paréntesis la posición que alcanzan en el corpus castellano y en el catalán los juegos y manifestaciones culturales mencionados hasta el momento según su índice de disponibilidad: “botifarra”/botifarra (28-15), *cucut i amagar* (539-69), “pica pared”/pica paret (153-58), “tocar y parar”/tocar i parar (221-80), *castells* (0-515), *esplai* (0-301), *sardana* (325-346), “pilla pilla” (13-17), “escondite” (9-28). Como se puede observar, son significativas las altas posiciones alcanzadas por las dos últimas lexías en castellano en ambos corpus. Por otro lado, también es curioso, por poner solo un ejemplo, que la palabra “fútbol” tenga mayor disponibilidad léxica en catalán que *futbolí*. Hablaremos, en el próximo capítulo, de los fenómenos de contacto lingüístico hallados en este centro de interés.

⁷⁰² Prueba de ello es que “ordenador”/ordinador alcanza la décima posición en castellano y la undécima en catalán, “internet”, por su parte, la vigésima y la vigesimosexta, respectivamente. Con respecto a esta última lexía, creemos que, en el panorama actual, alcanzaría un rango más elevado, al igual que “móvil”, que queda en la posición 67 en castellano y en la 120 en catalán. Recordamos que los datos son del año 2002.

⁷⁰³ Entre paréntesis incorporamos el rango que alcanza la lexía en castellano, en primer lugar, y, a continuación, en catalán, según el índice de disponibilidad.

⁷⁰⁴ En la actualidad, y según nuestra propia experiencia como profesora de secundaria, constatamos que existe una amplia mayoría de jóvenes catalanohablantes que tiende a utilizar el castellano en las redes sociales (*twitter, facebook...*), preferentemente cuando se da un uso literario de la lengua. Sería interesante estudiar las actitudes sociolingüísticas que explican este comportamiento, más aún cuando se da entre jóvenes con inclinaciones ideológicas nacionalistas y firmes defensores del uso del catalán, con lo que no deja de resultar paradójico.

recabado, sobre todo, anglicismos, muestra de la influencia de esta lengua, y de su cultura, en las formas de ocio adolescente (v.gr. Bartol 1998 o Casanova 2008).

En los restantes centros de interés, el catalán obtiene mejores resultados, aunque las diferencias son muy distintas en función del área analizada. Las distancias mayores se constatan en *El cuerpo humano* (25.54 en castellano frente a 28.85 en catalán), en *Comidas y bebidas* (24.21 y 26.49, respectivamente) y en *Ropa y complementos* (21.62 y 23.19). La diferencia existente en la primera de las áreas asociativas supone 3.31 palabras de media, que se actualizan de más cuando el informante completa la encuesta en lengua catalana bajo el mismo estímulo. Esta diferencia es la mayor y se traduce en casi 800 palabras (795) en la comparación de los totales⁷⁰⁵. Seguidamente, no es nada desdeñable la diferencia que se observa en *Comidas y bebidas*, de 2.28 palabras en el promedio de los informantes y, en cifras absolutas, de 546 palabras. Algo ya más ajustada es la distancia en *Ropa y complementos*, donde las lenguas se distancian por 1.57 palabras en el promedio, lo que supone casi 400 palabras en los totales (377).

Si nos fijamos en las características de estos tres centros (1, 5 y 2), para ver si alguna de ellas puede explicar este comportamiento, constatamos que, para empezar, son centros muy productivos, ya que los tres se sitúan entre las cuatro primeras posiciones en las dos lenguas. Como veremos más adelante (vid. 6.1.5.), también son centros muy cohesionados, sobre todo *El cuerpo humano* y *Ropa y complementos*, lo que significa que las asociaciones que realizan los informantes en torno a ellos son muy similares⁷⁰⁶. Además, podemos intuir que es esta considerable diferencia entre la producción en castellano y en catalán, sobre todo en los centros 1 y 5, la que provoca que los tres centros más productivos en las dos lenguas no coincidan, puesto que en catalán *El cuerpo humano* y *Comidas y bebidas* sobrepasan a *Animales*, que ostenta el primer rango en lengua castellana.

A continuación, encontramos otras áreas donde la diferencia entre el castellano y el catalán en cuanto a la productividad ya no es tan notable como en las anteriores, pero en las que el predominio del catalán continúa siendo considerable. De este modo, por

⁷⁰⁵ Bastardas (1986) también detecta este centro como el más favorable a la lengua catalana.

⁷⁰⁶ El número de vocablos o palabras distintas no parece influir, puesto que los tres centros ocupan diversas posiciones según este parámetro (vid. 6.1.4.). Nótese, por otra parte, que las disimilitudes entre las dos lenguas son siempre inferiores a las que se detectan en función de los centros de interés.

ejemplo, en *La cocina*, la diferencia con respecto al catalán es de 263 respuestas, lo que supone 1.1 palabras en el promedio. Seguidamente, con una distancia de casi una palabra entre las dos lenguas, siempre a favor del catalán, hallamos *La escuela* y *Trabajos del campo y del jardín*, que cuentan con una diferencia muy cercana a la unidad (0.96 y 0.87, respectivamente —*vid.* tabla 6.6—). La distancia entre las dos lenguas se reduce, aunque sigue siendo de más de media palabra a favor del catalán, en *Profesiones y oficios* y *Objetos colocados en la mesa para la comida*.

En los restantes casos, a pesar de sobrepasar el catalán al castellano en una prueba de producción controlada ante unos determinados enunciados, las diferencias son bastante menores y no parecen cobrar ya tanta importancia. Siempre por debajo de la media unidad en el promedio por informante entre las dos lenguas se sitúan: *Partes de la casa* (0.43), *Muebles de la casa* (0.4), *Animales* (0.3), *Colores* (0.29), *Medios de transporte* (0.24) y *El campo* (0.19).

Antes de abordar el análisis de las causas que pueden provocar que en algunas áreas léxicas los informantes produzcan un mayor número de palabras en una lengua que en otra, hemos de conocer el alcance estadístico de estas diferencias. Es decir, es necesario comprobar que las divergencias en el caudal de léxico disponible de las dos lenguas no se deben al azar⁷⁰⁷.

6.1.3. Análisis estadístico de la diferencia en la producción de palabras

Con la finalidad de conocer de entre las diferencias halladas en la producción de las dos lenguas cuáles tienen relevancia significativa, hemos llevado a cabo el test T de Student para muestras emparejadas y hemos comparado las respuestas de unos mismos informantes en catalán y en castellano, tanto en la media general de cada lengua como en cada centro de interés. A continuación, incluimos tres cuadros fruto del análisis estadístico realizado. En el primero (tabla 6.7), el de los estadísticos de muestras relacionadas, se puede hallar la media (en general y por cada centro de interés) con la desviación y el error típico. En el segundo (tabla 6.8), se plasman las correlaciones entre las muestras con sus niveles de significación. En la tercera tabla (la 6.9), de la cual ya

⁷⁰⁷ Hernández Muñoz (2006: 340) constata cómo una diferencia numérica global de casi una palabra puede no resultar significativa estadísticamente, hecho que refrenda la necesidad de utilizar programas estadísticos para ilustrar las diferencias meramente descriptivas.

podemos extraer conclusiones, se detalla la diferencia entre las medias, el error y la desviación típica, y, partiendo de un nivel de confianza del 95%, se señala el valor de la diferencia. Como veremos a continuación, dado que oscila entre 0.00 y 0.05, esta es significativa y extensible al conjunto de la muestra en casi la mitad de los centros de interés.

Tabla 6.7. Estadísticos de muestras relacionadas⁷⁰⁸.

		Media	N	Desviación típica	Error típico de la media
Par 1	Promedio	17.904	240	2.925	.1888
	Promedio	18.638	240	3.165	.2043
Par 2	Centro 1	25.596	240	5.4457	.3515
	Centro 1	28.850	240	5.4973	.3548
Par 3	Centro 2	21.621	240	4.7921	.3093
	Centro 2	23.192	240	4.9867	.3219
Par 4	Centro 3	14.167	240	3.9215	.2531
	Centro 3	14.596	240	3.8928	.2513
Par 5	Centro 4	12.396	240	3.6069	.2328
	Centro 4	12.800	240	3.7697	.2433
Par 6	Centro 5	24.213	240	5.3129	.3429
	Centro 5	26.488	240	6.0500	.3905
Par 7	Centro 6	13.204	240	3.9532	.2552
	Centro 6	13.854	240	4.0612	.2621
Par 8	Centro 7	15.379	240	4.8855	.3154
	Centro 7	16.475	240	5.1121	.3300
Par 9	Centro 8	21.167	240	4.4465	.2870
	Centro 8	22.125	240	5.0122	.3235
Par 10	Centro 9	11.083	240	3.5419	.2286
	Centro 9	11.050	240	3.4938	.2255
Par 11	Centro 10	21.350	240	5.0968	.3290
	Centro 10	21.250	240	5.5700	.3595
Par 12	Centro 11	18.230	239	5.8827	.3805
	Centro 11	18.406	239	5.7874	.3744
Par 13	Centro 12	16.846	240	3.8417	.2480
	Centro 12	17.088	240	4.3848	.2830
Par 14	Centro 13	8.854	240	3.3448	.2159

⁷⁰⁸ Los datos sombreados en la tabla son los relativos a la lengua castellana, los restantes, por tanto, son los del catalán. La *desviación típica* hace referencia a la media de las diferencias, en valor absoluto, entre cada valor y la media, mientras que el *error típico de la media* es la desviación típica de la distribución muestral de la media.

	Centro 13	9.717	240	3.4670	.2238
Par 15	Centro 14	26.021	240	5.0607	.3267
	Centro 14	26.321	240	5.8037	.3746
Par 16	Centro 15	16.308	240	4.5726	.2952
	Centro 15	15.713	240	4.8055	.3102
Par 17	Centro 16	19.621	240	4.7209	.3047
	Centro 16	20.367	240	5.0103	.3234
Par 18	Centro 17	18.354	240	3.8185	.2465
	Centro 17	18.646	240	4.0036	.2584

Tabla 6.8. Correlaciones de muestras relacionadas⁷⁰⁹.

		N	Correlación	Significación
Par 1	Promedio y Promedio	240	.667	.000
Par 2	Centro 1 y Centro 1	240	.575	.000
Par 3	Centro 2 y Centro 2	240	.493	.000
Par 4	Centro 3 y Centro 3	240	.470	.000
Par 5	Centro 4 y Centro 4	240	.591	.000
Par 6	Centro 5 y Centro 5	240	.559	.000
Par 7	Centro 6 y Centro 6	240	.440	.000
Par 8	Centro 7 y Centro 7	240	.451	.000
Par 9	Centro 8 y Centro 8	240	.560	.000
Par 10	Centro 9 y Centro 9	240	.524	.000
Par 11	Centro 10 y Centro 10	239	.527	.000
Par 12	Centro 11 y Centro 11	240	.510	.000
Par 13	Centro 12 y Centro 12	240	.455	.000
Par 14	Centro 13 y Centro 13	240	.405	.000
Par 15	Centro 14 y Centro 14	240	.406	.000
Par 16	Centro 15 y Centro 15	240	.485	.000
Par 17	Centro 16 y Centro 16	240	.519	.000
Par 18	Centro 17 y Centro 17	240	.490	.000

⁷⁰⁹ La correlación indica el grado de relación entre dos variables y en qué medida estas se relacionan. Si equivale a 0 es porque no hay relación, mientras que un valor de 1 indicaría que la relación es perfecta.

Tabla 6.9. Prueba de muestras relacionadas⁷¹⁰.

	Diferencias relacionadas					t	gl	Sig. (bilateral)
	Media	Desviación típica	Error típico de la media	95% Intervalo de confianza para la diferencia				
				Inferior	Superior			
Promedio-Promedio	-.733	2.495	.161	-1.051	-.416	-4.556	239	.000
C. 1- 1	-3.254	5.045	.325	-3.895	-2.612	-9.992	239	.000
C. 2 -2	-1.570	4.928	.318	-2.197	-.944	-4.938	239	.000
C. 3 -3	-.429	4.022	.259	-.940	.082	-1.653	239	.100
C. 4 -4	-.404	3.338	.215	-.828	.020	-1.87	239	.062
C. 5 -5	-2.275	5.377	.347	-2.958	-1.591	-6.554	239	.000
C. 6 -6	-.650	4.242	.273	-1.189	-.110	-2.373	239	.018
C. 7 -7	-1.095	5.241	.338	-1.762	-.429	-3.239	239	.001
C. 8 -8	-.958	4.462	.288	-1.525	-.390	-3.327	239	.001
C. 9 -9	.033	3.433	.221	-.403	.469	.150	239	.881
C. 10 -10	.100	5.204	.335	-.561	.761	.298	239	.766
C. 11 -11	-.175	5.776	.373	-.911	.560	-.470	238	.639
C. 12 -12	-.241	4.320	.278	-.791	.307	-.866	239	.387
C. 13 -13	-.862	3.717	.239	-1.335	-.389	-3.59	239	.000
C. 14 -14	-.300	5.953	.384	-1.057	.457	-.781	239	.436
C. 15 -15	.595	4.764	.307	-.010	1.201	1.937	239	.054
C. 16 -16	-.745	4.777	.308	-1.353	-.1383	-2.418	239	.016
C. 17 -17	-.291	3.952	.255	-.794	.210	-1.143	239	.254

Tras observar los resultados del análisis estadístico, podemos adelantar distintas conclusiones. En principio, la diferencia entre las dos lenguas es significativa, tanto en su promedio global por centro de interés (17.9 en castellano frente a 18.64 en catalán) como en ocho de las diecisiete áreas estudiadas, siempre a favor de la lengua catalana. En los tres centros en los que, como vimos en el apartado anterior, el castellano supera al catalán —*Calefacción e iluminación y formas de airear un recinto, La ciudad y, sobre todo, Juegos y distracciones*— la diferencia entre las dos lenguas no es significativa, puesto que se sitúa por encima de 0.05, lo cual quiere decir que no es posible refutar la hipótesis nula que postula que las diferencias se deben al azar. El resultado ya era, en cierto modo, esperable en relación con los dos primeros centros,

⁷¹⁰ El *intervalo de confianza*, en este caso fijado en el 95%, indica los valores mínimo y máximo aceptables para la diferencia entre las medias. De todos los valores de las tres tablas, el más importante es el de la significación (tabla 6.9) que nos da la probabilidad de rechazar la hipótesis nula cuando es verdadera.

donde —según vimos en la tabla 6.6— la distancia que separa las dos lenguas es insignificante. Sin embargo, en *Juegos y distracciones* ya era de poco más de media palabra y las razones que aportamos nos hacían pensar que podía resultar una diferencia extensible al universo. Finalmente, no es posible asegurar que la distancia entre las dos lenguas no se deba al azar, aunque el resultado obtenido en esta área (0.054) es muy cercano a la significación estadística. Por tanto, cabe señalar que los resultados apuntan claramente a que el caudal léxico de los informantes, analizados siempre en general y no en subgrupos según las distintas variables sociales, es mejor en lengua catalana, toda vez que, en los únicos centros donde el castellano supera al catalán, esta diferencia no tiene validez estadística.

Las ocho áreas en que la diferencia es significativa a favor del catalán son: 1. *El cuerpo humano*, 2. *Ropa y complementos*, 5. *Comidas y bebidas*, 6. *Objetos colocados en la mesa para la comida*, 7. *La cocina y sus utensilios*, 8. *La escuela*, 13. *Trabajos del campo y del jardín* y 16. *Profesiones y oficios*. Si recordamos los datos que ofrecimos en la tabla 6.6., donde se vislumbran las diferencias entre las dos lenguas, estas áreas ocupan los ocho primeros lugares, esto es, las categorías donde las divergencias son mayores han resultado ser significativas estadísticamente.

Al intentar hallar una explicación al comportamiento dispar que realizan unos mismos informantes, bilingües todos ellos y con un nivel de dominio similar del castellano y del catalán⁷¹¹, se nos plantean diversos interrogantes: ¿Qué lleva a un informante nacido en Lleida⁷¹², escolarizado en catalán y cuya lengua materna es el catalán o el castellano, a escribir casi 29 palabras bajo el rótulo *El cos humà* y, en cambio, solo 25 bajo *El cuerpo humano*? ¿Por qué la diferencia es mayor, y significativa estadísticamente, en esta área y en otras, como en *Colores*, no? ¿Por qué en algunos centros de interés, los menos, es el castellano el que aventaja al catalán? Hasta el momento, solo hemos intentado responder a la última de estas cuestiones,

⁷¹¹ Recordemos que, al finalizar el curso en el que se encuentran matriculados —2º de bachillerato— llevan a cabo un examen de las dos lenguas, dentro de las pruebas de la selectividad. El nivel de estos dos exámenes es prácticamente el mismo. Además, los datos que vimos sobre el censo lingüístico de 2001 para esta franja de edad (de 15 a 19 años) señalan que más del 90% de la población domina las cuatro destrezas lingüísticas en lengua catalana (*vid.* 3.4.2., gráfico 4.17).

⁷¹² Como indicamos en el capítulo metodológico, también admitimos aquellos informantes nacidos en otras provincias pero que llevaban 10 años, como mínimo, viviendo en Lleida. En la práctica, este porcentaje supone apenas el 6% de la muestra, del cual los nacidos fuera de Cataluña no alcanzan siquiera el 1%.

reconociendo el predominio del castellano en algunos ámbitos relacionados con el ocio juvenil (*vid.* 6.1.2).

A nuestro modo de ver, y teniendo en cuenta las características de la muestra y la situación sociolingüística de la provincia, vista ya ampliamente (*vid.* 4.3.2.), hay dos factores cuya importancia cabe analizar en este punto. El primero de todos ellos es la lengua materna del individuo, que consideramos, por su relevancia, como variable social independiente en este trabajo, y cuya incidencia cuantitativa analizaremos más adelante (*vid.* 6.2.5.). Recordemos que en la encuesta dedicamos una pregunta a esta cuestión, aclarando cuando fue necesario que la lengua materna es aquella en la que el sujeto piensa, razona, cuenta en voz baja, y, por supuesto, la que aprende en primer lugar. La muestra se divide en cuatro grupos según esta variable, de los cuales, la mayoría de informantes se adscribe a dos modalidades: castellanohablantes (el 27.5%) y, sobre todo, catalanohablantes (el 65.42%)⁷¹³. Como en este punto del análisis manejamos los datos globales de la muestra en su conjunto, los datos reflejan —a nuestro modo de ver— la influencia del catalán como lengua materna de la mayoría de los informantes en todas las áreas temáticas suscitadas, especialmente en las ocho que resultan significativas. El mayor peso numérico que tienen los catalanohablantes en el conjunto hace que, en general, los resultados sean más satisfactorios a la producción en lengua catalana⁷¹⁴.

El segundo factor que consideramos que tiene una crucial importancia es el programa educativo de inmersión lingüística que han seguido todos nuestros informantes. El contacto con la escuela tiene una gran transcendencia y nos interesa en la medida en que asegura un contacto continuado con la lengua catalana⁷¹⁵. Con respecto al castellano, este pasa a formar parte del currículum, como una asignatura semanal de tres horas, a partir de la enseñanza primaria, esto es, cuando el alumno tiene

⁷¹³A partir de este momento, nos referimos así a los dos subgrupos dentro de la muestra cuya lengua materna es el catalán (catalanohablantes) o el castellano (castellanohablantes). Como ya quedó claro en la metodología y en el capítulo dedicado a la situación social y lingüística de la provincia, todos los informantes son bilingües.

⁷¹⁴Más adelante (*vid.* 6.2.5.), analizaremos por separado la producción léxica de los dos grupos mayoritarios, catalanohablantes y castellanohablantes, y podremos estudiar la incidencia de la *lengua materna* en la producción por separado de las dos lenguas.

⁷¹⁵No hay ningún centro educativo en nuestra provincia que utilice el castellano como lengua vehicular de la enseñanza o que imparta sus enseñanzas en dicho idioma, de acuerdo con la normativa vigente (*vid.* 4.1).

seis años. Este hecho significa que, en nuestro estudio, contamos con dos perfiles de alumnado muy distintos según la relación que tienen con las dos lenguas: el castellanohablante, que empezó a tener contacto con el catalán, de forma continua y constante, a partir de su escolarización (en general, a partir de los tres años) realizada íntegramente en lengua catalana y, el catalanohablante, cuya inmersión en la escuela se realizó en su propia lengua materna y cuyo contacto con el castellano empezó a los seis años, como mínimo, con la asignatura de lengua española en el colegio. Es evidente que acabamos de realizar una simplificación de la realidad, puesto que en una situación como la nuestra de bilingüismo social extenso y estable, las circunstancias particulares respecto al uso de las dos lenguas en cada individuo pueden ser de lo más variadas. De otra parte, también la notable presencia de las dos lenguas en muchos ámbitos de la vida social del informante puede haber influido en el aprendizaje de las mismas, más allá del peso de la escolarización.

Otra variable que —en nuestra opinión— influye notablemente es el hábitat en el que vive el informante, puesto que, como ya explicamos, fueron mayoritariamente las ciudades las que absorbieron en los años de posguerra a los inmigrantes castellanohablantes, de lo que se deduce que hay un mayor uso del catalán en los núcleos poblacionales más pequeños y un mayor uso del castellano en la capital⁷¹⁶. Este hecho creemos que explica, entre otras cosas, la tímida ventaja del castellano en *La ciudad*. A pesar de que dicha diferencia no tiene relevancia estadística, es de notar que en este centro la producción en castellano alcance un mejor promedio, aunque las dos lenguas estén muy igualadas, y que en *El campo* o en *Trabajos del campo y del jardín*, áreas relacionadas con el ámbito rural, los resultados sean mejores en lengua catalana⁷¹⁷.

Por tanto, tal como hemos visto, la influencia de la lengua materna, la de la escuela y la del entorno serán fundamentales en el caudal de léxico disponible de las dos lenguas del alumno. De otra parte, desde la investigación de Hernández Muñoz (2006) sabemos que la disponibilidad se asocia a las palabras más *típicas*, más *familiares* y más *tempranamente aprendidas*. La primera de estas características es la responsable de que

⁷¹⁶ En la propia distribución de la población en la provincia según su lugar de nacimiento, ya se observa este hecho (*vid.* 4.3.2. gráficos 4.11 y 4.12). También tendremos muy en cuenta esta observación cuando estudiemos el peso de la variable *zona geográfica* (*infra* 6.2.3.).

⁷¹⁷ También hay que señalar que el léxico de estas áreas se trabaja en la asignatura de *Coneixement del medi natural, social i cultural* que se imparte en catalán, como ya hemos comentado.

sean las mismas palabras, o los mismos conceptos, los que se evoquen en los primeros lugares bajo cada centro de interés en las dos lenguas, puesto que la forma de categorizar es universal. Más interesante para las diferencias entre la producción de las dos lenguas es la familiaridad —que hace referencia a las veces en que el sujeto entra en contacto con una determinada palabra a lo largo del día (piensa en ella, la nombra, etc.)— y la edad de adquisición —que influye en el sentido de que las palabras aprendidas antes son más disponibles que las aprendidas años más tarde—. Como es evidente, esta última característica tiene una clara correlación con la lengua materna, puesto que, lógicamente, será anterior el aprendizaje de un concepto en la lengua materna que en la segunda lengua. El hecho de que toda la muestra ya haya sido escolarizada en catalán, desde los tres años, resta un poco de importancia a esta última característica en el caso de los castellanohablantes, y, por el contrario, puede acentuar las diferencias entre las dos lenguas en el caso de los catalanohablantes (*cf.* 6.2.5).

La familiaridad, por su parte, se relaciona con las lenguas que habitualmente emplea en sus intercambios el informante en cada ámbito. Por ejemplo, si utiliza el catalán en su casa, cada vez que vea, piense o entre en contacto con la palabra *forquilla*, utilizada fundamentalmente a la hora de la comida, lo hará en catalán, y no en castellano⁷¹⁸. Si con sus amigos, en cambio, emplea las dos lenguas indistintamente, puede evocar las mismas veces en los dos idiomas una palabra relacionada con un juego, admitida bajo el enunciado de *Juegos y distracciones*. Como ya vimos, en la situación sociolingüística actual, no hay una distribución rígida de los ámbitos de uso de cada lengua, con lo que será la propia experiencia vital del individuo la que determine la distribución de las dos lenguas en su esfera social (Bastardas 1986: 51)⁷¹⁹.

Pasemos ahora a analizar las categorías donde la diferencia entre las lenguas cobra una relevancia estadística, considerando que las razones que llevan al predominio del catalán pueden ser distintas en cada caso⁷²⁰. En los tres primeros estímulos —*El*

⁷¹⁸ Esta es la razón por la que no pocos informantes escriben bajo este centro “horquilla”, evidentemente, piensan en el concepto en catalán y realizan una traducción errónea.

⁷¹⁹ Baetens (1989: 22) afirma que, “de hecho, un hablante que tenga un alto nivel de competencia en dos lenguas, como por ejemplo un intérprete profesional, seguramente pasará como hablante nativo de las dos sin mostrar ningún rasgo de interferencia, pero la disponibilidad léxica en áreas semánticas determinadas será mayor en una lengua que en la otra.” La traducción del catalán es nuestra. *Vid.* 2.3.

⁷²⁰ El hecho de que en las restantes seis áreas en las que el catalán sobrepasa al castellano la diferencia no adquiera relevancia estadística, no significa que no se diera en otro informante tomado al azar, solo que no podemos atribuirle con seguridad a esta causa.

cuerpo humano, Ropa y complementos y Comidas y bebidas— vimos, además, cómo la diferencia entre las dos lenguas era la mayor de todas⁷²¹. Este hecho puede explicarse, en nuestra opinión, porque estas tres categorías están muy presentes en la cotidianidad, en el entorno más inmediato de los informantes, lo que provoca que el sujeto entre “en contacto” a menudo con los conceptos que se evocan en sus respectivos campos y lo haga, por tanto, en su propia lengua materna. La familiaridad por tanto favorece, aquí, el uso de la primera lengua en estas categorías semánticas.

De otra parte, también la edad de adquisición parece influir a favor de la lengua materna mayoritaria de la población en estas áreas. A pesar de que no contamos con datos empíricos para poder contrastar este hecho, intuitivamente parece que el léxico evocado en estos centros de interés se aprende antes que el relacionado con otras áreas como *Trabajos del campo y del jardín o Juegos y distracciones*⁷²². Es de suponer que, si la edad de adquisición influye en la disponibilidad de las palabras, este factor cobra mayor importancia en aquellas áreas donde predomine el léxico aprendido a una edad más temprana. En este sentido, incluso parece anterior la edad de adquisición del léxico relacionado con *El cuerpo humano* y con *Comidas y bebidas*, donde el catalán y el castellano alcanzan la mayor diferencia según vimos (*supra* tabla 6.6.), que la edad de adquisición del vocabulario relativo a la *Ropa y complementos*⁷²³. Por otro lado, la distribución en ámbitos según la experiencia que tenga en el uso cada informante de las dos lenguas creemos que explica muchas de las disimilitudes que se dan en la producción entre las dos lenguas. De esta forma, en los centros de interés más relacionados con la vida familiar, la influencia de la lengua materna es mayor. Esta también puede ser la causa de que la productividad sea mayor en lengua catalana en los tres centros analizados hasta el momento (1, 2 y 5), entendiendo que los hábitos relacionados con la higiene, la vestimenta o la comida, se asocian fuertemente a la experiencia vital que cada informante tenga en su propio hogar desde la infancia.

⁷²¹ Es curioso cómo ya en el estudio de Bastardas (1986) los catalanohablantes obtienen el mejor resultado en su propia lengua en *El cos humà* (*vid.* 4.1.).

⁷²² Obviamente no nos referimos a toda la nómina de vocablos evocados en cada centro sino a los que alcanzan las primeras posiciones. Cualquier hablante aprende antes “mano”, “agua” o “camiseta” que “segar” o “parchís”.

⁷²³ Contra esta argumentación, hallamos otros casos como *Medios de transporte, Animales* o *Colores*, donde la diferencia entre las lenguas no es significativa, a pesar de que son áreas de aprendizaje temprano. Sin embargo, los vocablos de *Medios de transporte* aprendidos tempranamente son realmente pocos y, de otro lado, el elenco de colores habitualmente usado es reducido. Préstese atención en este punto a las conclusiones a las que llegaremos, más adelante, tras analizar el peso de la variable *lengua materna* (*vid.* 6.2.5.).

También creemos que es la razón por la que se explican las diferencias a favor del catalán en *Objetos colocados en la mesa para la comida* y en *La cocina y sus utensilios*. En estos dos últimos casos existe un vínculo total entre estas áreas y el entorno propiamente familiar, con lo que las posibilidades de evocar conceptos como “tenedor”, “cuchillo” o “nevera” y “sartén” en otros entornos, y por tanto, en otra lengua distinta a la materna, son realmente reducidas o, en cualquier caso serán incomparables al peso que cobra la vinculación con la lengua normalmente utilizada en el hogar.

Es relativamente sencillo explicar por qué, en el centro *La escuela*, los informantes obtienen mejores resultados en catalán. Como ya vimos (*supra* 4.1.), gracias al programa de inmersión lingüística, el uso habitual del catalán en las escuelas como lengua vehicular de aprendizaje está plenamente implantado en Cataluña en la actualidad. Por ende, los resultados muestran este mayor uso del catalán en el entorno escolar. Recordemos que, en su investigación, Galindo (2006) concluye que el castellano se usa más que el catalán en los patios de las escuelas catalanas, pero que, sin embargo, el catalán se prefiere para ser usado con los maestros. Ya Boix (1993: 122) señala, años antes, este uso del catalán como lengua “alta” relacionada con las instituciones educativas y los objetivos formales y académicos. En este sentido, nuestros datos muestran el cambio tan importante que ha habido en Cataluña desde la consolidación de la normalización lingüística, porque en el estudio de Bastardas (1986) realizado con una metodología muy similar a la nuestra, en este mismo centro de interés impera, por el contrario, el castellano.

Como hemos comentado en este mismo apartado, el castellano se asocia más a las ciudades que a los pueblos, lo que consideramos que justifica la diferencia con valor estadístico detectada en *Trabajos del campo y del jardín*, donde, a pesar del enunciado, se actualizan principalmente verbos relacionados con el mundo rural como “segar”, “cultivar” o “cosechar”. Por el contrario, como también vimos, en *La ciudad* se obtienen resultados favorables a la lengua castellana.

Con respecto al área *Profesiones y oficios*, las razones que pueden llevar a su mayor fecundidad en catalán, creemos que se relacionan con el hecho de que el léxico que evoca se trabaja en la escuela. Algunos investigadores con anterioridad (Gallosó 2003a, Martínez Olmos 2008) ya han hecho alusión a la importancia que puede tener en

los estudios de disponibilidad la presencia de algunos centros de interés relacionados con el currículum escolar, y cuyo léxico, por tanto, se refuerza claramente en la escuela. Consideramos que el efecto de la escolarización en lengua catalana, más el peso de la lengua materna, provocan que la diferencia a favor del catalán sea significativa en esta área. Este hecho también justificaría la distancia en *El cuerpo humano*, *Comidas y bebidas*, *Ropa y complementos* o *La escuela*, puesto que son categorías, sobre todo las dos primeras, que se trabajan a lo largo de toda la primaria dentro de la asignatura de *Conocimiento del medio*. De hecho, el efecto de la escolarización más la influencia de la lengua materna, acentúa las diferencias a favor del catalán en la productividad de la mayoría de las categorías semánticas, muchas de ellas también relacionadas con el entorno familiar. No obstante, cabe señalar que las disimilitudes halladas no resultan significativas en algunas áreas donde también se dan las circunstancias apuntadas como en *Medios de transporte* o *Animales*⁷²⁴. Por el contrario, las áreas en las que predomina el castellano, *Calefacción*, *iluminación* y *formas de airear un recinto*, *La ciudad* y *Juegos y distracciones*, no responden a estas características: ni su léxico se aprende tempranamente, ni tienen un peso considerable en el currículum escolar ni su uso predomina en el hogar.

Para finalizar, es de justicia reconocer que, relacionando los estímulos donde la diferencia entre las dos lenguas es significativa y los parámetros que analizamos en este capítulo, no se observa ninguna relación entre los mismos, a no ser que sea la gran fecundidad de *El cuerpo humano*, *Ropa y complementos* y *Comidas y bebidas*, campos que, además, resultan muy cohesionados (*vid.* 6.1.5.)⁷²⁵. Sin embargo, también aparecen como significativas las diferencias halladas en áreas como *Trabajos del campo y del jardín*, muy poco productiva o cohesionada, u otras infecundas como *Objetos colocados en la mesa para la comida* o *La cocina y sus utensilios*. Tampoco parece existir ninguna relación con el número de vocablos, que trataremos a continuación, puesto que las diferencias son significativas tanto en centros muy ricos léxicamente (como *Comidas y bebidas* o *Profesiones y oficios*) como en otros muy pobres (como *Objetos colocados en la mesa para la comida*).

⁷²⁴ Estudiaremos de nuevo estas áreas al analizar los datos según la variable *lengua materna* (*vid.* 6.2.5.).

⁷²⁵ El mismo hecho de que en ellos predomine un léxico tempranamente aprendido puede explicar este comportamiento.

6.1.4. Total de vocablos por centro de interés

Pasemos ahora a ver las diferencias en la producción con respecto a los vocablos, esto es, al número de palabras diferentes que se recogen bajo cada centro de interés. Samper, Bellón y Samper Hernández (2003: 60) afirman que la cantidad de vocablos “informa de la homogeneidad o heterogeneidad de las respuestas de los encuestados, es decir, de la mayor o menor amplitud de las asociaciones de cada uno de los campos léxicos analizados”. Este parámetro depende de la naturaleza de cada área temática, que propicia un número mayor o menor de asociaciones, pero también se ve fuertemente influido por el número de informantes que componen la muestra, de forma que a más informantes, mayor número de vocablos (Bartol 2004: 47, Fernández Smith *et al.* 2008: 86). En nuestra investigación salvamos este inconveniente al trabajar con muestras idénticas, de modo que un mayor número de vocablos nos informará de la mayor riqueza léxica de un área temática concreta en una de las dos lenguas.

Dado que evalúa el comportamiento conjunto de un grupo de informantes ante un determinado estímulo, este índice tiene el inconveniente —como ya advertimos (*vid.* 6.1.)— de que un solo informante atípico puede modificarlo, si realiza asociaciones distintas a los demás porque tenga un mayor bagaje léxico en ese tema concreto o por cualquier circunstancia derivada de su propia experiencia vital⁷²⁶. Por otra parte, hasta el momento, ningún trabajo de nuestro ámbito, como tampoco hacemos nosotros, ha aplicado programas estadísticos a estos datos, con lo que no es posible conocer el alcance real de las diferencias que se observan según este parámetro. Por todas estas razones, la información que nos aporta este índice ha de tomarse con cierta cautela, puesto que no es tan fiable como la que nos ofrece el promedio de palabras, donde las diferencias individuales quedan compensadas.

Tabla 6.10. Total de vocablos por centro de interés en castellano y en catalán.

Rango	Centro de Interés	Castellano	Catalán	Diferencia
1	El cuerpo humano	386	364	22
2	Ropa y complementos	291	313	-22
3	Partes de la casa (sin los muebles)	238	258	-20
4	Muebles de la casa	263	274	-11
5	Comidas y bebidas	476	523	-47
6	Objetos colocados en la mesa para la comida	238	282	-44

⁷²⁶ Bartol (2001: 227) pone ejemplos concretos de cómo algunos informantes alteran el número de palabras diferentes y, por tanto, el índice de disponibilidad de algunas palabras. Por otro lado, recuérdese también lo dicho en la nota 684 de este mismo capítulo.

7	La cocina y sus utensilios	347	373	-26
8	La escuela (muebles y materiales)	420	444	-24
9	Calefacción, iluminación y formas de airear un recinto	297	286	11
10	La ciudad	678	658	20
11	El campo	727	748	-21
12	Medios de transporte	276	303	-27
13	Trabajos del campo y del jardín	434	418	16
14	Animales	347	434	-87
15	Juegos y distracciones	576	560	16
16	Profesiones y oficios	591	577	14
17	Colores	239	249	-10

Si observamos con atención la tabla anterior, y la comparamos con la del total y promedio de palabras (tablas 6.3 y 6.5, *supra*), constatamos, al igual que todos los investigadores que nos han precedido, que no se da, en ninguna de las dos lenguas, una correlación entre los campos más productivos con respecto al número de palabras y los que cuentan con un mayor número de vocablos. Los centros con mayor cantidad de palabras distintas son, en ambas lenguas, *El campo*, *La ciudad* y *Profesiones y oficios*, áreas cuya productividad alcanza distintos rangos (*El campo*, la novena posición, *La ciudad*, la quinta en castellano y la sexta en catalán, y *Profesiones y oficios*, la séptima). De otra parte, si comparamos las dos lenguas, además de la coincidencia estricta entre los rangos de estos tres centros, es notable la similitud en el número de vocablos que alcanza cada área temática: *El campo* obtiene 727 en castellano y 748 en catalán; *La ciudad* 678 y 658, y *Profesiones y oficios*, 591 y 577. En general, las cifras que obtiene, en las dos lenguas, cada campo semántico presentan una similitud mayor que las que alcanzaban los totales de palabras, donde la mayor diferencia entre las lenguas era de casi 800 palabras. Esto es lógico puesto que es mucho más difícil que haya un vocablo nuevo que una palabra nueva, toda vez que, en el primer caso, este no debe coincidir con los que hayan escrito el resto de sujetos encuestados.

Como acabamos de ver, la coincidencia es absoluta entre las dos lenguas en los centros con mayor espectro de asociaciones, en los centros menos ricos léxicamente, en cambio, los paralelismos no son tan exactos. En castellano, las tres categorías que cuentan con menos vocablos son *Colores*, *Partes de la casa* y *Objetos colocados en la mesa para la comida*, que obtienen resultados prácticamente idénticos (239, 238 y 238, respectivamente); en catalán, en cambio, en la misma posición hallamos *Muebles de la casa*, *Partes de la casa* y *Colores*, con cifras de vocablos más dispares entre sí (274, 258 y 249).

Si tenemos en cuenta la media en cada lengua por centro de interés, en castellano quedan por encima de esta cifra (401.41, *vid.* tabla 6.2) solo siete centros de interés: *El campo*, *La ciudad*, *Profesiones y oficios*, *Juegos y distracciones*, *Comidas y bebidas*, *Trabajos del campo y del jardín* y *La escuela*, mientras que en catalán sobrepasan la media (415.53) ocho estímulos, los cinco que ocupan los primeros lugares en castellano más *La escuela*, *Animales* y *Trabajos del campo y del jardín*, por este orden.

Como ya hemos indicado, las cifras que alcanza cada centro de interés aunque distintas son parecidas en los dos idiomas, e incluso la progresión que se sigue en el aumento de las asociaciones propias de cada estímulo verbal es muy similar (*vid.* tabla 6.11). En general, la distancia entre los centros con mayor o menor número de vocablos es, al igual que sucedía con los totales de palabras, mayor en catalán que en castellano, aunque aquí este hecho carece de relevancia puesto que la diferencia es de solo 10 palabras (499 en catalán, las que hay entre *El campo* y *Colores* y 489 en castellano, las que separan *El campo* de *Objetos colocados en la mesa para la comida* o *Partes de la casa*, que alcanzan idéntico número de palabras diferentes).

Tabla 6.11. Rango de los centros de interés en cada una de las lenguas según el número de vocablos (de mayor a menor)⁷²⁷.

Rango	Castellano		Catalán	
1	11. El campo	727	11. El campo	748
2	10. La ciudad	678	10. La ciudad	658
3	16. Profesiones y oficios	591	16. Profesiones y oficios	577
4	15. Juegos y distracciones	576	15. Juegos y distracciones	560
5	5. Comidas y bebidas	476	5. Comidas y bebidas	523
6	13. Trabajos del campo y del jardín	434	8. La escuela	444
7	8. La escuela	420	14. Animales	434
8	1. El cuerpo humano	386	13. Trabajos del campo y del jardín	418
9	7. La cocina y sus utensilios*	347	7. La cocina y sus utensilios	373
10	14. Animales*	347	1. El cuerpo humano	364
11	9. Calefacción, iluminación y formas de airear un recinto	297	2. Ropa y complementos	313
12	2. Ropa y complementos	291	12. Medios de transporte	303

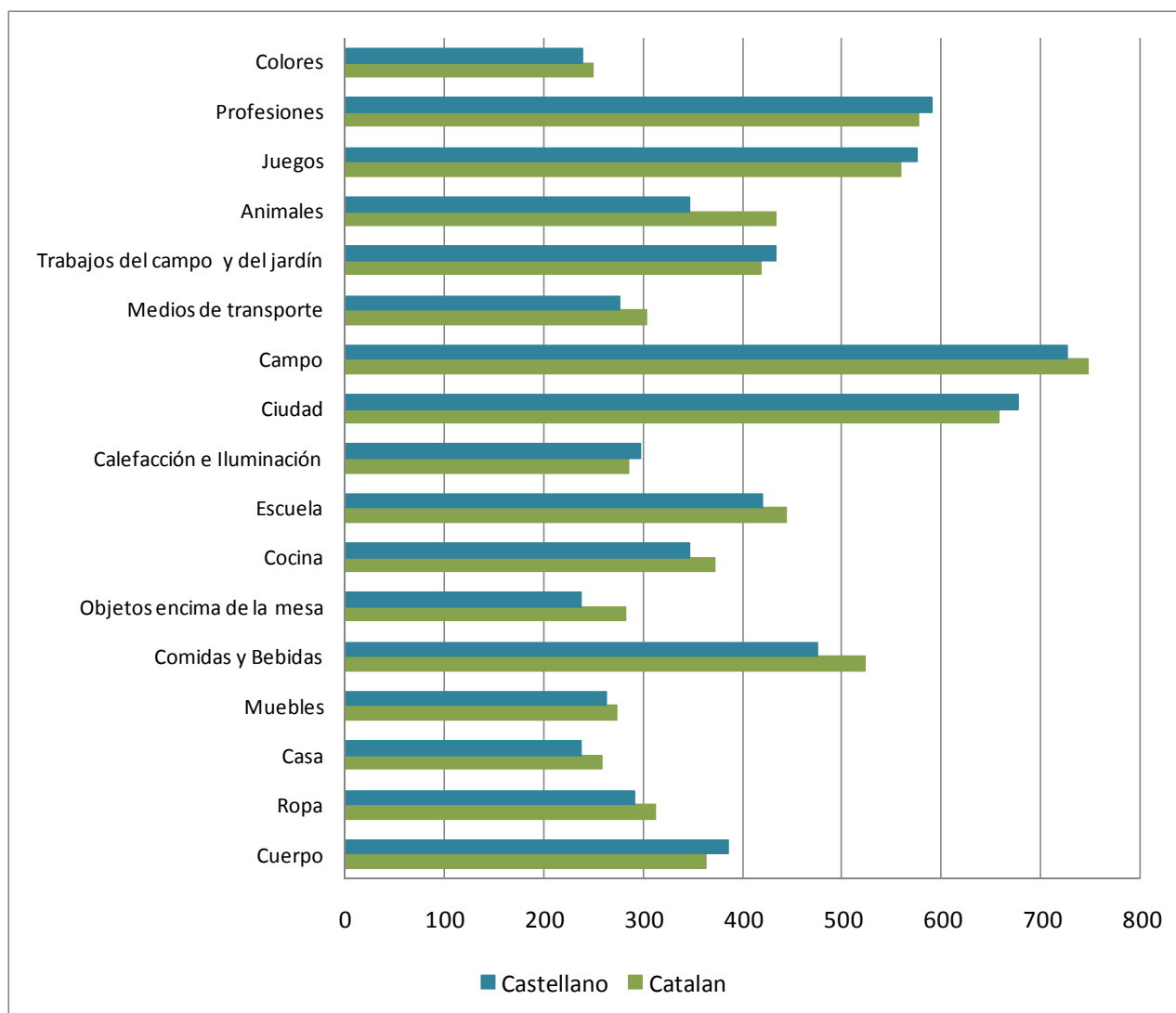
⁷²⁷ Se marcan con un asterisco los centros que obtienen el mismo número de vocablos, hecho, por otro lado, destacable, toda vez que se da en dos ocasiones en el corpus castellano. De otra parte, los centros que alcanzan el mismo rango se señalan de forma coloreada en la tabla; la media global, situada en 404.41 en castellano y en 415.53 en catalán se destaca con una doble raya.

13	12. Medios de transporte	276	9. Calefacción, iluminación y formas de airear un recinto	286
14	4. Muebles de la casa	263	6. Objetos colocados en la mesa para la comida	282
15	17. Colores	239	4. Muebles de la casa	274
16	3. Partes de la casa*	238	3. Partes de la casa	258
17	6. Objetos colocados en la mesa para la comida*	238	17. Colores	249

En la tabla anterior se observa la clasificación de los centros de interés en las dos lenguas según el número de vocablos registrados. Si nos fijamos con detalle, lo primero que llama la atención es que el orden no sigue unos patrones tan similares como los obtenidos en el ranquin según la productividad de los centros de interés⁷²⁸. Aquí las concomitancias se limitan a los cinco primeros puestos de la tabla y a la coincidencia, en noveno lugar, de *La cocina y sus utensilios* y, en decimosexto, de *Partes de la casa*. En las restantes áreas se observan diferencias o bien de una posición (*La escuela* ocupa la séptima posición en castellano y la sexta en catalán; *Ropa y complementos*, la duodécima en castellano y la undécima en catalán, etc.) o, más frecuentemente, de dos (*El cuerpo humano* es octavo en castellano y décimo en catalán; *Colores* es decimoquinto en castellano y último en catalán, etc.).

⁷²⁸ Lo mismo habían observado otros investigadores (v.gr. Gómez Molina y Gómez Devís 2004: 78) comparando los datos del castellano de distintas zonas: esto es, que la clasificación de los centros de interés según la producción de palabras es mucho más similar que la obtenida según el número de vocablos.

Gráfico 6.8. Total de vocablos por centro de interés en castellano y en catalán.



Si comparamos, ahora, las diferencias cuantitativas en cuanto al número de asociaciones realizadas bajo un mismo estímulo verbal entre las dos lenguas, lo primero que llama la atención es que, *grosso modo*, la situación parece ser más equilibrada que la que observamos respecto a la productividad léxica. Si el catalán supera en el número de palabras al castellano en la mayoría de las áreas (en catorce, concretamente), aquí la situación parece repartirse de forma algo más equitativa, a pesar de que sigue predominando la lengua catalana. En esta lengua el número de vocablos es mayor en once de las diecisiete áreas, mientras que el castellano descuella en seis. Cabe destacar que, en los centros en que el castellano obtiene un número de vocablos mayor, la diferencia es menor a la que se observa en dirección inversa. De media, los centros que tienen más vocablos en castellano superan al catalán en 16.5 palabras, en cambio, en sentido contrario, la lengua catalana sobrepasa al castellano en casi el doble,

concretamente con 30.8 vocablos de media. Creemos que esta tendencia no puede ser casual, sino que obedece, una vez más, al predominio del uso del catalán en nuestra provincia, así como a la mayor presencia en la muestra de alumnado cuya lengua materna es la catalana.

De otra parte, si observamos las áreas temáticas en las que el castellano sobrepasa al catalán, vemos que tres de ellas, *Calefacción, iluminación y formas de airear un recinto, La ciudad y Juegos y distracciones*, ya sobresalen con respecto a la producción de palabras en esta lengua, aunque, como vimos en el apartado anterior (*vid.* 6.1.3.), dicha diferencia carece de validez estadística. No obstante, las razones que dimos para argumentar por qué estos centros alcanzan mayor productividad en castellano también explican, a nuestro modo de ver, la superioridad respecto al número de vocablos: esto es, la preferencia por el uso de esta lengua en ámbitos relacionados con el ocio adolescente, los medios de comunicación o los mercados de consumo (en relación con el centro 15), el mayor uso del castellano en las ciudades fruto del asentamiento en estas del contingente migratorio de los años 60 (con respecto al centro 10) o la escasa incidencia de la lengua materna en un centro como el 9 que, además, es de los más improductivos.

También hay otras áreas, como *El cuerpo humano, Trabajos del campo y del jardín y Profesiones y oficios*, donde las asociaciones realizadas son más numerosas en castellano. El hecho de que esto suceda en campos donde la productividad en catalán es mayor, y, además, dicha diferencia tiene una significación estadística es sorprendente, aunque, como ya comentamos al inicio de este capítulo, los valores que miden las palabras y los vocablos son distintos⁷²⁹. Creemos que, en concreto, estas diferencias se explican porque el número de vocablos nos informa de la actuación global del conjunto de informantes, hecho que hace que, como señalamos al inicio de este apartado, los resultados sean fácilmente alterables por un solo sujeto. Los criterios de edición tampoco pueden haber influido, por cuanto hemos tenido un especial cuidado en que fueran exactamente los mismos en las dos lenguas (*vid.* 5.4). Finalmente, cabe añadir que la diferencia entre el número de vocablos en los centros donde impera el castellano oscila entre las 22 palabras de *El cuerpo humano* y las 11, de *Calefacción, iluminación*

⁷²⁹ Prueba de ello es el distinto orden que se obtiene al ordenar los centros de interés según cada parámetro (*cfr.* tabla 6.4. y 6.11).

y formas de airear un recinto, lo que no parece, para nada, unas cifras sobresalientes teniendo en cuenta que se calculan sobre la totalidad de los informantes. Lo único que consideramos que cabe destacar, en este punto, es que la diferencia a favor del castellano que se observa en centros como *El cuerpo humano* o *Trabajos del campo y del jardín*, provoca que los rangos de los centros no coincidan en las dos lenguas (vid. tabla 6.11)

La situación no es la misma con respecto a las categorías donde la riqueza léxica es mayor en lengua catalana. Aquí las diferencias oscilan entre los 87 vocablos de más que alcanza *Animales* y los 10 palabras que separan el centro *Colores* en castellano y en catalán. A pesar de que el índice más adecuado para las comparaciones es el promedio de palabras y de la posible existencia de informantes que realicen asociaciones inusuales, consideramos que la diferencia obtenida en relación con los vocablos en algunos centros ligados a la familia como *Comidas y bebidas*, *Objetos colocados en la mesa para la comida* o *La cocina y sus utensilios*, no indica otra cosa que el predominio de la lengua materna catalana en la mayoría de nuestros informantes. De otra parte, el peso de la escolarización, realizada íntegramente en la lengua vernácula —como hemos ido repitiendo—, puede estar condicionando la mayor riqueza léxica de algunas áreas que se trabajan en el currículum como la ya mencionada *Comidas y bebidas*, u otras como *Medios de transporte* o *Animales*⁷³⁰. Es curiosa la diferencia tan notable que se observa en este último centro, que tiene 87 palabras distintas más en lengua catalana, puesto que, si bien la productividad en torno a esta área era ligeramente superior también en esta lengua, dicha diferencia, como vimos, no era significativa estadísticamente. En los restantes centros las divergencias son menores, por ejemplo, en *Comidas y bebidas* y en *Objetos colocados en la mesa para la comida* es de 47 y 44 vocablos, respectivamente, a favor del catalán. En las siguientes áreas temáticas la diferencia oscila entre 20 y 30 palabras distintas: *Medios de transporte* (27), *La cocina y sus utensilios* (26), *La escuela* (24), *Ropa y complementos* (22), *El campo* (21) y *Partes de la casa* (20). Finalmente, en dos centros de interés, *Muebles de la casa* y *Colores*, la distancia es realmente tan modesta (11 y 10 vocablos respectivamente) que consideramos que sería irrelevante sacar alguna conclusión a partir de esta cifra.

⁷³⁰ Sin embargo, *El cuerpo humano*, que también se trabaja en la escuela, obtiene mejores valores en cuanto al número de vocablos en lengua castellana, lo cual no deja de ser sorprendente, aunque la diferencia sea de solo 22 palabras diferentes.

6.1.5. Densidad léxica e índice de cohesión de los centros de interés

A continuación, examinamos otros dos índices importantes para la descripción de la disponibilidad léxica en catalán y en castellano: la *densidad* y el *índice de cohesión*. Ambas magnitudes miden el grado de coincidencia existente entre las respuestas de los informantes, esto es, nos dan una idea del léxico que comparten los encuestados ante un mismo estímulo. El primer índice, la *densidad*, es resultado de dividir el total de palabras en cada uno de los centros de interés por el número de vocablos. En las primeras investigaciones hispánicas sobre disponibilidad fue el único parámetro usado para analizar la convergencia asociativa de las respuestas. Sin embargo, hoy en día, todos los investigadores coinciden en señalar que los valores obtenidos mediante el índice de cohesión resultan, claramente, más adecuados. Al igual que la densidad, el *índice de cohesión* es una fórmula matemática que relaciona el número de palabras y de vocablos con el fin de caracterizar la mayor o menor dispersión del léxico consignado en torno a los centros de interés.

Para obtener el nivel de cohesión de un determinado centro de interés se divide el promedio de respuestas de cada área asociativa por el número de vocablos de dicha área. En el caso, hipotético, de que todos los sujetos dieran exactamente las mismas respuestas, al calcular el índice de cohesión, este daría 1. Por tal razón, los centros que se acercan más a este valor son aquellos en los que la coincidencia entre los informantes es mayor —los denominados centros *compactos* o *cerrados*—, mientras que los centros que propician respuestas más variadas corresponden a áreas donde la convergencia entre los informantes es menor —los centros *abiertos* o *difusos*—⁷³¹. La densidad tiene el fuerte inconveniente de carecer de límites rigurosos, mientras que el índice de cohesión, por el contrario, siempre alcanza un valor entre 0 y 1. Además, como el dividendo que utiliza la densidad en su fórmula es el total de respuestas, este parámetro se ve directamente influido por el número de informantes que componen la investigación e imposibilita la comparación entre muestras desiguales.

⁷³¹ Echeverría (1991: 62), creador de la fórmula de cálculo de este índice, afirma al respecto: “nuestro índice de cohesión informa si el centro de interés es cerrado o compacto, o más bien abierto o difuso”. El autor ya utiliza este índice en su trabajo de 1987.

El número de informantes también influye en el índice de cohesión, pero de forma mucho más parcial e indirecta. Cuanto más crece el tamaño de la muestra, más vocablos se recogen, y más pequeño, por ende, es el índice de cohesión. Por consiguiente, este parámetro deja de ser efectivo cuando se emplea con subgrupos derivados de la estratificación de la muestra (Fernández Smith *et al.* 2008: 93), porque, en general, el estrato con menos miembros siempre tendrá el índice de cohesión mayor⁷³². Hernández Muñoz (2004: 61-62) y Fernández Smith *et al.* (2008: 97-98) llevan a cabo distintos análisis estadísticos que demuestran esta relación, inversamente proporcional, entre el número de vocablos y el índice de cohesión⁷³³.

De otra parte, la cohesión de las áreas temáticas está relacionada con la estructura interna de las lenguas, por eso se ha afirmado que, en las áreas más concretas, se observa un nivel de coincidencia más alto en las respuestas de los informantes (Carcedo 2001: 61). Desde una perspectiva similar, Echeverría *et al.* (1987: 57) ya señalan la importancia del número de clases lógicas que componen cada campo temático⁷³⁴. Más recientemente, Hernández Muñoz (2006: 155), desde un nuevo enfoque cognitivo, afirma que “el índice de cohesión es una manifestación cuantitativa de los efectos que produce en las respuestas de los hablantes el hecho de que los centros de interés sean tan diferentes en su concepción epistemológica”. Tras preguntarse qué razones estructurales llevan a que haya centros “compactos” o “difusos”, clasifica todos los centros de interés según su estructura interna y concluye que las áreas que generalmente obtienen un índice de cohesión bajo son aquellas que tienen estructuras relacionales o radiales (*El campo, La ciudad, La escuela, La cocina y sus utensilios, Calefacción e iluminación y Juegos y distracciones*) o son categorías *ad hoc* (*Objetos colocados en la mesa para la comida*), mientras que las que ostentan índices de cohesión altos, responden a categorías denominadas “naturales” en los estudios psicolingüísticos (*Partes del cuerpo, La ropa, Partes de la casa, Alimentos y bebidas, Medios de transporte y Animales*). El centro *Colores*, que Hernández Muñoz (2006) no

⁷³² Por ejemplo, el nivel sociocultural que esté compuesto por menos informantes ostentará un nivel mayor de cohesión por área asociativa que los restantes niveles (*vid.* 6.2.4).

⁷³³ Con anterioridad, Gómez Molina y Gómez Devís (2004: 87) ya habían constatado que: “a mayor índice de cohesión, menor número de vocablos, y a menor índice de cohesión, mayor número de palabras diferentes”.

⁷³⁴ “Ciertos centros de interés alcanzan un alto grado de coherencia semántica, en el sentido [de] que se les puede limitar a un número reducido de clases lógicas; otros, en cambio, presentan un débil grado de coherencia semántica, porque el número de clases lógicas que los componen es difícilmente delimitable” (Echeverría *et al.* 1987: 57).

analiza, sería una categoría natural, e incluso “bien definida”, siguiendo la clasificación que utiliza esta misma autora.

Hechas ya estas consideraciones, pasemos a analizar los datos obtenidos. Hemos de decir que, en el presente trabajo, incluimos los resultados con respecto a la densidad, con la finalidad de que puedan compararse con otros estudios que ofrecen también este parámetro, aunque —por las razones que hemos argüido al inicio de este apartado— nos centramos en el análisis del índice de cohesión. De cualquier modo, los resultados que obtenemos con estos dos índices son muy similares, en el sentido de que, aunque las cifras absolutas varían (*cf.* tablas 6.12 y 6.13), el rango que obtiene cada área según los dos parámetros en ambas lenguas es exactamente el mismo (*vid.* tabla 6.14)⁷³⁵.

Tabla 6.12. Densidad léxica por centro de interés en castellano y en catalán.

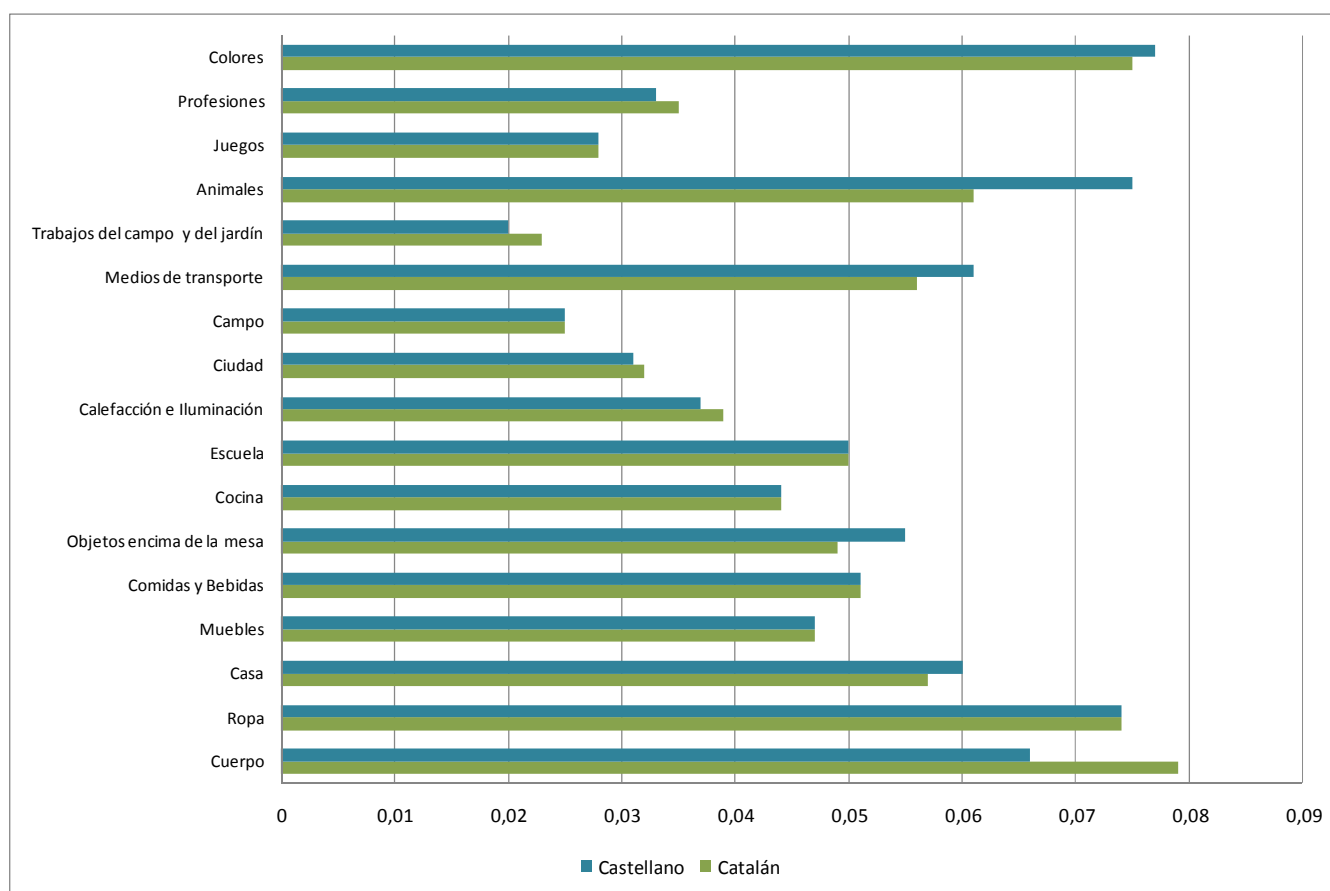
Nº	Centro de Interés	Castellano	Catalán
1	El cuerpo humano	15.88	19.02
2	Ropa y complementos	17.83	17.78
3	Partes de la casa (sin los muebles)	14.29	13.58
4	Muebles de la casa	11.31	11.21
5	Comidas y bebidas	12.21	12.15
6	Objetos colocados en la mesa para la comida	13.32	11.79
7	La cocina y sus utensilios	10.64	10.60
8	La escuela (muebles y materiales)	12.10	11.96
9	Calefacción. iluminación y formas de airear un recinto	8.96	9.27
10	La ciudad	7.56	7.75
11	El campo	6.01	5.90
12	Medios de transporte	14.65	13.54
13	Trabajos del campo y del jardín	4.90	5.58
14	Animales	18.00	14.56
15	Juegos y distracciones	6.80	6.73
16	Profesiones y oficios	7.97	8.47
17	Colores	18.43	17.97

⁷³⁵ Por esta razón, no incluimos la clasificación de los centros de interés según la densidad léxica, puesto que obtendríamos el mismo orden que refleja la tabla 6.14.

Tabla 6.13. Índice de cohesión por centro de interés en castellano y en catalán⁷³⁶.

Nº	Centro de Interés	Castellano	Catalán
1	El cuerpo humano	0.066	0.079
2	Ropa y complementos	0.074	0.074
3	Partes de la casa (sin los muebles)	0.060	0.057
4	Muebles de la casa	0.047	0.047
5	Comidas y bebidas	0.051	0.051
6	Objetos colocados en la mesa para la comida	0.055	0.049
7	La cocina y sus utensilios	0.044	0.044
8	La escuela (muebles y materiales)	0.050	0.050
9	Calefacción, iluminación y formas de airear un recinto	0.037	0.039
10	La ciudad	0.031	0.032
11	El campo	0.025	0.025
12	Medios de transporte	0.061	0.056
13	Trabajos del campo y del jardín	0.020	0.023
14	Animales	0.075	0.061
15	Juegos y distracciones	0.028	0.028
16	Profesiones y oficios	0.033	0.035
17	Colores	0.077	0.075

Gráfico 6.9. Índice de cohesión por centro de interés en castellano y en catalán.



⁷³⁶ Hemos calculado el índice de cohesión a partir de los datos anteriores (tablas 6.5 y 6.10), puesto que el programa *Dispolx* ofrece solo dos decimales y nos parecía que la información era insuficiente. En el análisis de las variables sociales, por el contrario, incluimos el índice tal como lo proporciona este programa informático, puesto que allí este valor no cobra tanta importancia como en el presente apartado, donde analizamos los datos globalmente.

Como se observa en las tablas y en el gráfico anterior, las áreas temáticas más compactas y más difusas coinciden en ambas lenguas, con lo que podemos ya adelantar que la caracterización que se obtiene de los centros, según este parámetro, parece ser constante por encima de la lengua en la que se completa la encuesta. No obstante, hay unas pequeñas diferencias que, a continuación, comentaremos. En castellano y en catalán, *Colores* es uno de los centros más cerrados del conjunto, toda vez que ocupa la primera posición en castellano y la segunda, en catalán (*vid.* tabla 6.14). Ostenta la tercera posición, en ambas lenguas, *Ropa y complementos*, donde también se observa una gran convergencia entre las respuestas de los informantes. Sin embargo, llama la atención la diferencia entre el centro *Animales*, segundo en castellano y cuarto en catalán, y *El cuerpo humano*, que es el centro más cohesionado en catalán y que destaca con la cifra más alta de todas las áreas asociativas (0.079), pero, en cambio, es el cuarto en lengua castellana. De cualquier modo, lo que explica los altos niveles de cohesión de estas cuatro áreas (1, 2, 14 y 17) es el hecho de que son categorías naturales e inclusivas, también denominadas *universales*, donde sus miembros se organizan en torno a un prototipo y “se relacionan con la etiqueta categorial según una relación de pertenencia estricta” (Hernández Muñoz 2006: 150)⁷³⁷.

De otra parte, si nos fijamos en los centros más abiertos o difusos, las coincidencias son aún mayores entre las dos lenguas. En ambas obtienen el índice de cohesión menor *Trabajos del campo y del jardín*, *El campo* y *Juegos y distracciones*. Este hecho es lógico, puesto que, como ya hemos indicado, la cohesión se relaciona con el número de vocablos y, como vimos en el apartado anterior (*vid.* 6.1.4.), estos tres estímulos ostentan valores muy elevados en esa magnitud. Concretamente, *El campo* ocupa el primer puesto en ambas lenguas, mientras que *Juegos y distracciones* alcanza la cuarta posición. El área 13, *Trabajos del campo y del jardín*, se queda en la sexta posición en castellano y en la octava en catalán; pero, si recordamos su escaso valor respecto a la productividad, se constata que las actualizaciones de los informantes, en

⁷³⁷ La autora se refiere en esta última afirmación a los centros “inclusivos”, donde es posible enunciar una proposición lógica del tipo “x es Y”, esto es, “*el pantalón es ropa*” (Hernández Muñoz 2006: 150). De otra parte, el primer centro de interés, enunciado en muchos trabajos como *Partes del cuerpo humano* es, según esta investigadora, una categoría no ya natural sino “bien definida”. Por otro lado, ya señalamos en un trabajo anterior (Serrano Zapata 2003b: 78) que las posibilidades semánticas que ofrecen algunas áreas temáticas son reducidas, puesto que no existe una infinidad de animales, ni de partes del cuerpo o de prendas de ropa —o de colores, añadiríamos ahora—.

torno a este estímulo, son escasas y, además, muy distintas. Las concomitancias entre las dos lenguas en los centros más difusos no quedan ahí, sino que se extienden a la parte final de la tabla (*vid.* tabla 6.14). Desde la décima posición hasta la decimocuarta también coinciden *Muebles de la casa*, *La cocina y sus utensilios*, *Calefacción, iluminación y formas de airear un recinto*, *Profesiones y oficios* y *La ciudad*. Recordemos que estas dos últimas áreas asociativas ya alcanzan, según el número de palabras diferentes, posiciones muy elevadas, concretamente la tercera y la segunda, respectivamente en las dos lenguas (*vid.* tabla 6.11). Casos más llamativos son los de los centros *Muebles de la casa* o *Calefacción, iluminación y formas de airear un recinto*, puesto que ambos se sitúan por debajo de la media con respecto al número de vocablos. El bajo índice de cohesión obtenido se explica porque su productividad es también muy escasa (*vid.* tabla 6.5).

Tabla 6.14. Rango de los centros de interés en cada una de las lenguas según el índice de cohesión (de mayor a menor)⁷³⁸.

Rango	Castellano		Catalán	
1	17. Colores	0.077	1. El cuerpo humano	0.079
2	14. Animales	0.075	17. Colores	0.075
3	2. Ropa y complementos	0.074	2. Ropa y complementos	0.074
4	1. El cuerpo humano	0.066	14. Animales	0.061
5	12. Medios de transporte	0.061	3. Partes de la casa	0.057
6	3. Partes de la casa	0.060	12. Medios de transporte	0.056
7	6. Objetos colocados en la mesa para la comida	0.055	5. Comidas y bebidas	0.051
8	5. Comidas y bebidas	0.051	8. La escuela	0.050
9	8. La escuela	0.050	6. Objetos colocados en la mesa para la comida	0.049
10	4. Muebles de la casa	0.047	4. Muebles de la casa	0.047
11	7. La cocina y sus utensilios	0.044	7. La cocina y sus utensilios	0.044
12	9. Calefacción, iluminación y formas de airear un recinto	0.037	9. Calefacción, iluminación y formas de airear un recinto	0.039
13	16. Profesiones y oficios	0.033	16. Profesiones y oficios	0.035
14	10. La ciudad	0.031	10. La ciudad	0.032
15	15. Juegos y distracciones	0.028	15. Juegos y distracciones	0.028
16	11. El campo	0.025	11. El campo	0.025
17	13. Trabajos del campo y del jardín	0.020	13. Trabajos del campo y del jardín	0.023

⁷³⁸ Las áreas que coinciden en el mismo rango se marcan de forma coloreada en la tabla.

De cualquier modo, es sorprendente la enorme semejanza entre la cohesión de los centros en ambas lenguas. En concreto, la coincidencia en la cifra exacta obtenida es total en nada menos que en siete estímulos: *Ropa y complementos* (0.074), *Muebles de la casa* (0.047), *Comidas y bebidas* (0.051), *La cocina y sus utensilios* (0.044), *La escuela* (0.050), *El campo* (0.025) y *Juegos y distracciones* (0.028). Esta circunstancia señala que, a pesar de las diferencias existentes en la productividad y riqueza léxica, la cohesión interna de las áreas temáticas presenta una gran similitud en los dos idiomas. Las diferencias, cuando se dan, parecen poco importantes —debido, en parte, a la propia naturaleza del índice de cohesión, que cuenta, como hemos dicho, con límites referenciales precisos—. Por esta razón, incluso la media del conjunto de áreas asociativas es la misma para las dos lenguas (0.049). Tampoco se aprecian diferencias entre la distancia que separa el área más difusa de la más compacta; en ambos casos es prácticamente idéntica (0.052 en castellano y 0.054 en catalán)⁷³⁹.

De otra parte, la diferencia entre los dos idiomas en la mayoría de los estímulos verbales es de solo unas milésimas, como sucede en *La ciudad* (0.031-0.032)⁷⁴⁰, en *Calefacción, iluminación y formas de airear un recinto* (0.037-0.039), en *Profesiones y oficios* (0.033-0.035) y en *Trabajos del campo y del jardín* (0.020-0.023), todas ellas más cohesionadas en catalán que en castellano. De forma inversa, en los estímulos *Partes de la casa* (0.060-0.057), *Objetos colocados en la mesa para la comida* (0.055-0.049), *Medios de transporte* (0.061-0.056), y *Colores* (0.077-0.075), las respuestas de los informantes muestran una mayor convergencia en castellano. Las divergencias más acusadas entre las dos lenguas se dan en *El cuerpo humano* (0.066-0.079) y en *Animales* (0.075-0.061). Por dicha razón, los dos estímulos, como hemos visto, ostentan rangos muy distintos en la clasificación de cada idioma. Para tratar de explicar este hecho, analizamos detalladamente estas dos áreas a continuación.

Con respecto al primero de los centros de interés, ya hemos visto que es un centro donde la productividad en lengua catalana es notablemente superior (*vid.* 6.1.2. y 6.1.3.) a la castellana. Además, el número de vocablos, que mantiene —como hemos indicado— una relación inversamente proporcional con la cohesión, es mayor en

⁷³⁹ De cualquier forma, es llamativo que siempre haya mayor divergencia entre las áreas léxicas en lengua catalana, tal como vimos que ocurría al analizar los totales de palabras, los promedios y los vocablos.

⁷⁴⁰ En este y en los restantes paréntesis del párrafo se incluye, en primer lugar, el valor alcanzado por el índice de cohesión en castellano y, a continuación, en lengua catalana.

castellano, con una diferencia de 22 vocablos. Estas dos circunstancias explican que este centro alcance en catalán un valor mucho más alto con respecto a la cohesión. Se constata, por ende, que el alumnado produce en lengua catalana un mayor número de palabras ante ese estímulo verbal, pero estas son más coincidentes que las que producen, ante el mismo estímulo, en lengua castellana, donde se observa una productividad menor pero una mayor dispersión en las respuestas.

Un caso distinto es el del área *Animales*, ya que la productividad en los dos idiomas es prácticamente equivalente. Concretamente, en castellano, el promedio de palabras por centro se sitúa en 26.02 y, en catalán, en 26.32, una diferencia de tres décimas que repercute escasamente en el valor del índice de cohesión. Lo que sí tiene relevancia es la diferencia en la cantidad de vocablos: en catalán hay 87 más que en castellano, lo que provoca que la cohesión sea menor en esa lengua. Es decir, a pesar de que el número de palabras que los informantes escriben bajo el estímulo es prácticamente idéntico, en catalán las respuestas son más variadas, lo que da lugar a que haya una menor coincidencia entre ellas y, por tanto, un índice de cohesión más bajo. La explicación que puede tener este comportamiento es que esta área se refuerza en el currículum escolar, lo que puede llevar a que los alumnos conozcan un mayor número de palabras distintas en ese idioma, a pesar de que la productividad alcance proporciones similares en las dos lenguas⁷⁴¹.

Si atendemos a la información que recoge la tabla 6.14, podemos clasificar los centros de interés en tres grandes grupos que coinciden, *grosso modo*, en las dos lenguas:

(a) los centros de cohesión alta, es decir, muy compactos o cerrados, cuyo índice supera el 0.6: *El cuerpo humano, Ropa y complementos, Animales, Colores* y, en lengua castellana, *Medios de transporte*;

(b) los centros con una cohesión media, cuyo índice oscila entre 0.4 y 0.6: *Partes de la casa, Muebles de la casa, Comidas y bebidas, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, La escuela*, y, en catalán, *Medios de transporte*; y, finalmente,

⁷⁴¹ En el capítulo 7, cuando tengamos información sobre los fenómenos de contacto lingüístico detectados en los dos corpus, daremos otra explicación a este mismo hecho (*cfr.* 7.3.)

(c) los centros dispersos o difusos, cuyo índice es inferior al 0.4: *Calefacción, iluminación y formas de airear un recinto, La ciudad, El campo, Trabajos del campo y del jardín, Juegos y distracciones y Profesiones y oficios.*

La diferencia que se observa en la clasificación respecto al centro 12, *Medios de transporte*, estriba en el mayor número de vocablos que obtiene esta área en catalán, dado que el promedio de palabras es similar en ambas lenguas. Como ya dijimos (*vid.* 6.1.4.) la mayor riqueza léxica de este centro de interés, en lengua catalana, puede relacionarse con el hecho de que es un área integrada en el currículum escolar, al igual que sucede en el centro 14, *Animales*, que —al contener un mayor número de palabras diferentes— obtiene un índice de cohesión menor en catalán⁷⁴².

Si nos fijamos en el orden de los centros de la tabla 6.14, llama la atención las grandes concomitancias en los rangos de las dos lenguas, que no solo coinciden en las ocho últimas posiciones sino también en *Ropa y complementos*, que ostenta el tercer puesto en la clasificación en los dos ránkines. En las restantes posiciones, las similitudes son mayores que las diferencias, toda vez que generalmente las categorías divergen en una posición, salvo los casos ya analizados.

6.1.6. Valoración de los resultados según las variables lingüísticas

A continuación, incluimos dos tablas como resumen de la incidencia cuantitativa que han tenido las variables lingüísticas, esto es, las diecisiete áreas temáticas en los resultados de las dos lenguas.

⁷⁴² Sin embargo, Martínez Olmos (2008: 168) señala una relación directa entre los centros más cohesionados y las áreas temáticas incluidas en el currículum de léxico escolar.

Tabla 6.15. Rangos de los centros de interés en las dos lenguas según distintos parámetros⁷⁴³.

Centro de interés	PALABRAS		VOCABLOS		COHESIÓN	
	CASTELLANO	CATALÁN	CASTELLANO	CATALÁN	CASTELLANO	CATALÁN
1. CUERPO	2	1	8	10	4	1
2. ROPA	4	4	12	11	3	3
3. CASA	13	13	16	16	6	5
4. MUEBLES	15	15	14	15	10	10
5. COMIDAS	3	2	5	5	8	7
6. OBJETOS	14	14	17	14	7	9
7. COCINA	12	11	9	9	11	11
8. ESCUELA	6	5	7	6	9	8
9. CALEFACCIÓN	16	16	11	13	12	12
10. CIUDAD	5	6	2	2	14	14
11. CAMPO	9	9	1	1	16	16
12. TRANSPORTE	10	10	13	12	5	6
13. TRABAJOS	17	17	6	8	17	17
14. ANIMALES	1	3	10	7	2	4
15. JUEGOS	11	12	4	4	15	15
16. PROFESIONES	7	7	3	3	13	13
17. COLORES	8	8	15	17	1	2

En la tabla anterior, se observa el rango que obtiene cada centro de interés en castellano y en catalán según los tres parámetros (promedio de palabras, vocablos e índice de cohesión) analizados hasta el momento. De esta forma, tenemos una visión conjunta de la relevancia que cobra cada área temática. Lo primero que podemos señalar es que la coincidencia mayor entre castellano y catalán se da en el índice relativo a las palabras. En este sentido, hasta diez estímulos verbales comparten el mismo rango. En cambio, las similitudes son menores en cuanto a los vocablos, donde solo comparten idéntica posición siete áreas asociativas. Como ya dijimos, estudios realizados con anterioridad —donde se comparan los resultados de diversas sintopías hispánicas— ya constatan este mismo hecho: la distribución de los centros según el número de palabras presenta una mayor semejanza que la que se obtiene mediante la cantidad de vocablos (Gómez Molina y Gómez Devís 2004: 78). Esto se debe a que las palabras, y por ende sus promedios, son índices más estables que los vocablos, que están más sujetos a las posibles variaciones individuales. En cuanto al índice de cohesión, este presenta bastantes similitudes —concretamente, nueve áreas alcanzan el mismo rango en

⁷⁴³ Las filas en las que hay coincidencia entre las dos lenguas en alguno de los parámetros se marcan de forma coloreada en la tabla. En la primera columna, se han sombreado los centros de interés en los que hay coincidencia entre las dos lenguas en al menos un valor de los tres analizados.

castellano y en catalán—, dado que este parámetro informa sobre la estructura interna de las lenguas y esta tiende a ser similar, al menos, en las dos lenguas de nuestra comunidad.

Por otro lado, hay que añadir que, incluso cuando no coinciden en la misma posición, hay mayor similitud entre los rangos de las palabras y del índice de cohesión que entre los de los vocablos. Así, por ejemplo, con respecto al promedio de palabras, salvo *Animales*, los restantes centros difieren únicamente en una posición (*El cuerpo humano* es segundo en castellano y primero en catalán, etc.). Por el contrario, en relación con los vocablos, las diferencias relativas a los rangos son mayores: en dos casos de tres posiciones (*Objetos colocados en la mesa sobre la comida* y *Animales*), en cuatro de dos posiciones (*El cuerpo humano*, *Calefacción, iluminación y formas de airear un recinto*, *Trabajos del campo y del jardín* y *Colores*), mientras que el resto dista en una única posición. Según el índice de cohesión, las lenguas muestran mayor semejanza que con respecto a los vocablos, pero se dan menos coincidencias que las observadas según el número de palabras (cinco áreas difieren en una posición, dos en dos, y una, en tres).

Si observamos las similitudes entre los rangos que obtiene cada centro de interés según los tres parámetros, podemos afirmar que, en general, no se observa ninguna correlación entre las palabras y los vocablos, a pesar de que algunos estímulos coinciden en quedar por encima de la media en las dos magnitudes y en ambas lenguas, como *Comidas y bebidas* o *La ciudad*; o por debajo, como *Partes de la casa*, *Muebles de la casa* u *Objetos colocados en la mesa para la comida*. No obstante, también se observan divergencias como el centro *Trabajos del campo y del jardín*, que queda último con respecto a las palabras y, en cambio, ocupa la sexta posición en castellano, y la octava en catalán, con respecto al número de vocablos. También en *Juegos y distracciones*, se observan estas discrepancias: en este caso, este campo ostenta un puesto bajo en productividad (undécimo y duodécimo, respectivamente) y, en cambio, la cuarta posición en palabras distintas.

Por el contrario, entre la cantidad de vocablos y el índice de cohesión sí puede reconocerse una relación inversa, como ya hemos visto en el apartado anterior (*vid.* 6.1.5.). De este modo, la mayoría de las áreas que alcanzan los primeros puestos con

respecto al número de palabras distintas, exhiben, por tanto, un índice de cohesión menor, y consiguientemente, quedan al final en este parámetro. Esto puede comprobarse, fácilmente, en la tabla 6.15, donde *El campo*, *La ciudad*, *Profesiones y oficios* y *Juegos y distracciones* encabezan el ranquin de vocablos y quedan en último lugar en relación con la cohesión. La tendencia contraria, aunque no se aprecia de forma tan clara, puede reconocerse en *Colores* y en *Ropa y complementos*.

Entre las palabras y el índice de cohesión no se aprecia ninguna correlación, a pesar de que, como ya sabemos, la cohesión utiliza el promedio de palabras en su fórmula. En algunos centros, como en *Trabajos del campo y del jardín* o en *La cocina y sus utensilios*, se da una concomitancia prácticamente total en los rangos; sin embargo, también se aprecian otras tendencias en campos como *Medios de transporte* o *Profesiones y oficios*, con lo que concluimos que no se puede reconocer un patrón claro de dependencia entre estas dos magnitudes.

En la siguiente tabla, se presenta el resumen de las diferencias halladas entre las dos lenguas en los tres parámetros analizados. Se señala, mediante un símbolo de confirmación, la lengua que ha arrojado un resultado mejor en cada parámetro. Es importante resaltar que, de los tres valores, el que mejor retrata el predominio de una u otra lengua en un contexto de bilingüismo estable como el nuestro, es, por su estabilidad, el promedio de palabras. Al respecto, las áreas en que, además, la diferencia entre las dos lenguas resulta significativa estadísticamente se marcan con un doble signo de confirmación.

Tabla 6.16. Resumen de las diferencias halladas entre las dos lenguas según distintos parámetros.

Centro de interés	Palabras		Vocablos		Índice de cohesión	
	Castellano	Catalán	Castellano	Catalán	Castellano	Catalán
1. CUERPO		✓✓	✓			✓
2. ROPA		✓✓		✓	=	=
3. CASA		✓		✓	✓	
4. MUEBLES		✓		✓	=	=
5. COMIDAS		✓✓		✓	=	=
6. OBJETOS		✓✓		✓	✓	
7. COCINA		✓✓		✓	=	=
8. ESCUELA		✓✓		✓	=	=

9. CALEFACCIÓN	✓		✓			✓
10. CIUDAD	✓		✓			✓
11. CAMPO		✓		✓	=	=
12. TRANSPORTE		✓		✓	✓	
13. TRABAJOS		✓✓	✓			✓
14. ANIMALES		✓		✓	✓	
15. JUEGOS	✓		✓		=	=
16. PROFESIONES		✓✓	✓			✓
17. COLORES		✓		✓	✓	

Como se aprecia, la preponderancia de la lengua catalana se constata tanto en la productividad (total y promedio de palabras) como en la riqueza léxica (vocablos). En la primera de estas dos magnitudes se confirma de forma más contundente, puesto que el catalán exhibe mejores resultados en catorce de las diecisiete áreas temáticas. En ocho de estas áreas (*El cuerpo humano, Ropa y complementos, Comidas y bebidas, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, La escuela, Trabajos del campo y del jardín y Profesiones y oficios*), además —tal como comprobamos en el apartado 6.1.3.—, la diferencia es significativa estadísticamente, esto es, extensible al universo estudiado. Cabe señalar que los tres centros de interés que muestran mejores resultados en lengua castellana —*Calefacción e iluminación y formas de airear un recinto, La ciudad y Juegos y distracciones*— también obtienen más vocablos en esta lengua, lo que viene a mostrar que son ámbitos en los que el español está más presente, aunque en ninguno de ellos la diferencia entre los promedios de palabras sea significativa. Los resultados de los campos léxicos *El cuerpo humano, Trabajos del campo y del jardín y Profesiones y oficios* no dejan de ser llamativos, puesto que estos estímulos obtienen más palabras en catalán pero más vocablos en castellano. Hay que notar, no obstante, que los vocablos son más inestables que los promedios de palabras —dado que se calculan sobre la totalidad de los informantes— y, además, la diferencia numérica en estas tres áreas a favor del castellano es modesta⁷⁴⁴.

En nuestra opinión, la gran relevancia que se observa de la lengua catalana en estos resultados se relaciona con el hecho de que la mayor parte de la muestra (el

⁷⁴⁴ Concretamente, es de 22 vocablos en *El cuerpo humano*, de 16 en *Trabajos del campo y del jardín* y de 14 en *Profesiones y oficios*, lo que no supone una gran diferencia si se tiene en cuenta que son datos calculados sobre el conjunto de los 240 informantes (*vid.* 6.1.5).

65.42%) tiene como lengua materna el catalán. Además, tampoco hay que menospreciar la influencia que puede haber tenido la escuela, donde se refuerza el léxico de algunas áreas —como *El cuerpo humano*, *Comidas y bebidas*, *Medios de transporte* o *Animales*— tanto en la productividad como en la riqueza léxica. De todos modos, el predominio de la lengua vernácula era esperable en nuestra provincia, ya que, tal como vimos en el apartado 4.3.2., esta se caracteriza, frente a las restantes provincias catalanas, por el gran peso que tiene la población autóctona en la misma y por unos altos índices tanto de uso como de identificación y de primera lengua del catalán.

El índice de cohesión muestra el resultado más equilibrado entre los dos idiomas, donde incluso muchos de los centros de interés alcanzan el mismo valor. Esto es lógico porque este parámetro no evalúa el comportamiento de los informantes sino la estructura interna de cada área asociativa (calculada según la relación entre palabras y vocablos). Por eso, en este punto, no se aprecian diferencias entre las dos lenguas, a pesar de que, como ya analizamos, algunos centros en concreto presentan valores más altos en una o en otra, como consecuencia de las diferencias observadas en las dos magnitudes anteriores (palabras y vocablos). Lo que sí se observa es la relación inversa, que ya dijimos que existía, entre vocablos e índice de cohesión. De esta forma, si observamos con detenimiento la tabla anterior, apreciamos cómo las áreas temáticas que tienen más vocablos en castellano (*El cuerpo humano*, *Calefacción e iluminación y formas de airear un recinto*, *La ciudad*, *Trabajos del campo y del jardín*, *Juegos y distracciones* y *Profesiones y oficios*) tienen un índice de cohesión, casi siempre, mayor en catalán. De forma contraria, los estímulos con más vocablos en catalán (*Ropa y complementos*, *Partes de la casa*, *Muebles de la casa*, *Comidas y bebidas*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios*, *La escuela*, *El campo*, *Medios de transporte*, *Animales* y *Colores*) tienen un índice de cohesión generalmente igual, o mayor, en castellano.

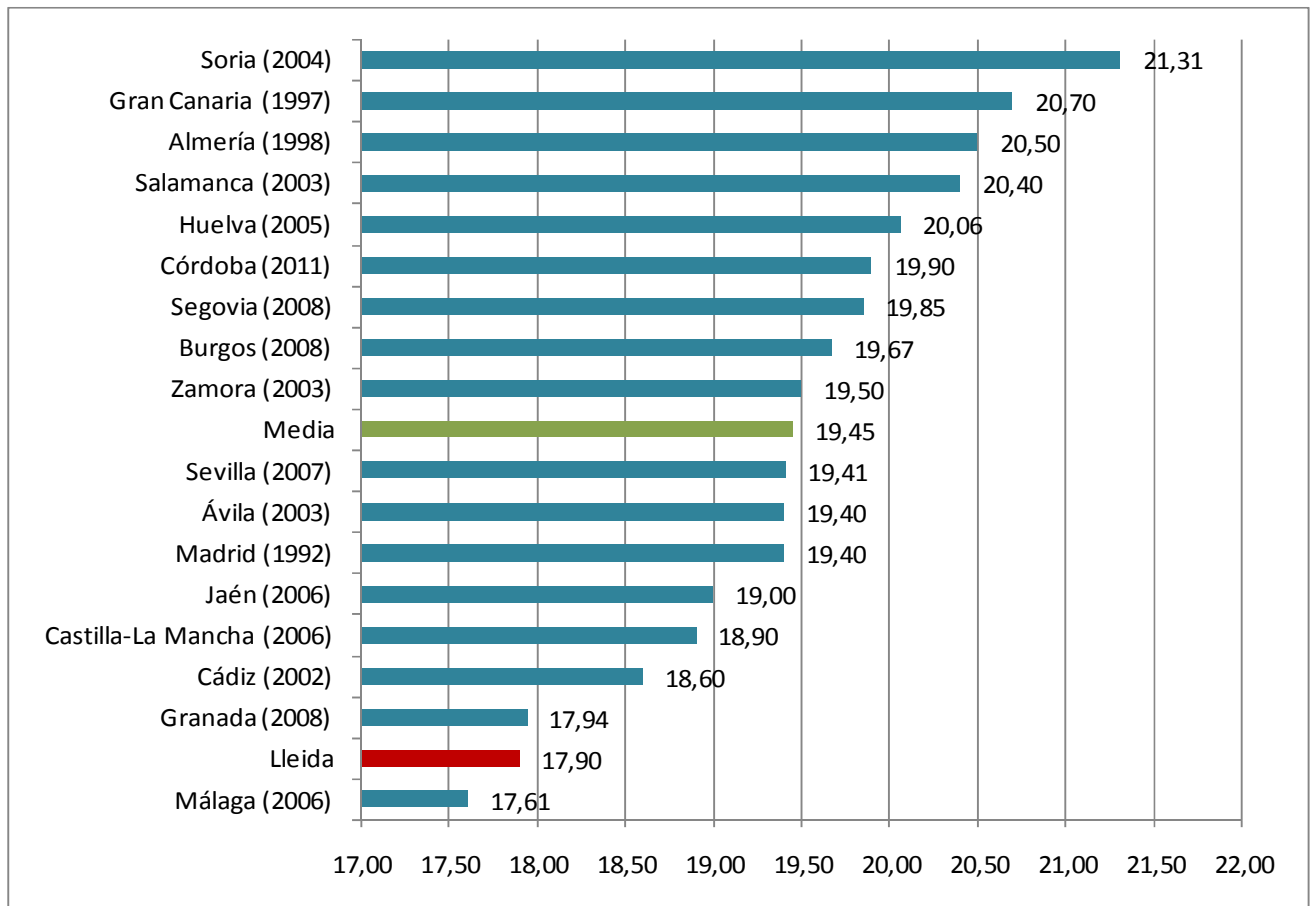
6.1.7. Comparación con otras sintopías

Es ya tradición, en los estudios sobre disponibilidad, tratar de cotejar los resultados obtenidos en la sintopía analizada con los realizados en otros lugares del mundo hispánico, para tratar de observar si se producen las mismas tendencias en cuanto a la producción de caudal léxico disponible. El establecimiento de unas pautas comunes metodológicas con respecto a la edad de los informantes, las condiciones de

reacción, los centros de interés y las variables sociales tenidas en cuenta, hacen posible esta comparación. El dato que se maneja, al respecto, es el promedio de palabras ya que los restantes parámetros —el total de palabras o de vocablos o el índice de cohesión— se encuentran, como hemos indicado repetidamente, influidos por el número de informantes que componen cada muestra.

Sin embargo, estas comparaciones no están exentas de inconvenientes. Hernández Muñoz (2004: 48) hace hincapié en el hecho de que las conclusiones interdialectales son siempre aproximativas, puesto que no hay que perder de vista que los materiales sobre los que se basan han sido previamente modificados por los investigadores, mediante los criterios de edición (*vid.* 5.4.). Algunas de las decisiones que se toman al respecto —como la relativa a las asociaciones que se admiten bajo cada estímulo— pueden tener consecuencias importantes en los resultados cuantitativos. Teniendo en cuenta esta consideración, comparamos, mediante el siguiente gráfico, la producción general de distintas sintopías monolingües españolas con el promedio global obtenido en lengua castellana en Lleida.

Gráfico 6.10. Promedio general de palabras en castellano en distintas sintopías españolas monolingües⁷⁴⁵



Tal como se plasma en el gráfico anterior, nuestros resultados se hallan por debajo de la mayoría de los estudios realizados en otras provincias. Aunque la distancia con las otras sintopías no es sustancial, sí que hay que reconocer que el alumnado de nuestra provincia actualiza una palabra y media menos (1.55) con respecto al promedio general, situado en 19.45⁷⁴⁶. El resultado de nuestra comunidad, sin embargo, es

⁷⁴⁵ Las medias de Salamanca, Zamora, Ávila, Córdoba, Cádiz, Almería, Gran Canaria y Madrid proceden de Arnal *et al.* (2004: 51) quienes, a su vez, parten de los datos de Samper, Bellón y Samper Hernández (2003: 59-60). Los promedios generales de Huelva, Málaga, Jaén, Sevilla, Burgos y Segovia los hemos calculado nosotros mismos: los de las tres primeras provincias andaluzas, a partir de la información que ofrecen Samper Padilla y Samper Hernández (2006: 29); el de Sevilla, partiendo de la información que ofrece la página web www.dispolex.com; en los casos de Burgos y Segovia, las medias proceden de los datos que brindan Fernández Juncal (2008) y Santiago Guervós (2008), respectivamente; finalmente, los datos de Soria proceden directamente de Bartol (2004) y los de Granada, de Pastor y Sánchez (2008). En los datos de Castilla-La Mancha (Hernández Muñoz 2006) hemos suprimido, a la hora de calcular la media, el centro 17, *La inteligencia*, por considerar que podía distorsionar los datos. Téngase en cuenta, de otra parte, que la fecha de publicación de los resultados —incluida en paréntesis en el gráfico— no siempre es inmediata a la recogida de los datos.

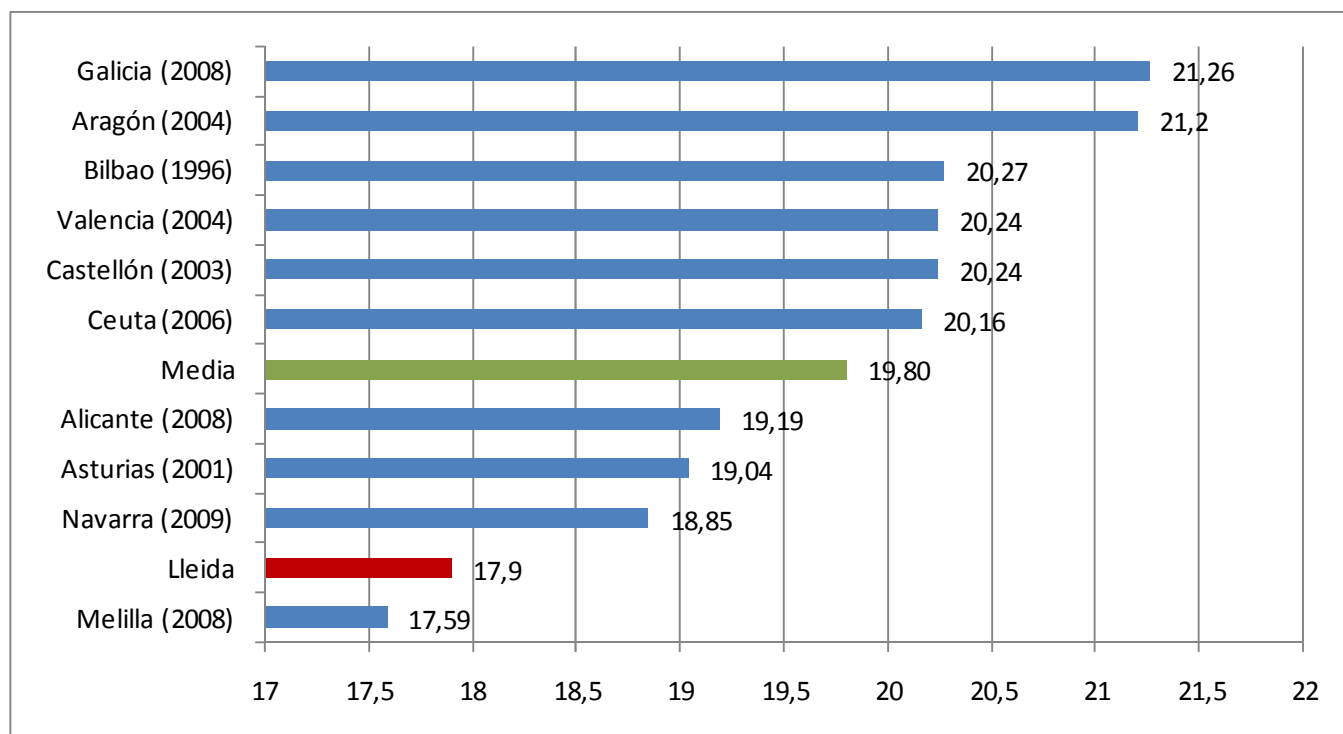
⁷⁴⁶ Cifra calculada a partir de los datos contenidos en el gráfico anterior, excluyendo nuestros propios datos.

prácticamente idéntico al de otra provincia monolingüe, Granada, e incluso supera al promedio obtenido en Málaga, lo que parece indicar que el caudal de léxico disponible es —con las salvedades que acabamos de mencionar sobre las limitaciones de este tipo de comparaciones— cuanto menos similar en nuestra provincia en relación con otros lugares donde se emplea un solo código lingüístico. Sí que hay que reconocer, no obstante, que nuestros resultados se sitúan, a todas luces, más cerca de las últimas posiciones que de las primeras.

Llegados a este punto, nos preguntamos si el hecho de trabajar con una muestra de informantes bilingües puede influir en estos datos, de forma que el dominio de dos lenguas tenga como consecuencia que los resultados relativos a la disponibilidad sean diferentes a los obtenidos en informantes monolingües. Con la finalidad de conocer si la misma tendencia se observa en otros territorios donde coexisten dos sistemas lingüísticos, llevamos a cabo el siguiente cotejo de nuestros resultados con los obtenidos en otros lugares donde el castellano convive con otra lengua dentro de España⁷⁴⁷. Debemos aclarar que, en el gráfico siguiente, hemos incluido todos los territorios españoles donde se pueden distinguir dos sistemas lingüísticos usados en una misma provincia o comunidad autónoma con una visión amplia del contacto de lenguas. Está fuera de toda duda que la situación sociolingüística de cada uno de estos lugares difiere —en algún caso enormemente— con la que se observa en nuestra provincia, y que dicha diferencia influye también en los resultados. No obstante, nos parece útil realizar esta comparación para vislumbrar posibles coincidencias y tendencias comunes.

⁷⁴⁷ Obviamente, nos referimos a que un porcentaje sustancioso de la población emplee otro código lingüístico en su interacción diaria, puesto que —como es sabido—, en la actualidad, las situaciones de multilingüismo social son parte de la cotidianidad en toda España, gracias a la inmigración reciente de otros países.

Gráfico 6.11 Promedio general de palabras en castellano en distintas sintopías españolas con más de una lengua en contacto⁷⁴⁸.



Tras analizar los datos, constatamos, al igual que ocurría en el gráfico anterior (6.10), el bajo promedio alcanzado por nuestra provincia que solo supera, en esta ocasión, la media general conseguida en Melilla. La distancia con el promedio global de todas las restantes sintopías es ahora mayor a la que se observa respecto a los lugares monolingües, concretamente de casi dos respuestas por centro de interés (1.9). Con la finalidad de comprobar si se pueden reconocer algunas tendencias similares en estos territorios donde conviven varias lenguas, incluimos la siguiente tabla (6.17) y gráfico (6.12):

⁷⁴⁸ Los datos proceden directamente de los propios autores (*vid.* nota siguiente donde se incluyen las referencias concretas), salvo en el caso de Galicia, Bilbao, Ceuta y Asturias, cuyo promedio ha sido calculado por nosotros a partir de la información que se ofrece en cada trabajo sobre el promedio de palabras en cada centro de interés. De otra parte, se excluyen nuestros propios datos para el cálculo de la media general, tanto en este gráfico como en la siguiente tabla. Téngase en cuenta también la siguiente nota al pie.

Tabla 6.17. Promedio de palabras por centro de interés en castellano en las provincias o comunidades autónomas españolas con más de una lengua en contacto⁷⁴⁹.

Nº	LL	ARA	NAV	BIL	AST	GAL	CAS	VAL	ALI	CEU	MEL	MEDIA
1	25.54	26.4	24.66	25.4	24.9	29.29	25.84	27.01	24.53	25.74	24.63	25.84
2	21.62	23.6	22.12	22.7	21.8	24.69	22.87	23.63	20.66	21.12	19.85	22.30
3	14.17	17	15.04	15	14.8	18.14	16.68	15.67	15.73	16.28	15.69	16.00
4	12.40	15.8	13.15	15.8	14.2	16.46	14.53	14.83	15.78	16.28	15.64	15.25
5	24.21	28.3	27.37	28.6	25.3	28.50	26.58	27.85	24.57	26.45	23.29	26.68
6	13.20	17.7	14.64	16.1	15	17.32	14.91	16.16	15.69	17.13	13.56	15.82
7	15.38	20.5	17.89	19.1	18.9	18.84	18.69	20.54	17.11	19.85	16.64	18.81
8	21.17	24.1	22.45	24.0	21.3	24.72	23.71	22.54	21.96	22.45	21.18	22.84
9	11.08	14	11.10	12.9	12.4	14.71	13.21	12.48	12.91	11.26	10.80	12.58
10	21.35	23.3	20.82	23.3	21.6	23.57	24.71	23.77	21.44	24.31	20.43	22.73
11	18.20	22.2	19.09	20.8	19	21.20	21.11	20.82	20.18	20.64	16.28	20.13
12	16.85	19.7	17.80	18.9	16.1	19.60	19.28	18.56	17.69	17.79	14.39	17.98
13	8.85	13.2	10.54	11.3	10.6	12.87	11.31	11.73	11.59	10.84	8.82	11.28
14	26.02	29.9	26.97	28.3	28.5	29.77	28.39	28.75	26.08	28.29	26.17	28.11
15	16.31	20.3	17.66	19.8	18.3	17.98	19.46	19.10	18.94	19.04	15.42	18.60
16	19.62	23.4	20.35	22.6	22	22.53	22.90	21.94	22.32	22.83	18.72	21.96
17	18.35	20.9	-----	20	----	21.87	19.86	18.73	-----	22.49	-----	20.64

En general, si observamos con detenimiento los datos de la tabla 6.17 y del gráfico 6.12, se aprecian las mismas tendencias ya detectadas en los estudios comparativos anteriores (Samper, Bellón y Samper Hernández 2003 y Samper Padilla y Samper Hernández 2006) con respecto a los centros más y menos productivos. De este modo, en casi todas las sintopías *Animales*, *Comidas y bebidas* y *El cuerpo humano* son, en este orden, las áreas más fecundas de la clasificación⁷⁵⁰ y, por el contrario, *Trabajos del campo y del jardín*, *Calentamiento e iluminación* y *formas de airear un recinto y Muebles de la casa*, como las más improductivas⁷⁵¹. Con respecto a los datos de Lleida, estos son, en la mayoría de estímulos, inferiores a los de las restantes sintopías y, tal como se observa en el gráfico anterior, se sitúan siempre por debajo de la media global

⁷⁴⁹ Fuentes: ARA= Aragón: Arnal *et al.* (2004), NAV= Navarra: Areta (2009), BIL= Bilbao: Etxebarria (1996b), AST= Asturias: Carcedo (2001), GAL= Galicia: López Meirama (2008), CAS= Castellón: datos facilitados amable y rápidamente por Manuela Casanova Avalos, directora de la investigación, VAL= Valencia: Gómez Molina y Gómez Devís (2004), ALI= Alicante: Martínez Olmos (2008), CEU= Ceuta: Ayora (2006) y MEL= Melilla: Fernández Smith *et al.* (2008). Es importante señalar que el centro 17 de la tabla siempre corresponde a la categoría *Colores*. De hecho, en la investigación ceutí se incluyen tres centros más: *El Mar*, *Defectos y cualidades físicos y morales* y *Acciones*, en Galicia, uno más: *El Mar* y en Navarra el centro 17, que no incluimos, corresponde a *Virtudes y defectos*. No obstante, hemos calculado la media de estas tres sintopías (datos del gráfico 6.11) sin tener en cuenta estos centros adicionales por considerar que podían distorsionar la media general.

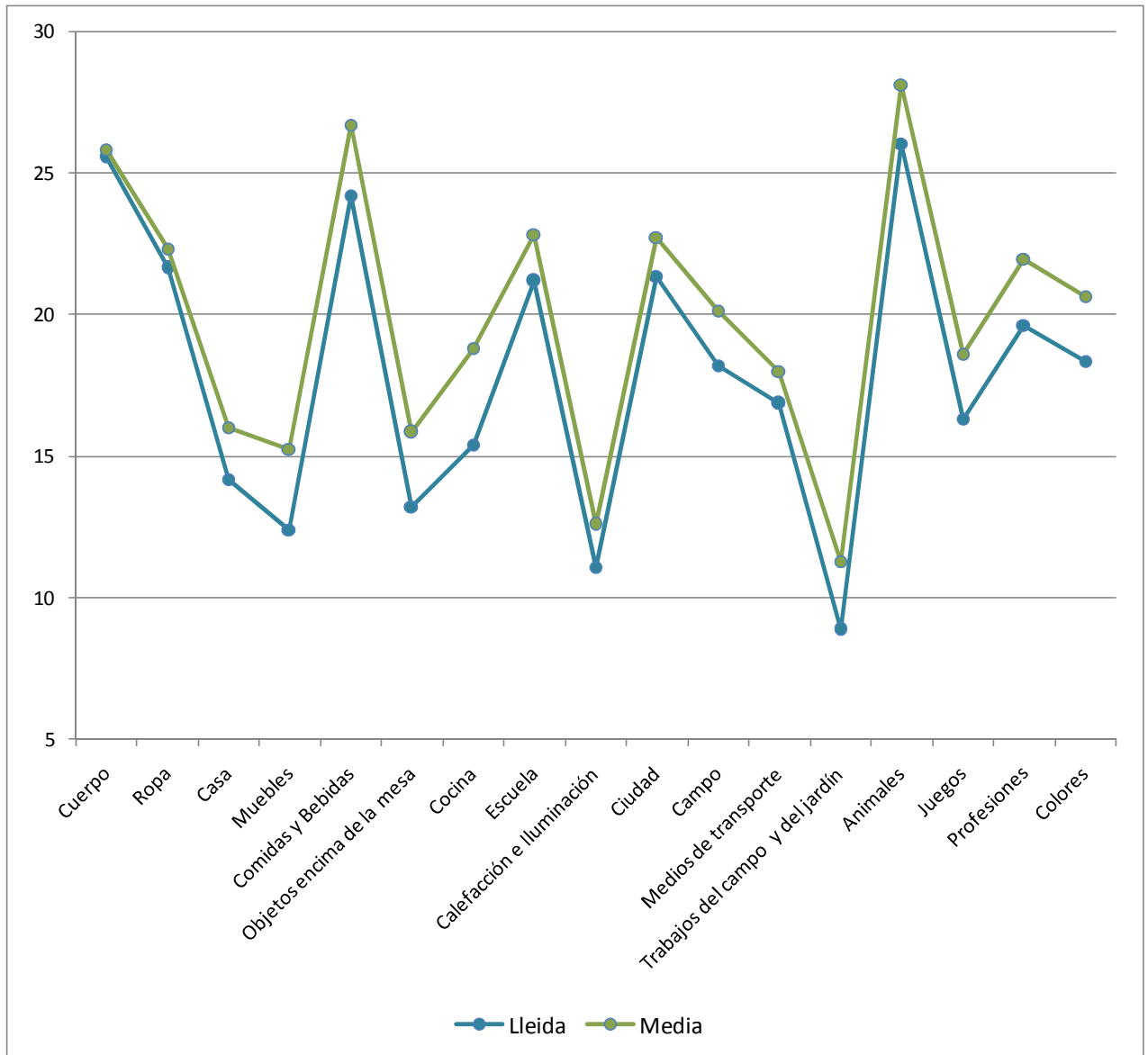
⁷⁵⁰ El orden de estas tres áreas varía en Bilbao, Navarra, Galicia y Melilla. Nuestros datos coinciden con los de Melilla y Galicia con respecto al orden de los tres primeros centros: (1) *Animales*, (2) *El cuerpo humano* y (3) *Comidas y bebidas*. De otra parte, recordemos que en muchos lugares el área léxica *El cuerpo humano* se sustituye por *Partes del cuerpo humano*.

⁷⁵¹ En Alicante y Melilla es *Objetos colocados en la mesa para la comida* el que ocupa el antepenúltimo lugar, mientras que en Bilbao ocupa esta posición *Partes de la casa*.

de cada centro de interés. No obstante, es llamativa la superioridad de *El cuerpo humano*, que despunta por encima del promedio recabado en Alicante, Bilbao, Melilla y Asturias, y la de *Ropa y complementos*, que sobrepasa a Alicante, Ceuta y Melilla⁷⁵². De otro lado, en *Medios de transporte*, los informantes leridianos alcanzan mejores resultados que los melillenses, una vez más, y que los asturianos. Además de estas diferencias, nuestros datos también son mejores en siete áreas asociativas que los recabados por Fernández Smith *et al.* (2008) en la ciudad de Melilla. De cualquier forma, tal como queda patente en los gráficos 6.11 y 6.12, en general, nuestros resultados son claramente inferiores.

⁷⁵² Una posible explicación a este hecho puede estar en la coletilla “y complementos”, que añadimos en nuestro trabajo en este centro, ausente en las restantes investigaciones.

Gráfico 6.12. Promedio de palabras por centro de interés en distintas sintopías españolas con más de una lengua en contacto (media) y en Lleida.



Con respecto a la comparación de los promedios por centro de interés con la media de los territorios donde convive más de una lengua (gráfico 6.12), las mayores diferencias se constatan en *La cocina y sus utensilios*, en *Muebles de la casa* y en *Objetos colocados en la mesa para la comida*, con una distancia de 3.11, 2.59 y 2.38 respuestas respectivamente. En cambio, las áreas donde las divergencias son menores son *El cuerpo humano* y *Ropa y complementos*, con 0.27 y 0.62 décimas de diferencia. En las restantes áreas, las diferencias siempre sobrepasan la unidad; nuestros datos, como ya hemos comentado, siempre se sitúan por debajo de la media.

En nuestra opinión, los bajos resultados obtenidos se explican porque, tal como hemos mencionado, las situaciones sociolingüísticas que se comparan en los gráficos anteriores (6.11 y 6.12 y en la tabla 6.17) son harto distintas y, salvo el caso de Bilbao, no hay ninguna otra sintopía, como la nuestra, donde la mayoría de los informantes que componen la muestra tenga una lengua materna distinta a la castellana⁷⁵³. Esta es la causa, en nuestra opinión, de que la mayor diferencia, con respecto a la media general, se dé en *Muebles de la casa*, *Objetos colocados en la mesa para la comida* o en *La cocina y sus utensilios*, áreas estrechamente relacionadas con el hogar y, por ende, con la lengua usada en ese ámbito. De la misma forma, en ninguna de las sintopías analizadas la lengua vernácula, u otra cualquiera distinta al español, tiene una extensión social ni siquiera similar a la del catalán en Lleida (*vid.* 4.3.2.). Asimismo, el nuestro es el único territorio donde, con independencia de su lengua materna, no hay ningún informante que se escolarice en lengua castellana sino que todos siguen un programa educativo que utiliza el catalán como única lengua vehicular. Como ya dijimos en el apartado relativo al marco sociolingüístico (*vid.* 4.1.), precisamente este hecho ha suscitado dudas sobre el nivel de castellano que adquiere el alumnado que estudia en Cataluña, aunque, hasta la fecha, los estudios realizados no han demostrado que el nivel de lengua española haya disminuido desde que se introdujeron los programas de inmersión lingüística (Sinner 2004: 18). Al respecto, y tras las conclusiones que pueden extraerse de la presente investigación, podemos afirmar que, si bien la productividad en lengua castellana alcanza un promedio claramente inferior al de otras provincias, este no dista mucho del promedio general (únicamente en 1.74 palabras, si calculamos la media de todas las comunidades analizadas en los gráficos 6.10 y 6.11)⁷⁵⁴.

Si tenemos en cuenta las circunstancias apuntadas, es decir, el hecho de que la mayoría de la muestra tiene como lengua materna el catalán y que ha sido escolarizada también en esta lengua, los resultados no dejan de ser favorables al castellano, ya que el

⁷⁵³ En Bilbao, los informantes cuya lengua materna es el euskera son 145 de un total de 245 (Etxebarria 1996b). Asimismo, en Castellón, la proporción de alumnado con valenciano o castellano como lengua materna es casi idéntica (106 y 104, respectivamente). En Galicia, más del 53% de la muestra se decanta por el castellano como lengua de uso habitual. Por otro lado, los informantes de Navarra cuya lengua materna es el euskera no alcanzan el 20% del total. En Valencia, los bilingües suponen solo el 26.2% y, en Ceuta y Melilla, las cifras de hablantes bilingües son aún inferiores, con el 10.90% y el 7.93% respectivamente. En Aragón, Asturias y Alicante no se ofrece información sobre la lengua materna de los informantes. Todos los datos anteriores se hallan en las referencias incluidas en la nota 749 excepto los de Castellón, que proceden de Blas y Casanova (2003b), y los de Navarra, que hemos extraído de Taberner (2008b: 559).

⁷⁵⁴ Esta se situaría en 19.64 respuestas por centro de interés.

promedio global, si bien es ciertamente bajo —tanto en relación con otros territorios monolingües hispánicos como, más aún, con respecto a algunos bilingües—, demuestra un caudal de léxico disponible, con casi 18 palabras por centro de interés, similar al de las restantes sintopías, e incluso superior a la media obtenida en la provincia de Málaga y en la ciudad de Melilla.

En general, no hay que perder de vista que nuestros resultados plasman el conocimiento del léxico disponible del español de una población bilingüe gran parte de la cual tiene el castellano como segunda lengua. Si recordamos la información que vertimos en el capítulo cuarto (*vid.* 4.2), precisamente este hecho, que el castellano no sea la lengua materna de buena parte de la población, es uno de los rasgos que caracterizan la variedad del español hablado en la ciudad de Lleida.

Por otro lado, en el apartado 6.2.5. (*infra*), comprobaremos si es cierta la hipótesis que hemos apuntado de que la lengua materna incide de forma cuantitativa en los resultados, de modo que la producción sería superior si aislásemos según esta variable a los informantes y comparásemos la producción que alcanza cada subgrupo, respectivamente, en su primera lengua, esto es, los informantes con lengua materna castellana en lengua española y los que tienen como lengua materna el catalán en catalana.

De cualquier forma, estos resultados han de tomarse con cautela, antes de extraer conclusiones tajantes a partir de ellos sobre la competencia léxica del alumnado en cualquiera de las dos lenguas estudiadas. Al problema de la posible disparidad en la edición de los datos, que ya hemos comentado, hay que sumar otro inconveniente importante y es que la propia metodología de la disponibilidad léxica, que utiliza la asociación controlada de palabras con un tiempo límite, no deja de ser una prueba “artificial” que nos informa sobre una parte de la competencia léxica del hablante⁷⁵⁵. Para poder extraer conclusiones acerca de si los informantes tienen más o menos vocabulario en unas u otras provincias, sería necesario complementar la información recabada mediante estas pruebas con otras que dieran fe de la actuación real del informante, esto es, si es capaz de incluir en un discurso las palabras evocadas en una encuesta de disponibilidad.

⁷⁵⁵ Recuérdese lo que afirma Hernández Muñoz (2006: 320) al respecto (*supra* p. 317 nota 681).

Hechas estas salvedades, pasemos, a continuación, a analizar los resultados en lengua catalana. Únicamente podemos comparar nuestros datos con los obtenidos por Llopis (2009) en Valencia, puesto que no se han realizado más estudios de disponibilidad léxica sobre esta lengua⁷⁵⁶.

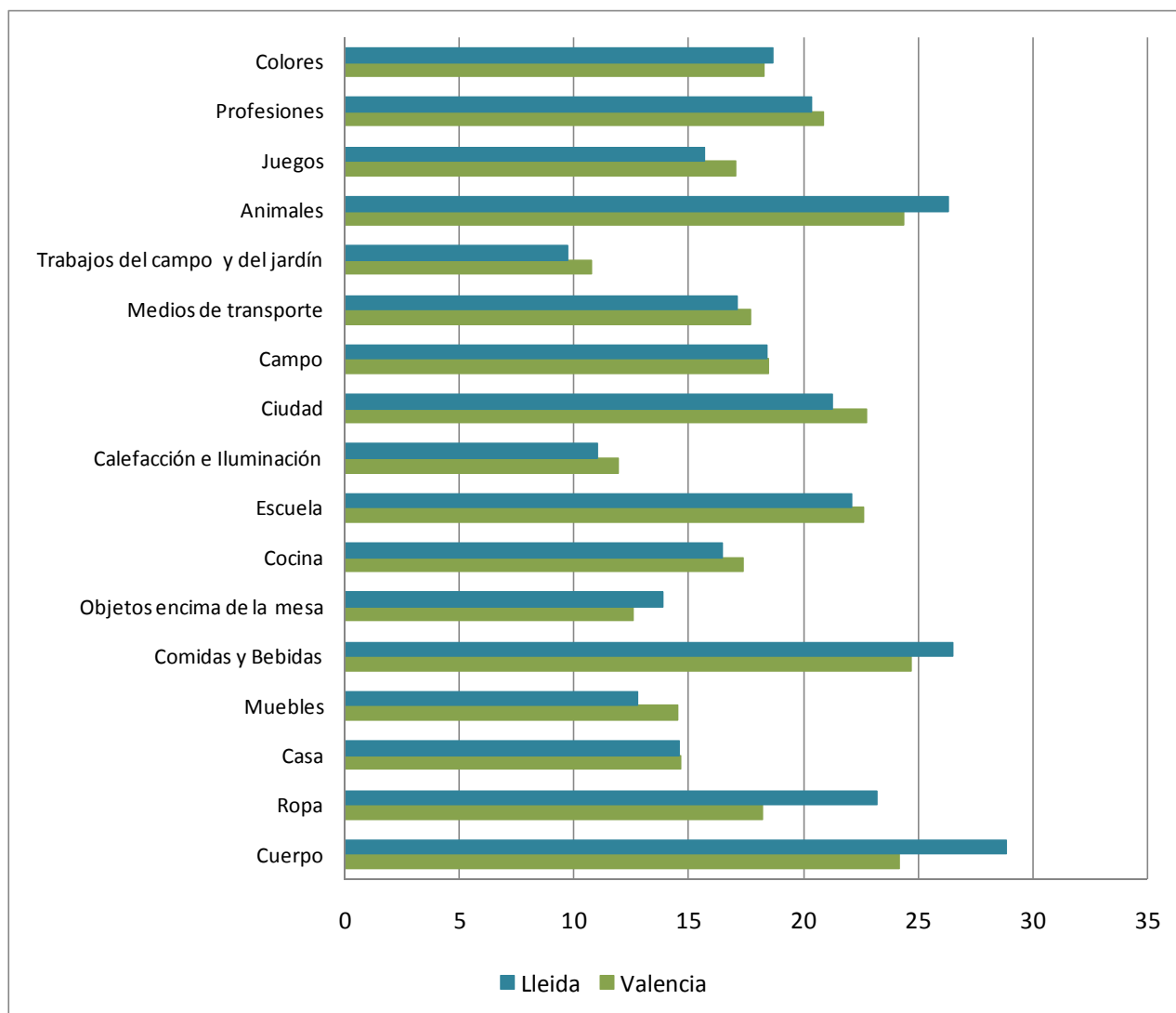
Tabla 6.18. Promedio de palabras en catalán en Valencia y en Lleida⁷⁵⁷.

Nº	Centro de Interés	Lleida	Rango	Valencia	Rango
1	El cuerpo humano	28.85	1	24.19	3
2	Ropa y complementos	23.19	4	18.24	9
3	Partes de la casa (sin los muebles)	14.60	13	14.62	13
4	Muebles de la casa	12.80	15	14.55	14
5	Comidas y bebidas	26.49	2	24.70	1
6	Objetos colocados en la mesa para la comida	13.85	14	12.55	15
7	La cocina y sus utensilios	16.48	11	17.35	11
8	La escuela (muebles y materiales)	22.13	5	22.61	5
9	Calefacción, iluminación y formas de airear un recinto	11.05	16	11.90	16
10	La ciudad	21.25	6	22.77	4
11	El campo	18.39	9	18.47	7
12	Medios de transporte	17.09	10	17.72	10
13	Trabajos del campo y del jardín	9.72	17	10.78	17
14	Animales	26.32	3	24.39	2
15	Juegos y distracciones	15.71	12	17.07	12
16	Profesiones y oficios	20.37	7	20.90	6
17	Colores	18.64	8	18.29	8
Media		18.64	----	18.30	----

⁷⁵⁶ Con la excepción del trabajo inédito de Gómez Devis (1995), cuyos datos no incluimos aquí dado que en este estudio se contempla la *edad* como variable social.

⁷⁵⁷ Los rangos que coinciden en las dos provincias analizadas se señalan de forma coloreada en la tabla.

Gráfico 6.13. Promedio de palabras por centro de interés en catalán en Lleida y Valencia.



Tal como se observa en la tabla y en el gráfico anteriores, los resultados en las dos sintopías son bastante cercanos. El promedio general alcanza un resultado similar, ligeramente superior en Lleida (18.60 frente a 18.30). En nuestra provincia se obtienen mejores cifras en seis áreas temáticas: *El cuerpo humano*, *Ropa y complementos*, *Comidas y bebidas*, *Objetos colocados en la mesa para la comida*, *Animales* y *Colores*; mientras que los resultados valencianos son superiores en las once áreas restantes: *Partes de la casa*, *Muebles de la casa*, *La cocina y sus utensilios*, *La escuela*, *Calefacción e iluminación y formas de airear un recinto*, *La ciudad*, *El campo*, *Medios de transporte*, *Trabajos del campo y del jardín*, *Juegos y distracciones* y *Profesiones y oficios*. Sin embargo, las mayores diferencias se detectan en *El cuerpo humano* y en *Ropa y complementos*, donde el alumnado leridano obtiene cuatro palabras y media de

más⁷⁵⁸. Las restantes divergencias son menos destacables y se sitúan, excepto en dos estímulos donde la diferencia es de apenas unas centésimas, en más de una palabra a favor del alumnado de Lleida o de Valencia, en cada caso. En general, consideramos que estos resultados muestran que el caudal de léxico disponible de los informantes de las dos sintopías es bastante similar, a pesar de que la situación sociolingüística favorece más el uso de la lengua vernácula en nuestra provincia⁷⁵⁹.

De otra parte, en la clasificación de las áreas asociativas según la productividad (tabla 6.18), se observan numerosas coincidencias: ocho áreas ostentan exactamente el mismo rango y las que difieren lo hacen en muy pocas posiciones⁷⁶⁰. Hay que reconocer que, en general, se aprecian las mismas tendencias ya señaladas con respecto a las restantes investigaciones realizadas sobre el castellano.

6.2. INCIDENCIA CUANTITATIVA DE LAS VARIABLES SOCIALES

En este segundo apartado, abordamos el análisis de la incidencia cuantitativa que tienen las cinco variables sociales elegidas (*sexo, tipo de centro, zona geográfica, nivel sociocultural y lengua materna*) sobre la producción en las dos lenguas. Para ello, y dada la gran cantidad de datos que poseemos, seguimos en todos los factores el siguiente orden: en primer lugar, cotejamos la media general obtenida por cada grupo social en las dos lenguas estudiadas; en segundo lugar, mediante los promedios de palabras de cada centro de interés, comparamos la producción lingüística en castellano y, seguidamente, en catalán de cada submuestra (por ejemplo, varones frente a mujeres). A continuación, ratificamos estos resultados con los obtenidos tras realizar el análisis de la homogeneidad de varianzas mediante ANOVA, de forma que detallamos si se refuta o no la hipótesis nula (esto es, si habrá o no alguna influencia sobre el léxico) tanto en la media global de cada lengua como en cada una de las áreas asociativas. Como ya dijimos en la introducción a este capítulo, todos los análisis realizados se encuentran debidamente clasificados en el tercer apéndice de este trabajo, razón por la cual solo incluimos en el análisis de cada variable social una tabla con el valor de la probabilidad (*p-value*) hallado.

⁷⁵⁸ Como ya hemos indicado con anterioridad, quizá el hecho de que el título del segundo centro de interés difiera levemente puede explicar esta llamativa diferencia.

⁷⁵⁹ Sin embargo, en el trabajo de Llopis (2009) solo se tienen en cuenta las comarcas valencianohablantes, hecho que creemos que incide en que los resultados sean tan semejantes.

⁷⁶⁰ El caso en el que se aprecia más diferencia es *Ropa y complementos*, que es cuarto en Lleida y ocupa el noveno rango, en cambio, en Valencia.

Tras este primer análisis, llevamos a cabo otro, basado simplemente en la descripción de las tendencias que observamos entre los promedios. En esta parte, nuestros resultados no cuentan con un respaldo estadístico que determine si las diferencias analizadas son o no extrapolables al conjunto estudiado, con lo que lo único que podemos determinar son posibles tendencias⁷⁶¹. De este modo, comparamos el promedio de palabras por centro de interés que alcanza cada grupo aislado, según la variable social tenida en cuenta, en la producción de cada idioma (por ejemplo, la producción de los varones en castellano y en catalán). Finalmente, incluimos la cantidad de vocablos recogida y el índice de cohesión obtenido en cada área por los distintos colectivos de sujetos, distribuidos según la lengua de la encuesta y la variable social en cuestión. Es importante señalar, que, en este punto —debido a los inconvenientes que ya hemos señalado sobre los totales de vocablos⁷⁶² y, por ende, de los índices de cohesión (*vid.* 6.1.4. y 6.1.5.)— el análisis se basa en dos premisas: comprobar que, efectivamente, el grupo más numeroso ostenta un mayor número de vocablos y, por tanto, un índice de cohesión menor, y en corroborar que se siguen las tendencias generales ya detectadas con respecto al número de palabras diferentes en cada lengua y en cada área temática, y el índice de cohesión respectivo⁷⁶³.

6.2.1. Variable social sexo

Tal como mostramos en el apartado metodológico, la muestra se divide en 93 varones (el 38.75% del total) y 147 mujeres (el 61.25%). La superioridad numérica de las mujeres repercute, como veremos, en algunos de los resultados. En primer lugar, nos fijaremos en la media de cada subgrupo en cada lengua. Como se aprecia en la siguiente tabla, las mujeres muestran una media superior a la que registran los varones, tanto en castellano (18.31 frente a 17.26), como en catalán (19.45 frente a 17.36). El mejor promedio en la producción lo ostentan las mujeres en lengua catalana, siendo también mayor la distancia, en esta lengua, con respecto a los varones (2.09 palabras). En

⁷⁶¹ A pesar de que estos mismos datos tomados en general, y no por submuestras, sí están analizados estadísticamente (*cf.* 6.1.3.).

⁷⁶² Hernández Muñoz (2006: 357) afirma: “los comentarios impresionistas sobre las cifras absolutas de vocablos tampoco permiten unas comparaciones rigurosas entre las variables sociales”.

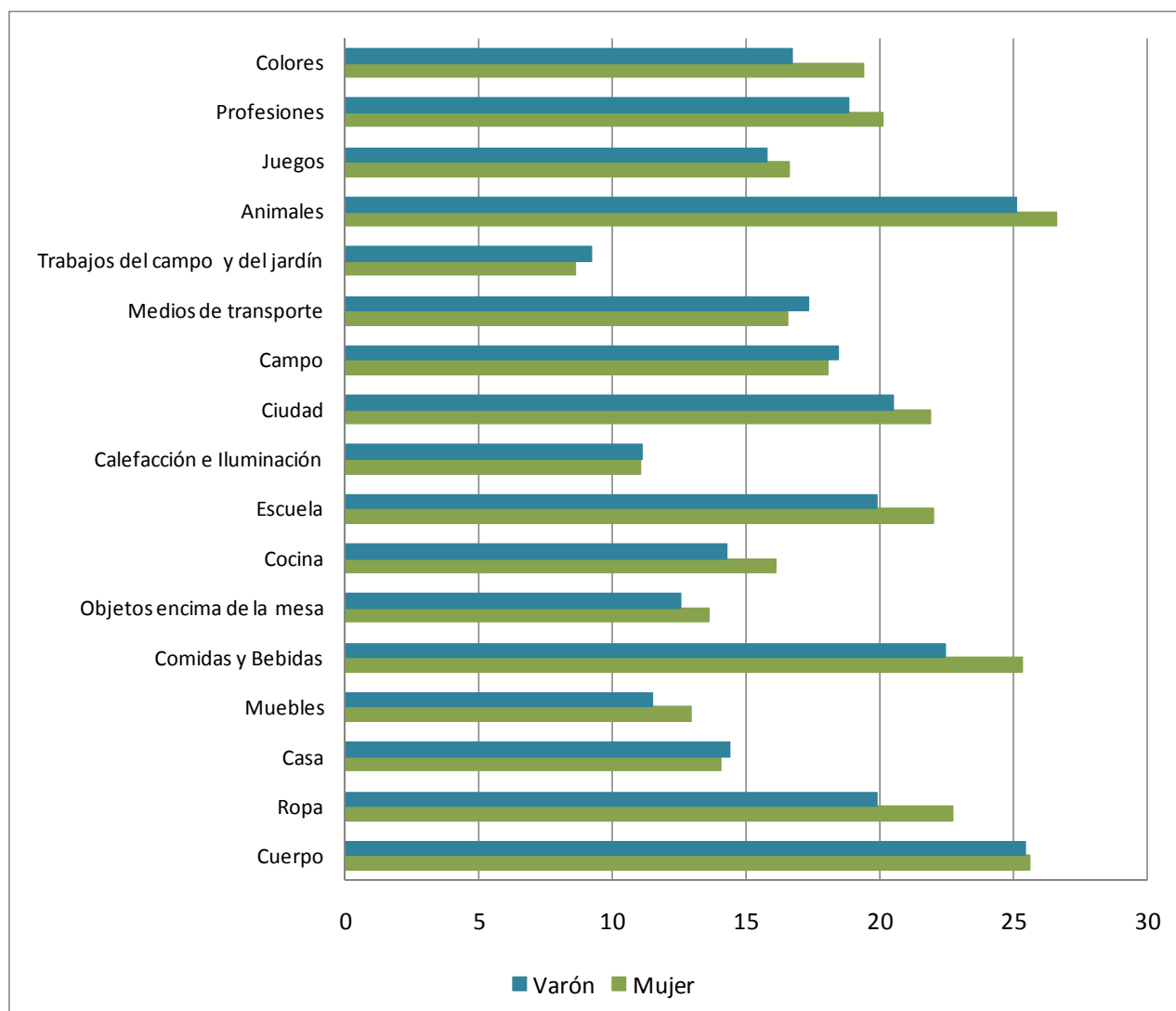
⁷⁶³ Como ya hemos advertido reiteradamente a lo largo del presente capítulo, no es posible comparar el número de vocablos —ni el índice de cohesión— si las muestras que se manejan son distintas, puesto que, a mayor número de informantes, mayor número de vocablos y, a más vocablos, menor índice de cohesión. Esta es la razón por la que, en los análisis de estos parámetros, los informantes se encuentran agrupados según cada variable social (*varones*> castellano/catalán// *mujeres*> castellano/catalán), puesto que solo son comparables de esta forma.

castellano, las diferencias en función del sexo también favorecen a las mujeres, pero en este caso son menores y se sitúan en torno a la unidad (1.05). La producción de los varones es la más similar del conjunto con tan solo una décima de diferencia (17.26 en castellano y 17.36 en catalán). En cambio, en el colectivo femenino se aprecia una divergencia más acentuada entre la producción global en lengua castellana (18.31) y la de la lengua catalana (19.45), sobrepasando la unidad (1.14).

Tabla 6.19. Promedio de palabras por centro de interés según la variable social sexo.

Nº	Centro de Interés	Castellano		Catalán	
		Varones	Mujeres	Varones	Mujeres
1	El cuerpo humano	25.45	25.59	28.41	29.13
2	Ropa y complementos	19.90	22.71	20.62	24.82
3	Partes de la casa (sin los muebles)	14.38	14.03	14.30	14.78
4	Muebles de la casa	11.52	12.95	12.13	13.22
5	Comidas y bebidas	22.46	25.32	23.77	28.20
6	Objetos colocados en la mesa para la comida	12.58	13.60	12.72	14.57
7	La cocina y sus utensilios	14.26	16.09	14.16	17.94
8	La escuela (muebles y materiales)	19.87	21.99	20.15	23.37
9	Calefacción, iluminación y formas de airear un recinto	11.13	11.05	10.47	11.41
10	La ciudad	20.52	21.88	19.31	22.48
11	El campo	18.42	18.07	17.44	18.99
12	Medios de transporte	17.33	16.54	17.03	17.13
13	Trabajos del campo y del jardín	9.23	8.62	9.80	9.67
14	Animales	25.11	26.60	24.65	27.38
15	Juegos y distracciones	15.80	16.63	14.75	16.32
16	Profesiones y oficios	18.82	20.13	18.83	21.34
17	Colores	16.70	19.40	16.60	19.93
Media		17.26	18.31	17.36	19.45

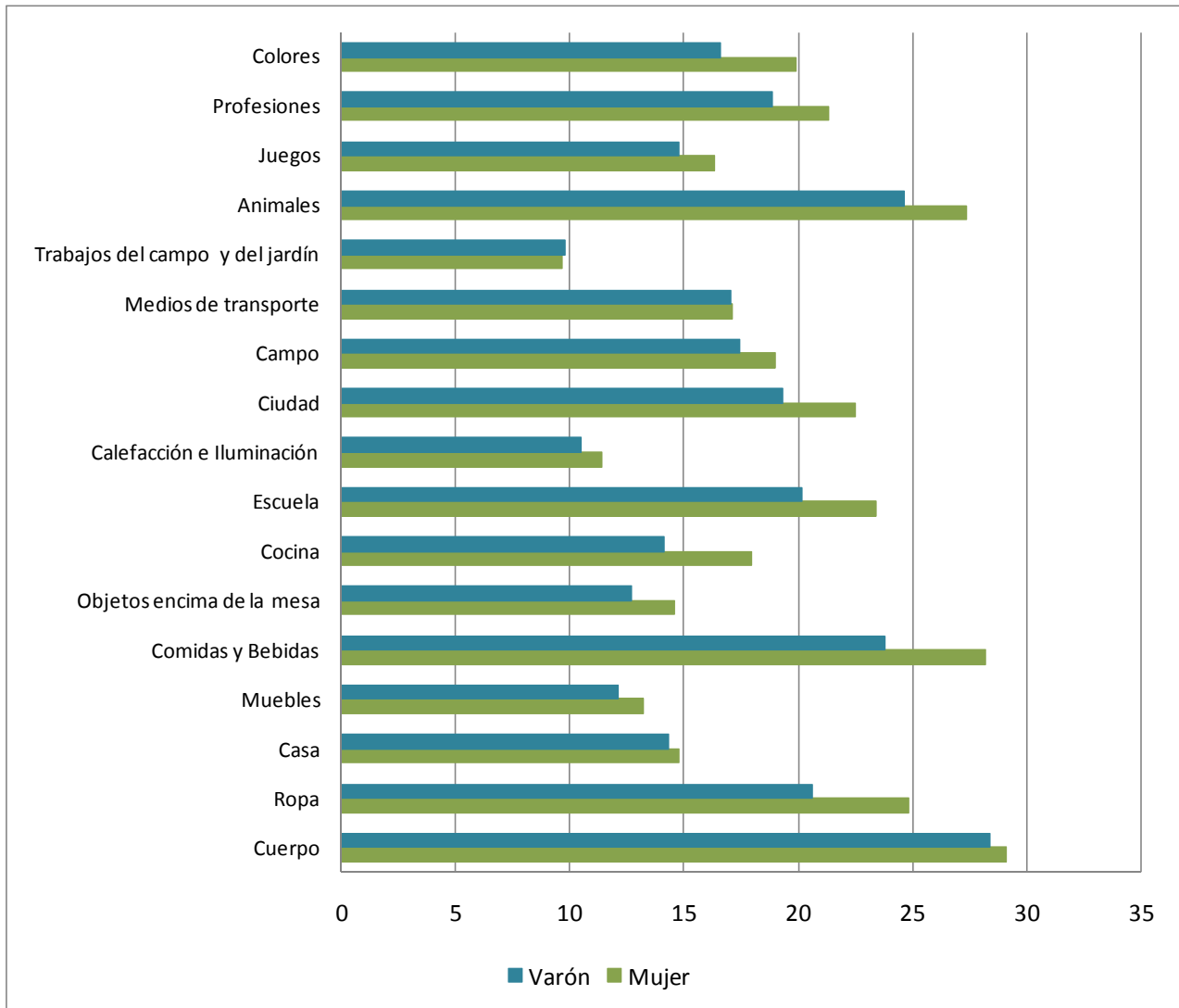
Gráfico 6.14. Promedio de palabras por centro de interés según la variable social sexo en lengua castellana.



Si nos ceñimos ahora al análisis de la producción en lengua castellana, donde hemos dicho que globalmente las mujeres alcanzan un resultado más favorable, advertimos que, en el desglose por áreas asociativas, en doce campos la producción femenina es superior (*El cuerpo humano, Ropa y complementos, Muebles de la casa, Comidas y bebidas, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, La escuela, La ciudad, Animales, Juegos y distracciones, Profesiones y oficios y Colores*), mientras que solo en cinco los varones exhiben mejores medias (*Partes de la casa, Calefacción, iluminación y formas de airear un recinto, El campo, Medios de transporte y Trabajos del campo y del jardín*). Es destacable que en cuatro áreas —*Ropa y complementos, Comidas y bebidas, La escuela y Colores*— las mujeres sobrepasen a los varones en más de dos palabras. Sin embargo, en dirección contraria,

las diferencias a favor del colectivo masculino no alcanzan la unidad en ningún caso. Pasemos a ver, ahora, las diferencias con respecto a la producción de los dos sexos en lengua catalana:

Gráfico 6.15. Promedio de palabras por centro de interés según la variable social sexo en lengua catalana.



Tal como se aprecia claramente en el gráfico 6.15, la producción de las mujeres también es superior en lengua catalana e, incluso, las diferencias son mayores —y más constantes— que las observadas en castellano. Las alumnas obtienen mejores resultados en todas las áreas, salvo en *Trabajos del campo y del jardín*, donde la media es superior, en poco más de una décima, en el colectivo masculino. Las divergencias son aquí más marcadas, puesto que en dos centros de interés llegan a superar las cuatro palabras

(*Ropa y complementos* y *Comidas y bebidas*), y en cuatro sobrepasan las tres unidades (*La cocina y sus utensilios*, *La escuela*, *La ciudad* y *Colores*).

A continuación, comprobamos si estas diferencias observadas entre los sexos en las dos lenguas tienen o no una significación estadística, esto es, si son predecibles. Tras realizar el análisis de la varianza pertinente, se obtiene un nivel de significación de 0.033 en lengua castellana y de 0.000 en catalana, de lo que se colige que el factor *sexo* es, en ambas lenguas, el causante de las diferencias observadas en la producción de palabras, puesto que la probabilidad se sitúa entre 0.000 y 0.05.

Tras el análisis de cada centro de interés en particular, y como puede deducirse de los datos de la tabla 6.20, llegamos a las siguientes conclusiones:

- De los doce centros donde los promedios obtenidos por el colectivo femenino son superiores en lengua castellana, dicha diferencia resulta estadísticamente significativa en ocho de ellos: *Ropa y complementos*, *Muebles de la casa*, *Comidas y bebidas*, *La cocina y sus utensilios*, *La escuela*, *La ciudad*, *Animales* y *Colores*.
- De las dieciséis áreas donde la productividad es mayor en las mujeres en lengua catalana, en trece de ellas la diferencia es extrapolable al universo estudiado: *Ropa y complementos*, *Muebles de la casa*, *Comidas y bebidas*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios*, *La escuela*, *Calefacción, iluminación y formas de airear un recinto*, *La ciudad*, *El campo*, *Animales*, *Juegos y distracciones*, *Profesiones y oficios* y *Colores*.

Tabla 6.20. Análisis estadístico de la diferencia según el sexo. Valor de la 'p' en las dos lenguas.

Nº	Centros de interés	Castellano	Catalán
1	El cuerpo humano	0.745	0.323
2	Ropa y complementos	0.000	0.000
3	Partes de la casa (sin los muebles)	0.511	0.352
4	Muebles de la casa	0.002	0.028
5	Comidas y bebidas	0.000	0.000
6	Objetos colocados en la mesa para la comida	0.052	0.001
7	La cocina y sus utensilios	0.004	0.000
8	La escuela (muebles y materiales)	0.000	0.000
9	Calefacción, iluminación y formas de airear un recinto	0.874	0.042
10	La ciudad	0.044	0.000

11	El campo	0.653	0.027
12	Medios de transporte	0.118	0.877
13	Trabajos del campo y del jardín	0.171	0.779
14	Animales	0.026	0.000
15	Juegos y distracciones	0.168	0.014
16	Profesiones y oficios	0.056	0.000
17	Colores	0.000	0.000

Estos resultados corroboran la importancia de esta variable en la estratificación social del léxico disponible en las dos lenguas de nuestra provincia. De otra parte, es llamativa la coincidencia que se da entre los estímulos verbales: además de que ocho áreas coinciden en la significación, en cuatro de ellas (*Ropa y complementos*, *Comidas y bebidas*, *La escuela* y *Colores*) la probabilidad se sitúa exactamente en la misma cifra (0.000) en los dos idiomas.

En principio, si nos fijamos en las tendencias generales, constatamos cómo, en pleno siglo XXI, se siguen observando diferencias en la producción según el rol social desempeñado tradicionalmente por cada sexo. De este modo, las alumnas alcanzan mejores resultados en áreas relacionadas con el papel de la mujer como ama de casa (*Muebles de la casa*, *Comidas y bebidas*, *Objetos colocados en la mesa para la comida*⁷⁶⁴, y *La cocina y sus utensilios*) y, de forma opuesta, los varones aventajan a las mujeres en categorías relativas a las labores del campo (*El campo* —únicamente en castellano— y *Trabajos del campo y del jardín* en las dos lenguas). Más allá de estos roles sexo-sociales, también se advierte el interés de cada sexo por determinados aspectos de la realidad —evidentemente mediatizado por los estereotipos de género tradicionales—, como se observa en el predominio femenino en las categorías de *Ropa y complementos* y *Colores* y en la superioridad masculina en *Medios de transporte*, aunque únicamente se dé en lengua castellana.

Sin embargo, como acabamos de ver, las únicas diferencias relevantes estadísticamente son las producidas a favor del colectivo femenino, y, de cualquier modo, el predominio de las mujeres en la productividad también se da en áreas aparentemente neutras, como *La escuela*, *La ciudad* o *Animales*, con lo que, a la vista de estos datos, y dado que se dan de forma constante en las dos lenguas, cabe relacionar

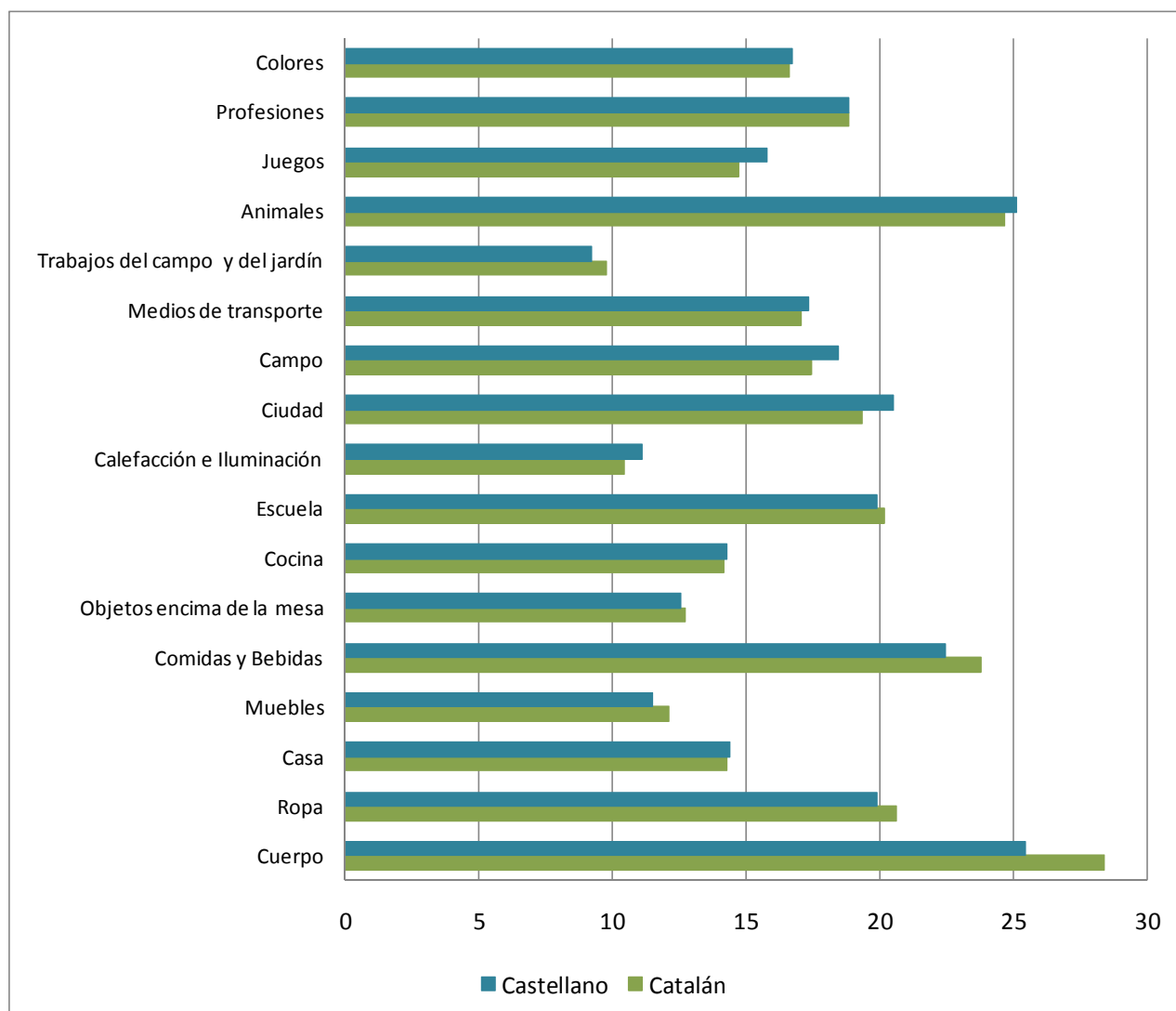
⁷⁶⁴ La diferencia, en este centro de interés, solo es significativa en lengua catalana, aunque en castellano esta área obtiene un resultado muy cercano a la significación estadística, como se observa en la tabla 6.20.

nuestros resultados con las teorías que otorgan una mayor competencia léxica a las mujeres con respecto al aprendizaje de una primera o segunda lengua (Lozano 1995, Calero 2007). De otra parte, en los estudios realizados con anterioridad, se obtienen resultados dispares con respecto a esta variable, en algunos casos a favor de los varones y, en otros, a favor de las mujeres, y tampoco faltan investigaciones donde la variable *sexo* no tiene incidencia alguna en el caudal de léxico disponible⁷⁶⁵.

Pasemos ahora a observar las diferencias según el *sexo* en la producción en las dos lenguas. Como explicamos en la introducción, no hemos sometido estos datos a análisis estadístico alguno, con lo que solo podemos describir las tendencias observadas sin extrapolar los datos al conjunto de la población.

⁷⁶⁵ Galloso (2003a) y Hernández Muñoz (2006) detectan una mayor productividad entre el colectivo masculino; por el contrario, y al igual que nosotros, González Martínez (2002) y Gómez Molina y Gómez Devís (2004) constatan índices más elevados en el colectivo femenino. Mateo (1998) o Carcedo (2001) no encuentran diferencias sustanciales según esta variable. Téngase en cuenta que no todas las investigaciones utilizan programas estadísticos que corroboren sus resultados.

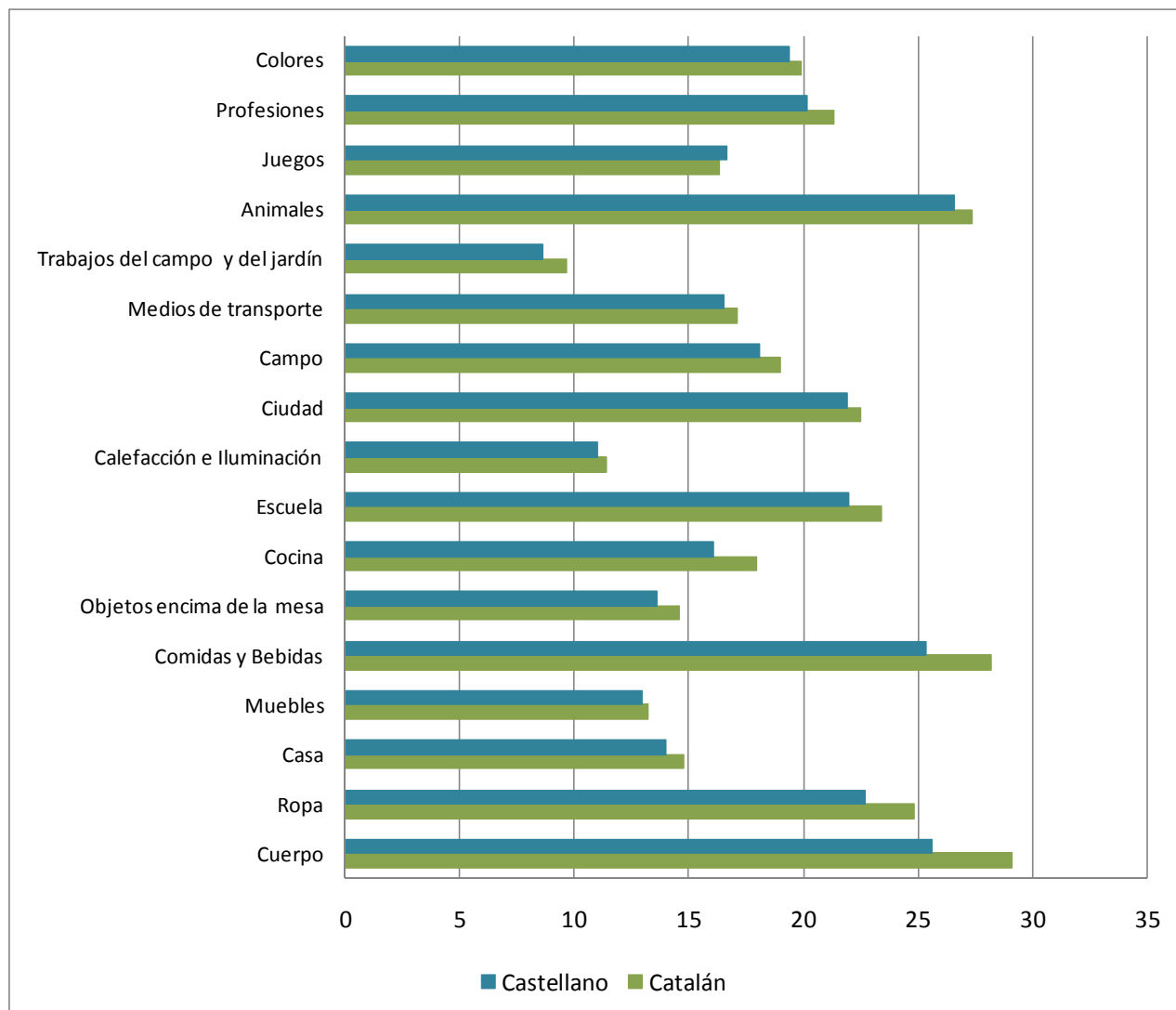
Gráfico 6.16. Comparación de la producción en castellano y en catalán de los varones.



En los varones, como ya hemos dicho, el promedio general en las dos lenguas es muy aproximado (17.26, en castellano, y 17.36, en catalán). Analizando centro a centro se observa cómo, en nueve áreas temáticas, los resultados son mejores en lengua castellana y, en ocho, en catalana. No obstante, las distancias —en todas estas áreas— son pequeñas, salvo en *El cuerpo humano*, donde en catalán se actualizan de media casi tres palabras más. Como dijimos al hablar de las diferencias en la producción en las dos lenguas (*vid.* 6.1.3.), la incidencia de la lengua materna, más el peso del currículum escolar, pueden explicar este comportamiento, que también se observa en las mujeres como veremos a continuación.

Los resultados referentes a la producción femenina en cada una de las dos lenguas son bastante diferentes, como puede verse en el gráfico 6.17.

Gráfico 6.17. Comparación de la producción en castellano y en catalán de las mujeres.



En las mujeres, se detectan mayores diferencias con respecto a la producción en cada idioma. Si en el colectivo masculino la situación estaba bastante equilibrada en las dos lenguas, en las mujeres, la productividad mayor en lengua catalana es constante en dieciséis centros de interés. El único campo léxico donde se obtienen mejores resultados en castellano es *Juegos y distracciones*, ámbito donde —como ya explicamos— se observa la influencia del castellano a través de los medios de comunicación y de las formas de ocio juvenil. Las divergencias mayores entre los dos idiomas se detectan ahora, al igual que sucedía en los cómputos globales, en *El cuerpo humano*, *Ropa y complementos* y *Comidas y bebidas* (cfr. 6.1.3.).

Cabe recordar que la tendencia, en los datos globales, era firme a favor de la lengua catalana salvo en tres categorías semánticas (*Calefacción, iluminación y formas de airear un recinto, La ciudad y Juegos y distracciones*). Pues bien, en el caso de los varones, además de en estos tres estímulos los resultados son superiores en seis áreas más en castellano (*Partes de la casa, La cocina y sus utensilios, El campo, Medios de transporte, Animales y Colores*); no obstante, entre las mujeres, únicamente *Juegos y distracciones* tiene un promedio mejor en español. El comportamiento del colectivo femenino, que —de forma contundente— muestra mejores medias en la lengua vernácula, nos ha llevado a preguntarnos si puede existir alguna relación entre esta variable y la *lengua materna*, de forma que hubiera, en la distribución de la muestra, más mujeres que varones cuya lengua materna fuera el catalán. Sin embargo, en el capítulo anterior (*vid.* 5.2.), cotejamos todas las variables y no se encontró relación alguna⁷⁶⁶, lo que viene a corroborar los resultados obtenidos hasta el momento, que indican la estratificación tan significativa de la variable *sexo* en nuestra investigación.

A continuación, incluimos los vocablos y el índice de cohesión según esta variable. Con respecto al número de palabras diferentes, las mujeres despuntan frente a los varones, puesto que son mayoría en la muestra. Este hecho sucede en todas las áreas menos en *Partes de la casa y Comidas y bebidas*, en lengua catalana, con una diferencia de 10 y 6 palabras respectivamente, y en *Medios de transporte, Trabajos del campo y del jardín y Animales*, donde los varones aportan más vocablos que las mujeres en ambas lenguas, a pesar de ser menos numéricamente, lo que puede relacionarse con el reconocimiento social del papel tradicional del varón en las tareas agrícolas o, como ya hemos comentado, con el mayor interés por determinados aspectos de la realidad de uno de los dos sexos⁷⁶⁷. En relación con las diferencias según el idioma de producción, los varones tienen más vocablos en castellano solo en seis áreas temáticas (*La escuela, Calefacción, iluminación y formas de airear un recinto, La ciudad, Trabajos del campo y del jardín, Juegos y distracciones y Profesiones y oficios*), exactamente igual que

⁷⁶⁶ Entre los varones, el 27.96% tiene el castellano como lengua materna y el 66.67%, el catalán. Los porcentajes entre las mujeres son muy similares: 27.21% frente a 64.63%. Recuérdese que también hay informantes englobados bajo las etiquetas de “ambas lenguas” o “aranés”.

⁷⁶⁷ La atracción que ejerce todo tipo de automóviles en el colectivo masculino, evidentemente influido por un estereotipo cultural, explica, a nuestro modo de ver, la diferencia entre el número de vocablos en *Medios de transporte*, que alcanza las 32 palabras en catalán y las 21 en castellano. La diferencia en *Trabajos del campo y del jardín* es de 44 y 11 palabras, respectivamente, y en *Animales*, de 3 y 33.

sucede en los resultados generales⁷⁶⁸. Entre las mujeres, en cambio, al igual que ocurría con los resultados relativos a los promedios de palabras, la diferencia a favor de la lengua catalana es constante en todos los centros, con excepción de *El cuerpo humano*. Se ha de tener en cuenta que las disimilitudes son, numéricamente, muy distintas y algunas carecen de importancia⁷⁶⁹; la mayor divergencia es la de *Animales*, que obtiene 78 vocablos más entre los varones y 48 entre las mujeres, en ambos casos a favor de la lengua catalana. En conclusión, en la comparación entre la producción y la riqueza léxica en cada una de las dos lenguas según la variable social *sexo*, se observa que, mientras los varones tienen unos resultados más igualados con respecto a las dos lenguas (17.26 frente a 17.36 en los promedios y una diferencia de 63 vocablos a favor, en las dos magnitudes, del catalán), las mujeres muestran un predominio claro de la lengua catalana (18.31 frente a 19.45 en la media y 256 vocablos más en lengua vernácula). Por lo demás, las tendencias que se observan en cuanto a los centros con mayor y menor número de vocablos son similares a las que vimos en los resultados generales⁷⁷⁰.

Tabla 6.21. Total de vocablos por centro de interés según la variable social *sexo*.

Nº	Centro de Interés	Varones		Mujeres	
		Castellano	Catalán	Castellano	Catalán
1	El cuerpo humano	251	253	323	284
2	Ropa y complementos	205	224	232	246
3	Partes de la casa (sin los muebles)	173	196	179	186
4	Muebles de la casa	174	178	211	215
5	Comidas y bebidas	366	404	381	398
6	Objetos colocados en la mesa para la comida	183	193	185	223
7	La cocina y sus utensilios	233	234	269	313
8	La escuela (muebles y materiales)	285	284	331	359
9	Calefacción, iluminación y formas de airear un recinto	194	179	200	214
10	La ciudad	457	412	482	488
11	El campo	468	481	529	561
12	Medios de transporte	217	230	185	209
13	Trabajos del campo y del jardín	299	276	255	265
14	Animales	275	353	272	320

⁷⁶⁸ Aunque no coinciden todos los centros de interés, la mayor parte de ellos sí, como es lógico (*vid.* 6.1.4.)

⁷⁶⁹ Por ejemplo, las que son de tan solo una palabra (*La cocina y sus utensilios* y *La escuela*, entre los varones, y *Profesiones y oficios*, entre las mujeres).

⁷⁷⁰ Salvo alguna pequeña excepción, por ejemplo, en lengua catalana *Comidas y bebidas* alcanza el tercer rango en los varones, cuando este mismo centro ostenta la quinta posición en los resultados generales. De la misma forma, *Medios de transporte* destaca por la baja posición que registra en los resultados relativos a las mujeres en lengua catalana, en relación con su rango general (*vid.* 6.1.4.)

15	Juegos y distracciones	374	340	419	424
16	Profesiones y oficios	406	381	452	453
17	Colores	150	155	189	192
Total		4710	4773	5094	5350

Tabla 6.22. Índice de cohesión por centro de interés según la variable social sexo.

Nº	Centro de Interés	Varones		Mujeres	
		Castellano	Catalán	Castellano	Catalán
1	El cuerpo humano	0.10	0.11	0.08	0.10
2	Ropa y complementos	0.10	0.09	0.10	0.10
3	Partes de la casa (sin los muebles)	0.08	0.07	0.08	0.08
4	Muebles de la casa	0.07	0.07	0.06	0.06
5	Comidas y bebidas	0.06	0.06	0.07	0.07
6	Objetos colocados en la mesa para la comida	0.07	0.07	0.07	0.07
7	La cocina y sus utensilios	0.06	0.06	0.06	0.06
8	La escuela (muebles y materiales)	0.07	0.07	0.07	0.07
9	Calefacción, iluminación y formas de airear un recinto	0.06	0.06	0.06	0.05
10	La ciudad	0.04	0.05	0.05	0.05
11	El campo	0.04	0.04	0.03	0.03
12	Medios de transporte	0.08	0.07	0.09	0.08
13	Trabajos del campo y del jardín	0.03	0.04	0.03	0.04
14	Animales	0.09	0.07	0.10	0.09
15	Juegos y distracciones	0.04	0.04	0.04	0.04
16	Profesiones y oficios	0.05	0.05	0.04	0.05
17	Colores	0.11	0.11	0.10	0.10

En relación con los índices de cohesión, estos resultan bastante similares en los dos sexos, a pesar de que los varones son menos en la muestra y, por tanto, es lógico que alcancen cifras superiores, como se observa en algunos centros (*Muebles de la casa*, *El campo* o *Colores*). Incluso es digno de mención que algunas categorías, como *Medios de transporte* o *Animales*, resultan más abiertas o difusas en el colectivo masculino, porque, como acabamos de ver, la actuación de los varones es más rica léxicamente. En general, se corroboran las tendencias en ambas lenguas ya detectadas con respecto a los centros más compactos y difusos al desglosar los resultados según esta variable social.

6.2.2. Variable social tipo de centro

Como ya explicamos en el capítulo metodológico, la muestra se divide en 60 informantes procedentes de colegios privados —todos ellos ubicados en la ciudad de Lleida— y 180 alumnos de institutos públicos —de los cuales 60 están en la capital y 120 en distintas poblaciones del resto de la provincia—. Por tanto, en frecuencia

relativa, el alumnado de los centros públicos es mayoritario en la muestra y supone el 75% del total, mientras que los informantes de instituciones privadas conforman el 25% restante. Al comparar los promedios alcanzados por cada colectivo, se observa, en primer lugar, que las diferencias no son tan notables como las que acabamos de ver con respecto al *sexo*. En segundo lugar, el mejor caudal de léxico disponible lo ostentan los informantes de institutos públicos en lengua catalana, con una media de 18.77 palabras por campo semántico. En contraste, el peor resultado lo obtiene este mismo alumnado, pero en lengua castellana, donde no alcanza las 18 palabras de media general (17.77). En los centros privados la actuación lingüística es más similar: los resultados son prácticamente idénticos en las dos lenguas (18.30 en lengua castellana y 18.25 en catalana).

Tabla 6.23. Promedio de palabras por centro de interés según la variable social tipo de centro.

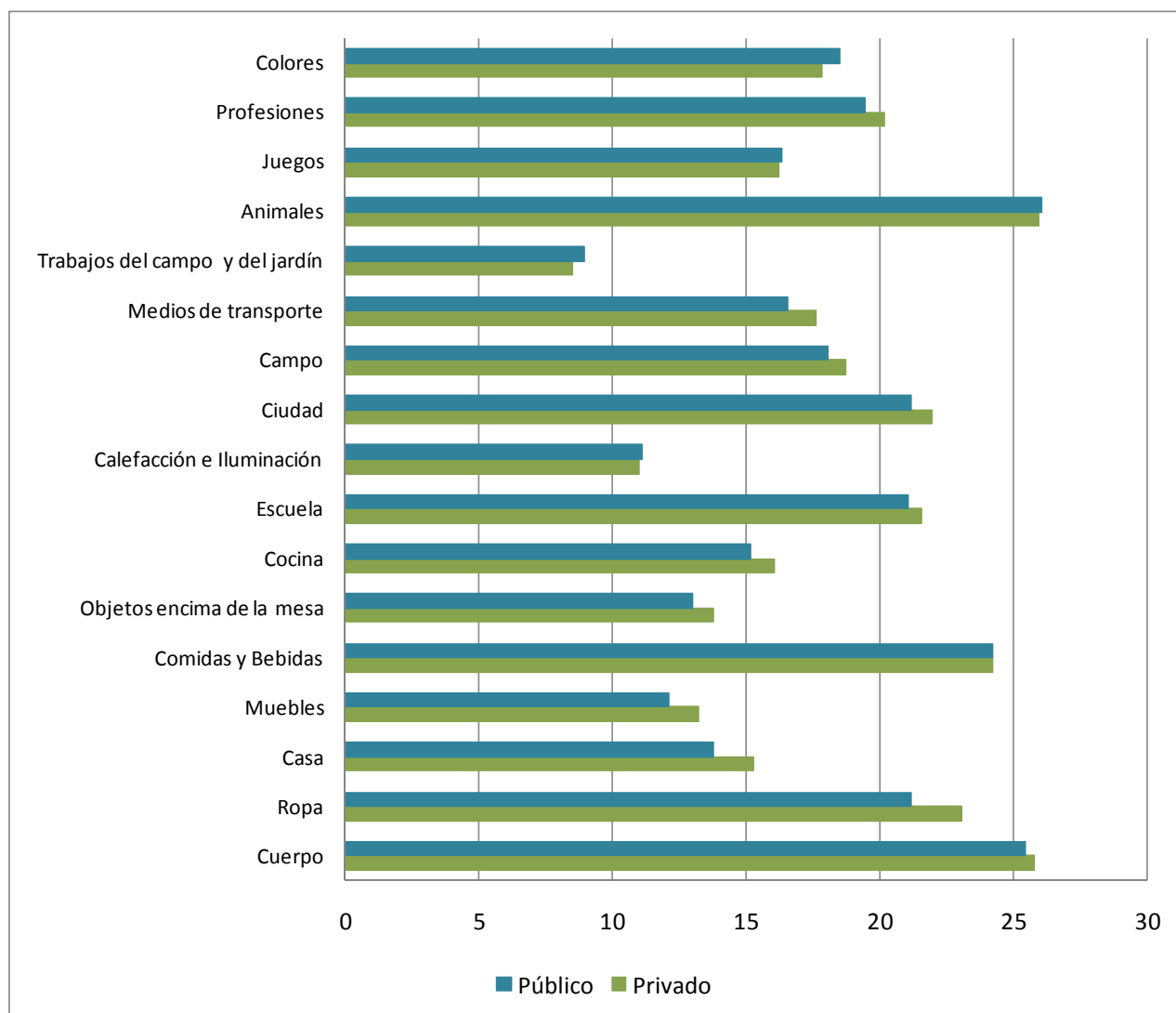
Nº	Centro de Interés	Castellano		Catalán	
		Público	Privado	Público	Privado
1	El cuerpo humano	25.45	25.80	28.76	29.13
2	Ropa y complementos	21.14	23.07	23.37	22.65
3	Partes de la casa (sin los muebles)	13.79	15.28	14.89	13.70
4	Muebles de la casa	12.11	13.25	12.97	12.28
5	Comidas y bebidas	24.20	24.25	26.58	26.20
6	Objetos colocados en la mesa para la comida	13.01	13.78	14.09	13.13
7	La cocina y sus utensilios	15.16	16.05	16.76	15.63
8	La escuela (muebles y materiales)	21.04	21.55	21.99	22.52
9	Calefacción, iluminación y formas de airear un recinto	11.11	11.00	11.31	10.27
10	La ciudad	21.15	21.95	21.09	21.72
11	El campo	18.03	18.72	18.62	17.72
12	Medios de transporte	16.58	17.63	17.26	16.60
13	Trabajos del campo y del jardín	8.97	8.52	10.04	8.73
14	Animales	26.04	25.97	26.71	25.17
15	Juegos y distracciones	16.34	16.22	15.52	16.30
16	Profesiones y oficios	19.43	20.18	20.37	20.37
17	Colores	18.52	17.85	18.79	18.18
Media		17.77	18.30	18.77	18.25

Tal como se observa en el gráfico 6.18, en los resultados relativos a la lengua castellana, los centros privados⁷⁷¹ alcanzan un resultado mejor que los públicos, aunque la diferencia no es muy grande entre los promedios de los dos tipos de instituciones educativas (17.77 frente a 18.30). En el desglose por áreas, los centros privados obtienen mejor promedio en doce centros de interés (*El cuerpo humano, Ropa y*

⁷⁷¹ Recuérdese que todos los centros privados son de titularidad concertada (vid. 5.2).

complementos, Partes de la casa, Muebles de la casa, Comidas y bebidas, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, La escuela, La ciudad, El campo, Medios de transporte y Profesiones y oficios). Las diferencias numéricas, como hemos dicho, son pequeñas; la mayor se da en *Ropa y complementos*, con casi dos palabras de distancia entre la producción de los dos grupos. Además, solo en cuatro categorías semánticas las divergencias entre las dos clases de centros escolares sobrepasan la unidad (*Ropa y complementos, Partes de la casa, Muebles de la casa y Medios de transporte*). En dirección contraria, los centros públicos obtienen mejores resultados en cinco áreas asociativas (*Calefacción, iluminación y formas de airear un recinto, Trabajos del campo y del jardín, Animales, Juegos y distracciones y Colores*); las diferencias, en estos casos, son aún menores.

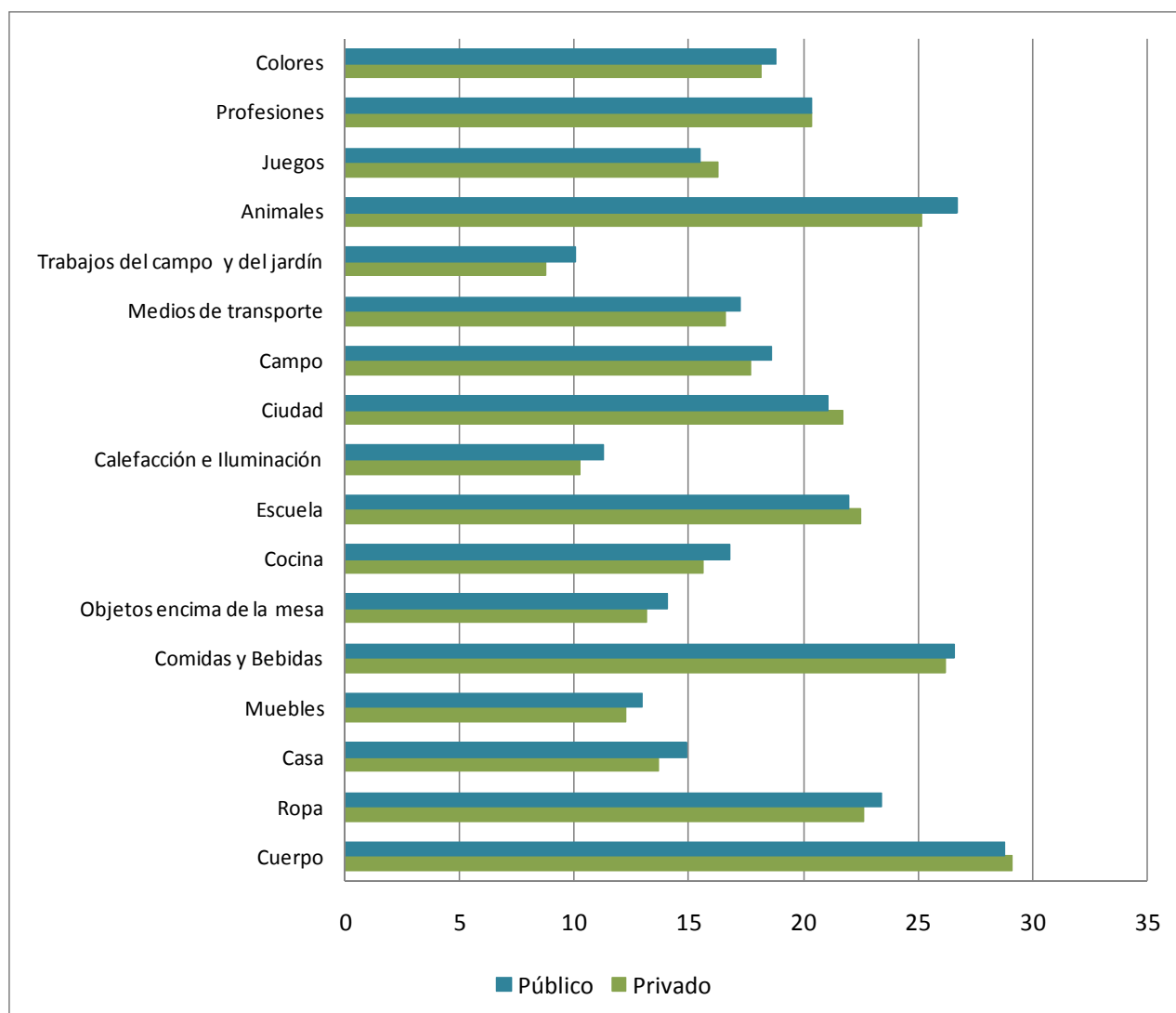
Gráfico 6.18. Promedio de palabras por centro de interés según la variable social tipo de centro en lengua castellana.



En la estratificación del corpus catalán en relación con la variable *tipo de centro*, se constata un patrón de comportamiento completamente distinto al observado en lengua castellana. En este punto, los centros públicos son los que sobrepasan en producción a los privados, aunque las diferencias son prácticamente idénticas en los promedios generales: 0.53 a favor de los centros privados en castellano, y 0.52 a favor de los públicos, en catalán. Concretamente, en doce áreas temáticas, los resultados son mejores en la enseñanza pública (*Ropa y complementos, Partes de la casa, Muebles de la casa, Comidas y bebidas, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, Calefacción, iluminación y formas de airear un recinto, El campo, Medios de transporte, Trabajos del campo y del jardín, Animales y Colores*), mientras que en los cuatro restantes el alumnado de los colegios concertados consigue mejores

promedios. Hay un centro de interés, *Profesiones y oficios*, cuyo resultado es exactamente el mismo en las dos lenguas. Las diferencias, como ya hemos indicado, no son importantes y solo en cinco casos sobrepasan la unidad (*Partes de la casa, La cocina y sus utensilios, Calefacción, iluminación y formas de airear un recinto, Trabajos del campo y del jardín y Animales*), todos ellos a favor del alumnado de los institutos de secundaria.

Gráfico 6.19. Promedio de palabras por centro de interés según la variable social tipo de centro en lengua catalana.



De otra parte, es curioso que, si contrastamos la información por centros de interés analizada hasta el momento, se observa que en algunas categorías semánticas la producción siempre es mayor en los colegios concertados (*El cuerpo humano, La escuela y La ciudad*) o en los institutos públicos (*Calefacción, iluminación y formas de*

airear un recinto, Trabajos del campo y del jardín, Animales y Colores), independientemente de la lengua en la que se complete la encuesta. Creemos que la distribución de los informantes en relación con la variable *zona geográfica*, cuyos resultados veremos a continuación, puede influir parcialmente en este hecho, de modo que el alumnado de los centros privados, todos ubicados en la capital, tenga un mayor caudal léxico en un área como *La ciudad*, y, de forma opuesta, el alumnado de los institutos públicos —situados en la ciudad pero también en la provincia— ostente mejores promedios en *Trabajos del campo y del jardín* o *Animales*, aunque también estas pueden ser tendencias relacionadas únicamente con la muestra concreta que analizamos.

Para conocer el alcance estadístico de todas estas diferencias, comparamos las medias de ambos grupos de sujetos en cada lengua mediante el test ANOVA. La significación global obtenida, para este factor social, es ‘p’= 0.230 para el castellano y ‘p’= 0.260 para el catalán, lo que nos advierte ya de que, en general, no es posible refutar la hipótesis nula, con lo que no podemos argüir que las diferencias relativas a la producción en función del *tipo de centro* al que acude el alumno se deban realmente a esta causa. Veamos qué sucede en el análisis concreto de cada centro de interés, recogido en la tabla 6.24.

Tabla 6.24. Análisis estadístico de la diferencia según el centro educativo.
Valor de la ‘p’ en las dos lenguas.

Nº	Centros de interés	Castellano	Catalán
1	El cuerpo humano	0.738	0.646
2	Ropa y complementos	0.007	0.332
3	Partes de la casa (sin los muebles)	0.011	0.039
4	Muebles de la casa	0.034	0.221
5	Comidas y bebidas	0.950	0.553
6	Objetos colocados en la mesa para la comida	0.191	0.113
7	La cocina y sus utensilios	0.220	0.141
8	La escuela (muebles y materiales)	0.442	0.486
9	Calefacción, iluminación y formas de airear un recinto	0.834	0.045
10	La ciudad	0.293	0.455
11	El campo	0.437	0.326
12	Medios de transporte	0.067	0.321
13	Trabajos del campo y del jardín	0.368	0.011
14	Animales	0.924	0.075
15	Juegos y distracciones	0.858	0.275
16	Profesiones y oficios	0.287	1.000
17	Colores	0.238	0.320

- De las doce categorías donde los promedios obtenidos por los centros privados son superiores a los públicos en lengua castellana, dicha diferencia resulta significativa en tres de ellas: *Ropa y complementos*, *Partes de la casa y Muebles de la casa*.
- De las doce áreas donde la productividad es mayor en los institutos públicos en lengua catalana, en tres de ellas la 'p' es inferior a 0.05: *Partes de la casa*, *Calefacción, iluminación y formas de airear un recinto* y *Trabajos del campo y del jardín*.

Es llamativa, de un lado, la presencia del tercer centro de interés, *Partes de la casa*, en ambas clasificaciones, y de otro lado, la significación del centro *Trabajos del campo y del jardín* a favor de los institutos públicos en lengua catalana que podría explicarse, como hemos comentado, por la presencia de alumnado ubicado en poblaciones rurales de la provincia solo en la muestra de los institutos públicos⁷⁷². Sin embargo, lo que queda demostrado es que, en nuestro corpus, el factor *tipo de centro* no discrimina el comportamiento léxico de los hablantes preuniversitarios, a pesar de que se observan dos tendencias opuestas en relación con esta variable: los centros privados alcanzan mejores promedios en lengua castellana, y los públicos, en catalán.

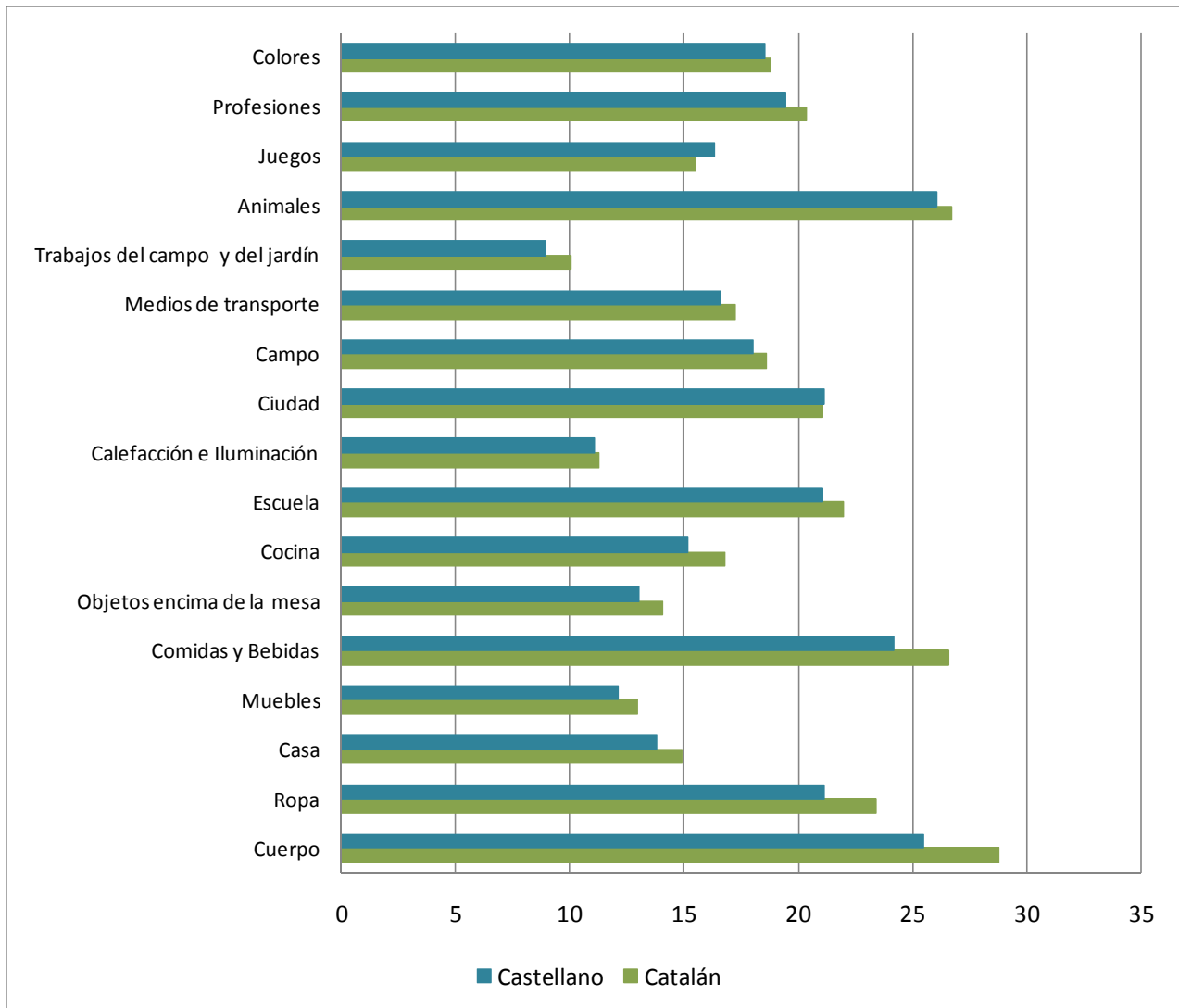
Los estudios que nos preceden hasta la fecha, en cambio, otorgan un valor discriminador a esta variable, siempre a favor de los centros privados o concertados, aunque en ocasiones las diferencias recabadas son mayores y, en otras, menos destacables. Mucho se ha discutido sobre la independencia de este factor que, como también sucede en nuestro caso (*vid.* 5.2, gráfico 5.9.), se relaciona, a menudo, con el *nivel sociocultural* de la familia. Sin embargo, hasta la fecha se ha comprobado no solo que es un factor importante en la estratificación del léxico disponible de algunas comunidades, sino que los resultados que arroja esta variable no son siempre equiparables a los obtenidos mediante la diferenciación social de los informantes. En algunas investigaciones realizadas tanto en castellano (Gómez Molina y Gómez Devís 2004) como en catalán (Llopis 2009) ha cobrado una gran relevancia, lo cual no deja de ser sorprendente si se tiene en cuenta que los contenidos que se imparten en las

⁷⁷² Esta área semántica también favorece al alumnado de los centros públicos en lengua castellana, aunque dicha diferencia no es significativa, como acabamos de ver.

instituciones educativas, independientemente de su titularidad, no pueden diferir demasiado. No es el caso de nuestra provincia, donde, como ha quedado evidenciado, esta variable no ejerce un efecto estadístico ni en el promedio global, ni en la mayor parte de las áreas asociativas en ninguna de las dos lenguas.

Pasemos ahora a evaluar los resultados según la producción de cada colectivo en cada una de las dos lenguas de la comunidad. Los centros públicos, como era de esperar, consiguen una mejor media en lengua catalana, con una diferencia exacta de una palabra (17.77 frente a 18.77). El catalán también impera en la mayoría de categorías semánticas, salvo en dos de ellas (*La ciudad* y *Juegos y distracciones*), que obtienen mejores promedios en castellano, tal como también se observa en los resultados globales (*vid.* 6.1.3.). Las mayores diferencias, en este sentido, se dan en *El cuerpo humano*, que supera las tres palabras a favor del catalán, y en *Ropa y complementos* y *Comidas y bebidas*, donde se sobrepasan las dos unidades. En estas áreas —como ya comentamos— creemos que se vislumbra el efecto de la lengua materna.

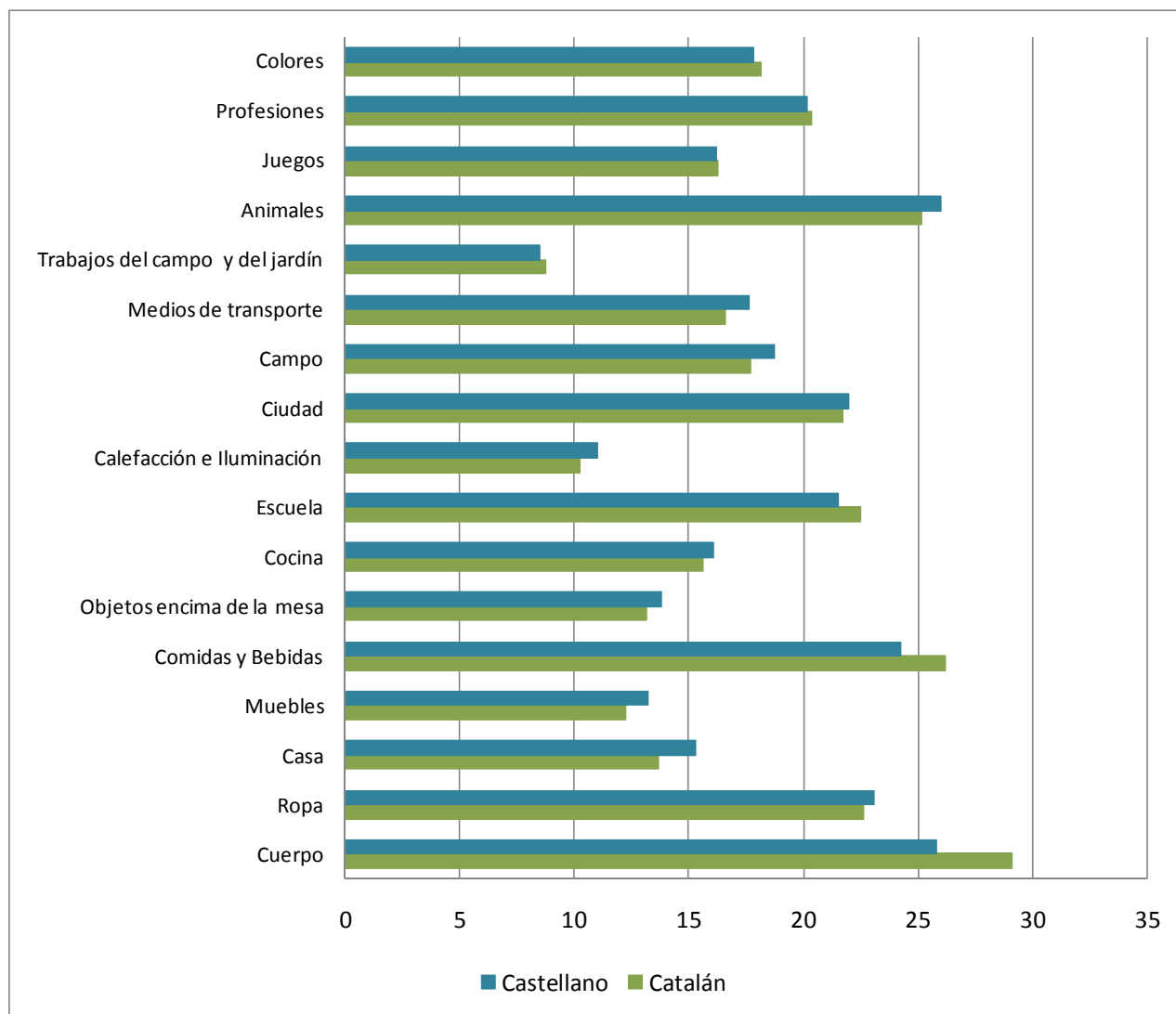
Gráfico 6.20. Comparación de la producción en castellano y en catalán de los informantes de los centros públicos.



En cuanto a la producción de los centros privados, el caudal de léxico disponible en las dos lenguas consigue un resultado muy cercano, con apenas unas centésimas de diferencia (18.30 frente a 18.25). El castellano obtiene un resultado mejor, aunque la diferencia entre las dos lenguas (0.05) es la menor de todas las analizadas en este capítulo. El alumnado de los colegios concertados alcanza mejores promedios en castellano en diez centros de interés (*Ropa y complementos, Partes de la casa, Muebles de la casa, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, Calefacción, iluminación y formas de airear un recinto, La ciudad, El campo, Medios de transporte y Animales*), mientras que en los siete restantes, muestra una media más

favorable al catalán⁷⁷³. La mayor disimilitud entre las dos lenguas, superando las tres palabras, se da en *El cuerpo humano*, al igual que sucedía en los centros públicos, siempre a favor de la lengua catalana.

Gráfico 6.21. Comparación de la producción en castellano y en catalán de los informantes de los centros privados.



A la vista de los resultados que acabamos de ver y dado que la variable *tipo de centro* no ejerce una discriminación estadística —más que en algunos centros de interés aislados—, consideramos que la explicación más posible al comportamiento dispar y opuesto de la muestra con respecto a las dos lenguas de la comunidad radica en la relación existente entre algunas de las variables de estratificación social. Tal como dijimos en el capítulo anterior (*vid.* 5.2.), al cotejar todas las posibles relaciones entre

⁷⁷³ Curiosamente, uno de ellos es *Juegos y distracciones*, donde globalmente, los resultados muestran la preponderancia del castellano, aunque la diferencia es realmente pequeña en este caso.

los distintos factores, advertimos que había una relación estrecha entre el alumnado en función del tipo de centro al que acudía y su lengua materna. De este modo, en nuestra muestra en concreto, hay un mayor número de alumnado cuya lengua materna es el castellano en los centros privados y un mayor número de informantes cuya lengua materna es la vernácula en las instituciones públicas (*cf.* 5.2., gráfico 5.11)⁷⁷⁴.

Si se confirma la significación estadística de la variable *lengua materna*, este hecho podría explicar que los resultados favorezcan a los centros privados frente a los públicos en castellano, y a los públicos frente a los privados en catalán. Desconocemos si esta distribución de nuestra muestra es un hecho casual, o si, realmente, hay más alumnado castellano hablante en los centros privados de la ciudad, dato que, por otro lado, es imposible obtener, puesto que no consta en ninguna fuente oficial. Otra posibilidad sería que se detectara un mayor uso de la lengua castellana en los centros concertados, pero, de un lado, la actual legislación establece las mismas directrices para todos los centros de enseñanza con respecto al catalán como lengua vehicular y, de otro, como explicamos en el capítulo anterior (*vid.* 5.3.), no observamos ninguna diferencia considerable al respecto cuando recopilamos información, en el momento de aplicar la encuesta, sobre el número de horas que recibía el alumnado en las dos lenguas cooficiales, ya que la proporción era prácticamente idéntica en todas las instituciones educativas donde pasamos las encuestas.

Tabla 6.25. Total de vocablos por centro de interés según la variable social tipo de centro.

Nº	Centros de interés	Centros públicos		Centros privados	
		Castellano	Catalán	Castellano	Catalán
1	El cuerpo humano	347	318	205	210
2	Ropa y complementos	248	289	186	184
3	Partes de la casa (sin los muebles)	212	233	135	124
4	Muebles de la casa	233	249	149	139
5	Comidas y bebidas	420	478	303	305
6	Objetos colocados en la mesa para la comida	215	248	138	157
7	La cocina y sus utensilios	303	332	187	193
8	La escuela (muebles y materiales)	376	397	202	209
9	Calefacción, iluminación y formas de airear un recinto	256	255	135	122
10	La ciudad	582	567	329	308

⁷⁷⁴ Concretamente, en los centros públicos, la relación de alumnado cuya lengua materna es el castellano y el catalán respectivamente es de 23.33% frente a 68.89%; en cambio, en los centros privados dicha relación es del 40% frente al 55%. Recuérdese que también hay informantes que se engloban en otras adscripciones lingüísticas (*ambas lenguas/aranés*).

11	El campo	616	676	352	324
12	Medios de transporte	241	256	148	165
13	Trabajos del campo y del jardín	364	365	181	150
14	Animales	309	399	227	236
15	Juegos y distracciones	495	488	293	274
16	Profesiones y oficios	512	499	324	319
17	Colores	207	225	126	113
Total		5936	6274	3620	3532

Con respecto a la riqueza léxica, tal como se observa en la tabla anterior, en cifras absolutas, los institutos superan en todos los centros de interés y en las dos lenguas a los colegios privados, como es totalmente lógico, dado que aportan tres veces más informantes. Dentro de cada subvariable, los centros públicos, que ostentan mejores promedios en lengua catalana, también tienen un número mayor de palabras distintas en lengua vernácula, tanto en el total (6274 frente a 5936) como en el análisis de cada una de las áreas asociativas. Concretamente, solo en cinco de ellas los resultados en castellano son superiores (*El cuerpo humano, Calefacción, iluminación y formas de airear un recinto, La ciudad, Juegos y distracciones y Profesiones y oficios*)⁷⁷⁵. Los centros concertados, que muestran, como acabamos de ver, una tímida ventaja en la producción en lengua castellana, tienen también en esta lengua un mayor número de vocablos, aunque la diferencia (de 88 palabras) a favor del castellano es un tanto menor a la observada en los centros públicos a favor del catalán⁷⁷⁶. En relación con las categorías semánticas, en diez se recaban más vocablos en castellano (*Ropa y complementos, Partes de la casa, Muebles de la casa, Calefacción, iluminación y formas de airear un recinto, La ciudad, El campo, Trabajos del campo y del jardín, Juegos y distracciones y Profesiones y oficios y Colores*) y, en cambio, solo en las siete categorías restantes se aportan más palabras distintas en catalán. Las divergencias más acusadas son las 90 palabras más que actualizan los estudiantes de los institutos públicos en lengua catalana bajo *Animales* y las observadas en el centro 11, *El campo*⁷⁷⁷.

⁷⁷⁵ Obsérvese que estas son las mismas áreas que, en los resultados generales de vocablos, favorecen al castellano. La única que no aparece es *Trabajos del campo y del jardín*, que —por una palabra— obtiene mejor resultado en catalán en este colectivo. Nótese que la diferencia es también de una palabra, a favor del castellano, en el centro 9.

⁷⁷⁶ Esta es, concretamente, de 338 palabras distintas más, pero recuérdese —para evaluar estas cifras correctamente— que el alumnado de los centros públicos es tres veces más (180 frente a 60).

⁷⁷⁷ Los informantes de los centros públicos aportan 60 palabras más en catalán, y los de los privados, 28 más en castellano. Téngase en cuenta la posible presencia de informantes con un comportamiento atípico en una o en las dos lenguas.

La clasificación de las áreas semánticas según este parámetro sigue las tendencias globales ya vistas: *El campo*, *La ciudad* y *Profesiones y oficios* ocupan siempre las primeras posiciones; en las últimas, en cambio, hay más disparidad aunque se siguen patrones muy similares a los generales.

Tabla 6.26. Índice de cohesión por centro de interés según la variable social tipo de centro.

Nº	Centros de interés	Centros públicos		Centros privados	
		Castellano	Catalán	Castellano	Catalán
1	El cuerpo humano	0.07	0.09	0.13	0.14
2	Ropa y complementos	0.09	0.08	0.12	0.12
3	Partes de la casa (sin los muebles)	0.07	0.06	0.11	0.11
4	Muebles de la casa	0.05	0.05	0.09	0.09
5	Comidas y bebidas	0.06	0.06	0.08	0.09
6	Objetos colocados en la mesa para la comida	0.06	0.06	0.10	0.08
7	La cocina y sus utensilios	0.05	0.05	0.09	0.08
8	La escuela (muebles y materiales)	0.06	0.06	0.11	0.11
9	Calefacción, iluminación y formas de airear un recinto	0.04	0.04	0.08	0.08
10	La ciudad	0.04	0.04	0.07	0.07
11	El campo	0.03	0.03	0.05	0.05
12	Medios de transporte	0.07	0.07	0.12	0.10
13	Trabajos del campo y del jardín	0.02	0.03	0.05	0.06
14	Animales	0.08	0.07	0.11	0.11
15	Juegos y distracciones	0.03	0.03	0.06	0.06
16	Profesiones y oficios	0.04	0.04	0.06	0.06
17	Colores	0.09	0.08	0.14	0.16

Con respecto al último de los parámetros analizados, el índice de cohesión, queda demostrado, con un simple vistazo a la tabla 6.26, que el menor número de informantes de los centros privados incide en que estos tengan siempre un índice de cohesión mayor. Si nos fijamos en las diferencias según la lengua de la encuesta, parece evidente que este valor nos informa sobre la constitución interna de los centros de interés, y de que esta no difiere según el idioma, puesto que se dan muchísimas coincidencias entre la cohesión de las dos lenguas según el estímulo analizado (*La escuela*, *Calefacción, iluminación y formas de airear un recinto*, *La ciudad*, *El campo*, *Juegos y distracciones*, *Profesiones y oficios*), al igual que sucedía al analizar los datos en su globalidad (vid. 6.1.5.).

6.2.3. Variable social zona geográfica

Como ya explicamos en el capítulo anterior (*vid.* 5.2), consideramos esta variable en función del lugar donde se ubicaba el centro al que acude el alumno. De este modo, la mitad exacta de la muestra (120 informantes) pertenece a distintos centros educativos ubicados en la capital, mientras que la otra mitad procede de institutos situados en distintos puntos de la provincia. El hecho de que sea exactamente el mismo número de informantes el que se distribuya en las dos variantes facilita la comparación de los datos que analizamos en el presente apartado, como se verá a continuación.

Al comparar los promedios alcanzados, se detecta cómo el mejor resultado lo obtiene el alumnado de procedencia rural en lengua catalana (con una media de 18.94), el peor dato lo obtienen exactamente los mismos alumnos, pero en lengua castellana (con 17.67 palabras de promedio). Por el contrario, la actuación lingüística de los informantes de la capital es mucho más homogénea, dado que la diferencia entre las dos lenguas es únicamente de un par de décimas (18.13 en castellano y 18.34 en catalán). De otra parte, en lengua castellana, consigue mejor media global el alumnado de los centros urbanos (18.13 frente a 17.67), mientras que, en catalán, son los informantes rurales los que ostentan un mejor promedio (18.94 frente a 18.34). La diferencia en estos dos últimos casos es similar, alrededor de la media palabra (0.46 en el primero y 0.6 en el segundo).

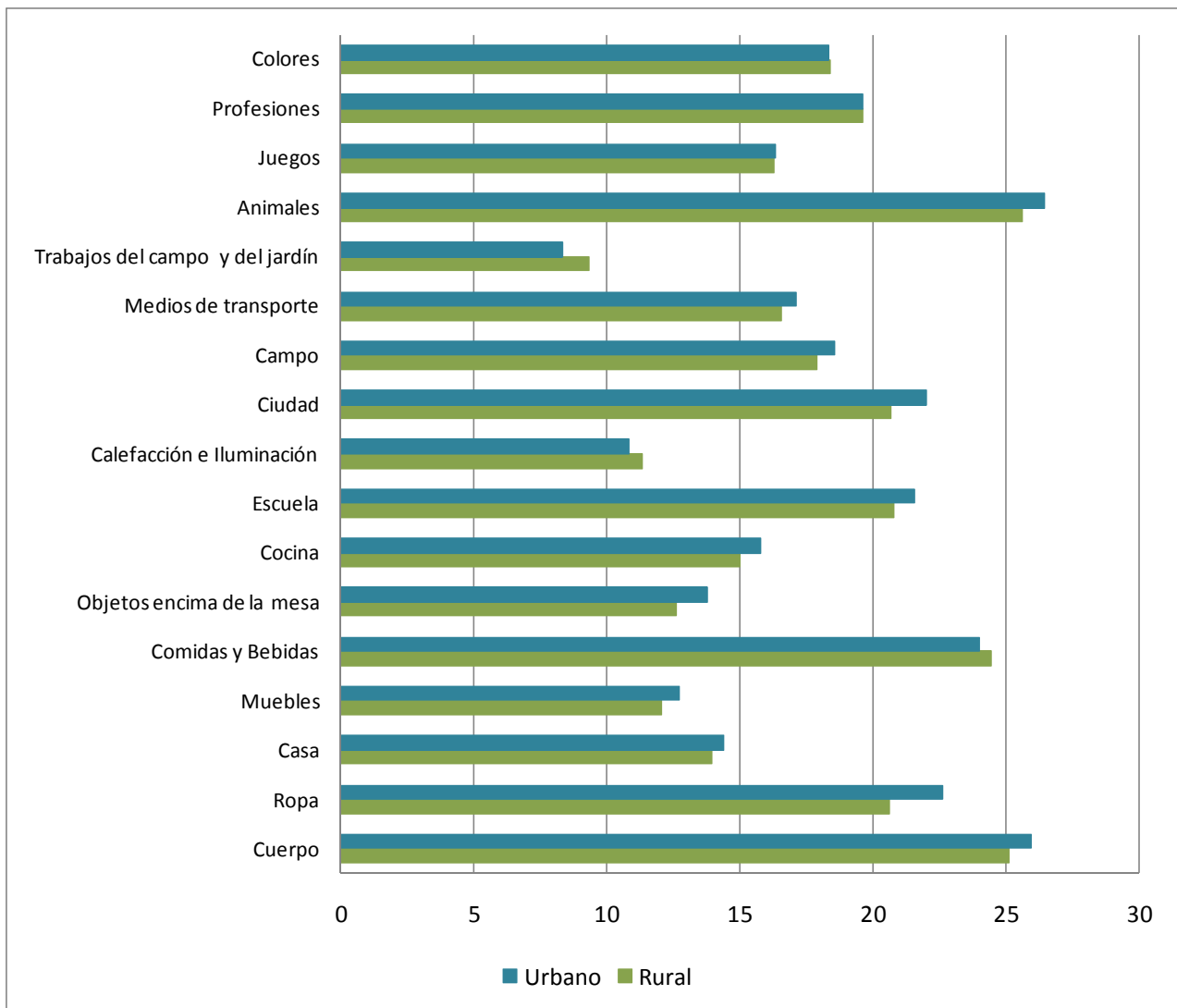
Tabla 6.27. Promedio de palabras por centro de interés según la variable zona geográfica.

Nº	Centros de interés	Castellano		Catalán	
		Urbano	Rural	Urbano	Rural
1	El cuerpo humano	25.95	25.13	28.74	28.96
2	Ropa y complementos	22.63	20.61	22.90	23.48
3	Partes de la casa (sin los muebles)	14.37	13.97	13.85	15.34
4	Muebles de la casa	12.73	12.06	12.50	13.10
5	Comidas y bebidas	23.98	24.45	25.92	27.06
6	Objetos colocados en la mesa para la comida	13.80	12.61	13.66	14.05
7	La cocina y sus utensilios	15.77	14.99	16.30	16.65
8	La escuela (muebles y materiales)	21.54	20.79	22.05	22.20
9	Calefacción, iluminación y formas de airear un recinto	10.81	11.36	10.42	11.68
10	La ciudad	22.02	20.68	21.86	20.64
11	El campo	18.53	17.88	18.11	18.68
12	Medios de transporte	17.11	16.58	16.88	17.30
13	Trabajos del campo y del jardín	8.36	9.35	9.12	10.32
14	Animales	26.43	25.61	25.51	27.13
15	Juegos y distracciones	16.32	16.29	15.72	15.70

16	Profesiones y oficios	19.62	19.63	19.89	20.84
17	Colores	18.32	18.38	18.44	18.84
Media		18.13	17.67	18.34	18.94

Con respecto a los resultados en lengua castellana, reflejados en el gráfico 6.22, los informantes urbanos consiguen mejores promedios en doce áreas temáticas: *El cuerpo humano, Ropa y complementos, Partes de la casa, Muebles de la casa, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, La escuela, La ciudad, El campo, Medios de transporte, Animales y Juegos y distracciones*. En este sentido, la mayor diferencia es la de *Ropa y complementos*, de poco más de dos palabras, y solo en dos estímulos más se supera la unidad (*Objetos colocados en la mesa para la comida y La ciudad*), puesto que, como hemos visto, la diferencia general entre la producción de los informantes de las dos zonas geográficas no es mucha (0.46). En cinco centros de interés son los informantes de centros rurales los que tienen mayor productividad (*Comidas y bebidas, Calefacción, iluminación y formas de airear un recinto, Trabajos del campo y del jardín, Profesiones y oficios y Colores*). En esta dirección, la mayor disimilitud se da, precisamente, en *Trabajos del campo y del jardín*, y es prácticamente de una palabra entre el promedio de los dos grupos sociales.

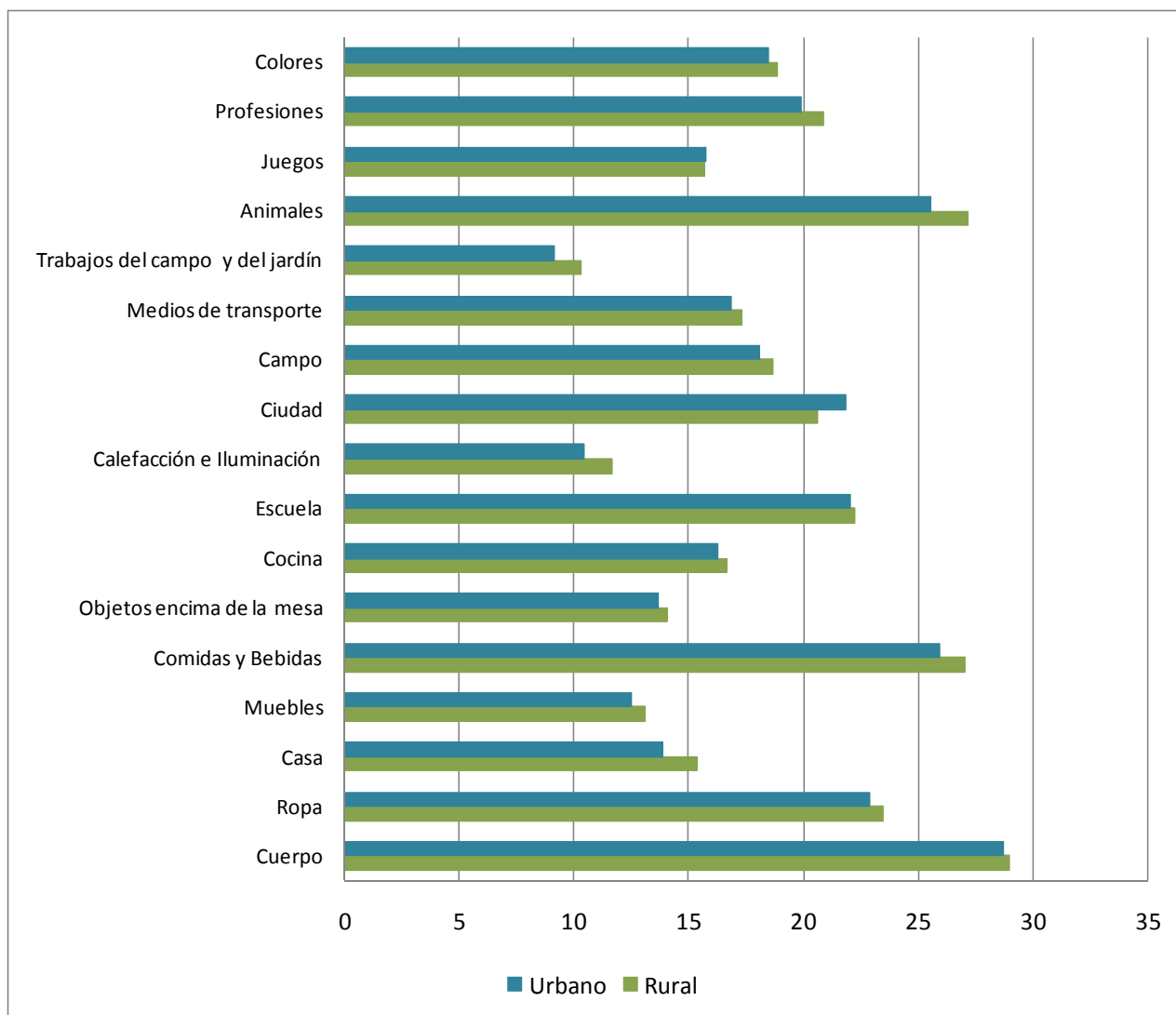
Gráfico 6.22. Promedio de palabras por centro de interés según la variable social zona geográfica en lengua castellana.



En cuanto a los resultados en lengua catalana, al igual que sucedía en la variable anterior *tipo de centro*, se constata un comportamiento totalmente opuesto de los dos colectivos, lo que nos da una idea de la importancia que adquiere la lengua en que se completa la encuesta. Ahora los mejores resultados los obtienen los sujetos procedentes de distintos puntos de la provincia frente a los de la capital, aunque la distancia es similar a la que se observa en lengua castellana (0.6). El promedio de palabras por centro de interés es mejor entre los informantes rurales en todas las categorías semánticas, salvo en *La ciudad* y *Juegos y distracciones*. Precisamente, en *La ciudad*, la diferencia —a favor del alumnado urbano— sobrepasa la unidad, mientras que en *Juegos y distracciones* es irrelevante, de apenas dos centésimas. En sentido contrario, a favor de los informantes rurales, solo en cinco centros de interés se supera la unidad:

Partes de la casa, Comidas y bebidas, Calefacción, iluminación y formas de airear un recinto, Trabajos del campo y del jardín y Animales.

Gráfico 6.23. Promedio de palabras por centro de interés según la variable social zona geográfica en lengua catalana.



Un hecho que nos parece muy interesante es la importancia que parece cobrar el entorno en el caudal de léxico disponible del informante. En nuestra opinión, no puede ser un hecho casual que —independientemente de la lengua en que se complete la encuesta— un centro como *La ciudad* otorgue siempre mejores resultados al alumnado ubicado en la capital, tanto en castellano como en catalán⁷⁷⁸. De modo contrario y tal como hemos visto, en *Trabajos del campo y del jardín* —donde se evoca, como vimos, principalmente acciones relacionadas con los trabajos agrícolas— también se observan

⁷⁷⁸ Recordemos que, como explicamos en el apartado 4.3.1, no existe ningún núcleo poblacional de importancia demográfica o social comparable con la capital en nuestra provincia.

diferencias, en ambas lenguas, a favor del alumnado rural. En estos dos centros de interés y en las dos lenguas, la distancia entre los promedios siempre supera la unidad, lo que ya es bastante si se tiene en cuenta que las divergencias globales que imprime esta variable, en los dos corpus, no son muy grandes⁷⁷⁹.

Tras realizar el análisis de ANOVA se comprueba cómo no es posible refutar la hipótesis nula en el promedio general en ninguna de las dos lenguas, dado que la probabilidad de significación es superior a 0.05 ($p = 0.223$ en castellano y $p = 0.142$ en catalán). Las diferencias que ejerce esta variable sobre los resultados podrían deberse a otras causas distintas a la distribución de la muestra en dos zonas geográficas, de igual modo los resultados no pueden extrapolarse al conjunto del universo estudiado.

Tabla 6.28. Análisis estadístico de la diferencia según la zona geográfica.
Valor de la 'p' en las dos lenguas.

Nº	Centros de interés	Castellano	Catalán
1	El cuerpo humano	0.315	0.761
2	Ropa y complementos	0.001	0.366
3	Partes de la casa (sin los muebles)	0.431	0.003
4	Muebles de la casa	0.148	0.218
5	Comidas y bebidas	0.490	0.144
6	Objetos colocados en la mesa para la comida	0.019	0.456
7	La cocina y sus utensilios	0.220	0.597
8	La escuela (muebles y materiales)	0.192	0.817
9	Calefacción, iluminación y formas de airear un recinto	0.230	0.005
10	La ciudad	0.040	0.091
11	El campo	0.387	0.471
12	Medios de transporte	0.291	0.664
13	Trabajos del campo y del jardín	0.021	0.016
14	Animales	0.351	0.030
15	Juegos y distracciones	0.955	0.968
16	Profesiones y oficios	0.741	0.142
17	Colores	0.906	0.450

Tras analizar la probabilidad de cada centro de interés en particular en relación con la variable *zona geográfica*, según vemos en la tabla 6.28, se constata que:

- De los doce centros de interés donde los promedios obtenidos en los centros urbanos son superiores a los públicos en lengua castellana, dicha diferencia resulta significativa en tres de ellos: *Ropa y complementos*, *Objetos colocados en la mesa para la comida* y *La ciudad*. Cabe recordar que, también en

⁷⁷⁹ Sin embargo, no se detecta tal diferencia en otros centros como *El campo*, donde sería esperable que el alumnado rural obtuviera también mejores promedios.

castellano, el alumnado de los centros rurales obtiene mejores promedios en cinco áreas temáticas. En una de ellas, *Trabajos del campo y del jardín*, dicha diferencia es significativa.

- De las quince categorías semánticas donde la productividad es mayor en los institutos rurales en lengua catalana, en cuatro de ellas es posible refutar la hipótesis nula: *Partes de la casa*, *Calefacción*, *iluminación y formas de airear un recinto*, *Trabajos del campo y del jardín* y *Animales*⁷⁸⁰.

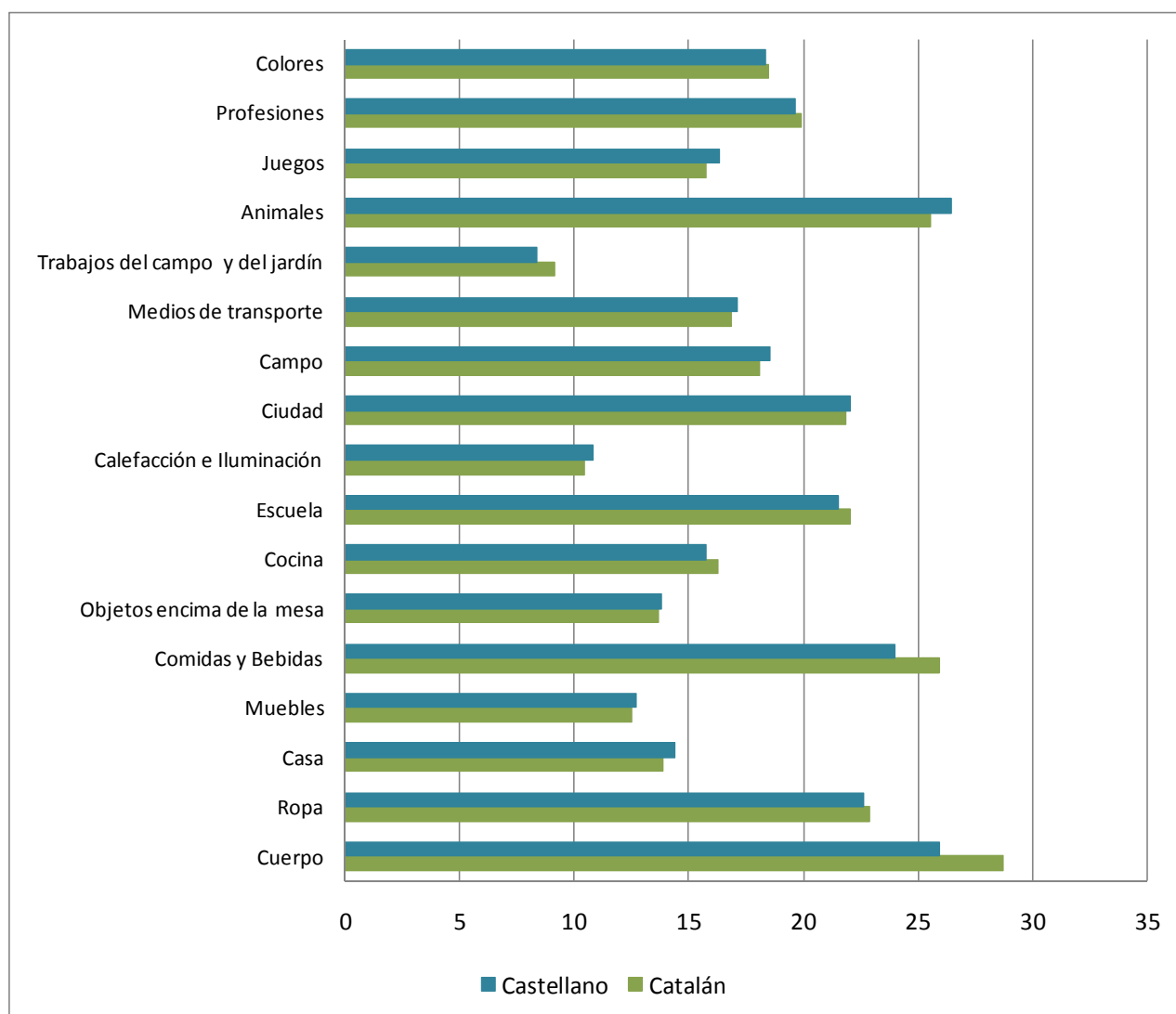
Cabe destacar que la diferencia es significativa, a favor del alumnado rural, en el centro *Trabajos del campo y del jardín* en las dos lenguas estudiadas. Este hecho creemos que confirma la influencia que adquiere el hábitat medioambiental en el léxico disponible de los estudiantes. De modo contrario, en lengua castellana, también tiene significación estadística el resultado superior del alumnado urbano en *La ciudad*. La importancia del entorno también puede influir en la diferencia significativa que se da en *Animales*, a favor del alumnado rural, aunque solo se observe en catalán.

Los estudios anteriores evidencian, con respecto a esta variable, que la productividad léxica es mayor, generalmente, en los informantes urbanos⁷⁸¹. Sin embargo, la importancia de este factor difiere según la comunidad que se analice y presenta resultados distintos según la forma en que se divida cada sintopia: en diversas zonas geográficas o, como en nuestro caso, diferenciando la capital del resto de la provincia. En este trabajo, como acabamos de comprobar, la influencia que ejerce en los resultados no es muy importante, ya que, ni en el promedio general ni en la mayor parte de centros de interés se corrobora la significación estadística de este factor social. Su valor es similar al que ejerce el factor *tipo de centro*, aunque un poco superior, ya que estratifica un centro de interés más en cada lengua (cuatro en cada lengua según la *zona geográfica* y, en cambio, tres según el *tipo de centro*). Ambas variables están por debajo de la relevancia que cobra en nuestro análisis el *sexo*, hasta ahora la variable más significativa.

⁷⁸⁰ Nótese que tres de estas áreas (3, 9 y 13) coinciden con las que obtiene resultados más favorables el alumnado de los centros públicos y en las que dicha diferencia es significativa (*vid.* 6.2.2.)

⁷⁸¹ No obstante, hay algún ejemplo donde se corrobora la tendencia contraria, como en Gómez Molina y Gómez Devís (2004).

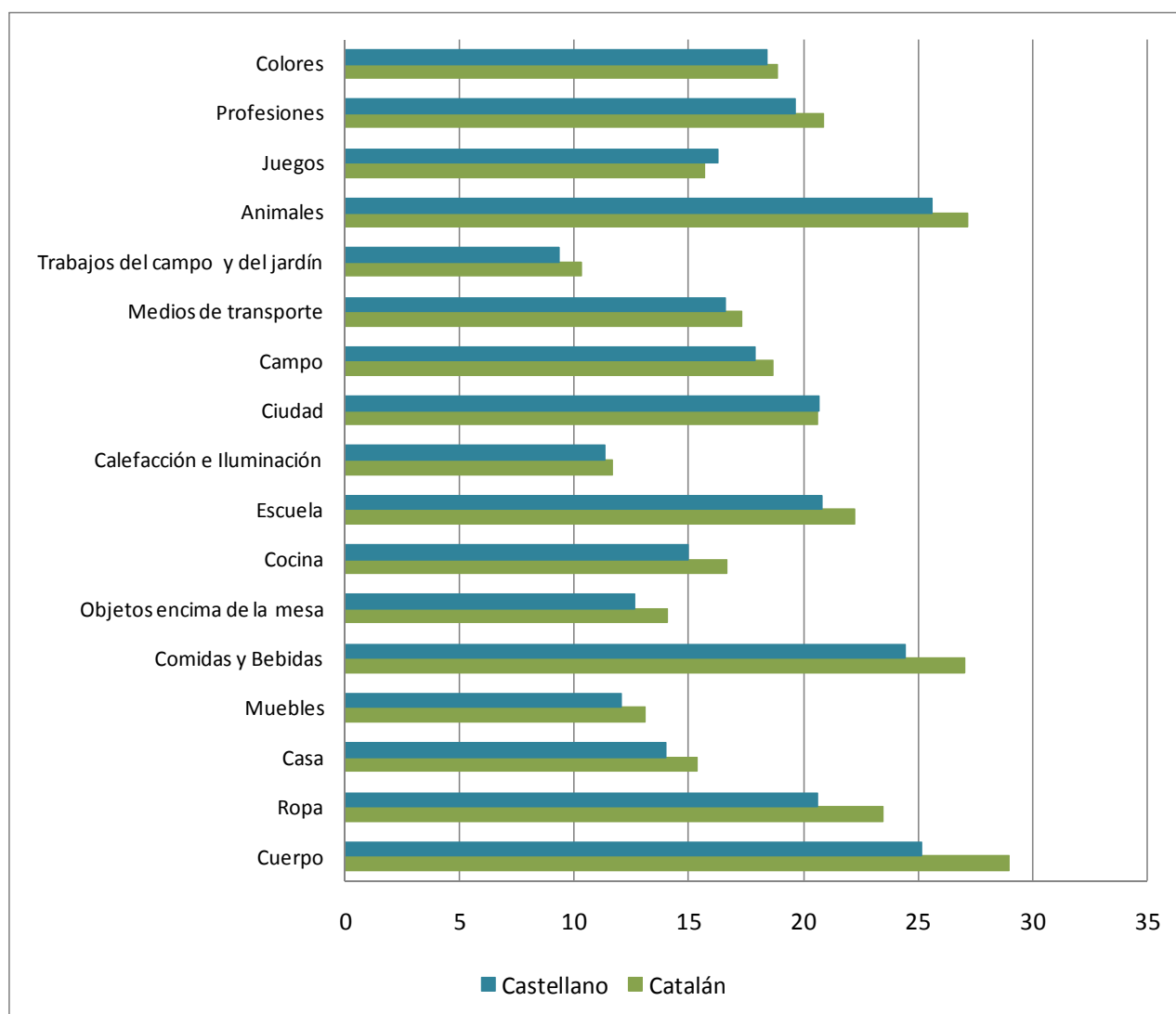
Gráfico 6.24 Comparación de la producción en castellano y en catalán de los informantes urbanos.



En el análisis de los resultados según la lengua de producción, se constata cómo el alumnado de los centros urbanos —esto es, procedente de la capital— produce unos resultados muy similares en ambas lenguas. Aunque el promedio general es dos décimas a favor del catalán (18.34 frente a 18.13), en el desglose por áreas temáticas se comprueba cómo la productividad léxica es mayor en nueve áreas en lengua castellana (*Partes de la casa, Muebles de la casa, Objetos colocados en la mesa para la comida, Calefacción, iluminación y formas de airear un recinto, La ciudad, El campo, Medios de transporte, Animales y Juegos y distracciones*); en las ocho restantes, por el contrario, el resultado es mejor en catalán. Las mayores diferencias se dan en el primero de los centros de interés (con casi tres palabras de más en lengua vernácula) y en el quinto, siguiendo las tendencias generales.

En cuanto al alumnado de los centros rurales, cuyos resultados se plasman en el gráfico 6.25, este muestra un predominio mucho mayor de la lengua catalana, con una distancia global de 1.27 respuestas. Su actuación lingüística es mejor en catalán en todas las áreas excepto en *La ciudad* y en *Juegos y distracciones*. Las mayores diferencias se observan, además de en *El cuerpo humano* y en *Comidas y bebidas* —al igual que sucedía en el comportamiento del alumnado urbano—, en *Ropa y complementos*. La diferencia en la producción en el primer centro de interés es muy llamativa, con casi cuatro palabras de más en catalán, mientras que en las otras dos áreas, la segunda y la quinta, se superan las 2.5 unidades. Es revelador que, tal como dijimos en el análisis global de los datos, ambos colectivos tengan más palabras en castellano en *La ciudad* y *Juegos y distracciones*, debido a la influencia del español en estos ámbitos (*cfr.* 6.1.3).

Gráfico 6.25. Comparación de la producción en castellano y en catalán de los informantes rurales.



Antes de continuar el análisis, nos gustaría constatar la relación que se observa entre los resultados obtenidos mediante esta variable y la anterior, el *tipo de centro*. En ambos casos, en la producción en lengua castellana, el alumnado de los centros privados (60 informantes) y urbanos (120, de los cuales 60 son los de los colegios concertados) consigue mejores resultados que los informantes de zonas rurales y centros públicos. De forma inversa, estos últimos obtienen en la producción en lengua catalana promedios más favorables que los centros urbanos y que los colegios privados.

También se observan similitudes en la producción de cada colectivo en las dos lenguas. La actuación lingüística de los informantes de los centros rurales y públicos es claramente a favor del catalán, con una distancia similar, en ambos casos, respecto al castellano (1.27 en el caso de los centros rurales y de una respuesta en el de los centros

públicos). En cambio, los informantes procedentes de instituciones educativas concertadas y de centros urbanos, en general, muestran una menor diferencia entre el caudal léxico de las dos lenguas. Aunque se observan tendencias opuestas —los urbanos consiguen mejores promedios en catalán y los privados en castellano—, la distancia entre los promedios generales que se obtienen en cada idioma es similar (0.21 los centros urbanos a favor del catalán y 0.05 los concertados a favor del castellano).

A nuestro modo de ver, e igual que sucedía en el apartado anterior, creemos que nuestros resultados se explican, de nuevo, por la relación que existe entre algunas de las variables sociales contempladas en este estudio. En el capítulo cuarto, observamos cómo la población de origen castellano hablante tiene un peso mayor en la ciudad de Lleida que en el resto de la provincia (se puede constatar, claramente, comparando los gráficos 4.11 y 4.12, *vid.* 4.3.2.). La misma realidad se comprueba en nuestra muestra, ya que, en el capítulo anterior, al cotejar las variables entre ellas, ya nos dimos cuenta de esta relación: mientras en los centros urbanos hay más catalanohablantes pero en una proporción similar a la de los informantes con lengua materna castellana, el predominio de los catalanohablantes en los centros rurales es abrumador, tal como se evidencia en el gráfico 5.10 (*vid.* 5.2). Si la mayor presencia de alumnado castellano hablante en los centros privados no contaba con una justificación clara, aquí se constata que la distribución de nuestra población evidencia la relación que existe en la propia provincia de Lleida, y es consecuencia, como ya explicamos (*vid.* 4.1.), de que el contingente migratorio llegado en los años de la posguerra tendió a instalarse en las ciudades, donde había una mayor demanda estable de mano de obra en el sector de los servicios y de la construcción. Por tanto, con todos estos datos sobre la mesa, y dado que la variable *zona geográfica* solo ejerce discriminación estadística en cuatro centros de interés, debemos esperar a corroborar la importancia del factor *lengua materna* para poder evaluar estos resultados en su globalidad.

Tabla 6.29. Total de vocablos por centro de interés según la variable social zona geográfica.

C.I.	Centros de Interés	Urbano		Rural	
		Castellano	Catalán	Castellano	Catalán
1	El cuerpo humano	287	288	300	261
2	Ropa y complementos	247	253	201	226
3	Partes de la casa (sin los muebles)	169	177	183	194
4	Muebles de la casa	204	203	187	200
5	Comidas y bebidas	395	396	349	411
6	Objetos colocados en la mesa para la comida	198	227	162	182
7	La cocina y sus utensilios	250	281	258	263
8	La escuela (muebles y materiales)	308	323	302	320
9	Calefacción, iluminación y formas de airear un recinto	199	180	195	213
10	La ciudad	473	458	465	446
11	El campo	512	515	482	517
12	Medios de transporte	201	221	193	212
13	Trabajos del campo y del jardín	270	245	284	282
14	Animales	279	324	268	326
15	Juegos y distracciones	417	390	395	402
16	Profesiones y oficios	445	435	412	414
17	Colores	176	172	169	182
Total		5030	5088	4805	5051

Con respecto a los vocablos, esto es, a la riqueza léxica que poseen los informantes según la zona geográfica en la que se adscriben, los resultados se plasman en la tabla anterior. El hecho de que el alumnado se distribuya en dos grupos idénticos numéricamente hace que todos los datos sean comparables entre sí. El mejor resultado lo ostentan, globalmente, los informantes urbanos ya que actualizan 262 palabras distintas más que los rurales (10118 frente a 9856, si sumamos los vocablos obtenidos en cada lengua). Si tenemos en cuenta la producción en las dos lenguas contempladas, los informantes urbanos en lengua catalana tienen la cifra más alta de palabras distintas evocadas; por el contrario, el peor resultado, al igual que sucede con respecto a la productividad, lo ostenta el alumnado rural en lengua española. Como también se observa en relación con los promedios globales de palabras, los informantes urbanos tienen un comportamiento más similar en los dos idiomas, ya que actualizan, en conjunto, solo 58 vocablos más en catalán. En cambio, entre los informantes rurales se aprecian más las diferencias en función de la lengua de la encuesta: en catalán, tienen mayor diversidad de vocabulario y aportan 246 palabras distintas más.

Por áreas asociativas, los informantes urbanos tienen más vocablos en diez áreas en catalán (*El cuerpo humano, Ropa y complementos, Partes de la casa, Comidas y bebidas, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, La escuela, El campo, Medios de transporte y Animales*) y en siete en castellano (*Muebles de la casa, Calefacción, iluminación y formas de airear un recinto, La ciudad, Trabajos del campo y del jardín, Juegos y distracciones, Profesiones y oficios y Colores*)⁷⁸². Las diferencias son distintas en todos los casos y en tres ocasiones —*El cuerpo humano, Muebles de la casa y Comidas y bebidas*— de tan solo un vocablo. Los informantes rurales actualizan en todos los centros más palabras distintas en catalán, salvo en *El cuerpo humano, La ciudad y Trabajos del campo y del jardín*, aunque en este último centro, la diferencia es realmente insignificante.

En general, comparando los resultados de los cuatro subgrupos en cada área temática, el alumnado rural consigue los mejores datos en catalán: en concreto, en seis centros de interés sus resultados descuellan frente a los de los restantes colectivos (*Partes de la casa, Comidas y bebidas, Calefacción, iluminación y formas de airear un recinto, El campo, Animales y Colores*). Los estudiantes procedentes de los centros ubicados en la capital consiguen las mejores cifras en cinco áreas en lengua catalana (*Ropa y complementos, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, La escuela y Medios de transporte*). Sin embargo, las diferencias son numéricamente muy distintas, con lo que es difícil valorar en conjunto la relevancia de estos datos. En cifras absolutas, las disimilitudes más llamativas se dan en *Animales y Comidas y bebidas*, donde los informantes rurales tienen 58 y 62 vocablos más en lengua catalana que en castellana, respectivamente; en *Objetos colocados en la mesa para la comida* es el alumnado urbano el que proporciona, en lengua vernácula, 65 palabras distintas más que los informantes rurales en lengua castellana. En global, las divergencias que se observan no son muy acusadas; impera —por encima de las diferencias entre cada uno de los colectivos— la misma tendencia ya constatada en los datos generales: que unos centros de interés propician más la evocación de términos que otros.

⁷⁸² Nótese que cinco de estos centros coinciden, como es lógico, con los que ostentan un mayor número de vocablos en castellano en los resultados generales (vid. 6.1.4.).

Con respecto a los rangos que alcanza cada área temática según el número de vocablos, también se repite aquí el patrón observado en los resultados globales. Solo se detectan algunas diferencias en las posiciones: por ejemplo, entre el alumnado urbano y en lengua castellana, *Trabajos del campo y del jardín* ocupa el noveno lugar, y no el sexto, como consta en la clasificación general; de modo similar, *Calefacción, iluminación y formas de airear un recinto* es decimocuarto, y no undécimo. También entre el alumnado urbano, pero en lengua catalana, *Trabajos del campo y del jardín* ocupa un lugar distinto al que tiene en el ranking general: es undécimo, cuando en la tabla clasificatoria de los datos globales en catalán es octavo. Las divergencias observadas en los restantes subgrupos son solamente de una o dos posiciones, con lo que no parecen cobrar importancia alguna.

Tabla 6.30. Índice de cohesión según la variable social zona geográfica.

C.I.	Índice de cohesión según la variable social 'zona geográfica'	Urbano		Rural	
		Castellano	Catalán	Castellano	Catalán
1	El cuerpo humano	0.09	0.10	0.08	0.11
2	Ropa y complementos	0.09	0.09	0.10	0.10
3	Partes de la casa (sin los muebles)	0.09	0.08	0.08	0.08
4	Muebles de la casa	0.06	0.06	0.06	0.07
5	Comidas y bebidas	0.06	0.07	0.07	0.07
6	Objetos colocados en la mesa para la comida	0.07	0.06	0.08	0.08
7	La cocina y sus utensilios	0.06	0.06	0.06	0.06
8	La escuela (muebles y materiales)	0.07	0.07	0.07	0.07
9	Calefacción, iluminación y formas de airear un recinto	0.05	0.06	0.06	0.05
10	La ciudad	0.05	0.05	0.04	0.05
11	El campo	0.04	0.04	0.04	0.04
12	Medios de transporte	0.09	0.08	0.09	0.08
13	Trabajos del campo y del jardín	0.03	0.04	0.03	0.04
14	Animales	0.09	0.08	0.10	0.08
15	Juegos y distracciones	0.04	0.04	0.04	0.04
16	Profesiones y oficios	0.04	0.05	0.05	0.05
17	Colores	0.10	0.11	0.11	0.10

Si se observa los índices de cohesión obtenidos por cada subgrupo social en cada área temática (tabla 6.30), es fácil concluir —como en las anteriores ocasiones— que las semejanzas entre el comportamiento lingüístico de los estudiantes en los dos idiomas son numerosas. Al respecto, recordemos que también en esta tabla todos los datos son comparables entre ellos, puesto que manejamos cuatro muestras idénticas en número y,

como ya sabemos, la oscilación en el tamaño de la muestra influye en los valores del índice de cohesión. Tal como se observa, las coincidencias son abundantes: en cuatro áreas temáticas —*La cocina y sus utensilios*, *La escuela*, *El campo* y *Juegos y distracciones*— todos los subgrupos obtienen exactamente el mismo índice. Asimismo, en la actuación lingüística de los informantes urbanos, siete áreas coinciden en el mismo valor, y en la de los rurales, nueve. Las similitudes no terminan aquí sino que, en dos áreas muy distintas según este parámetro, como son *Medios de transporte* y *Trabajos del campo y del jardín*, la cohesión de los dos colectivos —urbano y rural— es la misma en cada lengua analizada. De otra parte, en algunos centros de interés —como *Ropa y complementos* y *Objetos colocados en la mesa para la comida*— se constata una mayor convergencia asociativa entre el alumnado rural tanto en castellano como en catalán. Igual que sucede en los resultados generales, *El cuerpo humano* resulta claramente más cohesionado en lengua catalana y, en cambio, *Animales* es más compacto en castellano; las divergencias son mayores entre los informantes rurales. En las restantes categorías semánticas se observan diferencias escasas, normalmente de solo un punto entre las dos lenguas y entre los dos colectivos de informantes⁷⁸³. En general, se detectan los patrones ya comentados sobre los centros más compactos o difusos con algunas pequeñas diferencias, como por ejemplo, *Colores* —y no *El cuerpo humano*— es el área más cerrada del conjunto entre los datos de los informantes urbanos en lengua catalana.

6.2.4. Variable social *nivel sociocultural*

Según la cuarta variable social analizada, el *nivel sociocultural*, la muestra se divide en tres estratos: el nivel bajo (69 informantes, 28.75% del total), el nivel medio (135 informantes, 56.25%) y el nivel alto (36 informantes, 15%). Tal como se aprecia en la distribución, la mayoría de la muestra —poco más de la mitad— se engloba en el nivel medio, reflejando así la proporción real que tiene la clase media en el conjunto de la población (*cf.* 5.2.). De otra parte, tal como explicamos en el apartado metodológico, tuvimos en cuenta tres parámetros para determinar el nivel sociocultural del informante: la profesión y los ingresos de los progenitores, y el “nivel cultural de la familia”, concepto mediante el cual intentamos incorporar el bagaje cultural del propio

⁷⁸³ Téngase en cuenta, al evaluar todos los datos del presente capítulo referidos al índice de cohesión, que manejamos únicamente dos decimales y, por tanto, la cifra está redondeada. La única excepción la suponen los datos de la tabla 6.13, por las razones que allí damos.

informante y de su familia, y que supone una novedad metodológica en nuestra disciplina. Pues bien, a la vista de los resultados globales, visibles en la tabla 6.31 y en los gráficos 6.26 y 6.27, el estrato que alcanza el mejor resultado, tanto en lengua castellana como en catalana, es el nivel medio. De hecho, el promedio general más elevado corresponde al nivel medio en lengua catalana, que aporta 19.13 respuestas de promedio por centro de interés. Tras este resultado, la mejor cifra es 18.34, que obtiene tanto el estrato medio en lengua castellana como el estrato alto en lengua catalana. Los peores valores, por el contrario, son los del sociolecto bajo en castellano (16.99), seguidos de los conseguidos por este mismo colectivo en lengua catalana (17.84). Esto es, la distribución que se observa es idéntica en las dos lenguas: el nivel medio consigue el mejor promedio (18.34 en castellano y 19.13 en catalán), seguido del estrato alto (18.02 en castellano y 18.34 en catalán) y del bajo (16.99 frente a 17.84). Cabe destacar también que la distancia respectiva en cada lengua entre el peor y el mejor resultado —esto es, entre los niveles bajo y medio— es muy similar en los dos idiomas y, en ambos casos, sobrepasa la unidad (1.35 en castellano y 1.29 en catalán). Los resultados más cercanos entre sí son los que se dan en lengua española entre los niveles alto y medio, con tan solo 0.32 décimas de diferencia a favor de este último nivel. También en castellano se da la mayor distancia entre los resultados de cada grupo; como ya hemos dicho, el nivel bajo se separa del medio por 1.35 respuestas.

Tabla 6.31. Promedio de palabras por centro de interés según la variable social nivel sociocultural.

Nº	Centros de interés	Bajo		Medio		Alto	
		Castellano	Catalán	Castellano	Catalán	Castellano	Catalán
1	El cuerpo humano	23.78	27.26	25.92	29.31	27.47	30.17
2	Ropa y complementos	20.68	22.72	22.16	23.28	21.39	23.75
3	Partes de la casa (sin los muebles)	13.33	13.81	14.44	14.86	14.75	15.11
4	Muebles de la casa	11.64	12.06	12.84	13.44	12.19	11.81
5	Comidas y bebidas	23.30	25.78	24.96	27.20	23.17	25.17
6	Objetos colocados en la mesa para la comida	11.97	13.10	13.66	14.38	13.86	13.33
7	La cocina y sus utensilios	14.80	15.52	15.76	17.05	15.06	16.14
8	La escuela (muebles y materiales)	20.84	21.45	21.64	22.71	20.00	21.22
9	Calefacción, iluminación y formas de airear un recinto	10.42	10.55	11.24	11.26	11.78	11.22
10	La ciudad	20.13	20.67	21.87	21.71	21.72	20.64
11	El campo	16.45	17.25	19.35	19.24	17.28	17.42
12	Medios de transporte	15.49	15.81	17.38	17.61	17.44	17.61
13	Trabajos del campo y del jardín	8.65	9.48	8.65	9.82	10.00	9.78
14	Animales	25.26	25.41	26.47	27.05	25.78	25.33

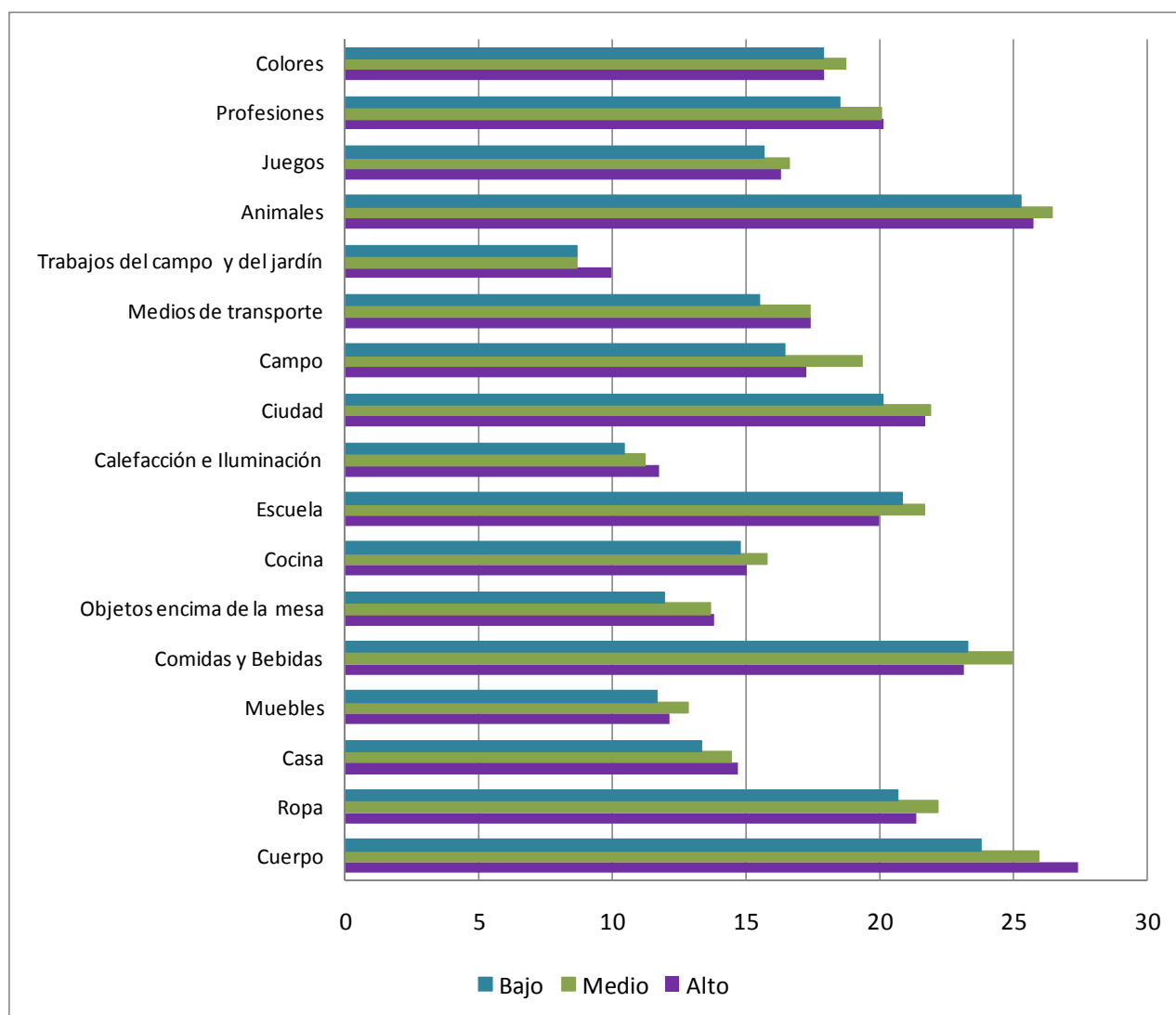
15	Juegos y distracciones	15.68	15.19	16.63	16.26	16.31	14.67
16	Profesiones y oficios	18.48	19.09	20.05	21.01	20.19	20.42
17	Colores	17.91	18.12	18.70	19.07	17.92	18.06
Global		16.99	17.84	18.34	19.13	18.02	18.34

Si reparamos en los datos que ofrece la tabla anterior, donde se aprecia la producción en cada una de las áreas temáticas, se observan distintas pautas en las dos lenguas analizadas. En castellano, el nivel medio obtiene los mejores resultados en diez categorías semánticas (*Ropa y complementos, Muebles de la casa, Comidas y bebidas, La cocina y sus utensilios, La escuela, La ciudad, El campo, Animales, Juegos y distracciones y Colores*), mientras que el estrato alto descuella en las siete restantes. En catalán, en cambio, el nivel medio muestra su preponderancia en todas las áreas de interés salvo en tres, *El cuerpo humano, Ropa y complementos y Partes de la casa*, en las que el nivel alto obtiene cifras más elevadas, y en una, *Medios de transporte*, donde coincide numéricamente con el resultado del estrato más alto. Hay que remarcar, asimismo, las coincidencias entre los dos idiomas, dado que en nueve áreas (*Muebles de la casa, Comidas y bebidas, La cocina y sus utensilios, La escuela, La ciudad, El campo, Animales, Juegos y distracciones y Colores*) el nivel medio obtiene los mejores resultados y en dos más (*El cuerpo humano y Partes de la casa*) es el nivel alto el que recaba los valores más altos, independientemente de la lengua en que se haya completado la encuesta.

Como ya hemos dicho, en las dos lenguas el estrato medio ostenta, generalmente, los mejores resultados y no el estrato alto, como acostumbra a suceder en la mayoría de estudios de disponibilidad. En lengua castellana, la mayoría de las veces en que esto ocurre, le sigue, a continuación, el estrato alto; salvo en dos ocasiones, en *Comidas y bebidas* y en *La escuela*, donde el nivel bajo supera al alto. Cabe destacar, asimismo, que en seis ocasiones —*El cuerpo humano, Partes de la casa, Objetos colocados en la mesa para la comida, Calefacción, iluminación y formas de airear un recinto, Medios de transporte y Profesiones y oficios*— se sigue el esquema clásico donde el mejor resultado lo ostenta el sociolecto alto, seguido del medio y, en tercer lugar, se sitúa el sociolecto bajo. Los datos relativos a la lengua catalana, difieren levemente: si bien en tres centros de interés, *El cuerpo humano, Ropa y complementos y Partes de la casa*, es posible reconocer el patrón lineal de estratificación, en los restantes centros, al igual que sucede en castellano, el nivel medio destaca frente a los

demás; excepto en un área, *Medios de transporte*, donde el resultado de los niveles medio y alto coincide. Al contrario de lo que sucede en lengua española —donde normalmente a este resultado le sigue el del nivel alto— en catalán, en siete ocasiones es el nivel bajo el que sigue al nivel medio (*Muebles de la casa, Comidas y bebidas, La escuela, La ciudad, Animales, Juegos y distracciones y Colores*), y en seis es el nivel alto el que sigue al medio (*Objetos colocados en la mesa para la comida, La cocina y sus utensilios, Calefacción, iluminación y formas de airear un recinto, El campo, Trabajos del campo y del jardín y Profesiones y oficios*). Así pues, es pertinente concluir que, en las dos lenguas, sobresale el nivel medio entre los diferentes niveles socioculturales pero, mientras que en castellano el esquema más frecuente es ‘nivel medio, alto y bajo’; en catalán, se alternan las combinaciones ‘medio, alto y bajo’ y ‘medio, bajo y alto’. También, como hemos indicado, hay veces en las que se sigue el patrón lineal de estratificación; esto sucede en más ocasiones en castellano que en catalán (seis frente a tres).

Gráfico 6.26. Promedio de palabras por centro de interés según la variable social nivel sociocultural en lengua castellana.



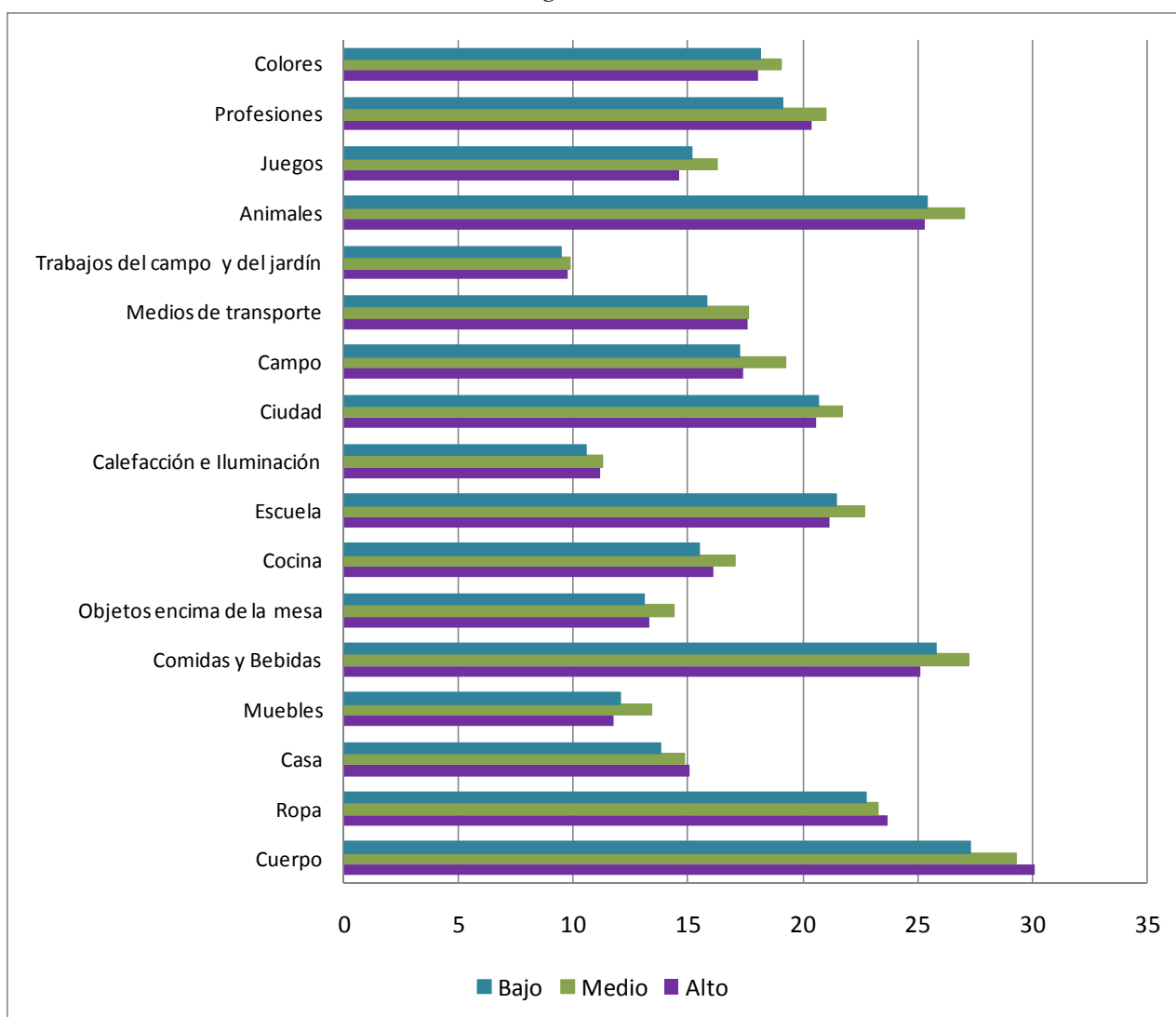
Pasemos ahora a analizar los resultados comparando los tres niveles sociales entre sí en lengua castellana. Como ya hemos dicho, el estrato medio predomina en la mayor parte de áreas temáticas, tal como se observa en el gráfico 6.26. Ahora bien, si contrastamos el comportamiento lingüístico de este colectivo con el del nivel bajo, la preponderancia del estrato medio es aún más evidente, toda vez que sobrepasa en todas las categorías semánticas al nivel bajo, menos en *Trabajos del campo y del jardín*, donde obtienen exactamente el mismo resultado (8.65) los dos grupos de informantes. La diferencia es constante a favor del nivel medio: en dos ocasiones, en *El cuerpo humano* y en *El campo*, se sobrepasan las dos respuestas de media —de hecho en el centro 11 casi se alcanzan las tres palabras (2.9)— y en nueve áreas (*Ropa y complementos*, *Partes de la casa*, *Muebles de la casa*, *Comidas y bebidas*, *Objetos*

colocados en la mesa para la comida, La ciudad, Medios de transporte, Animales y Profesiones y oficios) se supera la unidad. La desigualdad entre los promedios generales de estos dos estratos (1.35 palabras) es la mayor, como ya hemos indicado, de las que ejerce esta variable social en las dos lenguas y también la mayor de las provocadas por alguna de las cinco variables sociales sobre la producción en el léxico disponible en lengua castellana (*vid.* 6.2.6, gráfico 6.41).

Si comparamos los resultados de los niveles alto y medio, lo más relevante es, en contraste con el análisis anterior de los estratos bajo y medio, la poca distancia que media entre el resultado de los dos grupos (0.32 décimas). El comportamiento de estos dos colectivos en relación con su producción en lengua castellana es similar: en siete campos alcanza mejores cifras el nivel alto (*El cuerpo humano, Partes de la casa, Objetos colocados en la mesa para la comida, Calefacción, iluminación y formas de airear un recinto, Medios de transporte, Trabajos del campo y del jardín y Profesiones y oficios*), mientras que el medio impera en los diez restantes. La diferencia supera las dos palabras en tan solo un área temática (*El campo*), a favor del nivel medio, y la unidad, en cuatro categorías (*El cuerpo humano, Comidas y bebidas, La escuela y Trabajos del campo y del jardín*), a favor de uno u otro estrato. Los resultados son, varias veces, prácticamente idénticos (por ejemplo, en *Objetos colocados en la mesa para la comida, La ciudad, Medios de transporte, Profesiones y oficios...*) y no parece apreciarse, como hemos comentado, una diferencia destacable entre la producción en castellano de estos dos grupos de informantes.

Los estratos alto y bajo se separan por una palabra en el promedio general (1.03). En el comportamiento lingüístico de estos dos grupos se aprecia cómo el nivel alto supera siempre al bajo, salvo en dos centros de interés (*Comidas y bebidas y La escuela*) donde los resultados del nivel bajo son, sorprendentemente, mejores que los del nivel alto. La diferencia mayor de todas se da en el primero de los centros de interés, *El cuerpo humano*, con 3.69 palabras de más que actualiza el nivel alto frente al bajo. También en siete áreas se supera la unidad a favor del estrato alto (*Partes de la casa, Objetos colocados en la mesa para la comida, Calefacción, Iluminación y formas de airear un recinto, La ciudad, Medios de transporte, Trabajos del campo y del jardín y Profesiones y oficios*). Cabe mencionar que en las dos únicas ocasiones en que el estrato bajo supera al alto la desigualdad nunca alcanza una palabra.

Gráfico 6.27. Promedio de palabras por centro de interés según la variable social nivel sociocultural en lengua catalana.



En el gráfico 6.27, donde se plasman los resultados relativos a la lengua catalana para cada uno de los niveles socioculturales y en las diecisiete áreas analizadas, se advierte cómo el nivel medio obtiene, mayoritariamente, los mejores resultados, con una constancia superior a la observada en los datos sobre la lengua castellana. Comparando el nivel medio con el nivel bajo, se constata que el estrato medio destaca en todas las categorías semánticas, con una diferencia global (1.29) similar a la que había entre los mismos estratos en lengua castellana (1.35). La mayor disimilitud se aprecia en el campo 1, *El cuerpo humano*, donde se sobrepasan las dos unidades. Salvo en cuatro áreas (*Ropa y complementos*, *Calefacción, iluminación y formas de airear un recinto*, *Trabajos del campo y del jardín* y *Colores*), en todas las restantes, la diferencia entre los promedios de los dos colectivos siempre supera la unidad. En el cotejo entre el nivel medio y el alto se observa cómo, en general, sigue predominando el estrato medio, salvo

en las tres primeras áreas donde el alto obtiene mejores medias (*El cuerpo humano, Ropa y complementos y Partes de la casa*) y en *Medios de transporte*, donde, como ya hemos dicho, se obtiene un resultado idéntico. Las disimilitudes sobrepasan en una ocasión las dos respuestas (en *Comidas y bebidas*) y en ocho áreas la unidad (*Muebles de la casa, Objetos colocados en la mesa para la comida, La escuela, La ciudad, El campo, Animales, Juegos y distracciones y Colores*), siempre a favor del estrato medio. Vale la pena observar que la distancia entre los promedios globales de los dos estratos (0.79) es inferior a la observada entre los niveles bajo y medio (1.29).

Finalmente, si contrastamos los datos de los estratos bajo y alto, los resultados divergen en función del centro de interés que analicemos. Generalmente, el nivel alto obtiene cifras superiores, concretamente en diez centros de interés (*El cuerpo humano, Ropa y complementos, Partes de la casa, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, Calefacción, iluminación y formas de airear un recinto, El campo, Medios de transporte, Trabajos del campo y del jardín y Profesiones y oficios*), mientras que en los siete restantes el nivel bajo obtiene un resultado más favorable, lo cual no deja de ser llamativo. La distancia más acusada se da en el primero de los centros de interés, *El cuerpo humano*, con casi 3 respuestas (2.91) de más actualizadas por el estrato alto frente al bajo. En el resto de categorías semánticas, solo se supera cuatro veces la unidad en el cotejo de los dos promedios (*Ropa y complementos, Partes de la casa, Medios de transporte y Profesiones y oficios*), a favor siempre del estrato alto. La diferencia que existe entre los dos sociolectos en la media global es de media palabra, por tanto, es la más baja de las que se observan entre los distintos niveles sociales en la producción léxica en lengua catalana.

En conjunto, se observa que, tanto en catalán como en castellano, el mayor contraste se da entre los estratos medio y bajo. Este hecho se confirma en el desglose por centros de interés donde el predominio del estrato medio frente al bajo es constante en las dos lenguas (únicamente empatan en el centro 13 en castellano, mientras que en las restantes áreas, en las dos lenguas, el resultado es superior en el nivel medio). En la relación entre el nivel medio y el alto se advierte cómo en catalán la actuación lingüística difiere más que en castellano (0.79 frente a 0.32 en el promedio global). También en la lengua vernácula de la provincia el nivel medio, frente al alto, obtiene mejores resultados en la mayoría de áreas (salvo en tres), mientras que, en castellano, el

comportamiento entre estos dos grupos de informantes es más homogéneo a pesar de que predomina el estrato medio en diez categorías semánticas. En cuanto al cotejo de los niveles alto y bajo, en castellano se observa una mayor distancia en la media global (1.03 frente a 0.5); en esta lengua el sociolecto alto supera casi siempre al bajo, excepto en dos centros de interés. En catalán, donde la diferencia globalmente es menor, se constata que el comportamiento diverge en función del área analizada y el estrato alto consigue mejores cifras en diez ocasiones. En general, es fácil colegir, llegados a este punto del análisis, que —en las dos lenguas analizadas— la diferencia más firme y constante se da entre los estratos bajo y medio, mientras que el estrato alto se comporta de forma singular, consiguiendo, generalmente, peores resultados que el nivel medio pero mejores que los del nivel bajo.

El siguiente paso es comprobar si la superioridad léxica constatada, en las dos lenguas, del sociolecto medio frente a los otros dos estratos sociales tiene significación estadística. Tras realizar la comparación entre medias con el test ANOVA, se obtiene una significación de ‘p’ inferior a 0.05. Concretamente, en castellano ‘p’= 0.000 y en catalán ‘p’= 0.018, lo que nos hace refutar la hipótesis nula y aceptar que las divergencias entre las medias son ocasionadas por el factor *nivel sociocultural*. Dado que contamos con tres niveles distintos de estratificación social, hemos llevado a cabo una prueba de contraste *post-hoc*, la prueba de Tukey, para reconocer entre qué estratos es significativa la diferencia hallada. Tal como se observa en los resultados que proporciona esta prueba (véase el tercer apéndice), los niveles implicados en la significación con valor estadístico son el medio y el bajo.

Seguidamente, presentamos la tabla con los valores de ‘p’ referentes a las diferencias entre los promedios de palabras producidos por los sujetos encuestados en cada centro de interés.

Tabla 6.32. Análisis estadístico de la diferencia según el nivel sociocultural.
Valor de la 'p' en las dos lenguas.

Nº	Centros de interés	Castellano	Catalán
1	El cuerpo humano	0.004	0.012
2	Ropa y complementos	0.107	0.579
3	Partes de la casa (sin los muebles)	0.102	0.132
4	Muebles de la casa	0.075	0.010
5	Comidas y bebidas	0.048	0.104
6	Objetos colocados en la mesa para la comida	0.008	0.033
7	La cocina y sus utensilios	0.375	0.118
8	La escuela (muebles y materiales)	0.110	0.118
9	Calefacción, iluminación y formas de airear un recinto	0.132	0.373
10	La ciudad	0.061	0.349
11	El campo	0.002	0.033
12	Medios de transporte	0.002	0.016
13	Trabajos del campo y del jardín	0.083	0.795
14	Animales	0.257	0.086
15	Juegos y distracciones	0.376	0.118
16	Profesiones y oficios	0.057	0.034
17	Colores	0.291	0.182

Después de observar estos resultados, concluimos lo siguiente:

- En castellano, la diferencia entre los promedios resulta significativa en cinco centros de interés: *El cuerpo humano*, *Comidas y bebidas*, *Objetos colocados en la mesa para la comida*, *El campo* y *Medios de transporte*. Como vimos, comparando los tres estratos, el nivel medio es el que obtiene mejores resultados en dos de estas áreas (*Comidas y bebidas* y *El campo*), mientras que en las tres restantes (*El cuerpo humano*, *Objetos colocados en la mesa para la comida* y *Medios de transporte*) es el nivel alto el que lidera los promedios.
- En catalán, la probabilidad es inferior a 0.05 en seis centros de interés: *El cuerpo humano*, *Muebles de la casa*, *Objetos colocados en la mesa para la comida*, *El campo*, *Medios de transporte* y *Profesiones y oficios*. En los datos, el sociolecto medio obtiene resultados más favorables en todas estas áreas, salvo en la primera, *El cuerpo humano*, donde el nivel alto tiene un promedio superior y en *Medios de transporte*, donde los estratos medio y alto empatan.

Es digna de mención la coincidencia existente, entre las dos lenguas, en la significación estadística de cuatro áreas: *El cuerpo humano*, *Objetos colocados en la*

mesa para la comida, *El campo* y *Medios de transporte*⁷⁸⁴. Teniendo en cuenta que, en lengua castellana, solo en cinco áreas se detecta una diferencia significativa y, en catalán, en seis, parece mucha casualidad que cuatro áreas sean las mismas en los dos idiomas. Sin embargo, es difícil hallar una explicación que justifique globalmente este comportamiento, más allá de la superioridad léxica —tantas veces constatada en el mundo de la disponibilidad— de los sujetos pertenecientes a los estratos sociales más altos. En este caso, las razones que justifican las divergencias en estas cuatro áreas semánticas difieren, en nuestra opinión, según el área analizada. Si bien el léxico evocado en algunos de estos centros, como en *El cuerpo humano*, es —aparentemente— neutro; estudios realizados con anterioridad, ya han llamado la atención sobre cómo algunas voces de esta categoría semántica, principalmente tecnicismos, cultismos, así como léxico específico, son usadas solo por los informantes del sociolecto más alto (López Morales 1979: 180⁷⁸⁵). Algo similar podría suceder en *Medios de transporte*, dado que el léxico de ambos centros se trabaja en la escuela y es de suponer que el vocabulario de unos progenitores formados académicamente influya en el que posean —o sean capaces de evocar en un ejercicio de asociación controlada como es la disponibilidad— sus hijos. No obstante, las diferencias estadísticamente significativas también se extienden a otras áreas, como *Objetos colocados en la mesa para la comida*, categoría estrechamente relacionada con el hogar y donde el léxico recabado raramente contiene tecnicismos o cultismos, o *El campo*, un área con multitud de asociaciones posibles. Llegados a este punto, sin duda alguna, un análisis cualitativo que ahondara en las diferencias existentes entre los estratos medio y bajo en estas cuatro áreas temáticas nos ayudaría a encontrar una justificación a las disimilitudes cuantitativas halladas, empresa que esperamos abordar en un futuro.

Como ya hemos indicado, las investigaciones realizadas hasta la fecha (Samper Padilla y Hernández Cabrera 1997, Mateo 1998, Carcedo 2001, González Martínez 2002, Galloso 2003a, Bartol 2004, Gómez Molina y Gómez Devís 2004, Ayora 2006, etc.) coinciden en señalar, de forma casi unánime, la gran importancia del factor *nivel*

⁷⁸⁴ El centro *Profesiones y oficios*, que resulta significativo en lengua catalana, obtiene un valor en castellano cercano al de la significación estadística.

⁷⁸⁵ El académico cubano pone como ejemplos “tórax”, “parietal”, “ovarios”, “occipital”, etc.

sociocultural en el caudal de léxico disponible⁷⁸⁶. Refiriéndose a este factor social, Samper Padilla (2006: 117) afirma: “[e]s sumamente raro encontrar una comunidad donde no sea perceptible su incidencia; por otro lado, en muchas ocasiones constituye el factor que establece la discriminación social más acusada”⁷⁸⁷. En general, la mejor producción léxica se relaciona con los estratos altos de la sociedad, definidos habitualmente en función del nivel de instrucción y de la profesión de los progenitores (*vid.* 5.2.) y pone de relieve, una vez más, que la escuela no ejerce la influencia necesaria para paliar las diferencias entre los escolares de una u otra clase social ya que, a las puertas de entrar a la universidad, se continúan registrando diferencias notables en función del nivel social familiar.

En nuestro trabajo, se confirma también la importancia de esta variable social, aunque solo parcialmente, puesto que la diferencia global entre los promedios de cada nivel social solo es significativa entre los estratos medio y bajo. Como hemos visto, el contraste entre estos dos niveles es constante tanto en el desglose por áreas temáticas como en los promedios generales de las dos lenguas. Los procedimientos estadísticos aplicados confirman, asimismo, la diferencia entre los dos sociolectos. Sin embargo, ya en el análisis se intuía el comportamiento, cuanto menos peculiar, de los informantes del nivel más alto del espectro social, situado a caballo entre el estrato medio y el bajo en ambas lenguas. En castellano, recordemos que el nivel alto ostenta los mejores resultados en siete áreas y se mantiene muy próximo al estrato medio en el promedio general. En catalán, en cambio, solo lidera los resultados en tres ocasiones y su índice global difiere más con el del estrato medio, como hemos visto en el análisis puramente descriptivo.

Es importante recordar ahora que, como explicamos en el apartado metodológico con detalle (*vid.* 5.2.), hemos utilizado para la estratificación del nivel sociocultural un nuevo índice, creado por nosotros, que contaba con un parámetro más —el “nivel

⁷⁸⁶ Los trabajos en los que esta variable no cobra relevancia constituyen realmente excepciones dentro de la tendencia general (*v.gr.* Llopis 2009 o Prado *et al.* 2009, con respecto a la muestra en castellano), como ya señalamos (*vid.* 2.2.1).

⁷⁸⁷ López Morales fue el primero en llamar la atención sobre las diferencias léxicas que se producían, tanto cuantitativas como cualitativas, entre los distintos niveles socioculturales y relacionó —aunque con cautela— sus datos con la *teoría del déficit* de Bernstein. La hipótesis que en ese momento barajaba el autor, “la existencia de desniveles de disponibilidad léxica, estrictamente ligada a factores de índole socioeconómico” (López Morales 1979: 181) se ha visto, por ende, sobradamente corroborada en los trabajos realizados con posterioridad.

cultural”— además de los dos habitualmente empleados para determinar el nivel sociocultural del alumno, —esto es, los estudios y la profesión de los progenitores—. Después de ver los resultados, podemos afirmar, en principio, que la fórmula con los tres indicadores sociales que utilizamos por primera vez en este trabajo, resulta adecuada para establecer diferencias sociolectales, toda vez que la superioridad del nivel medio frente al bajo tiene significación estadística, lo cual indica, una vez más, la relevancia del condicionante sociocultural en el caudal de léxico disponible. No obstante, el comportamiento atípico del estrato alto nos hace reconsiderar la valoración del índice que hemos creado y cuya utilidad se ha demostrado en la oposición entre los niveles medio y bajo. En el capítulo anterior, comparamos la distribución de la muestra teniendo en cuenta únicamente los dos parámetros habituales o considerando los tres valores empleados en este análisis. Si bien en aquella ocasión (*vid.* 5.2), concluimos que la distribución de la muestra era más adecuada con nuestra clasificación, ahora no tenemos la certeza de que sea así.

Si hubiéramos utilizado el cálculo del nivel social “estándar” (establecido por Samper Padilla y Hernández Cabrera 1997: 230), el estrato bajo difícilmente modificaría su resultado (puesto que únicamente contaría con 4 informantes más y pasaría de 69 a 73); al nivel medio, por el contrario, se añadirían 17 sujetos, que sobre un total de 152, tampoco es probable que modificasen en gran medida los resultados actuales. El nivel alto es el que más variaría, toda vez que perdería más de la mitad de sus integrantes, pasando de 36 a 15, tal como vimos en el capítulo quinto, donde comparamos las dos clasificaciones (*vid.* tabla 5.5). Del comportamiento lingüístico de esos 15 individuos dependería que los resultados del nivel alto se revelaran como significativos con respecto a los de los otros niveles o no y, por tanto, que la distribución fuera mejor o peor considerando dos parámetros —como es habitual— o tres parámetros —como hemos hecho en nuestro estudio—. En cualquier caso, tampoco podemos tener la seguridad de que el nivel alto se comportara de forma distinta a la que hemos descrito en el presente análisis —ni que la actuación lingüística de los restantes estratos no variara— pero, como mínimo, existe esa posibilidad.

La nuestra no es la única investigación donde se observa una conducta peculiar por parte del sociolecto más alto de la escala. En trabajos recientes de nuestro ámbito, se ha dejado constancia de la existencia de patrones muy similares de estratificación social,

que estarían en consonancia con el mayor conservadurismo lingüístico de las clases medias detectado por Labov (1983) en los primeros trabajos de sociolingüística variacionista⁷⁸⁸. Por ejemplo, Martínez Olmos (2008: 204-205) en su investigación sobre el léxico disponible de Alicante observa cómo el nivel medio-medio supera al medio-alto en el promedio global (19.52 frente a 19.19) y sobrepasa al medio-bajo en 1.16 respuestas. Concretamente, el sociolecto medio-medio supera al medio-alto en diez centros de interés, mientras que en los seis restantes es el estrato medio-alto el que consigue el mejor resultado de los tres niveles. En el estudio del léxico disponible de la ciudad de Melilla, Fernández Smith *et al.* (2008) obtienen también resultados semejantes a los nuestros. Tras distribuir la muestra en cuatro niveles socioculturales, los sociolectos se ordenan del siguiente modo según su promedio global: lidera la clasificación el nivel medio-alto, seguido del alto, del medio-bajo y, en último lugar, del bajo. El comportamiento del estrato alto es tan sorprendente que sus informantes presentan promedios inferiores a los niveles medio-bajo en ocho centros de interés y bajo, en dos. Además, los autores comprueban, mediante el análisis ANOVA de los promedios, cómo el nivel alto, que catalogan como “atípico”, no presenta relaciones significativas con ningún otro nivel, mientras que estas sí se dan entre otros niveles (el nivel bajo en comparación con los niveles medio-bajo y medio-alto).

Vistos, pues, los trabajos reseñados, cabe argüir distintas explicaciones a este comportamiento del estrato alto. En primer lugar, en las distribuciones que se realizan sobre los sociolectos, el nivel alto suele ser el que está integrado por menos informantes. En Martínez Olmos (2008) son 35 los encuestados del nivel más elevado, el 7.60% de la muestra, en Fernández Smith *et al.* (2008) solo 15, el 4.76%, y en nuestra investigación 36, el 15% del conjunto poblacional, como ya hemos indicado. Esta cantidad —que suele ser aún menor cuando se divide esta variable en cuatro niveles como hacen Fernández Smith *et al.* (2008)— puede simplemente resultar inferior al número de sujetos necesarios para plasmar el comportamiento léxico de la clase social alta y, por tanto, no ser representativa del universo estudiado⁷⁸⁹.

⁷⁸⁸ El patrón curvilíneo constatado por este autor demuestra que las clases medias son las que más se aproximan a la norma o a lo considerado socialmente correcto. Esto podría traducirse en un interés mayor por ampliar el vocabulario disponible de sus hijos, y utilizar el dominio de la lengua como forma de promoción social.

⁷⁸⁹ Los mismos autores del léxico disponible de Melilla reconocen que habría que unir los dos grupos de nivel alto (el alto y el medio alto) en un solo estrato “por el bajo porcentaje de informantes que presenta el nivel superior y porque puede tratarse de una diferenciación poco relevante desde el punto de vista de la

Más allá de que el número de informantes sea representativo o no, nos parece interesante la justificación que da Martínez Olmos (2008: 424) a sus resultados, muy similares a los nuestros. Según esta autora, algunos planteamientos sociológicos actuales afirman que “se están borrando los límites de la clase social media-media y media-alta”. Dado que ya son tres los estudios realizados donde el estrato alto muestra un comportamiento singular, creemos que sería necesario profundizar en las diferencias entre los estratos sociales, a la luz de los conocimientos sociológicos actuales, puesto que todo parece indicar que el comportamiento atípico del sociolecto alto en relación con los otros niveles no es un hecho aislado. Las investigaciones que se realicen en el futuro corroborarán, o no, esta tendencia del estrato alto a desdibujarse en relación con el medio, o demostrarán la superioridad léxica del nivel medio frente a los otros niveles, en la línea de las primeras investigaciones sociolingüísticas (Labov 1983).

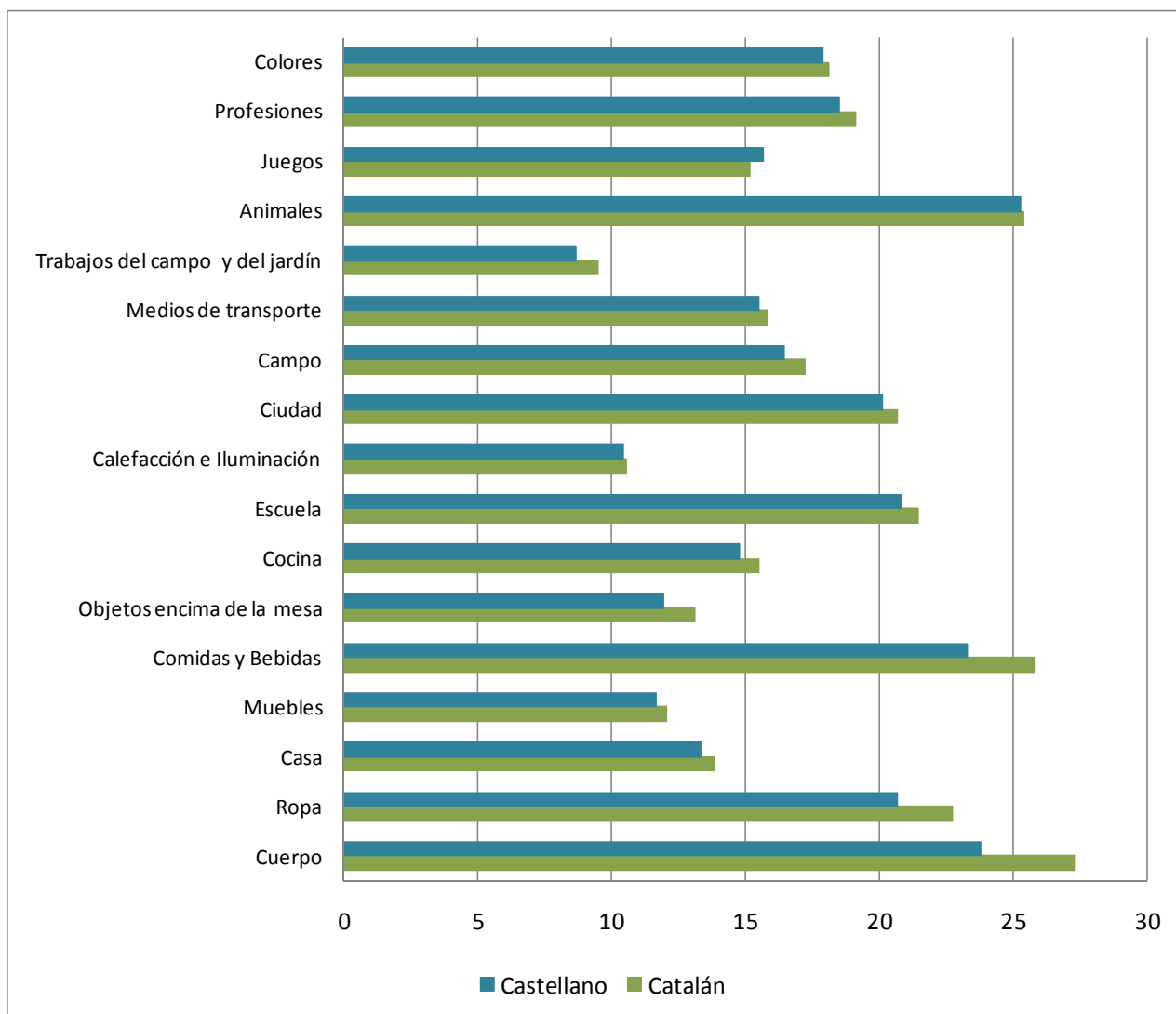
No obstante, también sería muy importante, para la distribución de la muestra en distintos niveles sociales, tener en cuenta las características concretas de la población estudiada, tal como se colige de las conclusiones del trabajo de Hernández Muñoz (2006). En este estudio, donde se divide la población analizada en cuatro niveles sociales (alto, medio-alto, medio-bajo y bajo) se observa, asimismo, un comportamiento atípico del estrato alto, que obtiene los mejores promedios tras el nivel medio-bajo y el medio-alto, que encabeza el ranquin. En un principio, ninguna de las diferencias entre los cuatro estratos sociales es significativa, pero la autora, tras realizar distintas pruebas agrupando a los informantes en dos o tres niveles, los clasifica de nuevo, según las peculiaridades propias de la muestra concreta que maneja, y llega a una distribución donde la clasificación en dos y en tres niveles sociales sí tiene significación estadística. La misma investigadora, por tanto, concluye que los resultados que se deducen de este factor social deben presentarse con cautela ya que “dependiendo de las peculiaridades del conjunto poblacional estudiado, las diferentes clasificaciones socioculturales serán más o menos relevantes a la hora de comparar la producción de los sociolectos” (Hernández Muñoz 2006: 345). Como ella misma demuestra, si se redistribuyen los

extracción sociocultural de estos informantes” (Fernández Smith *et al.* 2008: 78). También Martínez Olmos (2008: 424) cree que una posible solución sería considerar el grupo medio-alto y el medio-medio como uno solo en futuras investigaciones. Asimismo, en nuestro anterior trabajo exploratorio sobre el léxico disponible castellano en la ciudad de Lleida (Serrano Zapata 2003b, 2004), constatamos cómo los representantes del estrato alto, que se reducían a 5 miembros, arrojaban unos resultados sorprendentes, por debajo del nivel bajo en la mayoría de las ocasiones. Sin duda alguna, como indicamos entonces, el reducido tamaño de la muestra influía en el resultado.

informantes según las características del grupo poblacional estudiado, las diferencias que se produzcan pueden llegar a ser significativas.

Pasemos ahora a describir las disimilitudes observadas entre los sociolectos en función de la lengua en que se completa la encuesta. Los encuestados que pertenecen al nivel social bajo presentan unos resultados que otorgan un predominio claro del catalán, con una diferencia en el promedio global de casi una palabra (0.85). En el análisis por áreas (véase gráfico 6.28) se puede comprobar cómo la diferencia a favor de la lengua catalana es constante en todas las categorías semánticas salvo en una, *Juegos y distracciones*, donde la media es favorable al castellano en casi media respuesta más (0.49), siguiendo así la tendencia general ya descrita con anterioridad (*vid.* 6.1.2.). Como también se aprecia en los resultados globales, las divergencias mayores se dan en el primer centro (*El cuerpo humano*, con una diferencia de 3.48 palabras), en el quinto (*Comidas y bebidas*, con 2.48) y en el segundo (*Ropa y complementos*, con 2.04). Salvo en la sexta área, *Objetos colocados en la mesa para la comida*, donde en catalán se aportan 1.13 respuestas más, en las restantes nunca se supera la unidad entre el léxico disponible en cada uno de los dos idiomas.

Gráfico 6.28. Comparación de la producción en castellano y en catalán de los informantes del nivel bajo.

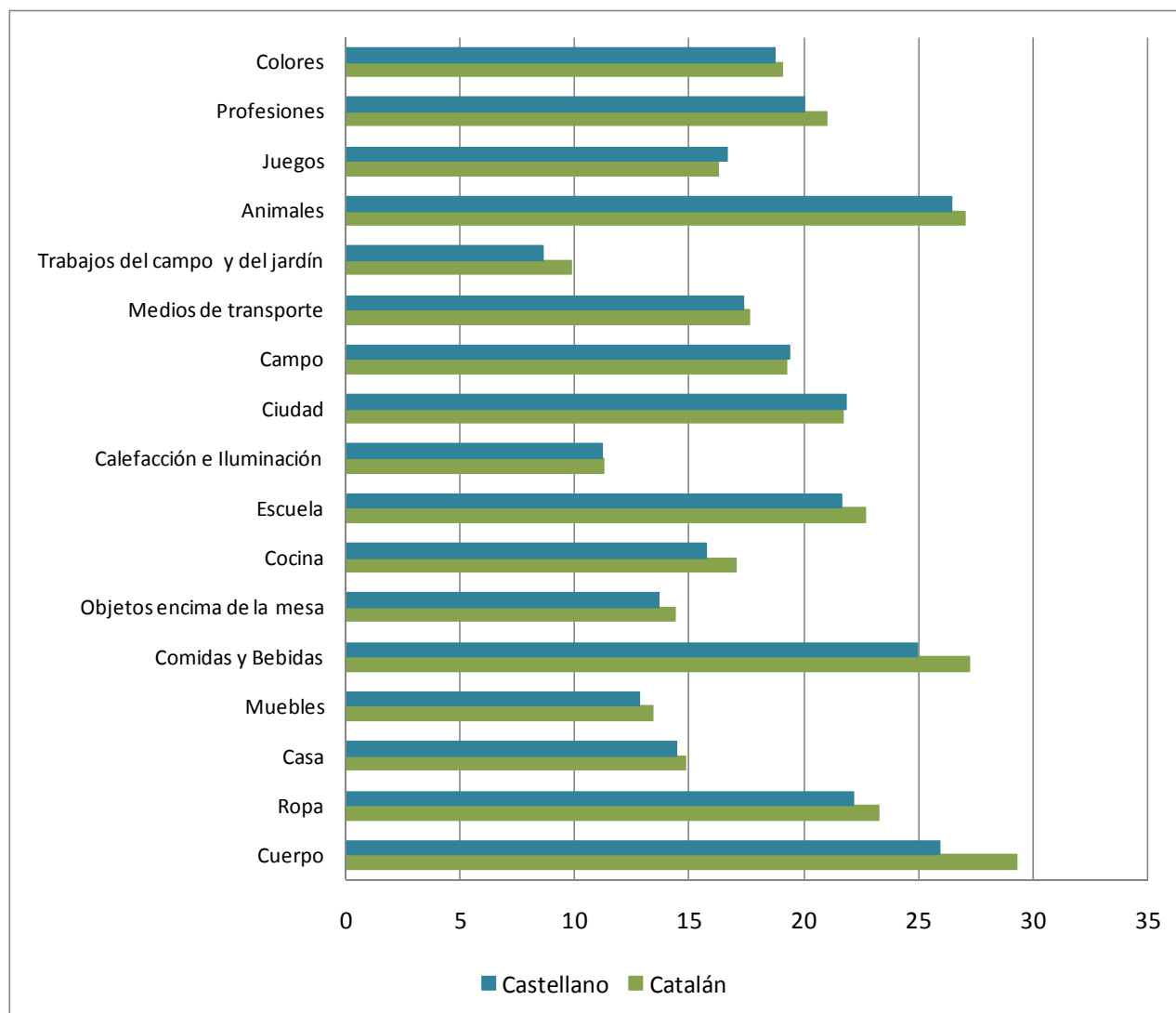


En la disponibilidad del estrato medio (véase gráfico 6.29), también se detecta el predominio del catalán, aunque con una diferencia global (0.79) un tanto inferior a la observada en el sociolecto bajo (0.85). En el análisis por áreas temáticas, los promedios en castellano son mejores en tres categorías, *La ciudad*, *El campo* y *Juegos y distracciones*, aunque cabe destacar que, en las dos primeras áreas, la diferencia es de poco más de una décima (0.16 y 0.11, respectivamente), con lo que este resultado parece irrelevante. En la mayoría de las áreas, por ende, en catalán se recaban mejores promedios. Las divergencias son mayores, siguiendo la tendencia general, en *El cuerpo humano*, con más de tres palabras de diferencia (3.39) y en *Comidas y bebidas*, con más de dos (2.24), al igual que sucede en el estrato bajo⁷⁹⁰. Solo en cuatro áreas más la

⁷⁹⁰ Sin embargo, la diferencia de más de dos palabras en *Ropa y complementos* que se da en el estrato bajo y que, como veremos a continuación, también se da en el alto, aquí es de la mitad (1.12).

diferencia sobrepasa la unidad (*Ropa y complementos, La cocina y sus utensilios, La escuela y Trabajos del campo y del jardín*). Es digno de mención que en alguna categoría, como *Calefacción, iluminación y formas de airear un recinto*, los resultados de las dos lenguas son prácticamente idénticos⁷⁹¹.

Gráfico 6.29. Comparación de la producción en castellano y en catalán de los informantes del nivel medio.

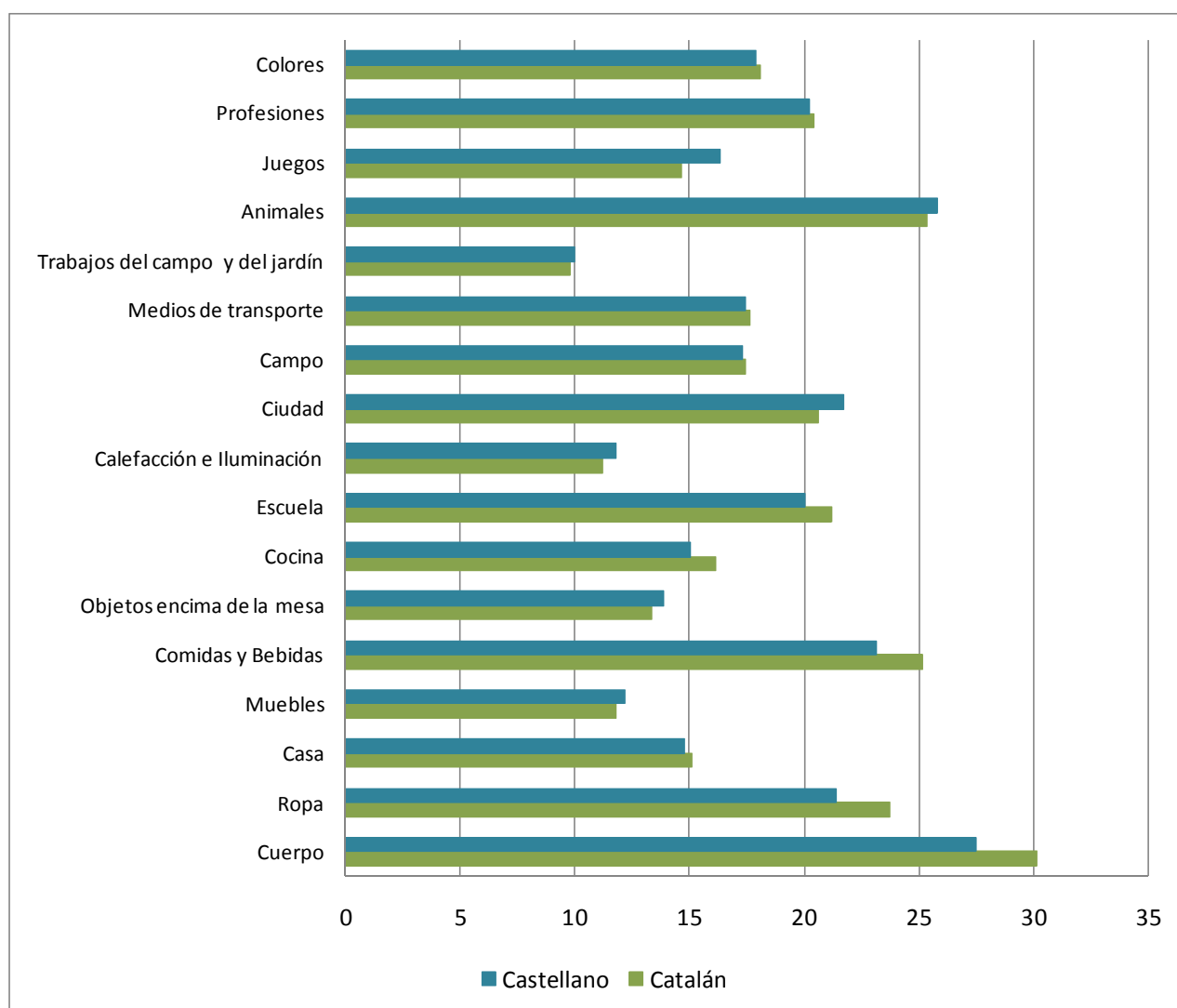


Finalmente, en los resultados del sociolecto alto (véase gráfico 6.30) se observa cómo, en la línea del análisis realizado hasta este momento, su comportamiento difiere del de los otros dos estratos y, en primer lugar, destaca la cercanía de los resultados obtenidos en las dos lenguas, únicamente diferenciadas por 0.32 respuestas. En el

⁷⁹¹ Este centro de interés es también el que presenta la menor diferencia entre las dos lenguas (0.13) en los informantes del sociolecto bajo.

análisis por centros de interés, también se detectan unos resultados que difieren de los vistos hasta ahora: en diez centros el catalán consigue mejores datos (*El cuerpo humano, La ropa, Partes de la casa, Comidas y bebidas, La cocina y sus utensilios, La escuela, El campo, Medios de transporte, Profesiones y oficios y Los colores*), pero en los siete restantes ostenta cifras más reducidas en comparación con el castellano. Entre estos campos se encuentran algunos de los aparecidos hasta el momento como favorables al castellano en los otros sociolectos —como *Juegos y distracciones y La ciudad*—, otros donde en los restantes niveles se detecta un resultado muy similar entre las dos lenguas —así sucede en *Calefacción, iluminación y formas de airear un recinto*—, pero también muchos otros como *Muebles de la casa, Objetos colocados en la mesa para la comida, Trabajos del campo y del jardín y Animales*. Al igual que en los estratos anteriormente analizados, son mayores las disimilitudes halladas en *El cuerpo humano* (2.7 respuestas de diferencia), *Ropa y complementos* (2.36) y en *Comidas y bebidas* (2), aunque en este sociolecto —el alto—, nunca se superan las tres respuestas de diferencia, como sucede en los análisis anteriores. En cuatro ocasiones más se supera la unidad: en dos de ellas a favor del castellano, en *La ciudad* y en *Juegos y distracciones*, y en dos más a favor del catalán, en *La cocina y sus utensilios* y en *La escuela*.

Gráfico 6.30. Comparación de la producción en castellano y en catalán de los informantes del nivel alto.



En resumen, en los tres sociolectos se observan, en mayor o menor medida, las tendencias generales que detallamos con anterioridad (*vid.* 6.1.2). Esto es, por un lado, la tendencia mayor al uso del castellano en algunas áreas relacionadas con el entorno urbano, —*La ciudad*— o el ocio juvenil —*Juegos y distracciones*— y, por el otro, la distancia tan acusada que existe entre la producción en castellano y en catalán en *El cuerpo humano*, *Ropa y complementos* y *Comidas y bebidas*. Cabe añadir, no obstante, que, de los tres estratos, es el nivel bajo el que se decanta con mayor claridad hacia la lengua catalana, tanto en el análisis por áreas como en el promedio general. El nivel medio sigue muy de cerca esta tendencia; por el contrario, el estrato alto muestra los resultados más similares entre las dos lenguas, con siete áreas donde la producción es superior en castellano y con un promedio global muy parecido en los dos idiomas.

Estos resultados son llamativos en cuanto apuntan a un mayor uso del catalán entre la clase baja y un uso más igualado entre la clase social alta, lo que, en principio, no parece tener una correspondencia con la realidad social de Lleida. Como ya hemos comentado en diversas ocasiones a lo largo de este trabajo, el castellano se asocia en Cataluña a una parte importante de la clase popular, toda vez que muchos de los que tienen esta lengua como materna son descendientes de la inmigración de los años 60, proveniente principalmente de Andalucía —y en Lleida también de Aragón— que se desplazó a Cataluña con la intención de buscar un puesto de trabajo. No obstante, como también dijimos (*vid.* 4.3.2.), en el sector terciario hay una parte importante de funcionariado que procede de territorios castellanohablantes instalado, principalmente, en la capital del *Segrià*. Otra posibilidad, como algún autor ya ha apuntado con anterioridad (Colón 1967: 203) es que la clase obrera esté mucho más expuesta al catalán que los empleados públicos quienes, más difícilmente, se dejan influir por el ambiente. Sin embargo, esta circunstancia parece improbable que aún se diera en sus descendientes, que son la población que estudiamos en este trabajo.

De otra parte, si relacionamos esta variable social con la *lengua materna*, como hemos hecho en anteriores ocasiones, los datos son aún más sorprendentes. Es precisamente el estrato bajo, que obtiene los resultados más favorables al catalán, el que mayor número de castellanohablantes contiene (un 31.88% frente al 25.93% del estrato medio y el 25% del alto). En sentido contrario, es el estrato alto el que más catalanohablantes aporta al total, con un 69.44% del total de integrantes de este sociolecto, aunque los datos, en las dos direcciones, son bastante similares (el estrato medio tiene un 67.41% de informantes cuya lengua materna es el catalán y el bajo, un 59.42%). En conclusión, sería necesario realizar un análisis estadístico para corroborar si estas diferencias entre el comportamiento de los estratos bajo y medio, por un lado, y el alto, por otro, en relación con las dos lenguas oficiales, son extrapolables al resto de población, y por tanto, tienen un reflejo real que valdría la pena analizar en profundidad, o, por el contrario, se deben al azar o a otras causas accidentales⁷⁹².

⁷⁹² La mayor tendencia al uso de la lengua vernácula por parte de la clase social más baja también puede explicarse por el deseo que sienten las clases populares de integrarse en la sociedad que las acoge y por la creencia de que sus hijos tendrán mayores posibilidades sociales si dominan la lengua de la comunidad de acogida. Las clases acomodadas, por el contrario, son más reacias a cambiar su lengua materna. Recordemos, como ya explicamos detalladamente (*vid.* 4.3.2), que hay un porcentaje de población de origen castellanohablante que se identifica con el catalán y opta por transmitir esta lengua a sus hijos.

Tabla 6.33. Total de vocablos por centro de interés según la variable nivel sociocultural.

Nº	Centros de interés	Bajo		Medio		Alto	
		Castellano	Catalán	Castellano	Catalán	Castellano	Catalán
1	El cuerpo humano	212	206	320	297	175	172
2	Ropa y complementos	177	182	229	263	145	152
3	Partes de la casa (sin los muebles)	131	131	189	216	110	107
4	Muebles de la casa	146	147	212	229	108	98
5	Comidas y bebidas	283	310	401	437	246	263
6	Objetos colocados en la mesa para la comida	153	159	195	245	110	102
7	La cocina y sus utensilios	191	202	282	306	148	146
8	La escuela (muebles y materiales)	242	249	332	334	185	199
9	Calefacción, iluminación y formas de airear un recinto	139	143	218	212	110	103
10	La ciudad	323	341	518	495	259	245
11	El campo	347	372	550	577	250	253
12	Medios de transporte	133	142	221	242	135	133
13	Trabajos del campo y del jardín	201	190	289	296	158	127
14	Animales	223	250	284	361	186	216
15	Juegos y distracciones	317	265	437	451	208	203
16	Profesiones y oficios	320	323	439	461	271	227
17	Colores	128	127	196	203	88	99
Global		3666	3739	5312	5625	2892	2845

Con respecto a la riqueza léxica, esto es, al número de vocablos emitido por cada sociolecto, lo primero que se observa es la fuerte influencia que tiene, como ya sabemos, el número de integrantes de cada nivel social en el total de vocablos recabado. De este modo, en todas las áreas temáticas y en las dos lenguas analizadas (véase tabla 6.33), los resultados siempre favorecen al estrato medio, por ser el que tiene más informantes (135), seguido del estrato bajo —el segundo grupo más numeroso (69)— y, en último lugar, del nivel social alto, compuesto por un número menor (36). La única excepción a esta tendencia son los datos recabados en *Medios de transporte*, donde el nivel bajo tiene en castellano 133 vocablos, exactamente el mismo número que ostenta el nivel alto en catalán, lo cual no deja de ser significativo, toda vez que los representantes del nivel bajo son casi el doble que los del alto. El sociolecto bajo, cuyos resultados en la producción se decantan claramente hacia el catalán, también ostenta una riqueza léxica mayor en esa lengua, con una diferencia en el total de 73 palabras diferentes (3739 frente a 3666). El catalán también predomina en la mayoría de centros de interés, únicamente *El cuerpo humano*, *Trabajos del campo y del jardín*, *Juegos y*

distracciones y *Colores* obtienen un mayor número de vocablos en castellano⁷⁹³. Los resultados entre las dos lenguas no difieren demasiado, con la única excepción de *Juegos y distracciones*, que recoge 52 palabras diferentes más en castellano.

En el estrato medio, se observan resultados semejantes con un predominio del catalán tanto en el total (5625 frente a 5312) como en la mayoría de áreas temáticas analizadas. Las únicas excepciones son ahora *El cuerpo humano*, *Calefacción*, *iluminación y formas de airear un recinto* y *La ciudad*, donde, aunque a corta distancia, el número de vocablos es mayor en lengua castellana, tal como sucede en las datos generales (vid. 6.1.4.). Es llamativa la diferencia hallada en *Animales*, con 77 palabras más evocadas en lengua catalana o en *Objetos colocados en la mesa para la comida*, con 50 más. En estos dos ejemplos creemos que se trasluce el peso de la lengua materna de la mayoría de los informantes, tanto en el caso de un área, como es *Animales*, cuyo léxico se aprende tempranamente, como en otra estrechamente relacionada con la lengua utilizada en el hogar, esto es, *Objetos colocados en la mesa para la comida*⁷⁹⁴.

En cuanto al sociolecto alto, este muestra, al igual que sucedía en los promedios globales y por área según el idioma de la encuesta, un comportamiento más equitativo en ambas lenguas: en once centros el número de vocablos recabado es mayor en castellano (*El cuerpo humano*, *Partes de la casa*, *Muebles de la casa*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios*, *Calefacción*, *iluminación y formas de airear un recinto*, *La ciudad*, *Medios de transporte*, *Trabajos del campo y del jardín*, *Juegos y distracciones* y *Profesiones y oficios*), mientras que en los seis restantes los datos son favorables al catalán. Asimismo, es el único estrato donde, en total, se obtiene más vocablos en castellano que en catalán, concretamente 47 más. Las disimilitudes más acusadas se dan en *Profesiones y oficios* y *Trabajos del campo y del jardín*, con 44 y 31 vocablos, respectivamente, a favor del castellano⁷⁹⁵. En cambio, a favor del catalán, hallamos, de nuevo, *Animales*, con 30 palabras distintas actualizadas de más en esta lengua, en comparación con el castellano. También es llamativa la casualidad de que en tres categorías (*Partes de la casa*, *Objetos colocados*

⁷⁹³ En algunas áreas, como *El cuerpo humano* o *Colores*, la diferencia es realmente mínima. De otro lado, en *Partes de la casa* se recaba exactamente el mismo número de vocablos en las dos lenguas. Nótese que, salvo *Colores*, el resto de centros de interés mencionados obtiene, en los datos generales, una ventaja a favor del castellano.

⁷⁹⁴ Recordemos que, en el sociolecto medio, los catalanohablantes son el 67.41% del total.

⁷⁹⁵ También en el ranquin general estos centros obtienen más vocablos en esta lengua (vid. 6.1.4.)

en la mesa para la comida y Calefacción, iluminación y formas de airear un recinto) se recaben exactamente el mismo número de vocablos —110— en lengua castellana.

En cuanto a la distribución de las áreas temáticas, según esta variable, en un ranquin de mayor a menor en función de las palabras distintas obtenidas, son pocas las diferencias que se detectan con respecto a la clasificación general del total de datos según este mismo parámetro (vid. 6.1.4.). En todos los casos, los cinco centros con mayor riqueza léxica en las dos lenguas —*El campo, La ciudad, Profesiones y oficios, Juegos y distracciones y Comidas y bebidas*— encabezan las clasificaciones, con la excepción del estrato alto en catalán donde *Animales* desplaza en quinta posición a *Juegos y distracciones*, relegado a la sexta. La mitad de las veces, esto es en el estrato medio en los dos idiomas y en el bajo en castellano, además, estas cinco áreas asociativas siguen incluso fielmente el orden constatado en la clasificación general (tal como las acabamos de enumerar unas líneas más arriba)⁷⁹⁶. En general, las coincidencias con la clasificación global son mayoritarias y las diferencias, cuando las hay, son sobre todo de una o dos posiciones⁷⁹⁷.

Tabla 6.34. Índice de cohesión por centro de interés según la variable social nivel sociocultural.

Nº	Centros de interés	Bajo		Medio		Alto	
		Castellano	Catalán	Castellano	Catalán	Castellano	Catalán
1	El cuerpo humano	0.11	0.13	0.08	0.10	0.16	0.18
2	Ropa y complementos	0.12	0.12	0.10	0.09	0.15	0.16
3	Partes de la casa (sin los muebles)	0.10	0.11	0.08	0.07	0.13	0.14
4	Muebles de la casa	0.08	0.08	0.06	0.06	0.11	0.12
5	Comidas y bebidas	0.08	0.08	0.06	0.06	0.09	0.10
6	Objetos colocados en la mesa para la comida	0.08	0.08	0.07	0.06	0.13	0.13
7	La cocina y sus utensilios	0.08	0.08	0.06	0.06	0.10	0.11
8	La escuela (muebles y materiales)	0.09	0.09	0.07	0.07	0.11	0.11
9	Calefacción, iluminación y formas de airear un recinto	0.07	0.07	0.05	0.05	0.11	0.11
10	La ciudad	0.06	0.06	0.04	0.04	0.08	0.08
11	El campo	0.05	0.05	0.04	0.03	0.07	0.07
12	Medios de transporte	0.12	0.11	0.08	0.07	0.13	0.13

⁷⁹⁶ En el nivel bajo en catalán solo coinciden los tres primeros centros con el orden general, ya que en cuarta y quinta posición se intercambian *Comidas y bebidas* y *Juegos y distracciones*. En el estrato alto, *Profesiones y oficios* lidera la clasificación en castellano y, en catalán, *Comidas y bebidas*.

⁷⁹⁷ El comportamiento más distinto, en general, quizá sea el del estrato bajo en castellano, donde, además de muchas distinciones de una o dos posiciones, se detectan tres diferencias de tres posiciones y una de cuatro con respecto a la clasificación general de los datos.

13	Trabajos del campo y del jardín	0.04	0.05	0.03	0.03	0.06	0.08
14	Animales	0.11	0.10	0.09	0.07	0.14	0.12
15	Juegos y distracciones	0.05	0.06	0.04	0.04	0.08	0.07
16	Profesiones y oficios	0.06	0.06	0.05	0.05	0.07	0.09
17	Colores	0.14	0.14	0.10	0.09	0.20	0.18

Con respecto al último de los valores analizados —la cohesión— se puede constatar en la tabla anterior (tabla 6.34), de forma muy clara, cómo, una vez más, el número de integrantes de cada sociolecto influye en el valor obtenido. De este modo, en el análisis de cada área asociativa, e independientemente de la lengua estudiada, siempre alcanza los índices más altos el nivel superior, ya que —al estar constituido por menos sujetos— tiene también un número menor de vocablos como acabamos de ver y, por ende, las posibilidades de que las respuestas coincidan son mayores. A continuación del estrato alto, registra el índice más elevado el bajo, y finalmente, el estrato medio, el grupo constituido por la cifra mayor de informantes⁷⁹⁸. Como en las anteriores ocasiones donde analizamos este valor por subgrupos sociales (*vid.* 6.2.1., 6.1.2 y 6.1.3.), se dan muchísimas coincidencias en los valores que alcanzan las dos lenguas (concretamente en el estrato bajo coinciden el mismo índice once áreas, en el medio, nueve y en el alto, seis). Con respecto al rango que ocupa cada centro de interés según el índice de cohesión, una vez más se corroboran las tendencias generales detalladas con anterioridad (*vid.* 6.1.5.) con algunas diferencias sin importancia, ya que los centros más abiertos y los más cerrados coinciden en todas las clasificaciones⁷⁹⁹. De otra parte, es difícil establecer el rango que obtiene cada centro de interés, puesto que muchas categorías, al redondear los decimales, coinciden en su valor, tal como se observa en la tabla.

6.2.5. Variable social *lengua materna*

Pasamos, finalmente, al análisis de la última variable social, la *lengua materna* del informante. El estudio de su posible influencia es uno de los puntos más interesantes

⁷⁹⁸ Los tres sociolectos se ordenan de forma contraria según el número de vocablos registrado, esto sucede porque, como ya explicamos (*vid.* 6.1.5.) el número de vocablos y el índice de cohesión guardan una relación inversamente proporcional.

⁷⁹⁹ Algunas de estas diferencias serían, por ejemplo, que en castellano en el estrato alto, *El cuerpo humano* y *Animales*, segundo y cuarto respectivamente, intercambian sus posiciones con respecto al ranquin general o cómo, en el estrato bajo, *Medios de transporte* alcanza la segunda posición, compartida con *Ropa y complementos*, cuando en el conjunto de los datos es el quinto. En lengua catalana, en el estrato bajo, lidera la clasificación *Colores*, y no *El cuerpo humano*. En la misma lengua y en el estrato alto, *Comidas y bebidas* obtiene una posición bastante baja, la duodécima, con respecto a su puesto general, el séptimo. De cualquier modo, las diferencias son poco importantes y es difícil valorarlas al manejar solo dos decimales.

de este trabajo, dada la situación de bilingüismo social en la que están inmersos todos nuestros encuestados. Para empezar, cabe recordar que al estratificar la muestra obtenemos cuatro grupos lingüísticos de tamaño muy distinto. En general, la mayor parte de sujetos se engloba, al igual que sucede en la población real de la provincia (*vid.* 5.2., gráfico 5.7), en el conjunto cuya lengua materna es el castellano o el catalán. El primer grupo, el castellanohablante, lo componen 66 informantes, el 27.50% del total; mientras que el segundo está integrado por 157 individuos, el 65.42% de la muestra. De otro lado, contamos con dos colectivos más: el de aquellos encuestados que incluyen tanto el catalán como el castellano como lengua materna —recogidos bajo la etiqueta de “ambas lenguas”— y el de los informantes cuya lengua materna es el aranés, toda vez que *Vielha e Mijaran* fue uno de los seis puntos de la provincia donde se recogieron datos⁸⁰⁰. El primero de estos dos grupos, el de “ambas lenguas”, está formado por 10 individuos, el 4.17% de la muestra; el segundo, aún menor, lo conforman 7 informantes, el 2.92% del total. El tamaño tan desigual de estos cuatro grupos influirá, como veremos a continuación, en los resultados.

En la siguiente tabla (tabla 6.35) se recogen los promedios generales y por centro de interés según la variable *lengua materna*. Con la finalidad de facilitar la comprensión de los datos, recordamos al lector que en la tabla se indica, en la parte superior, la variante de cada grupo social, esto es, la lengua materna bajo la que agrupamos cada conjunto de informantes (LM en la tabla). Por tanto, de modo similar al apartado anterior (*vid.* 6.2.4.), donde este lugar era ocupado por las variantes del nivel sociocultural “bajo”, “medio” y “alto”, ahora hallamos aquí las distintas afiliaciones lingüísticas, es decir, castellano, catalán, ambas lenguas (“ambas” en la tabla) y aranés. A continuación, cada colectivo de informantes se divide según los promedios obtenidos en cada uno de los dos corpus: el del léxico disponible en castellano (“castellano” en la tabla) y el del léxico disponible en catalán (“catalán” en la tabla).

⁸⁰⁰ Para más información sobre esta lengua y su situación sociolingüística actual en *la Vall d’Aran*, véase el final del capítulo cuarto.

Tabla 6.35. Promedio de palabras por centro de interés según la variable lengua materna.

Nº	Centros de interés	LM: Castellano		LM: Catalán		LM: Ambas		LM: Aranés	
		Castellano	Catalán	Castellano	Catalán	Castellano	Catalán	Castellano	Catalán
1	CUERPO	26.32	27.65	25.17	29.53	23.90	27.60	28.71	26.71
2	ROPA	22.91	22.12	20.79	23.46	23.10	24.40	26.00	25.57
3	CASA	15.02	13.14	13.67	15.22	15.10	14.70	16.00	14.29
4	MUEBLES	13.39	12.89	11.93	12.85	12.80	12.70	12.86	10.86
5	COMIDAS	24.95	25.56	23.54	26.83	25.50	25.10	30.43	29.43
6	OBJETOS	14.68	12.48	12.29	14.36	15.00	14.20	17.29	14.86
7	COCINA	16.50	14.64	14.59	17.17	18.00	17.70	18.86	16.57
8	ESCUELA	22.00	21.70	20.72	22.39	21.30	20.60	23.14	22.43
9	CALEFACCIÓN	11.30	9.89	10.92	11.55	11.20	11.10	12.43	10.57
10	CIUDAD	21.15	19.97	21.36	21.69	21.80	22.10	22.43	22.29
11	CAMPO	19.08	16.86	17.50	19.01	20.20	18.90	22.86	18.29
12	TRANSPORTE	16.45	16.24	16.99	17.45	18.30	17.40	15.29	16.57
13	TRABAJOS	8.92	8.73	8.73	10.06	10.20	10.80	9.14	9.86
14	ANIMALES	27.23	24.09	25.38	27.20	25.40	26.10	30.00	28.00
15	JUEGOS	16.14	14.92	16.10	15.93	18.60	16.50	19.43	17.14
16	PROFESIONES	19.71	19.11	19.27	20.77	21.50	20.30	23.86	23.29
17	COLORES	19.17	17.98	17.87	18.95	18.60	17.30	21.29	19.86
Global		18.52	17.53	17.46	19.08	18.85	18.68	20.59	19.21

A lo largo de este apartado, vamos a analizar los datos con exhaustividad y separadamente. Sin embargo, podemos adelantar ya algunas conclusiones que se deducen con solo echar una ojeada a los promedios generales. Entre todos los grupos de informantes, descuella la actuación lingüística del colectivo cuya lengua materna es el aranés, que alcanza los mejores resultados tanto en castellano, con un promedio de 20.59, como en catalán, con un promedio de 19.21. De otra parte, es muy significativo que los peores resultados los obtengan los catalanohablantes en el corpus castellano, con un 17.46 de promedio, y, de forma similar, los castellanohablantes en el corpus catalán, con 17.53 de media⁸⁰¹. En el corpus castellano, como ya hemos indicado, lideran la clasificación los hablantes de aranés (20.59), seguidos de los informantes que se decantan por las dos lenguas como maternas, que obtienen un promedio de 18.85 respuestas por centro de interés. En tercer lugar, y a corta distancia, les siguen los castellanohablantes, con una media de 18.52 y, en último lugar, se sitúan los catalanohablantes, que recaban 17.46 respuestas en global. En los datos relativos al

⁸⁰¹ Recordamos, una vez más que utilizamos los términos “castellanohablante” y “catalanohablante” para referirnos a las personas cuya lengua materna es el castellano y el catalán respectivamente. Como hemos subrayado en reiteradas ocasiones, todos nuestros informantes son bilingües, y por tanto, dominan ambas lenguas desde la infancia.

léxico disponible del catalán, se da una distribución un tanto distinta: los araneses ostentan, por muy pocas décimas, el primer puesto (con 19.21) y, a continuación, el mejor promedio lo alcanzan, como era esperable, los catalanohablantes (19.08), seguidos, en tercer lugar, de los hablantes que declaran tener las dos lenguas como maternas (18.68) y, en cuarto y último lugar, por los castellanohablantes (17.53). Cabe destacar que la distancia en cada lengua entre el peor y el mejor resultado es bastante mayor en castellano —donde se dan 3.13 palabras de diferencia entre el promedio global de los araneses y el de los catalanohablantes— que en catalán —donde solo se separan en 1.68 respuestas la media de los araneses de la de los castellanohablantes—. Si nos fijamos en los valores globales de la tabla, el comportamiento lingüístico más similar, o la menor diferencia entre cifras, es el de los araneses y los catalanohablantes en el corpus catalán que exhiben el resultado más cercano, con solo 0.13 décimas de diferencia (19.21 frente a 19.08). También cabe mencionar la similitud existente entre los datos recabados por el colectivo que tiene las dos lenguas oficiales como maternas en las dos lenguas (18.85 en castellano y 18.68 en catalán).

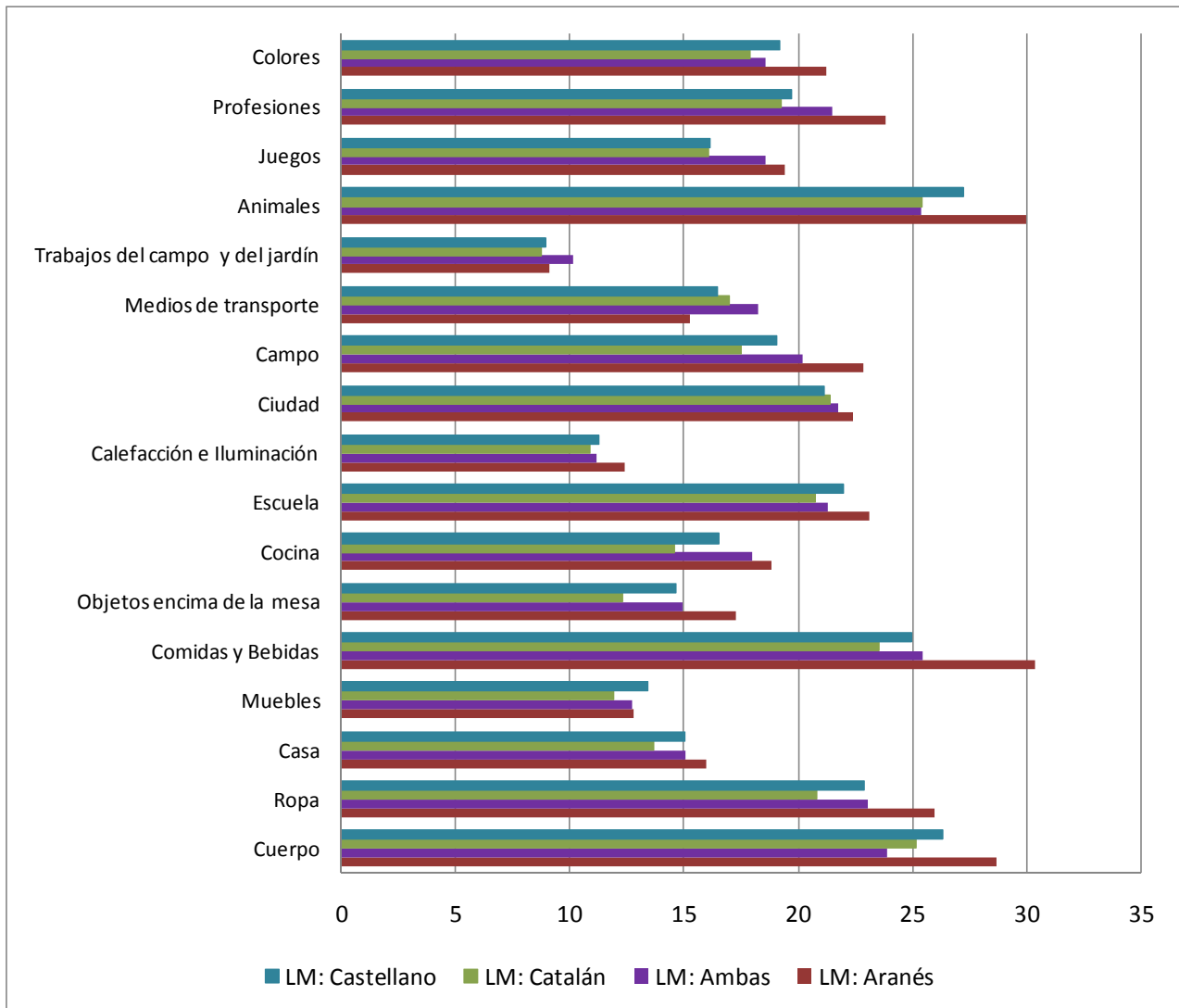
A continuación, analizamos únicamente los datos recabados en la encuesta en castellano. Para facilitar la lectura y dada la profusión de datos de la tabla 6.35, reproducimos, seguidamente, solamente los promedios de los cuatro grupos lingüísticos en lengua castellana, datos que extraemos de la tabla anterior, y que se reflejan en la tabla y en el gráfico siguientes:

Tabla 6.36. Promedio de palabras por centro de interés según la variable lengua materna en lengua castellana.

Nº	Centros de interés	LM: Castellano	LM: Catalán	LM: Ambas	LM: Aranés
1	El cuerpo humano	26.32	25.17	23.90	28.71
2	Ropa y complementos	22.91	20.79	23.10	26.00
3	Partes de la casa (sin los muebles)	15.02	13.67	15.10	16.00
4	Muebles de la casa	13.39	11.93	12.80	12.86
5	Comidas y bebidas	24.95	23.54	25.50	30.43
6	Objetos colocados en la mesa para la comida	14.68	12.29	15.00	17.29
7	La cocina y sus utensilios	16.50	14.59	18.00	18.86
8	La escuela (muebles y materiales)	22.00	20.72	21.30	23.14
9	Calefacción, iluminación y formas de airear un recinto	11.30	10.92	11.20	12.43
10	La ciudad	21.15	21.36	21.80	22.43

11	El campo	19.08	17.50	20.20	22.86
12	Medios de transporte	16.45	16.99	18.30	15.29
13	Trabajos del campo y del jardín	8.92	8.73	10.20	9.14
14	Animales	27.23	25.38	25.40	30.00
15	Juegos y distracciones	16.14	16.10	18.60	19.43
16	Profesiones y oficios	19.71	19.27	21.50	23.86
17	Colores	19.17	17.87	18.60	21.29
Global		18.52	17.46	18.85	20.59

Gráfico 6.31. Promedio de palabras por centro de interés según la variable lengua materna en lengua castellana.



Como ya hemos indicado y tal como se aprecia en el gráfico 6.31, el mejor resultado global lo ostenta el grupo aranés, que, concretamente, obtiene la cifra más alta en catorce de las diecisiete áreas temáticas. En las tres categorías semánticas en las que no descuella con el mejor promedio, es decir, en *Muebles de la casa*, *Medios de transporte* y *Trabajos del campo y del jardín*, los castellanohablantes lideran el ranking

en una ocasión; en *Muebles de la casa*, y en las otras dos, son los hablantes que optan tanto por el castellano como por el catalán los que obtienen los promedios más altos. A los resultados tan favorables de los araneses le sigue, en nueve ocasiones, de nuevo, el colectivo que se decanta por las dos lenguas oficiales que, de este modo, alcanza la segunda mejor cifra en *Ropa y complementos*, *Partes de la casa*, *Comidas y bebidas*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios*, *La ciudad*, *El campo*, *Juegos y distracciones* y *Profesiones y oficios*. La única vez en que los castellanohablantes obtienen la media más alta —en *Muebles de la casa* como hemos visto—, los hablantes de aranés les siguen a continuación, con solo media palabra de diferencia (13.39 frente a 12.86). De otra parte, en las escasas ocasiones en las que el grupo de ambas lenguas ostenta el primer puesto, se sitúan a continuación los araneses, en *Trabajos del campo y del jardín* y, de forma llamativa, los catalanohablantes en *Medios de transporte*.

Es importante destacar que los informantes cuya lengua materna es el catalán, obtienen, generalmente, el peor resultado. Concretamente, en catorce centros de interés, su promedio es el más bajo de los cuatro grupos lingüísticos. Por el contrario, los castellanohablantes solo ocupan una vez el último puesto, en *La ciudad*, donde todos los colectivos obtienen cifras no muy alejadas entre sí. También los araneses en una ocasión ostentan el promedio más bajo del conjunto, en *Medios de transporte*, esta vez sí, con una diferencia mayor, concretamente de tres palabras. En *El cuerpo humano* son los hablantes que declaran tener ambas lenguas como maternas los que ostentan el último puesto, con una diferencia bastante acusada con respecto al mejor promedio, en este caso, el de los araneses.

Los colectivos que exhiben un comportamiento más similar son los castellanohablantes y los hablantes con las dos lenguas oficiales como maternas. Estos dos grupos de informantes tienen una diferencia en el promedio global de 0.33 décimas. Además, en cuatro centros de interés, *Ropa y complementos*, *Partes de la casa*, *Objetos colocados en la mesa para la comida* y *Calefacción, iluminación y formas de airear un recinto*, la diferencia entre los dos resultados —que, excepto en el último centro mencionado, otorga el mejor puesto a los informantes de ambas lenguas como maternas— es realmente pequeña, de apenas décimas, o incluso, centésimas. Por el contrario, la distancia mayor se da entre los catalanohablantes —que, como hemos

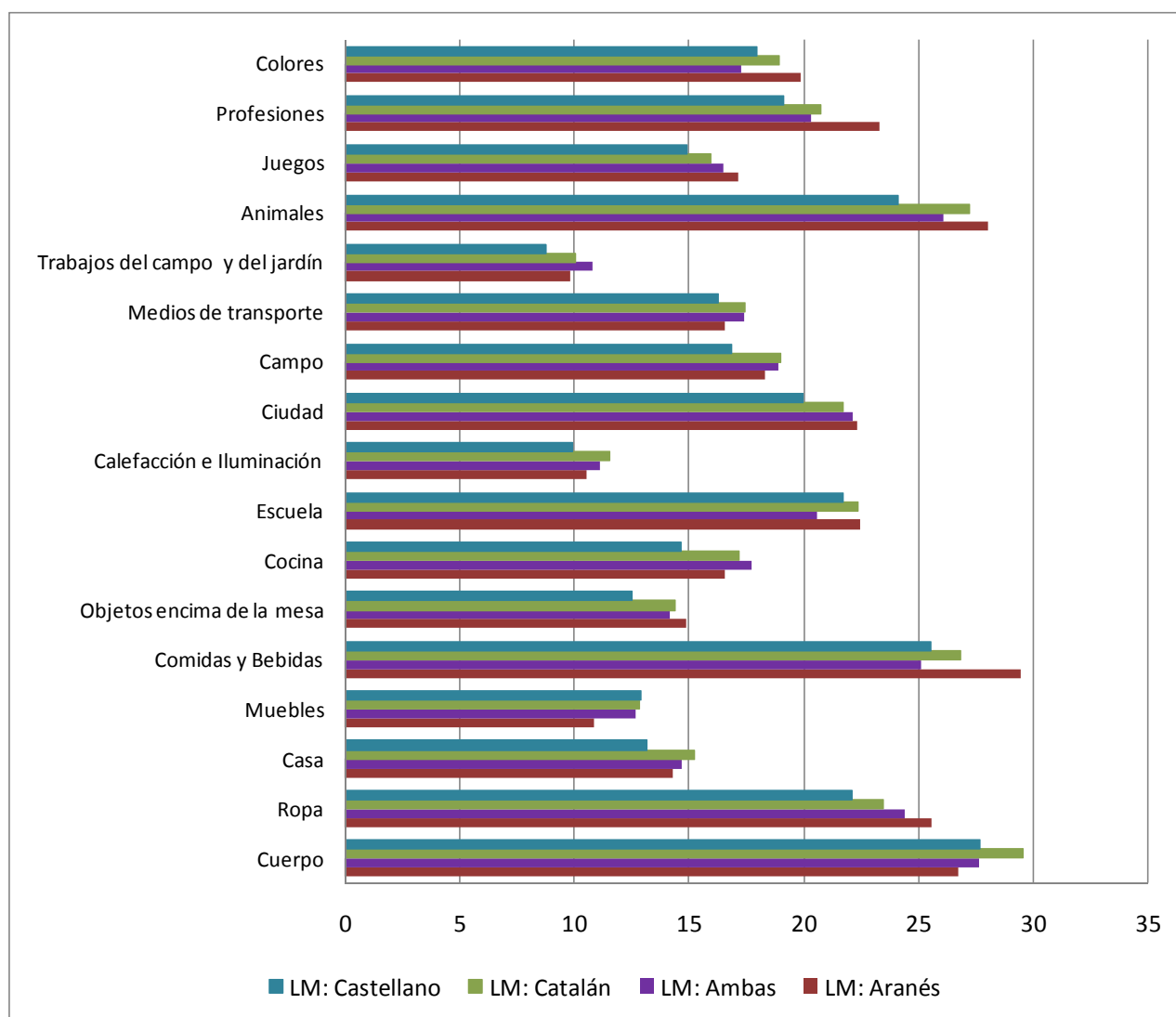
dicho, generalmente tienen el peor resultado— y los araneses, con sus promedios extraordinarios, de los que más adelante hablaremos. Estos dos grupos muestran una diferencia en el global de 3.13 respuestas, pero en el análisis por categorías semánticas, llegan a dibujarse unas divergencias realmente inusitadas de casi siete palabras (en *Comidas y bebidas*, concretamente de 6.89), de cinco o más (*Ropa y complementos*, *Objetos colocados en la mesa para la comida* y *El campo*), y no son pocos los centros en los que se supera las cuatro respuestas de diferencia (*La cocina y sus utensilios*, *Animales* y *Profesiones y oficios*). Intentaremos hallar, tras el siguiente análisis, una explicación que justifique estos resultados.

Seguidamente, analizamos los datos relativos a la productividad en lengua catalana. Al igual que hemos hecho con los datos del léxico disponible en castellano, reproducimos aquí parte de los datos de la tabla 6.35 —plasmados en el gráfico 6.32—, los relativos a la influencia de la variable *lengua materna* en el léxico disponible en lengua catalana:

Tabla 6.37. Promedio de palabras por centro de interés según la variable lengua materna en lengua catalana.

Nº	Centros de interés	LM: Castellano	LM: Catalán	LM: Ambas	LM: Aranés
1	El cuerpo humano	27.65	29.53	27.60	26.71
2	Ropa y complementos	22.12	23.46	24.40	25.57
3	Partes de la casa (sin los muebles)	13.14	15.22	14.70	14.29
4	Muebles de la casa	12.89	12.85	12.70	10.86
5	Comidas y bebidas	25.56	26.83	25.10	29.43
6	Objetos colocados en la mesa para la comida	12.48	14.36	14.20	14.86
7	La cocina y sus utensilios	14.64	17.17	17.70	16.57
8	La escuela (muebles y materiales)	21.70	22.39	20.60	22.43
9	Calefacción, iluminación y formas de airear un recinto	9.89	11.55	11.10	10.57
10	La ciudad	19.97	21.69	22.10	22.29
11	El campo	16.86	19.01	18.90	18.29
12	Medios de transporte	16.24	17.45	17.40	16.57
13	Trabajos del campo y del jardín	8.73	10.06	10.80	9.86
14	Animales	24.09	27.20	26.10	28.00
15	Juegos y distracciones	14.92	15.93	16.50	17.14
16	Profesiones y oficios	19.11	20.77	20.30	23.29
17	Colores	17.98	18.95	17.30	19.86
Global		17.53	19.08	18.68	19.21

Gráfico 6.32. Promedio de palabras por centro de interés según la variable lengua materna en lengua catalana.



Tal como se observa en el gráfico 6.32, también aquí destacan los promedios tan favorables obtenidos por el grupo aranés, que consigue no solo la mejor media global, con 19.21 respuestas, sino también la cifra más alta en nueve de las diecisiete áreas temáticas. En los ocho centros de interés en que los resultados mejores no son los de este colectivo, destacan los promedios recabados por los catalanohablantes, que en cinco áreas obtienen el mejor resultado (*El cuerpo humano*, *Partes de la casa*, *Calefacción, iluminación y formas de airear un recinto*, *El campo y Medios de transporte*). En otras dos categorías, *La cocina y sus utensilios* y *Trabajos del campo y del jardín* es el grupo que tiene las dos lenguas oficiales como maternas el que lidera la clasificación, y en una única ocasión, son los castellanohablantes los que alcanzan la

mejor media, concretamente, en *Muebles de la casa*, como también sucedía en castellano.

Con respecto al segundo puesto, este lo ostenta, casi siempre, el colectivo catalanohablante, que recordemos que en cinco ocasiones obtiene los mejores resultados. Solo en tres casos —*La ropa*, *La ciudad* y *Juegos y distracciones*— este grupo de informantes queda en tercer lugar, por detrás del aranés y del colectivo con ambas lenguas como maternas. En cuanto al grupo castellanohablante, este obtiene, en la mayoría de las áreas, el peor resultado del conjunto: solo una vez, como acabamos de ver, registra la mayor puntuación —en *Muebles de la casa*— y en otra —en *El cuerpo humano*— alcanza el segundo puesto. En las tres únicas áreas en que se sitúa en tercera posición, en *Comidas y bebidas*, *La escuela* y *Colores*, el resultado más bajo lo tienen los informantes que se decantan por las dos lenguas oficiales como maternas. Llama la atención los escasos promedios obtenidos por el grupo aranés en *El cuerpo humano* y en *Muebles de la casa*, por debajo de los restantes grupos lingüísticos.

En cuanto a las diferencias que se vislumbran entre los promedios generales, la mayor es la que separa el grupo más productivo, el aranés, del menos, el castellanohablante, y es de 1.68 respuestas. La actuación más similar es la del conjunto catalán y la del aranés, que solo distan en 0.13 en los datos globales. Como acabamos de ver, estos dos colectivos obtienen generalmente la mejor puntuación; en algunas áreas (como en *La escuela* o en *Trabajos del campo y del jardín*) solo se diferencian por unas décimas. De otra parte, las distancias que se observan entre el mejor y el peor resultado son inferiores a las cifras que hallamos en castellano. Aquí solo se superan las cuatro respuestas en dos áreas (*Comidas y bebidas* y *Profesiones y oficios*), las tres en tres áreas más (*Ropa y complementos*, *La cocina y sus utensilios* y *Animales*) y en el resto se aprecian diferencias de una o dos unidades.

A continuación, a modo de resumen, incluimos dos gráficos donde se puede ver, de forma escalonada, los promedios globales de cada grupo de informantes en las dos encuestas: en castellano (gráfico 6.33) y en catalán (6.34).

Gráfico 6.33. Promedios globales de los cuatro grupos lingüísticos en el corpus castellano.

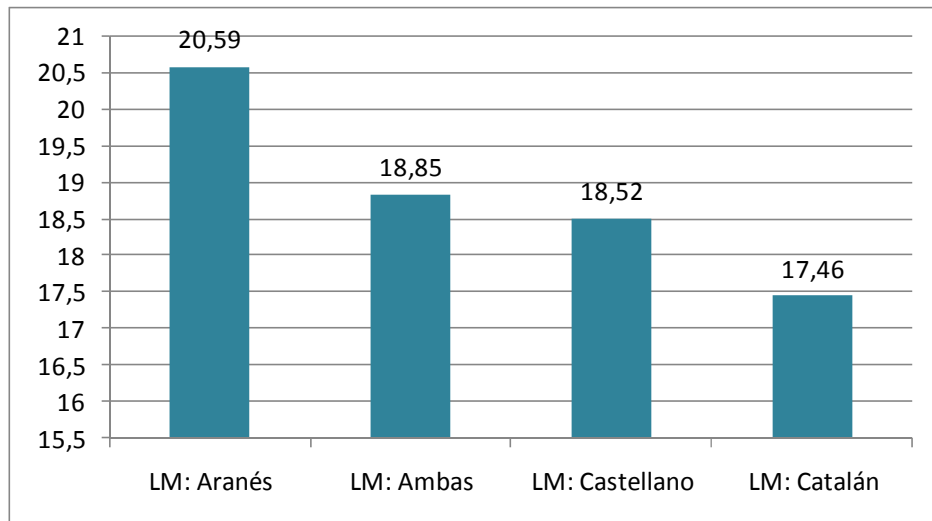
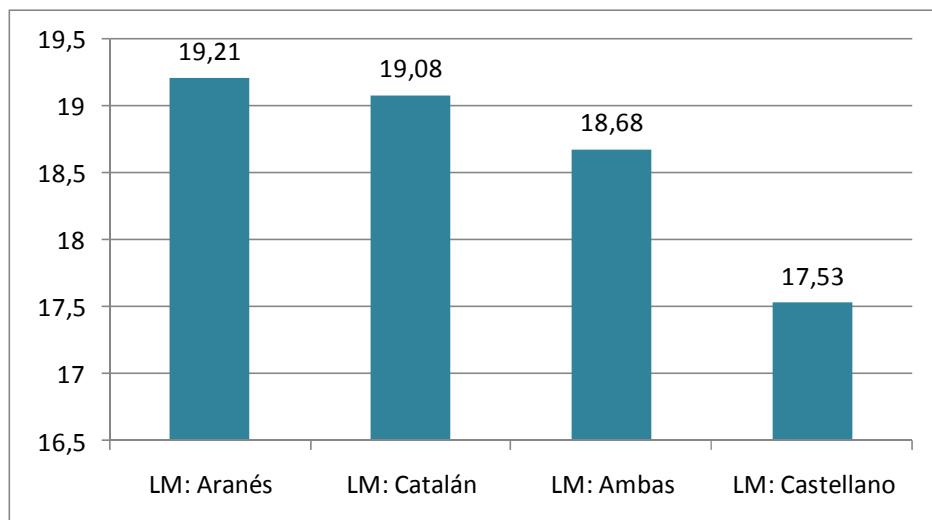


Gráfico 6.34. Promedios globales de los cuatro grupos lingüísticos en el corpus catalán.



Tal como se aprecia ilustrativamente, se dan algunas similitudes y diferencias entre el comportamiento léxico de los cuatro colectivos en las dos lenguas estudiadas. En ambos casos, el mejor resultado es el del grupo aranés, con una media que destaca, en castellano, en 1.74 palabras sobre el segundo mejor resultado —el del grupo de ambas lenguas como maternas— y en catalán, en cambio, con solo unas décimas (0.13) por encima del grupo catalanohablante. En el desglose por áreas, como hemos visto, también es más determinante la actuación del grupo aranés en castellano —donde tiene el mejor promedio en catorce áreas— que en catalán, donde solo lidera en nueve ocasiones el ranquin. Otro hecho remarcable es que en castellano al resultado del grupo aranés le sigue el del colectivo de “ambas lenguas”, hecho que no sucede en catalán,

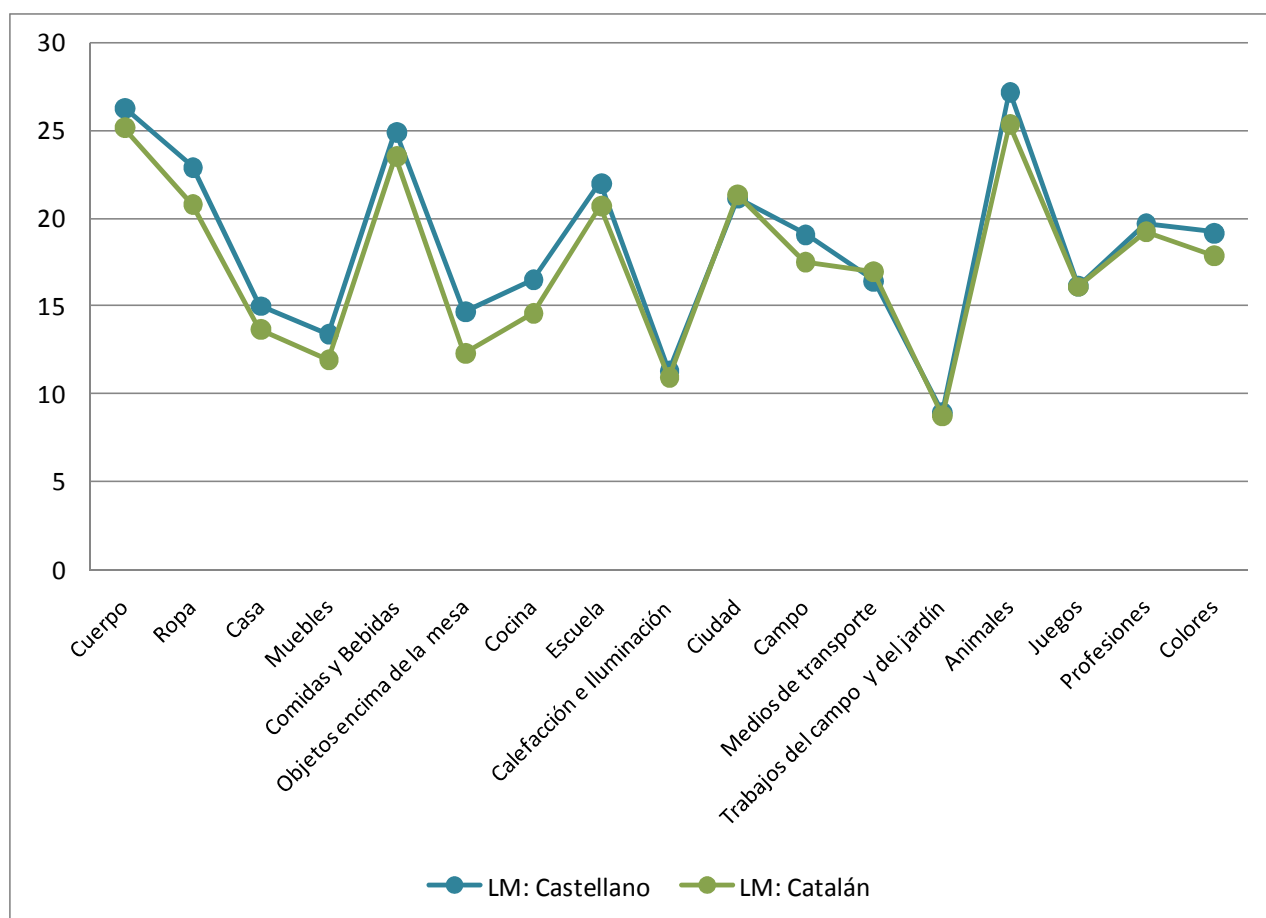
donde este grupo se sitúa en tercera posición. Creemos que la razón de esta diferencia radica en que este grupo de informantes, a pesar de incluir las dos lenguas oficiales como maternas, se decanta más por el uso del castellano que por el del catalán, tal como demuestra que, en la variable *lengua habitual* del informante —no contemplada en este análisis pero cuyos datos sí recopilamos mediante tres preguntas (*vid.* apéndices 1) — la mitad de los diez informantes se decante por el castellano, mientras que tres más lo hagan, de nuevo, por las dos lenguas y solo dos, en cambio, opten por el catalán. De cualquier modo, tendremos ocasión de corroborar en qué lengua muestra este colectivo los mejores resultados cuando comparemos su producción en las dos lenguas de la encuesta (*infra*). Finalmente, como ya hemos indicado, es significativo que en los dos corpus sea precisamente el grupo que tiene la otra lengua oficial como materna el que obtenga el peor resultado casi siempre, es decir, los castellanohablantes en catalán, y los catalanohablantes en castellano.

Llegados a este punto del análisis, nos hemos preguntado qué podía explicar los resultados tan favorables en las dos lenguas del grupo que tiene el aranés como lengua materna. Para ello, hemos profundizado en las características sociales de estos informantes. Como ya dijimos en la introducción a este apartado, este colectivo está formado por siete personas, todas ellas mujeres que, mayoritariamente, proceden del estrato medio (cuatro informantes) y alto (dos informantes). Creemos que estas dos circunstancias cobran una especial relevancia, porque —a estas alturas del análisis— sabemos que las diferencias detectadas en función de la variable *sexo* y del *nivel sociocultural* son significativas estadísticamente a favor de las mujeres y del estrato medio, precisamente. Esto explicaría el hecho de que el subgrupo aranés proporcione unos resultados tan buenos y destaque por encima de los restantes grupos lingüísticos, más allá de la posible incidencia cuantitativa del aranés como lengua materna. Además, dada la cantidad exigua de informantes que pertenecen a esta variante dentro de la variable *lengua materna* en esta investigación, no podemos asegurar que sea este factor, el de la *lengua materna*, y no otros, como el *sexo* o el *nivel sociocultural*, el que esté influyendo sobre los datos. De cualquier modo, todo apunta a que tanto el número de componentes de este grupo (7), como el que se recoge bajo la etiqueta de “ambas lenguas” (10)⁸⁰², es demasiado pequeño como para poder extraer conclusiones

⁸⁰² En este colectivo las variantes sociales se encuentran distribuidas de forma más equitativa: la mitad son varones y la mitad mujeres, también la mitad proceden del estrato bajo, y la mitad del estrato medio.

extrapolables al conjunto a partir de su comportamiento, tanto sobre su productividad como sobre su riqueza léxica. Como ya explicamos en la metodología, los sujetos que manifiestan tener las dos lenguas como maternas son también pocos en el conjunto total de la población de Lleida. De igual manera, los que declaran tener el aranés como lengua materna también están lejos de ser mayoría en nuestra provincia y tampoco abundan en el universo concreto estudiado, los estudiantes de 2º de bachillerato de la provincia (*vid.* 4.3.2). Su presencia en nuestra muestra es puramente anecdótica y se explica por la imposibilidad que se nos presentó de unificar estos informantes con ninguna de las otras variantes lingüísticas, puesto que el aranés es una lengua completamente distinta del castellano o del catalán. De cualquier modo, la presencia de estos dos grupos —totalmente minoritarios en la muestra y obteniendo, uno de ellos, unos resultados sorprendentemente favorables— provoca que no se pueda valorar adecuadamente las diferencias entre los castellanohablantes y los catalanohablantes, mayoría en la población estudiada y también en el universo real. Por esta razón, a continuación, vemos y analizamos los resultados que se exponen en los gráficos y tablas que acabamos de ver (*supra*, tablas 6.36 y 6.37 y gráficos 6.31 y 6.32) pero únicamente con los datos de estos dos subgrupos, castellanohablantes y catalanohablantes, que suponen el 92.92% del total. Cabe destacar que los análisis estadísticos que siguen a continuación también los hemos realizado únicamente teniendo en cuenta estos dos grupos de informantes, cuya lengua materna es el castellano o el catalán, y hemos prescindido de los datos de los otros dos colectivos lingüísticos, el de ambas lenguas como maternas y el del aranés. Empezamos, como hemos hecho siempre, por los resultados relativos a la lengua castellana:

Gráfico 6.35. Promedio de palabras por centro de interés según la variable lengua materna en lengua castellana (solo LM: castellano y LM: catalán).



Tal como era previsible, al comparar la actuación de estos dos grupos de informantes (gráfico 6.35), se comprueba claramente cómo los castellanohablantes obtienen mejores resultados. Este hecho se da de forma constante en todos los centros de interés, con la excepción de *La ciudad* y *Medios de transporte*, donde las cifras son, sin una gran distancia —0.21 en el primer caso y 0.54 en el segundo—, favorables a los catalanohablantes. Como ya hemos indicado anteriormente (*vid.* 6.1.3.), el léxico evocado bajo el estímulo de *La ciudad* se asocia al influjo del castellano, debido a que la presencia de castellanohablantes es mayor en el entorno urbano que en el rural, lo que podría explicar que los catalanohablantes tuvieran un promedio algo superior incluso al de los castellanohablantes en este centro de interés⁸⁰³. Sin embargo, en *Medios de transporte* no habíamos detectado hasta el momento una posible influencia del castellano, aunque la superioridad del grupo cuya lengua materna es el catalán podría

⁸⁰³ Recordemos que en el análisis global de las diferencias según la lengua en que se completa la encuesta, este centro ostenta mejores promedios en lengua castellana, aunque la diferencia —como vimos— no es significativa (*vid.* 6.1.2. y 6.1.3).

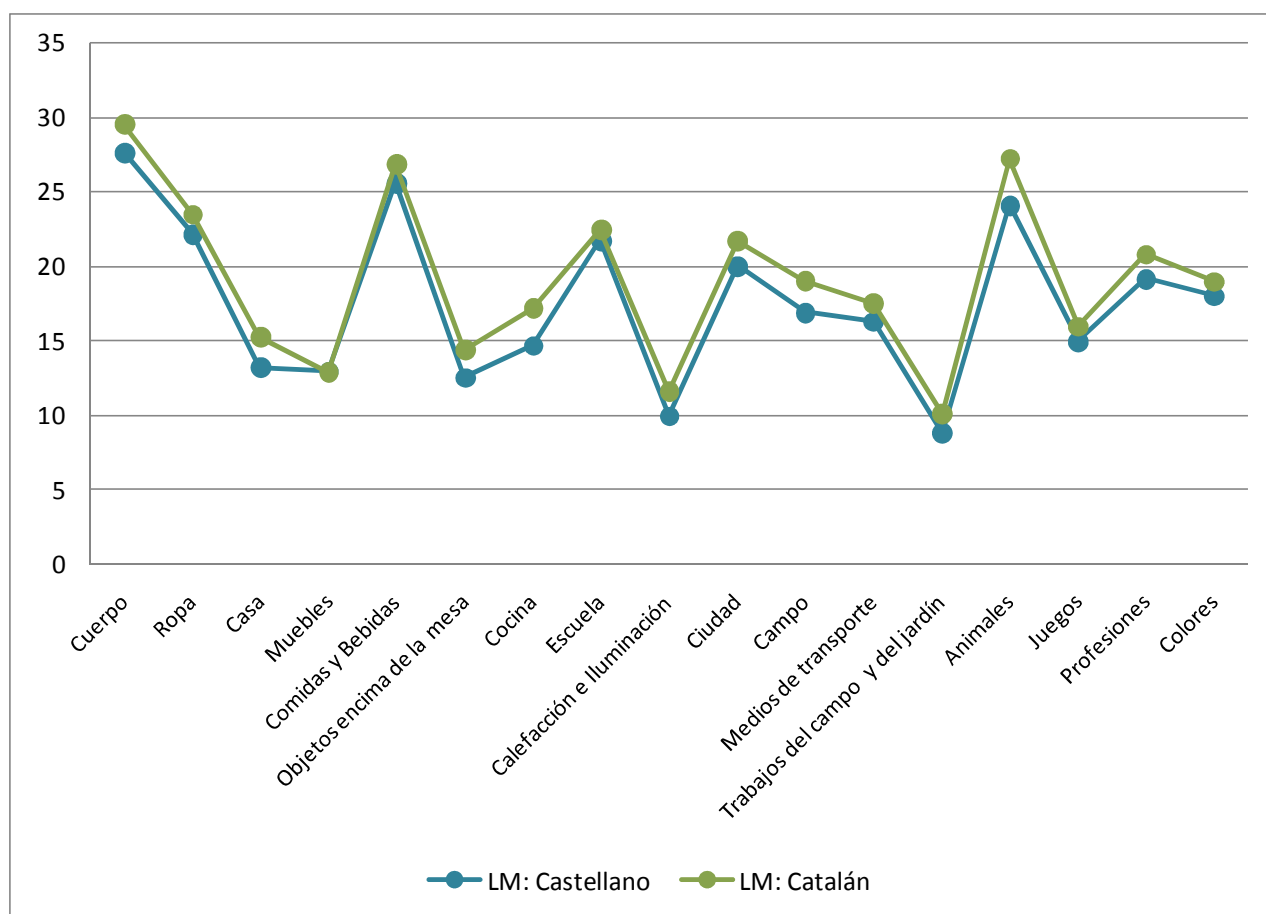
justificarse, quizá, porque muchas de las asociaciones llevadas a cabo en este estímulo son similares a las que se realizan bajo *La ciudad*, a la vez que también influiría en el léxico de este campo los mercados de consumo, donde el castellano tiene mayor presencia que el catalán. Sin embargo, estas razones no parecen ser suficientes para explicar que los catalanohablantes actualicen hasta media respuesta más que los castellano hablantes en esta categoría semántica. En cualquier caso, debemos esperar a los análisis estadísticos para comprobar si esta diferencia es significativa y, por tanto, merece la pena profundizar en las causas de tal disimilitud.

En las restantes áreas temáticas se observa, de modo constante, la superioridad productiva de los castellano hablantes. Es decir, el colectivo que completa la encuesta en su propia lengua materna tiene más palabras en el mismo tiempo de reacción que el que completa la encuesta en su segunda lengua. Las diferencias más acusadas, sin ser cantidades tan alejadas como veíamos al comparar los cuatro grupos lingüísticos entre sí en lengua castellana, se dan en *Ropa y complementos* y en *Objetos colocados en la mesa para la comida*, con 2.12 y 2.39 palabras de más, respectivamente, siempre a favor del grupo cuya lengua materna es el español. En muchos otros centros de interés, la diferencia supera la unidad: *El cuerpo humano*, *Partes de la casa*, *Muebles de la casa*, *Comidas y bebidas*, *La cocina y sus utensilios*, *La escuela*, *El campo*, *Animales y Colores*. Cabe señalar que la distancia que se observa en un centro como *Juegos y distracciones* es realmente pequeña, con 4 centésimas a favor del colectivo castellano hablante⁸⁰⁴. También en *Trabajos del campo y del jardín* la diferencia a favor del castellano es muy pequeña, de 0.19 décimas. Solo en dos centros más, *Calefacción, iluminación y formas de airear un recinto* y *Profesiones y oficios*, la diferencia es de menos de una palabra. De cualquier modo, es remarcable la mejor productividad léxica de los castellano hablantes en casi todas las áreas.

A continuación presentamos y analizamos los mismos datos pero relativos al corpus catalán.

⁸⁰⁴ Al analizar las diferencias del total de la muestra según la lengua de la encuesta, el resultado ya era superior en este campo en lengua castellana. Como ya explicamos, en este centro de interés se observa la influencia del castellano en el ocio juvenil a través de los mercados de consumo (*vid.* 6.1.2 y 6.1.3).

Gráfico 6.36. Promedio de palabras por centro de interés según la variable lengua materna en lengua catalana (solo LM: castellano y LM: catalán).



De modo similar a lo que sucedía en lengua castellana, también aquí se observa, con mayor contundencia si cabe, la superioridad léxica del grupo que completa la encuesta en su lengua materna. En todos los centros de interés el resultado de los catalanohablantes es mejor salvo en *Muebles de la casa*, donde por pocas centésimas el promedio es más elevado en el colectivo castellanohablante. La diferencia más acusada es la del centro 14, *Animales*, de más de tres palabras de diferencia (3.11). Las disimilitudes superan las dos unidades en tres áreas más: *Partes de la casa*, *La cocina y sus utensilios* y *El campo*. En las restantes áreas, salvo en *La escuela*, en *Colores* y en *Muebles de la casa* —como ya hemos comentado— las diferencias entre los dos colectivos siempre superan la unidad. Si comparamos los resultados en las dos lenguas de la encuesta, llama la atención cómo los castellanohablantes obtienen siempre la cifra más alta en *Muebles de la casa*, independientemente de la lengua de la encuesta, aunque por muy poca distancia, como hemos visto, en lengua catalana. De igual modo, los catalanohablantes muestran una superioridad productiva, en los dos idiomas, en *La*

ciudad y en *Medios de transporte*, aunque también por una exigua diferencia, sobre todo en el primero de los centros mencionados en la encuesta en castellano.

Antes de intentar hallar una causa que explique estas coincidencias, comprobamos, al igual que hemos hecho en las anteriores variables, si las diferencias detectadas son estadísticamente significativas mediante la comparación de los promedios de cada grupo lingüístico con ANOVA. Para realizar este análisis, dada la escasez de individuos de la muestra que manifiestan tener el aranés o las dos lenguas oficiales como maternas, excluimos los datos de estos informantes y realizamos todos los cálculos estadísticos sobre los restantes individuos, un total de 223, esto es, el 92.92% de la población estudiada. Tras realizar el análisis de la varianza pertinente, se obtiene un nivel de significación del 0.012 en lengua castellana y de 0.001 en catalana, del que se colige que las diferencias provocadas por la variable *lengua materna*, en ambas lenguas, son predecibles y extrapolables al universo estudiado, puesto que el nivel de probabilidad se sitúa entre 0.000 y 0.05.

Tabla 6.38. Análisis estadístico de la diferencia según la lengua materna.
Valor de la 'p' en las dos lenguas.

Nº	Centros de interés	Castellano	Catalán
1	El cuerpo humano	0.185	0.006
2	Ropa y complementos	0.002	0.062
3	Partes de la casa (sin los muebles)	0.015	0.000
4	Muebles de la casa	0.006	0.942
5	Comidas y bebidas	0.067	0.158
6	Objetos colocados en la mesa para la comida	0.000	0.002
7	La cocina y sus utensilios	0.007	0.001
8	La escuela (muebles y materiales)	0.050	0.353
9	Calefacción, iluminación y formas de airear un recinto	0.469	0.001
10	La ciudad	0.941	0.038
11	El campo	0.066	0.012
12	Medios de transporte	0.351	0.063
13	Trabajos del campo y del jardín	0.893	0.008
14	Animales	0.014	0.000
15	Juegos y distracciones	0.951	0.146
16	Profesiones y oficios	0.515	0.022
17	Colores	0.018	0.097

Tras el análisis de cada centro de interés en particular, y como puede deducirse de los datos recogidos en la tabla 6.38, llegamos a las siguientes conclusiones:

- De los quince centros donde los promedios obtenidos por los castellanohablantes son superiores en lengua castellana, dicha diferencia resulta estadísticamente significativa en siete de ellos: *Ropa y complementos*, *Partes de la casa*, *Muebles de la casa*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios*, *Animales* y *Colores*. En cuanto al centro 8, *La escuela*, este se sitúa justo en el límite de la significación estadística (0.050).
- De las dieciséis áreas donde la productividad es mayor entre los catalanohablantes en lengua catalana, en diez de ellas —*El cuerpo humano*, *Partes de la casa*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios*, *Calefacción, iluminación y formas de airear un recinto*, *La ciudad*, *El campo*, *Trabajos del campo y del jardín*, *Animales* y *Profesiones y oficios*— es posible refutar la hipótesis nula y concluir que las diferencias son causadas por la variable *lengua materna*, y no por otros factores accidentales.

En primer lugar, es interesante analizar la diferencia que existe entre los promedios globales de los dos subgrupos. Como hemos visto, los resultados que tienen significación estadística siempre favorecen al grupo que completa la encuesta en su propia lengua materna. No obstante, la distancia que existe en el corpus castellano entre la media global de los dos grupos lingüísticos, 1.06, es inferior a la que se detecta en el corpus catalán, situada en 1.55 respuestas. En nuestra opinión, esta disimilitud se debe a que los dos perfiles de alumnado (el castellanohablante y el catalanohablante) tienen una relación distinta con respecto a las dos lenguas cooficiales, tal como explicamos al analizar la producción en cada lengua en global (*vid.* 6.1.3.), lo que provoca, en última instancia, que tengan las habilidades comunicativas desarrolladas en distinto grado en cada lengua. Recordemos que, de un lado, tenemos el grupo cuya lengua materna es el castellano, escolarizado desde los tres años, o incluso antes, en la segunda lengua, el catalán⁸⁰⁵. De otro, están los sujetos cuya lengua materna es la vernácula y cuyo contacto con el castellano empieza, con seguridad, a los seis años, en primero de educación básica, con una asignatura semanal de tres horas. Sin embargo, estos

⁸⁰⁵ Aunque la escolarización no es obligatoria hasta los seis años, es muy poco frecuente que una familia no lleve antes al colegio a sus hijos. Si con anterioridad a los seis años acuden a una *escola bressol*, la inmersión también se realiza de forma íntegra en lengua catalana.

informantes pueden tener contacto con el castellano antes, y de hecho es muy probable que lo tengan, según las propias circunstancias vitales de cada individuo⁸⁰⁶. Del mismo modo, el contacto de los castellanohablantes con el catalán puede no limitarse a la escuela, sino que lo habitual, en nuestra provincia, es que se extienda a otros ámbitos, aunque también son las condiciones particulares las que determinan el mayor o menor peso de la lengua catalana en la vida social de cada individuo. Estos dos perfiles de alumnado explican, a nuestro modo de ver, las diferencias que se observan en la producción léxica.

Como hemos visto, en el corpus catalán las diferencias se acentúan, y mientras que el límite inferior es prácticamente similar en los dos corpus (17.46 los catalanohablantes en el corpus castellano y 17.53 los castellanohablantes en el corpus catalán), la cifra que alcanzan los catalanohablantes en su propia lengua materna es superior a la de los castellanohablantes en la suya (19.08 frente a 18.52). En nuestra opinión, esto sucede porque los informantes catalanohablantes refuerzan el léxico de su lengua materna a través de la escuela, donde se trabaja el vocabulario de varias de las áreas de la encuesta. De otra parte, la media que alcanzan los catalanohablantes en el corpus castellano (17.46) es muy similar a la de los castellanohablantes en el corpus catalán (17.53), lo que nos da una idea de la extensión social del castellano o de la coexistencia de las dos sistemas lingüísticos en la sociedad catalana, toda vez que parece difícil alcanzar este resultado si el contacto con el castellano se limita a una asignatura de tres horas semanales. En resumen, mientras la actuación lingüística del colectivo castellanohablante es más similar en las dos encuestas (18.52 y 17.53, respectivamente) en catalán se detectan más diferencias a favor de la lengua materna de los informantes (17.46 y 19.08, respectivamente). Por otro lado, el hecho de que cada colectivo obtenga los mejores resultados en su propia lengua materna también nos indica que el léxico que se evoca a través de los diecisiete centros de interés está ligado al entorno básico y a la familia antes que al currículum escolar, dado que, de modo

⁸⁰⁶ Principalmente, la presencia de castellanohablantes en su entorno próximo: ya sea en el barrio, en su grupo de amigos o incluso en la familia, entendida en sentido amplio. También tendrá una gran incidencia el entorno (rural o urbano) donde se desarrolle. Como hemos indicado reiteradamente, el uso del catalán es más frecuente en los núcleos pequeños y, el del castellano, en las ciudades. Tampoco hay que desdeñar la influencia de los medios de comunicación, sobre todo de la televisión, donde la oferta en castellano es claramente superior, a pesar de que la encuesta de usos lingüísticos del año 2003 daba un uso aproximado de las dos lenguas en dicho medio (*vid.* 4.3.2., gráficos 4.25 y 4.26).

contrario, no se observarían diferencias en función de la lengua materna, puesto que todos se escolarizan en la misma lengua.

En segundo lugar, si ahondamos ya en las diferencias halladas en función de las variables lingüísticas, vale la pena comentar que los únicos centros de interés donde los catalanohablantes alcanzan mejores promedios en castellano (*La ciudad y Medios de transporte*) y, de forma inversa, en los que los castellanoahablantes obtienen mejores cifras en catalán (en *Muebles de la casa*), la probabilidad es superior a 0.05, dato que indica que estas diferencias carecen de significación estadística. Por el contrario, las que sí tienen relevancia son, como ya hemos indicado, las referidas a siete centros en castellano y a diez en catalán. Aunque algunos de estos centros coinciden —concretamente cuatro de ellos, tal como analizamos a continuación—, hay varios cuyos datos solo son significativos en uno de los dos idiomas, ya sea en castellano (*Ropa y complementos, Muebles de la casa y Colores*) o en catalán (*El cuerpo humano, Comidas y bebidas, La escuela, Trabajos del campo y del jardín y Profesiones y oficios*). Creemos que esta divergencia es debida, nuevamente, a que los dos colectivos estudiados se relacionan de distinto modo con cada lengua. Si nos fijamos con atención, las áreas que solo son significativas para los castellanoahablantes se vinculan más a la lengua usada en el hogar, mientras que en las diferencias de los catalanohablantes se observa, en mayor medida, el peso del currículum escolar.

Como ya explicamos (*vid.* 6.1.3.), desde la investigación de Hernández Muñoz (2006) sabemos que la disponibilidad se asocia a las palabras más típicas de una categoría, las más familiares y las más tempranamente aprendidas. La lengua materna influye, de forma decisiva, en estos dos últimos factores, dado que se suele entrar en contacto, pensar o nombrar una palabra antes en lengua materna que en una segunda lengua⁸⁰⁷ y siempre se aprenden antes las palabras en la primera lengua que en la segunda⁸⁰⁸. Creemos que este último factor, la edad de adquisición, influye en la diferencia que se observa, en el corpus catalán, en centros como *El cuerpo humano* o *Animales* —que también aparece en el corpus castellano—, áreas cuyo léxico fundamental se aprende en la más tierna infancia. De modo inverso, esta misma

⁸⁰⁷ Aunque esta última característica puede depender del uso que realice un hablante concreto de las dos lenguas en función del ámbito en el que esté o con quién se relacione.

⁸⁰⁸ También puede haber excepciones según el tipo de bilingüismo individual que se dé.

característica puede estar incidiendo en la mayor productividad de *Ropa y complementos* y de *Colores*, en el corpus castellano.

Con respecto a los resultados del catalán, sin embargo, hay que tener también en cuenta que tanto el vocabulario de *El cuerpo humano* como el de *Animales* se trabaja en la escuela, al igual que el de muchas otras categorías donde la diferencia observada es significativa (*La ciudad, El campo, Trabajos del campo y del jardín, Profesiones y oficios* y, parcialmente, *Calefacción, iluminación y formas de airear un recinto*). El léxico de estos centros de interés se refuerza, fundamentalmente, en la asignatura de *Coneixement del medi natural, social i cultural*, presente en todos los cursos de la educación primaria obligatoria.

Cabe destacar que tres centros de interés —*Partes de la casa, Objetos colocados en la mesa para la comida y La cocina y sus utensilios*— presentan una diferencia significativa, siempre a favor del grupo que completa la encuesta en su propia lengua materna, en los dos corpus. La estrecha relación que guardan estas áreas con el hogar provoca que su vocabulario casi siempre se actualice en la lengua usada en este ámbito que, salvo alguna excepción que pueda existir, es siempre la lengua materna. Por tanto, la segunda de las características halladas por Hernández Muñoz (2006), esto es, la familiaridad, influye aquí en el sentido de que el informante entra en contacto con las palabras (o las nombra o piensa en ellas) en la lengua usada en el hogar, ya que el vocabulario de estas áreas raramente se enuncia fuera del ámbito estricto de la familia. Esta es la causa por la que, tal vez, las disimilitudes se extienden, en castellano, a otras áreas como *Muebles de la casa y Ropa y complementos*.

Es llamativo que, como ya hemos indicado, en un área como *Animales* se sigan produciendo diferencias a favor de la lengua materna de cada colectivo en los dos corpus. Esto sucede a pesar de que el vocabulario que se aprende en esta categoría tiene una importante presencia en el currículum escolar. Al respecto, Hernández Muñoz (2006) demuestra que *Animales*, en castellano como L1, comprende palabras aprendidas antes que las pertenecientes a otras categorías (como *Partes del cuerpo humano, Ropa, Muebles* o *Inteligencia*), lo que puede explicar, en nuestro caso, que los resultados siempre favorezcan al colectivo que completa la lengua en su propia lengua materna en este centro de interés. Sería necesario contar con más estudios que profundizaran en la

importancia de este factor, la edad de adquisición, sobre el léxico disponible y, en general, sobre los factores cognitivos que intervienen en la adquisición tanto de una primera lengua (tal como realiza Hernández Muñoz 2006) como de una segunda, en un caso como el nuestro de bilingüismo social extenso con hablantes bilingües desde la infancia⁸⁰⁹.

Como comentamos en su momento (*vid.* 6.2.2 y 6.2.3), los resultados que obtuvimos al analizar las variables *tipo de centro* y *zona geográfica* parecían estar influidos por la variable *lengua materna*, cuya incidencia acabamos de analizar. Como vimos, la diferencia que se observa, tanto en el promedio general como en la mayoría de las categorías semánticas, en la productividad léxica en función de las variantes “público/privado” y “urbano/rural” carece de relevancia estadística. No obstante, también se aprecia cómo los resultados son mejores, en el corpus castellano, entre los informantes de las instituciones privadas frente a los de los centros públicos, y entre los alumnos de los centros educativos ubicados en la capital frente a los de los situados en distintos puntos de la provincia. En catalán, se da la tendencia inversa y los estudiantes procedentes de centros públicos y de instituciones situadas fuera de la capital obtienen mejores promedios. Como ya apuntamos entonces, en la distribución de la muestra, hay una mayor presencia de informantes cuya lengua materna es el catalán en los centros públicos que en los privados, y en los rurales frente a los urbanos⁸¹⁰. A la vista de la discriminación que ejerce la variable *lengua materna*, las diferencias que se observan y que acabamos de comentar, tanto en el corpus castellano como en el catalán, se deben, con mucha probabilidad, a la incidencia de este factor, que actúa favoreciendo siempre al alumnado que completa la encuesta en su propia L1⁸¹¹.

⁸⁰⁹ Hernández Muñoz *et al.* (2014) analizan los factores cognitivos que intervienen en el español como segunda lengua y observan cómo estos son distintos a los que influyen como primera lengua. En el caso de una segunda lengua parece determinante la edad de adquisición de la L2, la tipicidad y el grado de cognición. En su estudio, Hernández Muñoz *et al.* (2014) utilizan una muestra de 43 estudiantes de español como lengua extranjera que son hablantes nativos de inglés. Es necesario indicar esta circunstancia ya que, a diferencia del presente trabajo donde todos los informantes son bilingües desde la infancia y donde la segunda lengua cumple una función social en la comunidad en la que viven, en su caso los informantes son adultos que han aprendido el español entre los 17 y los 23 años.

⁸¹⁰ Recordemos que, de cualquier modo, los catalanohablantes son mayoría en todos los subgrupos (*vid.* 6.2.2. y 6.2.3.).

⁸¹¹ No se observa, por el contrario, que las variables *nivel sociocultural* o *sexo*, que también han mostrado una discriminación estadística en los resultados, tengan relación con la distribución de informantes según las variables *zona geográfica* o *tipo de centro*.

De otra parte, también es interesante recordar ahora las diferencias que se aprecian en la producción comparando nuestra sintopía con otras del mundo hispánico, tanto monolingües como bilingües. Si recordamos los datos del apartado 6.1.7. (gráficos 6.10 y 6.11), el promedio general obtenido en lengua castellana (17.90) está claramente por debajo de la media alcanzada en otras comunidades, tanto en las que se maneja un solo código lingüístico, donde se observa una diferencia con respecto a la media de 1.55 respuestas, como aquellas, como la nuestra, donde coexisten dos sistemas lingüísticos, donde la distancia es aún mayor, de 1.9 palabras⁸¹². Sin embargo, como también predijimos entonces, nuestros resultados indican que la mayoría de la muestra tiene una lengua materna distinta a la de la encuesta, hecho que solo sucede en otro territorio como es Bilbao. Por eso, estábamos en lo correcto cuando consideramos que los resultados hubieran sido mejores si hubiéramos tenido en cuenta, únicamente, los datos de los castellanohablantes en su propia lengua materna. En tal caso, la diferencia con respecto a la media no alcanzaría la unidad con respecto a los territorios monolingües (sería de 0.93 respuestas) y la superaría levemente en el caso de los bilingües (sería de 1.28). Nuestros resultados, situados en 18.52, aunque seguirían por debajo de la media general de comunidades monolingües y bilingües, superarían ahora, además de a Málaga y a Melilla, los datos recabados en la provincia de Granada. De modo inverso, si hubiéramos valorado solo los resultados aportados por los catalanohablantes —situados en 19.08—, al comparar nuestros datos con los registrados en Valencia —que alcanzan las 18.30 palabras de promedio—, continuaríamos superando a los valencianos, aunque con una distancia mayor a la que se observa considerando todos los datos (0.78 frente a 0.3), lo cual parece lógico teniendo en cuenta que la extensión social del catalán es mucho mayor en nuestra provincia que en la de Valencia.

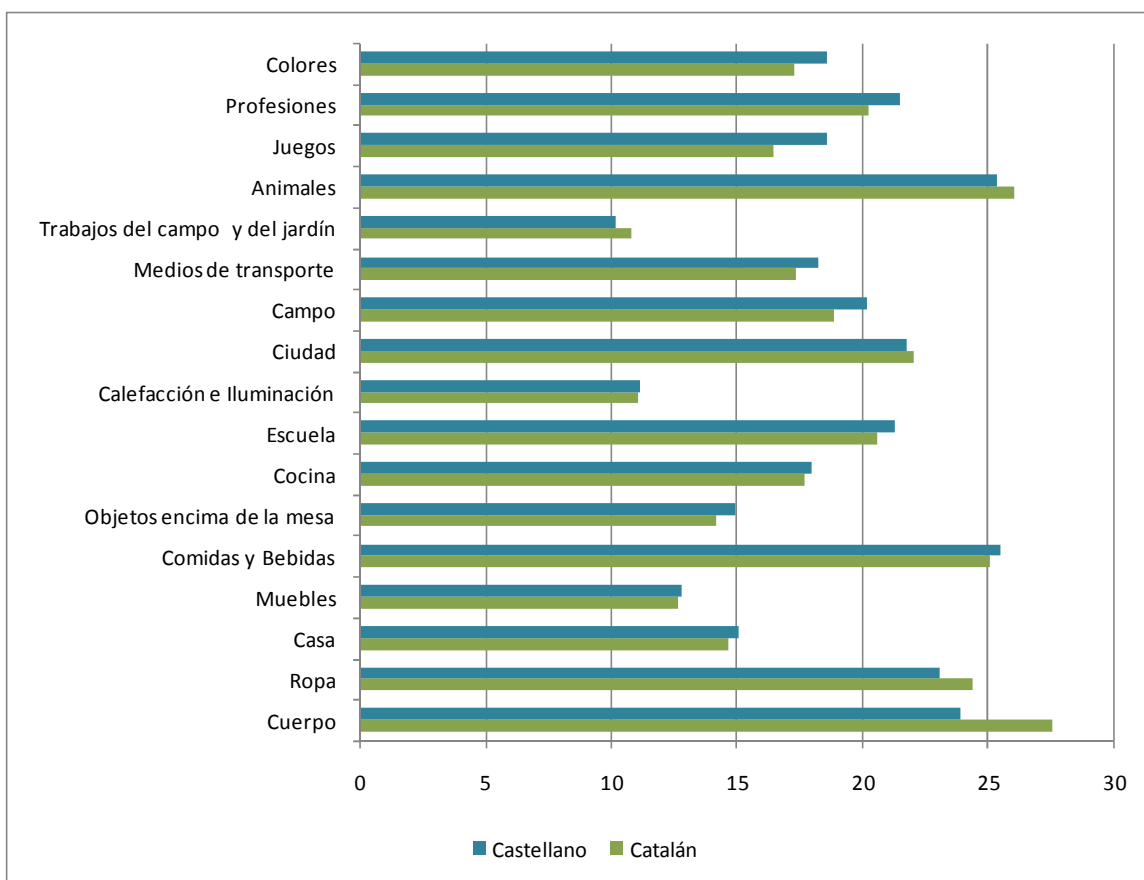
La gran relevancia que tiene la variable *lengua materna* en nuestra investigación está en consonancia con los estudios de disponibilidad realizados hasta la fecha en territorios bilingües que la han incluido como variable. Generalmente, este factor ha mostrado, como también se ve en nuestros datos, que los informantes que completan la encuesta en su propia lengua materna obtienen mejores resultados, ya sea en castellano (Blas y Casanova 2003b, Ayora 2006, Fernández Smith *et al.* 2008 o Areta 2009) o en catalán (Gómez Devís 1995, Llopis 2009). La única excepción sería precisamente nuestro trabajo exploratorio en la ciudad de Lleida (Serrano Zapata 2003b y 2004)

⁸¹² En el primer caso, la media se sitúa en 19.45 y, en el segundo, en 19.80.

donde precisamente el alumnado con lengua materna catalana obtenía mejores resultados en el léxico disponible en castellano. La realización del presente análisis, con un volumen mayor de informantes —que integran a los 48 que usamos en aquella ocasión—, y el uso de paquetes estadísticos nos han permitido llegar a resultados contrarios, más acordes con la realidad⁸¹³.

A continuación, vamos a analizar las diferencias de cada colectivo lingüístico según la lengua en que completa la encuesta. Es decir, comparamos los resultados que obtienen los cuatro colectivos lingüísticos al aplicarles la prueba en castellano o en catalán. Integramos, en este análisis, a los colectivos cuya lengua materna es el aranés o las dos lenguas oficiales de la comunidad, a pesar de que, como ya hemos advertido, sería deseable contar con más representantes de estos dos grupos. Empezamos, precisamente, por estos informantes.

Gráfico 6.37. Comparación de la producción en castellano y en catalán de los informantes cuya lengua materna es el castellano y el catalán (ambas lenguas).

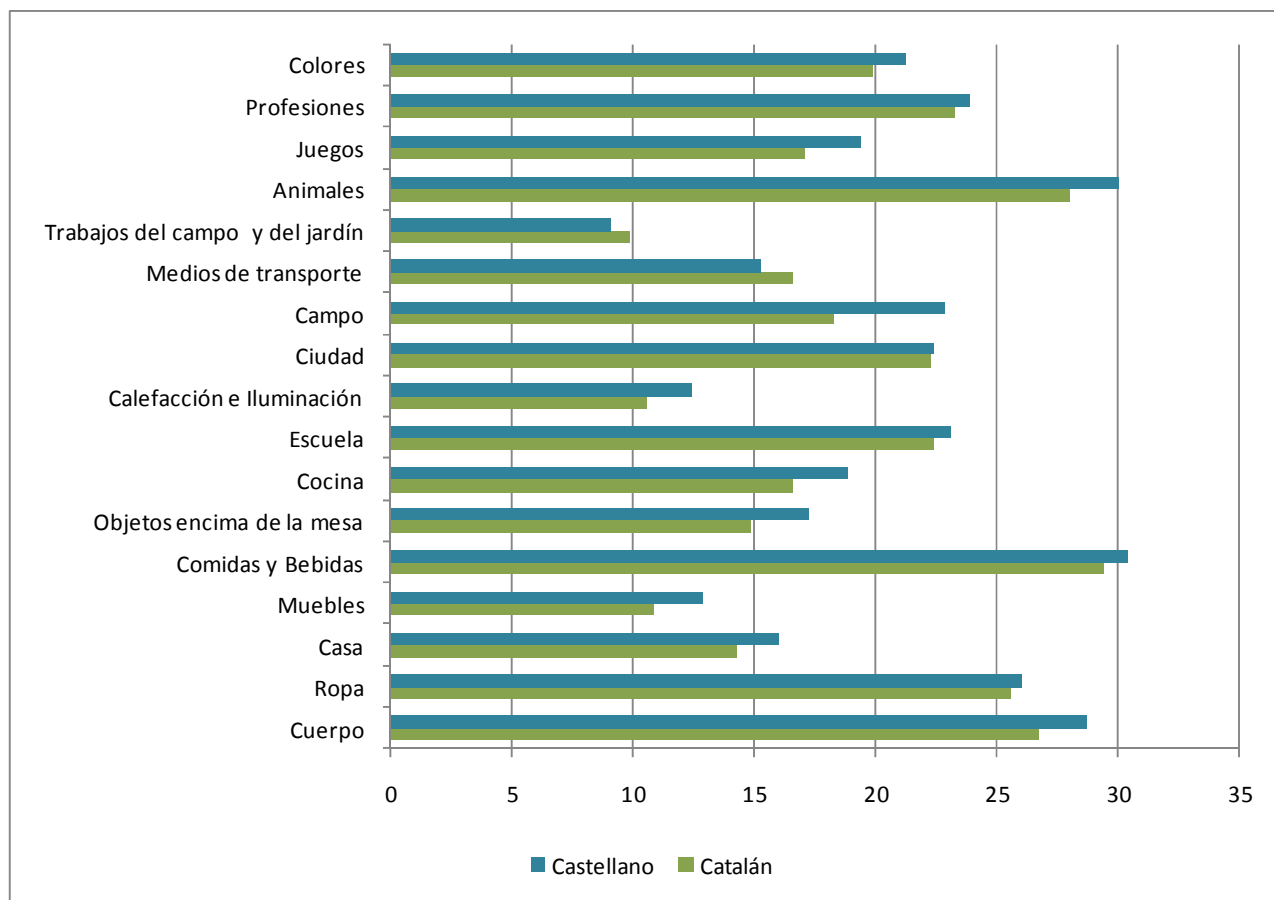


⁸¹³ Otra variable frecuentemente analizada en comunidades bilingües es la *lengua habitual* que, salvo algunas excepciones, no parece influir demasiado sobre la productividad, al menos comparándola con otras variables (*vid.* 3.4.3).

Tal como se aprecia en el gráfico anterior (gráfico 6.37), la producción en las dos lenguas del colectivo que declara tener ambas como maternas es bastante similar. De hecho, el promedio global de los dos idiomas (18.85 en castellano frente a 18.68 en catalán) es el más cercano de los cuatro grupos lingüísticos. Sin embargo, se observa una ligera preferencia hacia el castellano, puesto que, de los diecisiete campos semánticos, en doce los resultados son mejores en esa lengua, aunque hay que reconocer que en muchas áreas (v.gr. *Muebles de la casa, La cocina y sus utensilios, Calefacción, iluminación y formas de airear un recinto...*) la diferencia es de solo décimas. La preferencia hacia el castellano se justifica, a nuestro modo de ver, porque como ya hemos comentado, la mitad de los informantes que forman este grupo se decantan por el castellano como lengua de uso habitual.

Los centros en los que la productividad es superior en catalán son: *El cuerpo humano, Ropa y complementos, La ciudad, Trabajos del campo y del jardín y Animales*. Salvo la notable distancia de *El cuerpo humano*, donde en catalán se enumeran casi 4 palabras más (3.7) y la de *Ropa y complementos*, donde se sobrepasa la unidad (1.3), en las restantes áreas las diferencias no son destacables. En las categorías donde los resultados favorecen al castellano, la mayor disimilitud se detecta en *Juegos y distracciones*, con 2.1 respuestas. En las restantes áreas, solo en tres se supera la unidad (*El campo, Profesiones y oficios y Colores*), toda vez que, como hemos dicho, los resultados apuntan antes hacia una similitud entre la producción de las dos lenguas que hacia divergencias considerables.

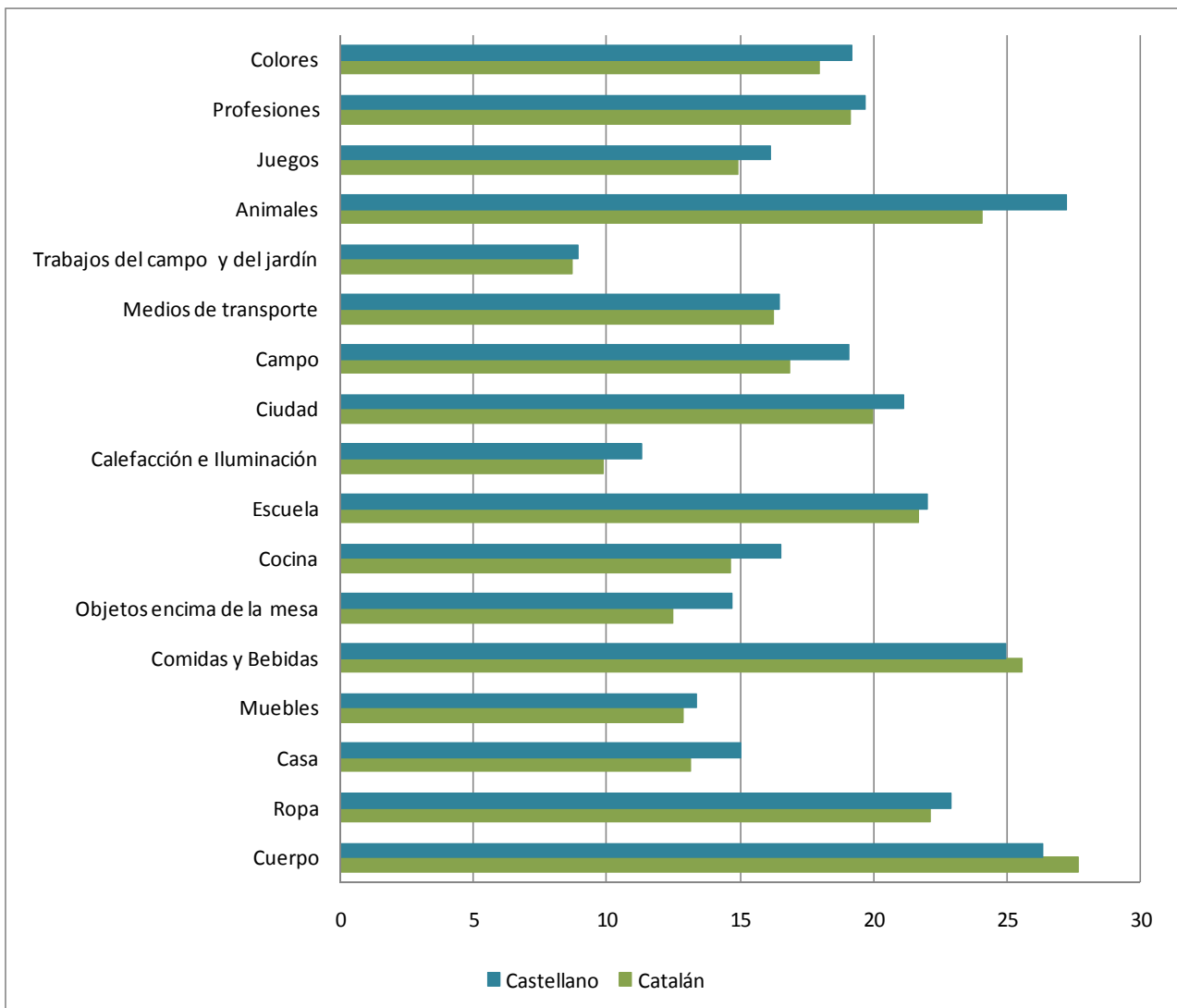
Gráfico 6.38. Comparación de la producción en castellano y en catalán de los informantes cuya lengua materna es el aranés.



El colectivo que tiene como lengua materna el aranés, cuyos resultados, como vimos, eran muy buenos en comparación con los restantes colectivos, se decanta hacia el castellano, puesto que obtiene, en esta lengua, mejores promedios en prácticamente todas las áreas (gráfico 6.38). Las únicas excepciones son *Medios de transporte* y *Trabajos del campo y del jardín*, donde por 1.28 respuestas y 0.72 respectivamente, los resultados son más favorables en lengua catalana. No obstante, las diferencias que se observan en sentido contrario, es decir cuando en castellano se consiguen cifras más elevadas, son notablemente superiores. La más destacable es la de *El campo*, con 4.57 palabras más en español; de modo similar, en *El cuerpo humano*, *Muebles de la casa*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios*, *Animales y Juegos y distracciones* se superan las dos unidades de diferencia y en *Partes de la casa*, *Comida y bebidas*, *Calefacción, iluminación y formas de airear un recinto* y *Colores* se sobrepasa la unidad. Solo en las cuatro áreas restantes —*Ropa y complementos*, *La escuela*, *La ciudad* y *Profesiones y oficios*— la disimilitud es de tan solo unas décimas siempre a favor del castellano. A pesar de la escasa representatividad de este colectivo,

nuestros datos parecen corroborar, en cierta medida, la polarización que explicamos que existía en *la Vall d’Aran* en torno a las lenguas más usadas por la población (*cfr.* 4.3.2, gráfico 4.28). Como explicamos en aquella ocasión, las lenguas preferidas en el territorio son el castellano y el aranés, mientras que el catalán queda relegado a una tercera posición. También es destacable, de otra parte, la presencia de otras lenguas, distintas a las tres mencionadas, en esa parte del territorio.

Gráfico 6.39. Comparación de la producción en castellano y en catalán de los informantes cuya lengua materna es el castellano.



En la comparación de la producción de los castellanohablantes en las dos lenguas (gráfico 6.39), se detecta, como era de esperar, una mayor productividad léxica en la lengua materna de los informantes. Concretamente, los resultados de este colectivo son superiores en castellano en todos los centros, excepto en *El cuerpo humano* y en

Comidas y bebidas, donde —sorprendentemente— los informantes alcanzan mejores promedios en catalán; en el primer centro con una diferencia de 1.33 y, en el quinto, de 0.61. Salvo en estos dos casos, en las restantes áreas los promedios son siempre superiores en castellano. Es llamativa la diferencia que se aprecia en *Animales*, donde los informantes actualizan de media 3.14 palabras más en su lengua materna. También en *El campo* la divergencia es destacable, de más de dos unidades (2.22), al igual que en *Objetos colocados en la mesa para la comida* (2.2), un centro que, como hemos ido repitiendo, se relaciona estrechamente con la lengua utilizada en el hogar. La misma influencia se observa en *Partes de la casa y La cocina y sus utensilios*, donde la diferencia casi alcanza las dos respuestas (1.88 y 1.86). Solo en cuatro centros más, las disimilitudes superan la unidad (*Calefacción, iluminación y formas de airear un recinto, La ciudad, Juegos y distracciones y Colores*), siempre a favor de la lengua castellana.

La diferencia entre el promedio global que los castellanohablantes tienen en castellano (18.52) y en catalán (17.53) es, prácticamente, de una respuesta. Debido a la importancia de este último dato, analizamos esta diferencia con el test T de Student (tablas 6.39, 6.40 y 6.41) y comparamos la media del grupo cuya lengua materna es el castellano en las dos lenguas. De las tablas que incluimos a continuación, se concluye que la divergencia hallada es estadísticamente significativa, a favor de la lengua castellana, puesto que el valor de ‘p’ se sitúa en 0.001, es decir, por debajo de 0.05.

Tabla 6.39. Estadísticos de muestras relacionadas (informantes castellanohablantes).

		Media	N	Desviación típica	Error típico de la media
Par 1	Promedio	18.5249	66	3.138	.386
	Promedio	17.5151	66	3.217	.396

Tabla 6.40. Correlaciones de muestras relacionadas (informantes castellanohablantes).

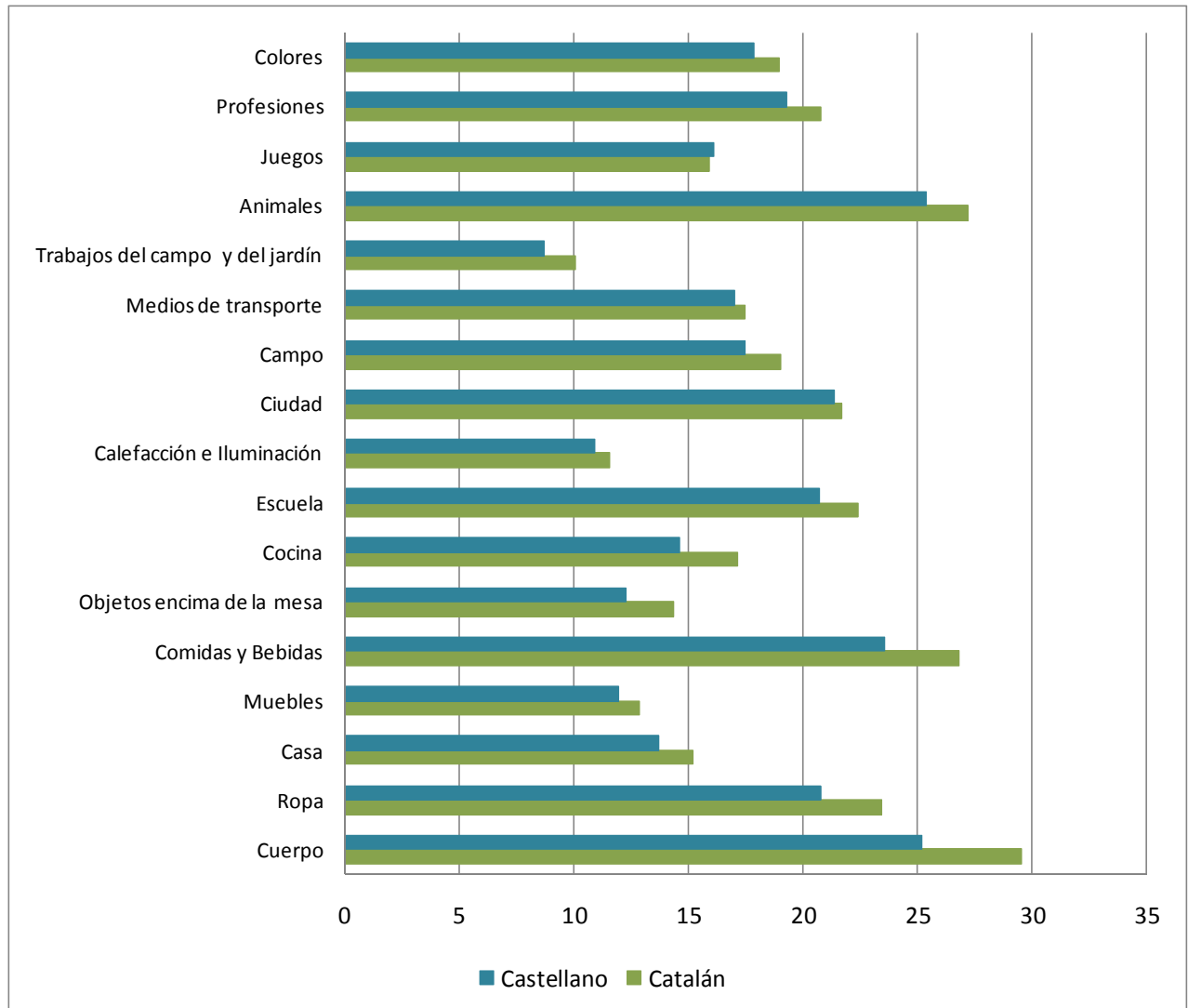
		N	Correlación	Sig.
Par 1	Promedio y Promedio	66	.722	.000

Tabla 6.41. Prueba de muestras relacionadas (informantes castellanohablantes).

	Diferencias relacionadas					t	gl	Sig. (bilateral)
	Media	Desviación típica	Error típico de la media	95% Intervalo de confianza para la diferencia				
				Inferior	Superior			
Par Promedio - 1 Promedio	1.009	2.370	.291	.427	1.592	3.460	65	.001

Finalmente, analizamos los resultados del colectivo mayoritario en la población estudiada, esto es, el que tiene el catalán como lengua materna. Su actuación lingüística es la que presenta una mayor diferencia en función de la lengua de la encuesta (gráfico 6.40). Los promedios en todos los centros de interés son favorables al catalán salvo en *Juegos y distracciones*, donde por 0.17 décimas, la media es superior en castellano. La diferencia más acusada es la de *El cuerpo humano*, con 4.36 respuestas. A esta cifra, sigue la que se detecta en *Comidas y bebidas*, de 3.29 respuestas, y las de los centros *Ropa y complementos*, *Objetos colocados en la mesa para la comida* y *La cocina y sus utensilios* —áreas relacionadas con el entorno inmediato y la lengua del hogar— donde se superan siempre las dos respuestas. En muchas más se sobrepasa la unidad (*Partes de la casa*, *La escuela*, *El campo*, *Trabajos del campo y del jardín*, *Animales*, *Profesiones y oficios* y *Colores*) y, en cambio, son minoría las que no llegan a sobrepasar este límite (*Muebles de la casa*, *Calefacción, iluminación y formas de airear un recinto*, *La ciudad* y *Medios de transporte*).

Gráfico 6.40. Comparación de la producción de los informantes cuya lengua materna es el catalán.



Como ya es sabido, para poder extrapolar estas diferencias al universo estudiado y comprobar cuáles resultan significativas, debemos acompañar este análisis, puramente descriptivo, de las pruebas estadísticas correspondientes. Por este motivo, al igual que con el colectivo castellanohablante, comparamos las medias de los catalanohablantes en los dos idiomas (17.46 en castellano y 19.08 en catalán). Tal como se observa en las tablas que siguen (tablas 42, 43 y 44), al estudiar el comportamiento concreto de este colectivo en las dos encuestas, se obtiene un valor de $p = 0.000$, esto es, por debajo de 0.05, el límite de la significación estadística. Este valor nos indica que la lengua en que se completa la encuesta es un factor que determina el número de respuestas que actualizan los informantes.

Tabla 6.42. Estadísticos de muestras relacionadas (informantes catalanohablantes).

		Media	N	Desviación típ.	Error típ. de la media
Par 1	Promedio	17.464	157	2.709	.216
	Promedio	19.083	157	3.044	.242

Tabla 6.43. Correlaciones de muestras relacionadas (informantes catalanohablantes).

		N	Correlación	Sig.
Par 1	Promedio y Promedio	157	.725	.000

Tabla 6.44. Prueba de muestras relacionadas (informantes catalanohablantes).

	Diferencias relacionadas					t	gl	Sig. (bilateral)
	Media	D. típica	Error típico de la media	95% Intervalo de confianza para la diferencia				
				Inferior	Superior			
Par 1 Promedio - Promedio	-1.618	2.157	.172	-1.959	-1.278	-9.402	156	.000

A la luz de estos resultados y conociendo que la diferencia entre los colectivos, al menos en sus promedios globales, tiene significación estadística, vale la pena detenernos, de nuevo, en las divergencias que se aprecian según el grupo lingüístico analizado⁸¹⁴. De un modo similar a como veíamos que ocurría al analizar la influencia de la lengua materna en cada corpus por separado, se aprecia que la diferencia entre el caudal léxico de los catalanohablantes en las dos lenguas es mayor (1.62) a la que se observa entre el colectivo castellanohablante (0.99), de la misma forma que el promedio del grupo catalán en su propia lengua materna es superior (19.08) al que alcanzan los castellanohablantes en la suya (18.52). Una vez más, creemos que la causa de esta diferencia radica en que el grupo catalán ha sido escolarizado en su propia lengua materna, razón que explica que alcance un resultado superior al realizar la encuesta en esta lengua, puesto que parte del léxico que aparece en la prueba de disponibilidad se consolida en la escuela. Por el contrario, el grupo castellanohablante no alcanza un promedio tan elevado en su primera lengua, aunque llega a un nivel de dominio del léxico disponible de los dos idiomas más similar.

⁸¹⁴ A partir de este momento, nos centramos únicamente en los resultados de los dos grupos mayoritarios: castellanohablantes y catalanohablantes.

Tras analizar los resultados área por área según cada colectivo lingüístico, entendemos mejor algunas de las diferencias que hallamos cuando analizamos los datos en global de la producción lingüística en los dos idiomas (*vid.* 6.1.2 y 6.1.3.). Como vimos entonces, la diferencia mayor entre la producción léxica se daba en las categorías *El cuerpo humano*, *Comidas y bebidas* y *Ropa y complementos*, por este orden. Con los datos actuales sobre la mesa, podemos concluir que los dos primeros centros tienen un promedio mejor en catalán porque, incluso el colectivo castellanohablante, presenta mejores resultados en catalán que en castellano, suponemos que debido al peso de la escolarización en lengua vernácula. Sin embargo, si comparamos la diferencia en los promedios en estas dos categorías, se ve claramente cómo es, principalmente, el colectivo catalanohablante el que ostenta una cifra más elevada en su lengua materna que en castellano (con una diferencia de 4.36 palabras en *El cuerpo humano* y de 3.29 en *Comidas y bebidas*). Creemos que, en este caso, puede contribuir a esta divergencia tanto el hecho de que el léxico de estas áreas, como hemos indicado repetidamente, se aprende tempranamente, como que, sobre todo el de *El cuerpo humano*, se trabaja durante las distintas etapas escolares.

La situación es del todo distinta con respecto al centro 2, *Ropa y complementos*. A pesar de que, como vimos (*vid.* 6.1.2., tabla 6.6.), la distancia entre la producción en castellano y en catalán, a favor del catalán, era una de las mayores, al analizar los datos por colectivos lingüísticos se observa cómo esta diferencia concierne únicamente al colectivo catalanohablante, puesto que el castellanohablante obtiene un resultado superior en su propia lengua materna. También llama la atención que un centro como *Animales*, cuya producción en las dos lenguas, en los datos globales, es muy similar (26.02 en castellano y 26.32 en catalán), al desglosar los resultados por grupos lingüísticos, cada colectivo tiene, claramente, más respuestas en su primera lengua, sobre todo el alumnado castellanohablante, que actualiza 3.14 palabras más. Esta diferencia es llamativa si tenemos en cuenta que esta área tiene una gran presencia en el currículum escolar, que es seguido, como ya hemos indicado repetidamente, íntegramente en catalán. Una posible explicación a este comportamiento estaría, de nuevo, en la edad de adquisición del vocabulario de esta categoría, uno de los factores cognitivos que intervienen en la disponibilidad. Recordemos que Hernández Muñoz (2006) demuestra que el léxico de este centro se aprende antes que el de otros, como *El*

cuero humano, por ejemplo, donde hemos visto que los castellanohablantes aportan más palabras en su segunda lengua⁸¹⁵.

El hecho de que cada colectivo ostente mejores promedios en su propia lengua materna también se observa en algunas otras áreas, además de en *Animales*, como en *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios* o *El campo*. Sin embargo, en otras se observa un comportamiento dispar: mientras uno de los dos colectivos se decanta claramente por su lengua materna, el otro obtiene un resultado aproximado en los dos idiomas. Esto sucede por ejemplo, en *La escuela* o en *Trabajos del campo y del jardín*, donde los castellanohablantes tienen mejores promedios en castellano pero por muy pocas décimas y, en cambio, los catalanohablantes muestran de forma más firme la superioridad léxica en su propia lengua materna. De modo contrario, los castellanohablantes se decantan claramente por su primera lengua en *Juegos y distracciones*, donde los catalanohablantes también ostentan un promedio más alto en esta lengua, pero por décimas de diferencia. Las razones que explican estas divergencias son distintas en cada caso: en *La escuela* creemos que se debe a la presencia del catalán como lengua vehicular de la enseñanza en todos los niveles educativos, en *Trabajos del campo y del jardín*, a la preferencia por el uso de la lengua vernácula en los núcleos rurales, y como ya hemos señalado en diversas ocasiones, en *Juegos y distracciones* se observa la influencia del castellano en las distintas formas de ocio juvenil. No obstante, cabe recordar que este análisis por centros de interés es meramente descriptivo. Sería necesario evaluar los datos de todas las áreas mediante un análisis estadístico que nos indicase qué diferencias, de las enumeradas, son significativas y, por tanto, es necesario profundizar en las causas que las explican, asunto que dejamos para estudios posteriores.

Con respecto a la riqueza léxica, esto es, al número de vocablos recogido por cada grupo lingüístico en las dos lenguas, los datos se plasman, a continuación, en la tabla 6.45. Desde un punto de vista global, los grupos que muestran una riqueza mayor de vocabulario en lengua castellana son los castellanohablantes, de un lado, y los araneses, de otro, colectivo que, como hemos visto, también exhibe en castellano, de

⁸¹⁵ Sin embargo, cabe recordar que en esta misma área —*El cuerpo humano*— analizando el corpus castellano y el catalán por separado y distribuyendo a los informantes según la lengua materna, cada grupo obtiene mejores promedios en su propia lengua (*cfr. supra*, dentro de este mismo apartado).

forma constante, una mayor productividad léxica. Los resultados globales son favorables a la lengua catalana en el caso de los catalanohablantes y en el de los hablantes que eligen las dos lenguas oficiales como maternas, aunque por muy poca distancia en este último caso (únicamente de 45 vocablos). Hay que recordar que este grupo se compone de tan solo 10 informantes, aunque el grupo de los araneses es menor (7 informantes) y presenta, en el global, una diferencia de 141 palabras distintas más en castellano, lo que nos indica de nuevo, la preferencia de este colectivo por esta lengua.

Por encima de estas diferencias, lo que se aprecia en la tabla 6.45 es, como hemos ido repitiendo a lo largo de este apartado, la relación tan evidente que existe entre el número de informantes y el total de vocablos recabado en cada área. Por esta razón, al igual que sucedía en la variable *nivel sociocultural* (vid.6.2.4), siempre el grupo que ostenta un mayor número de vocablos es el catalanohablante (mayoría en la muestra), seguido del castellanohablante, el segundo grupo más importante numéricamente. A continuación, se sitúa el colectivo que tiene ambas lenguas como maternas, y el último lugar lo ocupa el grupo menos cuantioso, esto es, el colectivo aranés. Las únicas excepciones a esta tendencia se dan en *Partes de la casa*, donde los araneses y los informantes con ambas lenguas como maternas registran el mismo número de vocablos en castellano (53) y en *Colores*, donde los resultados de los araneses son, tanto en castellano como en catalán, mejores que los del grupo de ambas lenguas, aunque por poca diferencia.

Tabla 6.45. Total de vocablos por centro de interés según la variable social lengua materna.

Nº	Centro de interés	LM: Castellano		LM: Catalán		LM: Ambas		LM: Aranés	
		CAST	CAT	CAST	CAT	CAST	CAT	CAST	CAT
1	CUERPO	239	212	304	279	104	120	84	80
2	ROPA	197	186	233	259	98	111	73	71
3	CASA	143	136	190	215	53	70	53	42
4	MUEBLES	157	159	208	219	51	57	42	37
5	COMIDAS	323	324	377	441	142	146	120	105
6	OBJETOS	152	155	192	223	72	67	58	48
7	COCINA	220	207	259	297	93	98	75	58
8	ESCUELA	234	228	330	376	107	97	82	70
9	CALEFACCIÓN	154	135	220	240	55	52	40	34
10	CIUDAD	356	334	512	505	125	138	79	85
11	CAMPO	406	355	546	607	125	119	102	77
12	TRANSPORTE	152	164	223	248	65	63	49	47
13	TRABAJOS	221	177	311	321	65	62	41	35
14	ANIMALES	232	262	284	364	113	122	97	90

15	JUEGOS	309	270	455	452	116	108	91	77
16	PROFESIONES	342	334	477	466	128	130	97	91
17	COLORES	143	139	193	194	51	48	54	49
Global		3980	3777	5314	5706	1563	1608	1237	1096

Si comparamos la actuación de cada colectivo en las dos lenguas, observamos cómo, al igual que sucedía con los promedios de palabras, los castellanohablantes tienen mayor riqueza léxica en su lengua materna, tal como se manifiesta tanto en el total (donde se da una diferencia de 203 vocablos con respecto al catalán) como en la mayoría de los centros de interés. Concretamente en doce, la cifra de vocablos es superior en castellano, mientras que en los cinco restantes (*Muebles de la casa*, *Comidas y bebidas*, *Objetos colocados en la mesa para la comida*, *Medios de transporte* y *Animales*), se obtiene una cantidad mayor en catalán⁸¹⁶. Las mayores disimilitudes, a favor siempre del castellano, se dan en *El campo* (de 51 vocablos), en *Trabajos del campo y del jardín* (de 44) y en *Juegos y distracciones* (de 39).

En cuanto a los catalanohablantes, estos también actualizan, con una diferencia de 392 unidades, más vocablos en su lengua materna, lo que se plasma en todas las áreas salvo en *El cuerpo humano*, *La ciudad*, *Juegos y distracciones* y *Profesiones y oficios*, donde por muy pocos vocablos —excepto en el primero de los centros de interés curiosamente⁸¹⁷— tienen mejores resultados en castellano. La superioridad de vocabulario en su lengua materna se aprecia, en mayor grado, en *Animales*, con 80 palabras distintas más⁸¹⁸, en *Comidas y bebidas*, con 64 y en *El campo*, con 61.

Por otro lado, el colectivo que opta por las dos lenguas como maternas, tiene un total de vocablos superior en lengua catalana, aunque con una distancia de solo 45 palabras. En general, al igual que la producción en las dos lenguas era prácticamente similar (con una diferencia de 0.17 en los promedios), también aquí hallamos que en

⁸¹⁶ No obstante, es justo indicar que salvo en *Medios de transporte* y en *Animales*, en las restantes áreas se observan diferencias insignificantes de 1, 2 ó 3 vocablos. De otra parte, es sorprendente que en *Animales* los castellanohablantes tengan más vocablos en catalán (30 más), ya que —como hemos visto— precisamente este colectivo se decanta claramente hacia su lengua materna, con una diferencia de más de 3 respuestas, en la productividad en torno a esta área. Quizá esta diferencia se deba a la producción de fenómenos de contacto lingüístico, dado que este centro parece propiciarlos (*cfr.* 7.3.).

⁸¹⁷ Otro dato llamativo en torno al centro 1, *El cuerpo humano*, es que tanto castellanohablantes como catalanohablantes tienen más vocablos en castellano, pero mejores promedios de palabras en catalán.

⁸¹⁸ Nótese cómo en este centro se obtienen más vocablos en catalán, independientemente del colectivo que estudiemos. La abundante presencia de transferencias lingüísticas en torno a esta área puede explicar este hecho, tal como analizamos en el capítulo siguiente (*vid.* 7.3.).

algunas áreas el número de vocablos es superior en castellano (*Objetos colocados en la mesa para la comida, La escuela, Calefacción, iluminación y formas de airear un recinto, El campo, Medios de transporte, Trabajos del campo y del jardín, Juegos y distracciones y Los colores*) y en otros, en catalán (*El cuerpo humano, Ropa y complementos, Partes de la casa, Muebles de la casa, Comidas y bebidas, La cocina y sus utensilios, La ciudad, Animales y Profesiones y oficios*). Las cifras, en general, se aproximan bastante, valga como ejemplo que la mayor diferencia es de 16 palabras (en *El cuerpo humano*) a favor del catalán⁸¹⁹.

Finalmente, en los 7 informantes cuya lengua materna es el aranés se detecta, de modo semejante a como ocurría al evaluar la productividad, una preferencia clara por el castellano, tanto en el promedio global (con 141 palabras de más) como en prácticamente todas las áreas analizadas. La única categoría donde los resultados son mejores en catalán es *La ciudad*, donde por 6 vocablos se supera la cifra obtenida en castellano. En las restantes, la cifra de palabras distintas recabadas es siempre mayor en castellano con una superioridad constante⁸²⁰.

Al comparar los centros que alcanzan las mejores y las peores posiciones con la clasificación general que obtuvimos de estos mismos datos (*vid.* 6.1.4.), se comprueba cómo hay un subgrupo —el de los catalanohablantes en catalán— cuya disposición de centros de interés coincide en todas las áreas con la clasificación general (*vid.* tabla 6.11), lo cual es, en parte, lógico puesto que es el grupo mayoritario. Por otro lado, tanto en los castellanohablantes como en los catalanohablantes, *El campo, La ciudad y Profesiones y oficios* encabezan la clasificación en ambas lenguas, exactamente igual que sucede en el ranquin general⁸²¹. En los grupos menos numerosos, el de ambas lenguas como maternas y el de los araneses, en cambio, lidera la clasificación, tanto en castellano como en catalán, *Comidas y bebidas*, quinto en el ranquin general. Por el contrario, *El campo* queda relegado a distintas posiciones: segundo en la producción de

⁸¹⁹ Téngase en cuenta que en varias categorías la diferencia a favor de un idioma u otro es de menos de 4 respuestas y que los informantes que forman este grupo son solo 10.

⁸²⁰ En *El campo*, se aprecia la mayor disimilitud (de 25 vocablos). Recuérdese al evaluar las diferencias entre estos dos últimos grupos que el tamaño tan reducido de la muestra incide en que los resultados puedan estar influidos por un informante atípico.

⁸²¹ Sin embargo, los centros que ocupan la cuarta y la quinta posición en la tabla general, *Juegos y distracciones* y *Comidas y bebidas*, también las ocupan en las dos clasificaciones de los catalanohablantes pero, en cambio, invierten su orden en las de los castellanohablantes.

los araneses en castellano e incluso sexto en la de los informantes de ambas lenguas en catalán. En general, mientras que en los dos grupos mayoritarios se aprecian las tendencias generales de la muestra, en los dos colectivos menores se observan muchas diferencias, lo cual se debe, muy probablemente, a la cercanía de las cifras, que provoca que por muy pocos vocablos se modifique la clasificación, o a la presencia de hablantes cuyo comportamiento lingüístico difiere del de la mayoría.

Tabla 6.46. Índice de cohesión por centro de interés según la variable social lengua materna.

Nº	Centro de interés	LM: Castellano		LM: Catalán		LM: Ambas		LM: Aranés	
		Castellano	Catalán	Castellano	Catalán	Castellano	Catalán	Castellano	Catalán
1	CUERPO	0.11	0.13	0.08	0.11	0.23	0.23	0.34	0.33
2	ROPA	0.12	0.12	0.09	0.09	0.24	0.22	0.36	0.36
3	CASA	0.11	0.10	0.07	0.07	0.28	0.21	0.30	0.34
4	MUEBLES	0.09	0.08	0.06	0.06	0.25	0.22	0.31	0.29
5	COMIDAS	0.08	0.08	0.06	0.06	0.18	0.17	0.25	0.28
6	OBJETOS	0.10	0.08	0.06	0.06	0.21	0.21	0.30	0.31
7	COCINA	0.07	0.07	0.06	0.06	0.19	0.18	0.25	0.29
8	ESCUELA	0.09	0.10	0.06	0.06	0.20	0.21	0.28	0.32
9	CALEFACCIÓN	0.07	0.07	0.05	0.05	0.20	0.21	0.31	0.31
10	CIUDAD	0.06	0.06	0.04	0.04	0.17	0.16	0.28	0.26
11	CAMPO	0.05	0.05	0.03	0.03	0.16	0.16	0.22	0.24
12	TRANSPORTE	0.11	0.10	0.08	0.07	0.28	0.28	0.31	0.35
13	TRABAJOS	0.04	0.05	0.03	0.03	0.16	0.17	0.22	0.28
14	ANIMALES	0.12	0.09	0.09	0.07	0.22	0.21	0.31	0.31
15	JUEGOS	0.05	0.06	0.04	0.04	0.16	0.15	0.21	0.22
16	PROFESIONES	0.06	0.06	0.04	0.04	0.17	0.16	0.25	0.26
17	COLORES	0.13	0.13	0.09	0.10	0.36	0.36	0.39	0.41

El último parámetro analizado, el de la cohesión, presenta una distribución por subgrupos lingüísticos influida también por el número de informantes (véase tabla 6.46). De este modo, el colectivo formado por menos miembros, el aranés, —que, por tanto, también obtiene un número menor de vocablos— alcanza un índice de cohesión superior al del colectivo que tiene ambas lenguas como maternas quien, a su vez, supera en todas las áreas y en los dos idiomas a los castellanohablantes, y estos, a los catalanohablantes, mayoría en la muestra. Al respecto, el único valor que no cumple esta condición es el de *El cuerpo humano*, que alcanza entre los castellanohablantes en castellano, el mismo valor (0.11) que entre los catalanohablantes en catalán⁸²². Por otro lado, igual que ocurría al analizar este parámetro en los distintos subgrupos según las

⁸²² No obstante, como ya comentamos (*vid.* 6.1.5.), este centro siempre es más cerrado en catalán que en castellano, tal como se observa en la tabla 6.46 al comparar los dos colectivos principales.

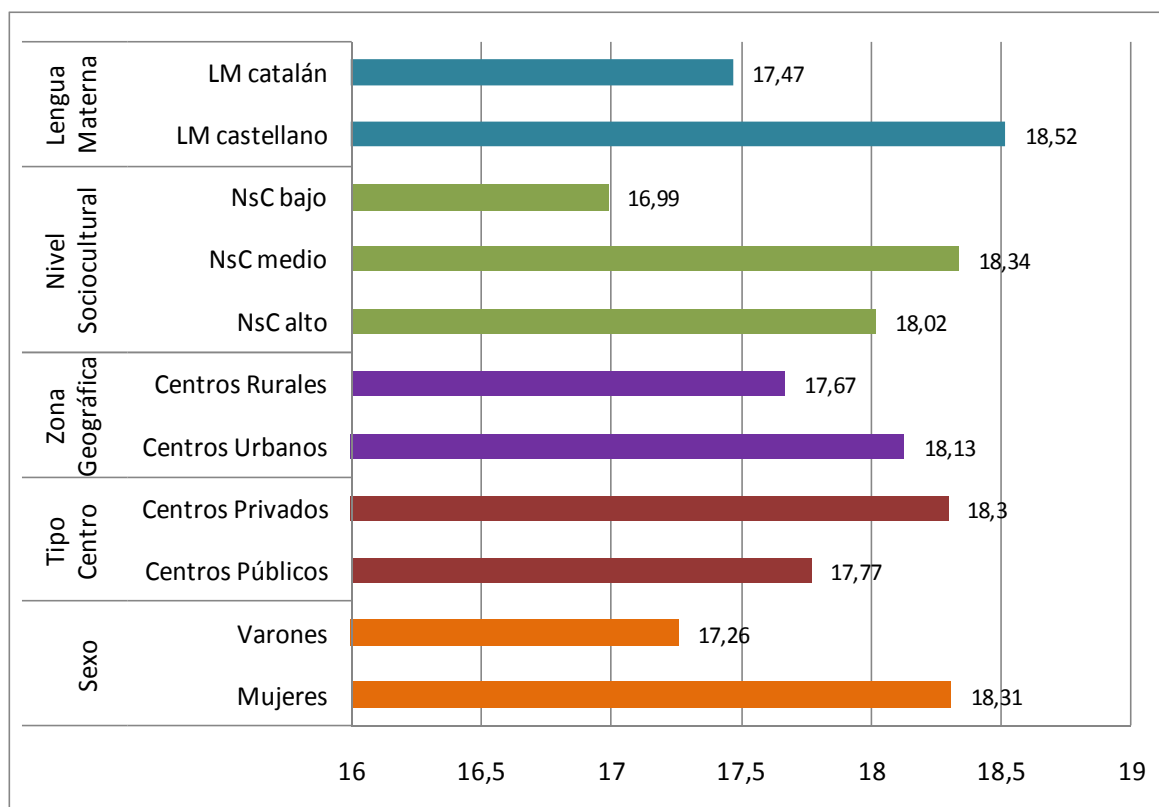
variantes sociales, hay numerosas coincidencias entre el valor que alcanza el índice en las dos lenguas, en este caso, sobre todo en los grupos mayoritarios, el castellanohablante y el catalanohablante, donde ocho áreas y trece, respectivamente, obtienen exactamente el mismo valor⁸²³. Con respecto a la ordenación de las áreas en un ranquin según este valor, hay que decir que la mayor parte de categorías, en las dos lenguas y en los cuatro subgrupos, coinciden con el orden general que otorga a unos centros la cualidad de abiertos o difusos y a otros, la de cerrados o compactos (*vid.* 6.1.5., tabla 6.14). Sin embargo, se aprecian más disimilitudes en los rangos de las áreas de los dos grupos menores, debido a que hay una mayor variabilidad en las cifras.

6.2.6. Valoración de los resultados según las variables sociales

Una vez analizadas las cinco variables independientes de este estudio, dada la ingente cantidad de datos manejados hasta el momento, consideramos adecuado realizar, como colofón a este apartado, un resumen tanto de las diferencias halladas según los distintos factores sociales en los dos corpus lingüísticos como de su relevancia estadística, así como de las divergencias constatadas en la producción en las dos lenguas de los distintos colectivos derivados de la muestra general. Empezamos por las diferencias que se observan, en el corpus castellano, según los cinco factores sociales contemplados.

⁸²³ El hecho de que en los colectivos integrados por menos miembros, el de ambas lenguas y el de los araneses, se aprecie una mayor variabilidad y, por ende, una menor coincidencia se explica porque al obtener valores más elevados del índice de cohesión, con dos decimales por encima de 0, es más improbable que haya coincidencia entre las cifras. Como ya dijimos (*vid.* 6.1.4), es difícil tanto establecer los rangos —puesto que muchas áreas coinciden—, como valorar las diferencias que tienen que ver con este parámetro —debido al redondeo de los dos decimales—.

Gráfico 6.41. Resumen de las diferencias en la producción en lengua castellana según las distintas variables sociales.



Tal como se deduce del gráfico anterior (gráfico 6.41), la distancia entre los promedios de cada grupo social es la siguiente:

- según la variable *sexo*: 1.05;
- según la variable *tipo de centro*: 0.53;
- según la variable *zona geográfica*: 0.46;
- según la variable *nivel sociocultural*: 1.35 (entre los niveles medio y bajo), 0.32 (entre los niveles medio y alto) y 1.03 (entre los niveles bajo y alto); y
- según la variable *lengua materna*⁸²⁴: 1.06.

De entre las cinco variables analizadas, destaca la importancia de las tres que se han revelado como significativas según el análisis efectuado mediante el SPSS, esto es, el *sexo*, la *lengua materna* y el *nivel sociocultural* (entre los niveles medio y bajo). Todas estas variables determinan una diferencia de más de una respuesta entre los

⁸²⁴ Prescindimos, en este resumen, de los resultados de los dos grupos lingüísticos menores (araneses e informantes con ambas lenguas como maternas) por su escasa representatividad y dado que no hemos sometido estos datos a ningún programa estadístico que los valide (*vid.* 6.2.5.). Si los incluyéramos, el promedio de los araneses (20.59) sería el más alto, y no el de los castellanohablantes, como afirmamos a continuación.

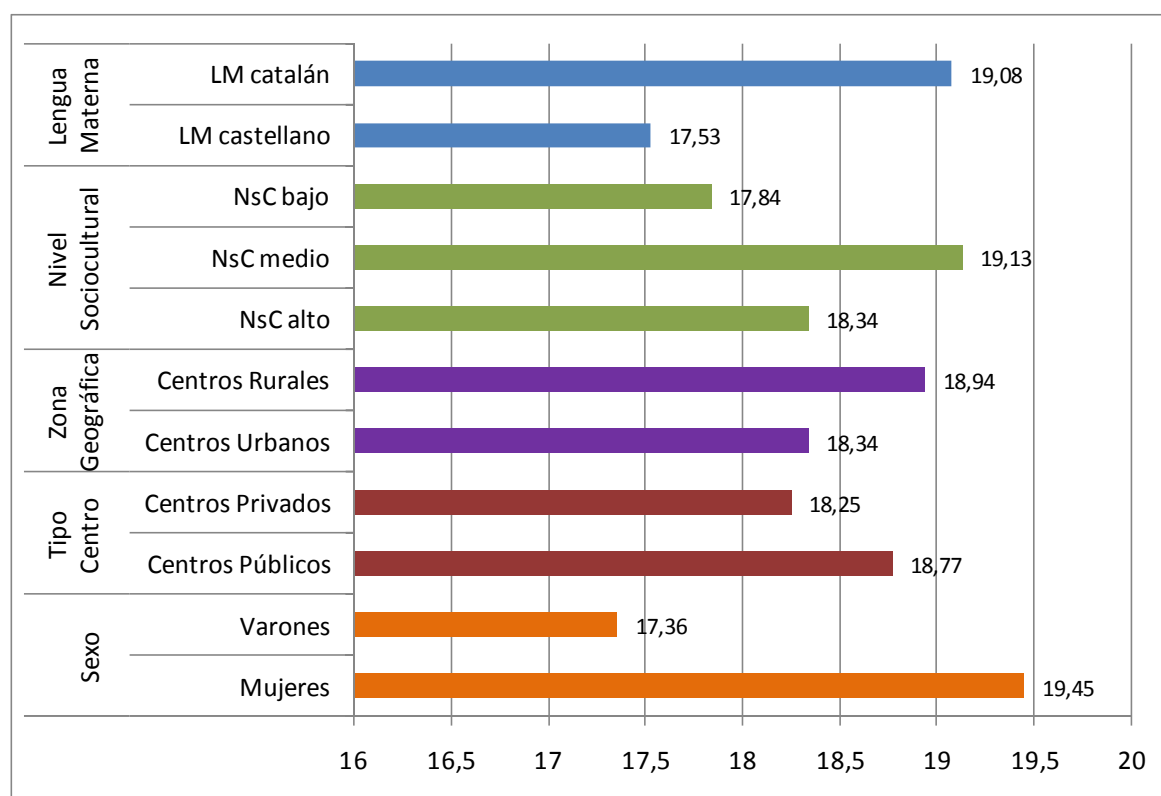
promedios generales de los distintos colectivos (varones/mujeres, estrato bajo/medio y castellanohablantes/catalanohablantes). Por el contrario, las dos variables que no han presentado significación estadística —el *tipo de centro* y la *zona geográfica*— presentan diferencias por debajo de la unidad, de aproximadamente media respuesta en ambos casos (0.53 en el *tipo de centro* y 0.46 en *zona geográfica*). La divergencia más acusada entre las distintos subgrupos sociales es la que efectúa el factor *nivel sociocultural*, entre los estratos medio y bajo (1.35); la disimilitud que producen los otros dos factores significativos, el *sexo* y la *lengua materna*, es prácticamente idéntica (1.05 y 1.06). El resultado más aproximado, en cambio, se aprecia entre los estratos medio y alto, que distan en solo 0.32 décimas. En general, los informantes cuya lengua materna coincide con la de la encuesta, esto es, los castellanohablantes, alcanzan el mejor promedio de todos, con 18.52 respuestas por centro de interés, lo que confirma, una vez más, la importancia de esta variable social. Por el contrario, los representantes del nivel social bajo consiguen el peor resultado con una media de 16.99.

A continuación, analizamos las diferencias que se observan en el corpus catalán, datos que se plasman en el gráfico 6.42. Con respecto a la distancia entre los promedios, los valores se corresponden con los siguientes:

- a) según la variable *sexo*: 2.09;
- b) según la variable *tipo de centro*: 0.52;
- c) según la variable *zona geográfica*: 0.6;
- f) según la variable *nivel sociocultural*: 1.29 (entre los niveles medio y bajo), 0.79 (entre los niveles medio y alto) y 0.5 (entre los niveles bajo y alto); y
- d) según la variable *lengua materna*⁸²⁵: 1.55.

⁸²⁵ Téngase en cuenta lo dicho en la nota anterior sobre la ausencia de los promedios de los araneses y los informantes con ambas lenguas como maternas. En la estratificación del corpus catalán, sin embargo, no habría variación con respecto al mejor y peor promedio aunque estos datos se incluyeran.

Gráfico 6.42. Resumen de las diferencias en la producción en lengua catalana según las distintas variables sociales.



Igual que en el corpus castellano, destaca la estratificación que producen las tres variables significativas —*sexo*, *nivel sociocultural* y *lengua materna*—. Los restantes factores sociales, por el contrario, no ejercen unas disimilitudes tan destacables y se perfilan, igual que en castellano, por debajo de la unidad (0.6 *zona geográfica* y 0.52 *tipo de centro*). Por el contrario, las tres variables significativas superan la unidad de diferencia en los promedios generales (1.29 entre los niveles medio y bajo y 1.55 entre catalanohablantes y catalanohablantes); especialmente llamativo es el caso de la variable *sexo*, donde las medias se distancian en dos respuestas (2.09) siendo, por ende, la divergencia mayor según algún factor social que se da en los dos corpus analizados⁸²⁶. De hecho, las mujeres son el colectivo con mayor productividad en catalán, con 19.45 palabras de media; mientras que precisamente los varones alcanzan el peor resultado, con 17.36 respuestas. Los grupos cuyas medias se acercan más entre sí son el nivel sociocultural bajo y el alto, separados por 0.5 décimas.

⁸²⁶ Recordemos que prescindimos, en este apartado, de los datos de araneses y hablantes de ambas lenguas, cuya significación estadística no hemos comprobado.

A la luz de estos resultados, llama la atención cómo la estratificación social es mayor en el corpus catalán que en el castellano. Como acabamos de comentar, las distancias entre los distintos grupos sociales en catalán son algo más acusadas (2.09 sexo, 1.55 lengua materna) mientras que en castellano son menores (1.05 sexo, 1.06 lengua materna.). Del mismo modo, la distancia entre los grupos con promedios más similares es menor en castellano que en catalán (0.32 frente a 0.5). Estos datos son aún más interesantes si recordamos que en el estudio de la variación en función de los centros de interés —esto es, según las variables lingüísticas—, también había una mayor variabilidad en el corpus catalán que en el castellano, al comparar en cada lengua —por separado— los centros más y menos productivos y con mayor y menor riqueza léxica (*vid.* 6.1.1, 6.1.2. y 6.1.4.). Estas dos circunstancias apuntan a que se da, en general, una mayor variación en el corpus catalán que en el castellano, tanto analizando los datos en función de las variables lingüísticas como en función de los factores sociales. Una posible explicación a este hecho puede residir, como ya hemos indicado en reiteradas ocasiones, en la distinta relación que tienen los dos grupos mayoritarios de la muestra (es decir, los que tienen como lengua materna el catalán o el castellano) con su segunda lengua. Como ya comentamos, el hecho de que los catalanohablantes refuercen el léxico de su propia lengua materna a través de la escuela provoca que alcancen mejores resultados en catalán que los castellanohablantes en castellano (*cf.* 6.2.5.).

Tabla 6.47. Valor de la 'p' en cada corpus según las variables sociales analizadas.

Variable social	Probabilidad Castellano	Probabilidad Catalán
SEXO	0.033	0.000
TIPO DE CENTRO	0.230	0.260
ZONA GEOGRÁFICA	0.223	0.142
N.SOCIOCULTURAL	0.000	0.018
LENGUA MATERNA	0.012	0.001

Tabla 6.48. Resumen de la significación de las diferencias halladas en cada uno de los dos corpus (número de centros de interés donde $p < 0.05$).

Variable social	Significación de las diferencias			
	Castellano		Catalán	
	Significativas	NO	Significativas	NO
SEXO	8	9	13	4
TIPO DE CENTRO	3	14	3	14
ZONA GEOGRÁFICA	4	13	3	14
N.SOCIOCULTURAL	5	12	6	11
LENGUA MATERNA	7	10	10	7

En cuanto a la valoración estadística de nuestros resultados, sintetizamos en las dos tablas anteriores (tabla 6.47 y tabla 6.48), la incidencia que han mostrado las cinco variables contempladas en esta investigación. Como ya hemos indicado a lo largo de este apartado, al analizar la diferencia entre los promedios globales se comprueba cómo tres factores extralingüísticos —*sexo*, *nivel sociocultural* y *lengua materna*— presentan una significación por debajo de 0.05 (tabla 6.47) y, por tanto, su influencia es extrapolable a la población estudiada. En el análisis concreto de las diecisiete áreas temáticas (tabla 6.48), se puede ver, una vez más, cómo la estratificación que ejercen las distintas variables es mayor en catalán que en castellano. En concreto, en el corpus catalán un total de 35 áreas presentan covariaciones significativas con algún factor social, mientras que en castellano esta cantidad se reduce a 27. En todas las variables sociales el número de áreas afectado es mayor en catalán, salvo en *zona geográfica*.

En conjunto, según el número de categorías semánticas que tienen inferencia estadística, hay dos factores sociales bastante significativos, el *sexo*, que influye en 8 áreas en castellano y en 13 en catalán, y la *lengua materna*, cuya incidencia se extiende a 7 centros en castellano y a 10 en catalán. Hay otra variable, el *nivel sociocultural*, que presenta una influencia media sobre el caudal léxico obtenido, y ejerce discriminación estadística en 5 y 6 áreas respectivamente en cada idioma. Finalmente, hay dos factores, *tipo de centro* y *zona geográfica*, cuyo valor significativo es menor, dado que inciden solo en la producción de 3 categorías (*tipo de centro*) y 4 en castellano y 3 en catalán (*zona geográfica*).

Tabla 6.49. Relación de las variables sociales que inciden en cada uno de los centros de interés en las dos lenguas analizadas.

Nº	Centros de interés	Diferencias significativas			
		Castellano	Nº	Catalán	Nº
1	CUERPO	NSC	1	NSC, LM	2
2	ROPA	SEXO, TC, ZG, LM	4	SEXO	1
3	CASA	TC, LM	2	TC, ZG, LM	3
4	MUEBLES	SEXO, TC, LM	3	SEXO, NSC	2
5	COMIDAS	SEXO, NSC	2	SEXO	1
6	OBJETOS	ZG, NSC, LM	3	SEXO, NSC, LM	3
7	COCINA	SEXO, LM	2	SEXO, LM	2
8	ESCUELA	SEXO, LM*	2	SEXO	1
9	CALEFACCIÓN		0	SEXO, TC, ZG, LM	4
10	CIUDAD	SEXO, ZG	2	SEXO, LM	2
11	CAMPO	NSC	1	SEXO, NSC, LM	3
12	TRANSPORTE	NSC	1	NSC	1
13	TRABAJOS	ZG	1	TC, ZG, LM	3
14	ANIMALES	SEXO, LM	2	SEXO, ZG, LM	3
15	JUEGOS		0	SEXO	1
16	PROFESIONES		0	SEXO, NSC, LM	3
17	COLORES	SEXO, LM	2	SEXO	1

Si analizamos ahora los centros de interés más permeables a la variación sociolectal en las dos lenguas, lo primero que constatamos es que las áreas que presentan una mayor covariación con los distintos factores extralingüísticos no coinciden. Con la excepción de *Objetos colocados en la mesa para la comida* —centro de interés con un elevado grado de inferencia estadística en ambas lenguas— son muchas las categorías que presentan valor significativo en catalán según varios factores sociales y, en cambio, escaso o nulo en castellano. Valga como ejemplo el caso de *Calefacción, iluminación y formas de airear un recinto, Profesiones y oficios* o, en menor medida, *El campo y Trabajos del campo y del jardín*. La situación inversa se observa en *Ropa y complementos*, que muestra relevancia estadística en cuatro variables en castellano y, en cambio, solo en una en catalán. También es destacable que en lengua catalana no haya ninguna área temática que no se vea influida por algún factor social; en castellano, por el contrario, esta situación sí se da en varias categorías (*Calefacción, iluminación y formas de airear un recinto, Juegos y distracciones y Profesiones y oficios*). Siendo conscientes de que cada lengua tiene una estratificación social distinta en función del área considerada, los centros que presentan mayor incidencia según algún factor extralingüístico en conjunto en los dos idiomas son *Objetos colocados en la*

mesa para la comida, Ropa y complementos, Partes de la casa, Muebles de la casa y Animales.

De otra parte, si clasificamos todas las áreas en función de las correlaciones estadísticas que se observan con las variables sociales, podemos establecer los siguientes grupos:

(a) Áreas que presentan una covariación social significativa con tres o cuatro factores sociales:

- *Objetos colocados en la mesa para la comida* en las dos lenguas,
- *La ropa y Muebles de la casa* en lengua castellana y
- *Partes de la casa, Calefacción, iluminación y formas de airear un recinto, El campo, Trabajos del campo y del jardín, Animales y Profesiones y oficios* en catalán.

(b) Áreas que presentan una discriminación estadística menor y que reciben la influencia de solo dos factores sociales:

- *La cocina y sus utensilios y La ciudad* en las dos lenguas,
- *Partes de la casa, Comidas y bebidas, La escuela*⁸²⁷, *Animales y Colores*, en castellano y
- *El cuerpo humano y Muebles de la casa* en catalán.

(c) Áreas influidas por uno o ningún factor social:

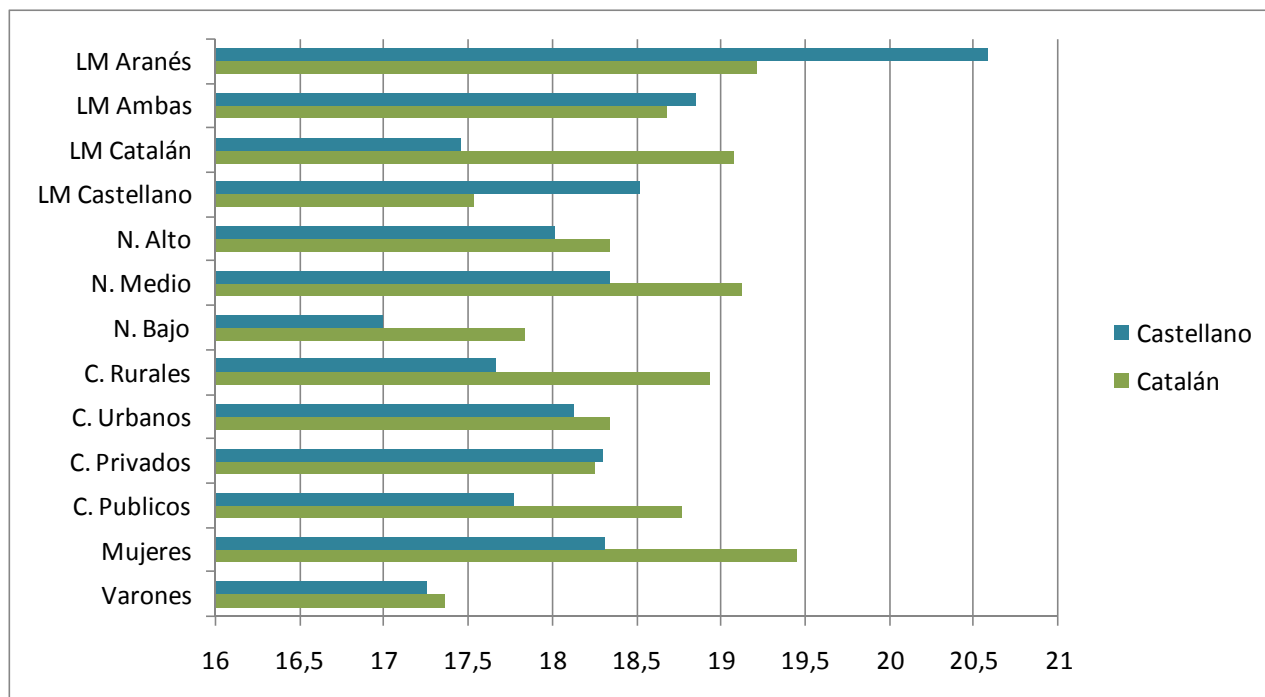
- *Medios de transporte* en ambas lenguas,
- *El cuerpo humano, Calefacción, iluminación y formas de airear un recinto, El campo, Trabajos del campo y del jardín, Juegos y distracciones y Profesiones y oficios*, en castellano y
- *Ropa y complementos, Comidas y bebidas, La escuela, Juegos y distracciones y Colores* en catalán.

En último lugar, pasamos a resumir las diferencias observadas en cada colectivo social con respecto a su producción de léxico disponible en las dos lenguas de la comunidad objeto de estudio. Es necesario aclarar, como ya hemos comentado a lo largo de este apartado, que estos datos no han sido sometidos a ninguna prueba de fiabilidad estadística que indique su alcance, con lo que solo podemos describir las tendencias constatadas sin extrapolar los resultados al conjunto del universo, con la

⁸²⁷ Como ya comentamos, este centro se halla en el límite de la significación estadística con respecto a la lengua materna ($p=0.05$), razón por la cual lo hemos marcado con un asterisco en la tabla 6.49.

única excepción de la variable *lengua materna*. Los resultados de los dos colectivos mayoritarios según este factor, es decir, el castellano hablante y el catalanohablante, han sido analizados mediante la prueba T de Student y han obtenido inferencia estadística en la diferencia global entre los promedios (*vid.* 6.2.5.).

Gráfico 6.43. Resumen de las diferencias en la producción en castellano y en catalán de cada grupo de informantes según las variables sociales.



Tal y como se observa en el gráfico anterior, la mayor parte de colectivos produce un caudal de léxico disponible superior en lengua catalana. Por el contrario, el grupo de informantes procedente de los centros privados y el que tiene el castellano, el aranés o ambas lenguas oficiales como maternas, exhibe una mayor cantidad de respuestas en lengua castellana. Es decir, se observan diferencias en función de la lengua de la encuesta según el factor social *lengua materna*, donde cada colectivo produce más palabras en su propia lengua materna y al agrupar a los informantes según el *tipo de centro*, donde —sin grandes distancias— los informantes de los centros concertados se decantan por el castellano mientras que los de las instituciones públicas lo hacen por la lengua vernácula. En las restantes variables, *sexo*, *zona geográfica* y *nivel sociocultural* todos los subgrupos muestran una mayor productividad de léxico disponible en catalán.

No obstante, las diferencias según las distintas variables son más o menos destacables. La mayor distancia entre los promedios es la que se observa en los catalanohablantes, que aportan 1.62 respuestas más al rellenar la encuesta en su lengua materna. Seguidamente, los hablantes de aranés, con una distancia de 1.38, muestran su preferencia por el castellano, como ya hemos comentado (*vid.* 6.2.5.). También por encima de la unidad, se sitúa la disimilitud entre los promedios, a favor del catalán en ambos casos, de los informantes de los centros rurales (1.27) y de las mujeres (1.14). Si bien no deja de sorprendernos la predilección tan marcada de este último colectivo en comparación con sus compañeros varones, la opción por el catalán de los informantes rurales se explica por la mayor presencia de alumnado cuya lengua materna es la catalana en este colectivo, como ya explicamos (*vid.* 6.2.3 y 6.2.5.). La diferencia entre la producción en cada una de las dos lenguas también alcanza justamente la unidad en el caso de los informantes procedentes de centros públicos y en el de los castellanohablantes (1 y 0.99, respectivamente). La relación entre las variables *lengua materna* y *tipo de centro* (y *zona geográfica*) justifican este resultado toda vez que, aunque los catalanohablantes imperan en todas las submuestras derivadas de la muestra general (salvo en las que emanan de la *lengua materna*, lógicamente), la presencia de castellanohablantes y catalanohablantes es más similar en los centros privados y en los urbanos que en los públicos y rurales, donde claramente predominan los informantes cuya lengua materna es la vernácula (*vid.* 5.2. gráficos 5.10 y 5.11).

Por el contrario, los resultados más aproximados entre las dos lenguas se dan entre el alumnado de los centros concertados (0.05), entre los varones (0.1), entre los hablantes con ambas lenguas como maternas (0.17) y entre los informantes procedentes del estrato más alto del espectro social (0.32). Estos alumnos se decantan bien a favor del castellano, en el caso de los centros concertados y en el de ambas lenguas como maternas, o del catalán, como los varones y los informantes del nivel alto. Sin embargo, la distancia entre los promedios se sitúa siempre en apenas unas décimas, con lo que concluimos que el caudal léxico disponible de estos colectivos es muy similar en los dos idiomas. Llama la atención que la variable *nivel sociocultural* sea la única que no establezca ninguna diferencia por encima de la unidad, dado que la distancia entre las medias en los hablantes del nivel bajo es de 0.85 respuestas y en los del medio, de 0.79. Como ya hemos indicado, sería necesario someter estos datos a pruebas de significación estadística que permitieran comprobar cuáles de estas disimilitudes tienen inferencia

estadística, más allá de la distancia que se pueda observar entre los promedios de los distintos colectivos.

Tabla 6.50. Comparación de la producción y la riqueza léxica de los distintos subgrupos en las dos lenguas de la encuesta.

VARIABLE SOCIAL	Palabras		Vocablos	
	Castellano	Catalán	Castellano	Catalán
VARONES		✓		✓
MUJERES		✓		✓
C. PÚBLICOS		✓		✓
C. PRIVADOS	✓		✓	
C. URBANOS		✓		✓
C. RURALES		✓		✓
N. BAJO		✓		✓
N. MEDIO		✓		✓
N. ALTO		✓	✓	
LM CASTELLANO	✓		✓	
LM CATALÁN		✓		✓
LM AMBAS	✓			✓
LM ARANÉS	✓		✓	

Para finalizar, presentamos una tabla comparativa con los resultados de cada subgrupo social relativos a la producción de palabras y al número de vocablos distintos en cada lengua (tabla 6.50). En la tabla, se marca con un símbolo de confirmación el idioma que, en global, recaba un promedio superior de respuestas y una cantidad mayor de vocablos. Se ha de tener en cuenta que, como acabamos de ver al comentar el gráfico 6.43, las diferencias entre la producción de un idioma u otro en cada subgrupo son distintas y, en algunos casos, revisten de más o menos importancia— aunque no sepamos el alcance que tienen estadísticamente—. La misma situación se da con respecto a las palabras diferentes, de hecho, en este caso aún es más difícil valorar el alcance de la diferencia entre la cantidad de vocablos de más que evocan los informantes en una u otra lengua⁸²⁸. No obstante, consideramos que el cotejo de estas

⁸²⁸ Este parámetro, como hemos indicado repetidamente, se ve influido por el número de informantes, que varía en cada uno de los subgrupos sociales, y también por la posible presencia de encuestados cuya actuación lingüística difiera de la norma habitual (*vid.* 6.1.4.).

dos magnitudes es apropiado para describir las tendencias de cada subgrupo respecto al léxico disponible en cada lengua.

Es fácil deducir, con solo echar una ojeada a la tabla 6.50, que los dos parámetros habitualmente usados para retratar la disponibilidad de una comunidad de habla suelen ir de la mano. Es decir, generalmente, los informantes agrupados según cada variable social que presentan una superioridad léxica en la producción de palabras en un idioma, también actualizan más vocablos, o palabras distintas, en esa misma lengua. Esto sucede en la mayoría de subgrupos: los varones, las mujeres, los informantes de centros públicos, de centros urbanos, de centros rurales, los de nivel bajo y medio y, finalmente, los catalanohablantes evocan más palabras y más vocablos en lengua catalana que en lengua castellana, en un mismo tiempo de reacción, siguiendo así la tendencia general a favor de la lengua vernácula que hemos visto a lo largo de todo este capítulo. Por el contrario, los informantes de los centros privados, que por poca distancia ostentan una media superior en castellano, también tienen en esa lengua una cantidad mayor de vocablos y lo mismo sucede con el alumnado castellanohablante y aranés: ambos colectivos se decantan, claramente, en las dos magnitudes por el castellano.

Otros grupos de informantes muestran un comportamiento más dispar entre las dos lenguas, como los encuestados del nivel sociocultural alto que, por pocas décimas, exhiben una media de palabras superior en catalán y, sin embargo, tienen mayor riqueza léxica en castellano. De forma inversa, el colectivo que tiene ambas lenguas como maternas, también por escasa distancia, muestra predilección por el castellano en cuanto a la productividad pero obtiene más vocablos en catalán. De cualquier modo, estas pequeñas divergencias dan fe de la actuación lingüística similar de estos dos colectivos con respecto a las dos lenguas de la comunidad objeto de estudio.

Tras haber analizado con todo detalle las diferencias cuantitativas con respecto a las variables lingüísticas y a las sociales, pasamos ahora al siguiente capítulo, donde describimos y clasificamos el gran número de fenómenos de contacto lingüístico hallados en los dos corpus, como prueba de la coexistencia de los dos sistemas lingüísticos estudiados en nuestra provincia.

7. ESTUDIO DE LOS FENÓMENOS DE CONTACTO LINGÜÍSTICO

Uno de los objetivos de este trabajo es el análisis de los trasvases lingüísticos que se producen entre las dos lenguas estudiadas a partir de los elementos concretos detectados en los dos diccionarios finales (*vid.* capítulos 10 y 11). Por este motivo, en el segundo capítulo (*vid.* 2.3.1), realizamos una revisión epistemológica de las principales consecuencias del contacto lingüístico, centrándonos para ello en los fenómenos originados por el contacto continuado de dos sistemas lingüísticos o por el uso de más de una lengua en un mismo territorio, como sucede con el castellano y el catalán en Lleida (*vid.* 4.3.2). En el presente capítulo, pretendemos clasificar y analizar, en conjunto, estos elementos lingüísticos no sin antes especificar —tal como realizamos en el siguiente apartado— las ventajas y limitaciones que presenta el estudio de las transferencias⁸²⁹ a través del léxico disponible y las dificultades concretas con las que nos hemos encontrado a la hora de realizar esta labor.

7.1. INTRODUCCIÓN

Las pruebas asociativas de la disponibilidad constituyen un campo ideal de aparición de fenómenos de contacto lingüístico. Ello es lógico si consideramos que la principal categoría gramatical que se recaba en estos listados es el sustantivo y, en menor medida, verbos y adjetivos. Tal como indica la mayoría de trabajos sobre la interferencia léxica, los nombres son la categoría más fácil de transferir de una lengua a otra. En concreto, si dos lenguas tienen una estructura sintagmática similar —como sucede en nuestro caso— las categorías léxicas se transfieren con el siguiente orden de preferencia: nombres, verbos, adjetivos, adverbios, preposiciones e interjecciones (Baetens 1989: 93)⁸³⁰. Además, como vimos en el tercer capítulo (*vid.* 3.1), Mackey (1971a, vol. II: 16) ya afirma, en los años 70, que la disponibilidad es un instrumento adecuado para detectar los préstamos, dado que aborda un sector del vocabulario donde estos suelen manifestarse⁸³¹.

⁸²⁹ Utilizamos este término, durante este capítulo, como sinónimo general de *fenómeno de contacto lingüístico*. Sobre este término y sus diferentes acepciones, remitimos a lo dicho en el capítulo 2 (*vid.* 2.3.1).

⁸³⁰ El orden de las categorías gramaticales sufre ligeras variaciones según el autor, aunque los sustantivos y los verbos siempre ocupan las primeras posiciones. Véanse las reticencias de Weinreich (1996: 98-99) a esta afirmación.

⁸³¹ «Puisque la disponibilité est une manifestation de la langue et non seulement de la parole, c'est l'instrument tout à fait indiqué pour déceler les emprunts, non seulement parce que le sujet bilingüe fait l'inventaire de ses connaissances d'une langue à la fois, mais aussi parce que la disponibilité touche surtout

Asimismo, las dos lenguas que analizamos —castellano y catalán— presentan una gran proximidad, tanto estructural como genética, lo que constituye uno de los factores más favorecedores de la influencia mutua (Weinreich 1953 *apud* Blas Arroyo 1993a: 30). Como ya vimos en el capítulo cuarto (*vid.* 4.1), en la situación estable de bilingüismo social extenso y equilibrado en la que nos encontramos, dado el elevado número de bilingües y su conocimiento aproximado de las dos lenguas⁸³², las transferencias entre ambas lenguas, además emparentadas, son más que esperables⁸³³.

De otra parte, los índices matemáticos obtenidos en los diccionarios finales del léxico disponible de las dos lenguas (*vid.* capítulos 10 y 11) nos permiten conocer, de forma rigurosa, qué lugar ocupan los fenómenos de contacto detectados con relación a los restantes elementos evocados y, sobre todo, cuál es el porcentaje de aparición de estos términos. Este parámetro resulta especialmente interesante puesto que nos permite distinguir entre los trasvases lingüísticos que responden a meras actualizaciones individuales —por desconocimiento u olvido del vocablo apropiado en la lengua de la encuesta— de las transferencias ya extendidas —que presentan un cierto grado de integración social—. Además, en el mismo diccionario también es posible descubrir, en la mayoría de las ocasiones, el término equivalente en la lengua en la que se completa la encuesta, lo que hace que podamos comparar tanto la frecuencia de estos dos vocablos como la posición que ocupan según su índice de disponibilidad⁸³⁴. En nuestro caso, también tenemos la ventaja de contar con los datos de las dos lenguas, con lo que es posible ampliar el cotejo y estudiar el valor de los términos sinónimos en los dos corpus.

Cabe aclarar en este punto —tal como hicimos también en el capítulo anterior— que, debido a que la cantidad de transferencias lingüísticas recabadas en esta

le secteur du vocabulaire où les emprunts se manifestent d'abord, à savoir, le vocabulaire concret" (1971a, vol. II: 16).

⁸³² Tal y como hemos corroborado en el análisis cuantitativo, al menos en la franja de edad que manejamos en esta investigación (*cf.* capítulo 6).

⁸³³ Véase al respecto Silva-Corvalán (1989: 179). En el capítulo cuarto, además, explicamos por qué la situación actual se muestra propicia a que los elementos lingüísticos se transfieran en las dos direcciones (del castellano al catalán y del catalán al castellano, *vid.* 4.1).

⁸³⁴ Posibilidades que ya aplicó Mackey (1971a, 1976) en su trabajo pionero sobre el francés de Canadá.

investigación no es precisamente escasa, al igual que los datos generales que manejamos, nos ceñimos en el presente capítulo a su clasificación y al análisis del valor general que alcanzan en relación con distintos parámetros. Las posibilidades de análisis son mucho mayores, como resulta evidente, pero exceden, a nuestro modo de ver, los límites de este trabajo. Una de las más interesantes, y que esperamos acometer en un futuro cercano, es conocer el perfil social que evoca un número mayor de elementos de este tipo en cada una de las dos lenguas⁸³⁵. Por el mismo motivo, tampoco abordamos aquí las transferencias que proceden de otras lenguas distintas al castellano y al catalán, claramente minoritarias en nuestros corpus en comparación con las procedentes de las dos lenguas cooficiales de la comunidad⁸³⁶.

A pesar de las presumibles ventajas que acabamos de enumerar, el estudio de los fenómenos de contacto lingüístico a través de la disponibilidad no está exento de inconvenientes⁸³⁷. El principal de todos ellos, bajo nuestro punto de vista, es consecuencia de la propia metodología utilizada para la obtención de los diccionarios de léxico disponible, toda vez que la producción artificial de palabras tras proporcionar un estímulo al informante, no supone un uso real de la lengua. A pesar de que, habitualmente, se defina el léxico disponible como *el caudal léxico utilizable en una situación comunicativa dada*, hoy sabemos que no es tan obvio como puede parecer *a priori* que el hablante vaya realmente a usar esa palabra en un discurso real. Al respecto, Hernández Muñoz (2006: 406) afirma:

“El léxico disponible es un hecho intralingüístico, ya que conlleva reflexiones sobre la lengua desligadas del uso. Por ejemplo, el hecho de que un hablante escriba una palabra en la encuesta no nos asegura que sea capaz de utilizar eficientemente esa palabra en el discurso cotidiano, sino sólo que en su sistema de conexiones semántico-léxicas esta palabra está vinculada de alguna manera a la etiqueta del centro de interés en el que se incluye.”

Además de la imposibilidad de saber si el hablante incluiría o no esa determinada forma en un discurso espontáneo real, en el fragmento anterior se deja entrever otra dificultad añadida: el hecho de que los vocablos que se evocan mediante este ejercicio controlado con tiempo límite están condicionados por las “etiquetas”, esto

⁸³⁵ Especialmente interesante sería el análisis de las transferencias halladas según la variable *lengua materna* del informante, a la vista de los resultados cuantitativos obtenidos (*vid.* 6.2.5.).

⁸³⁶ Recordamos al lector que estas voces se consignan en cursiva en los diccionarios finales cuando se registran de este modo en el DRAE o cuando no se recogen en él y, con respecto al catalán, cuando no aparecen en los diccionarios ni en las fuentes terminológicas normativas, es decir, en el DIEC o en el TERMCAT (*cf.* 5.4.1).

⁸³⁷ Llopis (2009: 302-303) también apunta algunas de las limitaciones del test de disponibilidad léxica en este sentido.

es, por los nombres de cada centro de interés que se ofrece al informante. Al realizar el análisis de las transferencias halladas, hemos comprobado este hecho, dado que por nuestros conocimientos como integrantes de la propia comunidad estudiada sabemos de la existencia de algunos préstamos, ya constatados además en la bibliografía sobre el tema, que, sin embargo, no aparecen en nuestras encuestas, o lo hacen con un porcentaje de aparición muy bajo, debido a que se alejan del estímulo proporcionado al informante⁸³⁸. Como sabemos, la característica que más influye en la disponibilidad de un término es precisamente la tipicidad, ya que las palabras más típicas de una categoría son las más disponibles (Hernández Muñoz 2006). Este es el motivo por el que, en la clasificación de los elementos de contacto lingüístico, no tomemos el índice de disponibilidad como criterio, a pesar de que sí lo hemos usado en el análisis concreto de los préstamos recabados (*vid.* 7.4.).

Aunque la disponibilidad no pueda tomarse como medida del uso real, sí es verdad que constituye una manifestación del léxico que tiene en potencia un individuo⁸³⁹ y que, por tanto, podría llegar a utilizar en una situación determinada y bajo unas condiciones concretas. En nuestra opinión, es significativo que un informante incluya en su listado una forma de otra lengua, más aún cuando completa la encuesta en su propia lengua materna y cuando ha recibido instrucciones estrictas sobre la lengua en que debe rellenar la prueba (*vid.* 5.3). Como veremos a continuación, no solo se incluyen términos de la segunda lengua —o primera según el caso— en los dos diccionarios, sino que el hecho de que algunas de estas unidades tengan un porcentaje de aparición superior a la forma adecuada en la lengua de la encuesta nos indica la extensión social que pueden llegar a tener.

A nuestro modo de ver, las limitaciones expuestas hasta el momento no impiden que el estudio de los fenómenos de contacto lingüístico pueda realizarse a partir de los

⁸³⁸ Esto sucede, por ejemplo, con el término “chafardear”, ampliamente extendido en el castellano hablado en nuestra provincia que, por el contrario, solo cuenta con una mención en el corpus castellano dentro de la categoría *Juegos y distracciones*. Esta misma limitación será la causa de que un mismo vocablo pueda tener una consideración diferente en función del centro de interés en el que aparezca, ya que el porcentaje de aparición variará en función del campo semántico en el que se inserte la unidad.

⁸³⁹ En palabras de Hernández Muñoz (2006: 226): “[c]ada vez es más evidente que la disponibilidad léxica no es meramente una medida de la facilidad de ‘uso’ de una palabra en el discurso espontáneo en torno a un tema, sino que es una consecuencia muy clara de unas condiciones experimentales determinadas, que constituyen una manifestación particular, y sólo particular, del léxico en potencia de un individuo”. La misma autora glosa algunos inconvenientes de las pruebas de disponibilidad léxica como método de acercarse al lexicón mental de un hablante bilingüe (Hernández Muñoz 2006: 124).

datos de la disponibilidad, solo apuntan a que sería necesario cotejar los resultados recabados mediante este tipo de pruebas con los obtenidos utilizando otros métodos de recogida de datos. De este modo se podría dar fe de la integración, o no, de las unidades detectadas en el discurso espontáneo real de los informantes. Esta información, además, resultaría fundamental para distinguir entre algunos fenómenos de contacto concretos, como entre el cambio de código y el préstamo, ya que podríamos comprobar de este modo si las unidades transferidas siguen los patrones de la lengua donante o los de la lengua receptora al integrarse en el discurso. No obstante, la similitud estructural de las dos lenguas puede que impidiera también secundar este criterio, el de la integración lingüística, para diferenciar entre trasvases lingüísticos. Gómez Molina (1999-2000: 317) afirma al respecto: “dada la afinidad genética de ambas lenguas románicas (castellano y catalán) la integración lingüística —fonológica, gramatical o semántica— no puede actuar en muchos casos como criterio determinante en la diferenciación del préstamo *versus* cambio de código”.

De otra parte, una de las dificultades concretas que hemos tenido a la hora de abordar el estudio de las transferencias ha sido, precisamente, la de diferenciar qué elementos pertenecen a cada una de las dos lenguas estudiadas y, por ende, cuáles pueden considerarse fenómenos de contacto lingüístico. En principio, consideramos, claro está, como propias de cada lengua las lexías según su inclusión o no en los diccionarios normativos de cada lengua (el DRAE en el caso del castellano y el DIEC o la base TERMCAT para el catalán). En los compuestos sintagmáticos con una forma homófona en las dos lenguas (como “cuarto”/*quarto* en *quarto de rentar*), hemos estimado, por encima de la encuesta que estuviera completando el informante, que todo el sintagma se halla en el mismo idioma, como resulta, por otra parte, lógico. Al igual que nos sucedió al editar los listados finales (*vid.* 5.4) no es tan fácil, como puede parecer *a priori*, diferenciar a qué idioma pertenece cada término. Las dos lenguas, emparentadas genéticamente, comparten muchos étimos; además de que existen numerosas lexías que, sin tener nada que ver en ocasiones, coinciden con formas existentes en la otra lengua. Nos hemos encontrado con esta circunstancia en reiteradas ocasiones al analizar unidades simples que son correctas en los dos idiomas, como por ejemplo *cama*, que aparece en el corpus castellano en el primero de los centros de

interés (*El cuerpo humano*)⁸⁴⁰. Es evidente que el hablante se refiere al término catalán equivalente a ‘pierna’, puesto que el significado del castellano no se corresponde con el centro de interés. El mismo fenómeno hemos observado —y, por tanto, hemos secundado el mismo criterio— en las numerosas lexías agramaticales producidas por una generalización incorrecta de las normas de derivación de una de las dos lenguas. Siguiendo con el mismo ejemplo, también en *El cuerpo humano* en el corpus castellano aparece *mugrón** (de *mugró*, ‘pezón’), palabra existente en castellano, pero cuyo significado se aleja de las asociaciones propias del primero de los centros de interés. En este caso concreto, es indudable que el informante ha empleado, erróneamente, una regla que frecuentemente sirve para diferenciar términos en catalán y en castellano, como es la añadidura de la -n en la terminación -ó de los sustantivos (*camió* / ‘camión’, *peó* / ‘peón’, *neó* / ‘neón’, etc.)⁸⁴¹.

Por tanto, el primero de los criterios utilizados para decidir a qué lengua pertenece una determinada forma ha sido basarnos en el significado que atribuye el informante al término en cuestión, sirviéndonos para ello del estímulo verbal en el que se halla el vocablo. Para poder utilizar este criterio, otras veces no ha sido suficiente el hecho de conocer el centro de interés donde se enumera la forma y, por esta razón, con la finalidad de conocer qué significado otorga el informante al término en cuestión, hemos recurrido, en primer lugar, a las fuentes de autoridad de cada uno de los dos idiomas (DRAE, CREA, GDLC, DIEC y TERMCAT), a los mismos datos que los estudiantes nos proporcionan en la otra lengua estudiada —fuente valiosísima de información—; a los recuentos de léxico disponible llevados a cabo en otros lugares en español y en catalán y, finalmente, a nuestra propia experiencia como integrantes de la comunidad de habla estudiada y como profesoras de lengua del colectivo concreto analizado.

En segundo lugar, existe un conjunto de términos que son homógrafos en las dos lenguas. Por ejemplo, “*cadira*” aparece en la encuesta en castellano (en *Muebles de la*

⁸⁴⁰ Al igual que en el capítulo 5, transcribimos los ejemplos referidos al castellano entre comillas y los del catalán, en cursiva, independientemente del corpus (castellano o catalán) en el que hayan sido encontrados. Cuando una palabra esté en cursiva y entrecorrida significa que es homógrafa en las dos lenguas. Los significados se transcriben con comillas simples y los términos agramaticales se marcan con un asterisco al final de la palabra, como es práctica habitual en los estudios lingüísticos.

⁸⁴¹ Otros ejemplos, similares a *cama* o *mugrón** serían *barbeta* (1), *cazón** (7) o *truja* (11) en el corpus castellano, y “faja” (2), “borrego”, *deport** (11) o *delfi** (14) en el catalán. Entre paréntesis hemos indicado el número del centro de interés donde se ha encontrado cada término.

casa) y existe tanto en castellano como en catalán. De forma general, hemos considerado, ante un caso así, que el término pertenece a la lengua en que se está completando la encuesta⁸⁴². Sin embargo, en este ejemplo en concreto, la palabra *cuenta*, en castellano, con una marca lexicográfica que nos informa de que está actualmente en desuso. Por este motivo, y dada la juventud de nuestros informantes, hemos considerado que todos los homógrafos que cuentan con una marca, diacrónica o de uso, que nos informa sobre la caducidad del término en una de las dos lenguas, pertenecen a la otra lengua estudiada (en el ejemplo de “*cadira*”, al catalán), ya que creemos que los informantes la aprenden a través de esa lengua y así es como la incorporan a su lexicón mental⁸⁴³. Incorporamos, a continuación, algunos de los casos que hemos hallado en esta situación, tanto en el corpus castellano como en el catalán, y también ofrecemos información sobre el significado del término —según el centro de interés en el que se ha encontrado— y el número de menciones recibido en los dos corpus para justificar, así su inclusión en la otra lengua.

Tabla 7.1. Voces marcadas lexicográficamente como en desuso o antiguas en el corpus castellano.

Palabra	C.I.	Marca DRAE	Definición DRAE	Menciones corpus castellano	Menciones corpus catalán
armilla	2	ant.	‘brazalete’	12	25
cadira	4 ⁸⁴⁴	desus.	‘silla’	6	232
escriptor	16	desus.	‘escritor’	3	22
jaqueta	2 ⁸⁴⁵	desus.	‘chaqueta’	6	119
misto	9	p.us.	‘cerilla’	3	12
solitud	11	ant.	‘carencia de compañía’	1	0
tallar	13	ant.	‘cortar o tajar’	1	58

⁸⁴² Idéntico problema y solución constata Casanova (2013) en su análisis de los valencianismos detectados en su corpus.

⁸⁴³ Por otro lado, estamos convencidos, sin haberlo comprobado, de que en el ejemplo anterior, y en los que se citan a continuación, cualquier hablante bilingüe castellano/catalán al que le preguntáramos, desconocería por completo que ese término, con su significado correspondiente, —o cualquiera de los que se presentan en la tabla 7.1— son normativos en lengua castellana.

⁸⁴⁴ En *La escuela* y en *Juegos y distracciones* también se incluye esta forma y recibe una mención en cada área.

⁸⁴⁵ En *La escuela* se incluye la forma híbrida *jaquetas** que actualiza un informante.

Tabla 7.2. Voces marcadas lexicográficamente como en desuso o antiguas en el corpus catalán.

Palabra	C.I.	Marca GDLC ⁸⁴⁶	Definición GDLC	Menciones corpus catalán	Menciones corpus castellano
despensa	3 ⁸⁴⁷	ant.	‘rebot’	26	52
margarita	11	ant.	‘margarida’	1 ⁸⁴⁸	15
paloma	14 ⁸⁴⁹	ant. i dial.	‘colom’ ⁸⁵⁰	8	27

Hemos aplicado el mismo criterio cuando alguna de las formas agramaticales creadas por los informantes, como por ejemplo *ferrero**, existía en las fuentes lexicográficas con las marcas citadas (“antiguo”, “poco usado” o “en desuso”). Al igual que en el ejemplo de *mugrón**, hemos clasificado estos términos como *interferencias*, es decir, como palabras inexistentes en las lenguas estudiadas y creadas, como ya hemos comentado, por una generalización incorrecta de las normas de derivación de alguna de las dos lenguas (en el caso anterior, proviene de *ferrer*, ‘herrero’)⁸⁵¹.

Cabe decir que no somos los primeros que nos enfrentamos al conflicto que se plantea a la hora de estudiar y, sobre todo, de diferenciar qué elementos pertenecen a cada lengua en el contacto castellano-catalán. Otros autores, como Hernández García (1998a: 214) o Vann (2008: 280-281), también destacan la dificultad de distinguir entre las transferencias lingüísticas y los casos, como los anteriores, en los que una forma lingüística arcaica se preserva, puesto que ambas circunstancias se deben al contacto entre dos lenguas que comparten el origen de la mayor parte de su léxico. También Atenza *et al.* (1997: 581) observan cómo “palabras recogidas en el diccionario como

⁸⁴⁶ Incluimos solo la marca hallada en el GDLC, ya que ninguno de estos términos cuenta con marca lexicográfica alguna en el DIEC o en el TERMCAT. De hecho, *despensa* solo se registra en el GDLC y *margarita* y *paloma*, con los significados que aparecen en la tabla, también se recogen únicamente en el diccionario de *l’Enciclopèdia Catalana*.

⁸⁴⁷ Este término también aparece en *Muebles de la casa* y *La cocina y sus utensilios*, con una y dos menciones respectivamente.

⁸⁴⁸ La forma híbrida *margarites** recibe tres menciones en este mismo centro de interés.

⁸⁴⁹ También aparece en *El campo* donde lo actualizan 2 informantes.

⁸⁵⁰ La palabra tiene otros significados, como ‘bolet’, que podrían encajar en el centro 11, pero no en el 14, donde es evidente que los informantes la usan con el significado de *colom*. De otra parte, el término también equivale a *papallona*, aunque el GDLC lo marca como dialectal.

⁸⁵¹ Algunos ejemplos de estas voces serían, en el corpus castellano, *fustero** (‘tornero’, ‘carpintero’) o *soterráneo** (‘subterráneo’), en desuso según el DRAE y, en el catalán, *camisó** (‘peça de roba blanca amb coll [...]’) con la marca *antic* en el GDLC.

arcaísmos o pertenecientes al registro culto aparecen en los escritos de los alumnos sin conciencia de esas notas”.

En tercer lugar, cabe señalar que hemos considerado, excepcionalmente, como palabras catalanas algunos términos homógrafos que aparecían en el corpus castellano y que se encuentran incluidos en el DRAE sin marca alguna. Estos poquísimos casos los hemos clasificado de este modo tras haber cotejado las menciones que recibe el término en los dos corpus y haber comprobado su total ausencia en los léxicos disponibles de otras zonas hispánicas⁸⁵². En nuestra opinión, la escasez de ejemplos en el CREA también prueba su poco uso en lengua castellana, además de que, por nuestra propia experiencia como hablantes bilingües, damos fe de que estos términos se emplean únicamente al hablar en lengua catalana⁸⁵³. A continuación, incluimos una tabla ilustrativa con estos ejemplos, todos ellos del corpus castellano:

Tabla 7.3. Voces con entrada en el DRAE consideradas catalanas en este estudio.

Palabra	C.I.	Definición DRAE	CREA ⁸⁵⁴	Menciones Corpus castellano	Menciones Corpus catalán
mocador	2-15 ⁸⁵⁵	‘moquero’	0	4-4	72-13
poma	5 ⁸⁵⁶	‘manzana’	6	3	109
pruna	5	‘ciruela’	2	2	10
destral	7-11	‘hacha’	3	1-1	3-2
botiga ⁸⁵⁷	10	‘botica’	4	6	136

⁸⁵² De los ejemplos que veremos a continuación, únicamente hemos hallado *pruna* en el léxico disponible de Valencia (Gómez Molina y Gómez Devís 2004) y en el de Alicante (Martínez Olmos 2008), cuya presencia puede explicarse, al igual que en nuestro caso, por el contacto con el catalán. Hemos comprobado la ausencia de estas voces, con la citada excepción, en el léxico disponible de Asturias (Carcedo 2001), de Cádiz (González Martínez 2002), de Aragón (Arnal *et al.* 2004), de Soria (Bartol 2004), de Valencia (Gómez Molina y Gómez Devís 2004), de Castilla La Mancha (Hernández Muñoz 2006), de Ceuta (Ayora 2006), de Galicia (López Meirama 2008) y de Alicante (Martínez Olmos 2008).

⁸⁵³ En estos casos, contrariamente a los que hemos comentado con anterioridad, el significado del término en castellano sí encaja con el centro de interés, dado que coincide con el que tiene el término en catalán.

⁸⁵⁴ Téngase en cuenta, al valorar la presencia en el CREA de estas voces, que, la mayoría de veces, hay ejemplos en los que estos términos se incluyen en sintagmas emitidos en otras lenguas. Sucede especialmente con *cinema* (en inglés) y *natura* (en latín), dada la abundancia de las expresiones *home cinema*, *free cinema* y *contra natura*. También suele haber casos de sintagmas en catalán que se recogen citados en obras castellanas como en uno de los dos casos en que aparece “*pruna*” en el CREA: “Después vino *la lluna*, *la pruna*... y pronto la abuela [...]”, en referencia a la canción de cuna tradicional catalana.

⁸⁵⁵ En *La escuela* también aparece con una sola mención.

⁸⁵⁶ En *Trabajos del campo y del jardín* se recoge el híbrido *pomas** en el compuesto “recoger *pomas**”, que nombra un informante.

cinema	10-15	‘cine’	168	4-2	54-59
natura	11 ⁸⁵⁸	‘naturaleza’	240	1	29
manobre	16	‘obrero [...]’	0	8	16

En cuarto lugar, el criterio que hemos aplicado en el caso de los extranjerismos que coinciden con la adaptación normativa de ese mismo préstamo en la otra lengua estudiada ha sido el de considerar la forma como propia de esta última lengua y, por tanto, como transferencia en la otra lengua analizada. Esto ocurre, en el corpus castellano, con algunos galicimos como *bidet*, *bufet* o *tiquet* y también con algún anglicismo como *handbol*⁸⁵⁹; en dirección contraria hallamos “puenting”, en lugar del normativo *salt de pont*, ausente por completo en nuestros listados.

Para finalizar, hay que aclarar que nos ceñimos al análisis de los fenómenos léxicos originados por el contacto lingüístico, aunque también hemos detectado en el corpus castellano algunos rasgos morfológicos, como el mayor uso del sufijo “-ete/a”, que también se deben al contacto del castellano con el catalán en Lleida (*v.gr.* “collarete”, “cesteta del pan”, “molinete de café”, “placeta”, etc.). Asimismo, otra de las líneas de estudio posibles en el futuro sería analizar el peso que alcanzan en el corpus castellano las voces regionales, esto es, las propias del castellano hablado en nuestra provincia. Muchas de las entradas de nuestro diccionario aparecen en el DRAE con una marca diatópica (*Cat.*, *Ara.* o *Val.*) o diacrónica (*del cat.*), que nos informa de la extensión dialectal o del origen catalán del término. Ejemplos del primer caso —donde abundan las palabras marcadas únicamente como aragonesismos— son “bróquil”, “tocino”, “madrilla”, “riera”, “torre”, “fiema”, “leñar”, “madrilla”, “sargantana” o “botiguero”, y del segundo, “payés”, “rape”, “mosén”, “alioli”, “butifarra”, “escalivada”, “masía”...) ⁸⁶⁰.

⁸⁵⁷ Esta palabra aparece con una marca dialectal en el DRAE que nos informa de su uso en Aragón, Cataluña y Navarra. Hablaremos, seguidamente, de este tipo de casos.

⁸⁵⁸ En *Animales* también aparece en el sintagma “animal en la *natura*” que menciona un informante.

⁸⁵⁹ En estos casos también se podría considerar que estas formas son términos dialectales del castellano, que evidencian cómo se adapta el préstamo en esta parte del territorio hispánico, aunque incluso en este caso, resulta claro que la opción escogida refleja la influencia de la fonética catalana. Galindo (2006: 173) adopta el mismo criterio con respecto a los extranjerismos en su trabajo sobre el uso del castellano y del catalán a la hora del recreo en las escuelas catalanas.

⁸⁶⁰ Como excepción, la palabra “*botiga*” ha sido considerada como catalana (*cfr.* tabla 7.3. y la explicación anterior a la tabla). De otra parte, la mayor presencia de estas voces en nuestra provincia podría comprobarse fácilmente mediante el cotejo con los restantes léxicos disponibles ya publicados.

7.2. CLASIFICACIÓN DE LOS FENÓMENOS DE CONTACTO LINGÜÍSTICO

A continuación, presentamos los fenómenos de contacto lingüístico que hemos recogido en los dos corpus. Para analizarlos, hemos utilizado algunos de los criterios establecidos en la bibliografía clásica, como la frecuencia de uso o el desplazamiento de sinónimos de la lengua prestataria (Gimeno y Gimeno 2003: 110). Como ya hemos indicado, es posible observar en nuestros datos el porcentaje de informantes que incluyen una determinada unidad léxica y también si esta desplaza, o no, los sinónimos de la lengua receptora, factor que hemos tenido en cuenta en el análisis concreto de los préstamos (*vid.* 7.4.). Nos hemos basado en el grado de integración social, medido a través de las menciones de cada término, y en la noción de gramaticalidad/agramaticalidad para diferenciar entre los distintos fenómenos. Aunque pensábamos tener en cuenta también el índice de disponibilidad, finalmente lo hemos descartado en esta primera clasificación, dado que el orden de las palabras en los listados viene determinado por la relación de estas con el estímulo dado, como ya hemos comentado en la introducción. Por el contrario, el porcentaje de aparición de cada palabra —que se corresponde con la tercera columna de los diccionarios finales (*vid.* capítulos 10 y 11)— indica cuántos informantes, sobre cien, han incluido esa forma en sus listados, independientemente de la posición en que la hayan actualizado. De este índice es posible extraer, mediante una sencilla regla de tres, cuántas menciones ha recibido el término en cuestión, cantidad que coincide, generalmente, con el número de informantes que han incluido el término, dado que, como ya explicamos en los criterios de edición (*vid.* 5.4.), las entradas repetidas por un mismo encuestado han sido eliminadas en la edición final de los datos⁸⁶¹.

Hechas estas salvedades, a continuación definimos los fenómenos de contacto lingüístico que hemos hallado en nuestros listados⁸⁶²:

(a) **Interferencias**: Designamos bajo esta denominación todas aquellas lexías que resultan agramaticales, por no ser correctas en ninguna de las dos lenguas analizadas, y que, al ser propias del idiolecto, carecen de integración social, tal como determina el bajo porcentaje de informantes que las actualizan.

⁸⁶¹ El único caso en que no coincidirían estos dos parámetros sería cuando un mismo informante incluye una transferencia en una lexía simple y, luego, la misma en una combinación sintagmática, aunque estos casos los hemos considerado a todos los efectos como unidades distintas en el presente capítulo, salvo una excepción que más adelante comentaremos (*vid.* nota 872).

⁸⁶² Nótese la semejanza de nuestra clasificación con la de Llopis (2009), a quien, en parte, seguimos.

Concretamente, hemos considerado de este modo las voces agramaticales cuyo porcentaje de aparición era inferior al 5%, lo que corresponde a, como máximo, 11 menciones del término⁸⁶³.

(b) Préstamos en vías de integración: Son aquellas palabras que aparecen fruto del desconocimiento que posee el hablante del término apropiado en la lengua en la que está cumplimentando la encuesta. Su aparición en los datos viene determinada por un nivel relativo de integración social, que indica que un número considerable de encuestados actualizan el término en cuestión, lo que podría indicar que se trata de un fenómeno en movimiento en la comunidad de habla. Estas voces alcanzan un porcentaje de aparición de entre el 5 y el 15%, lo que equivale a decir que han sido mencionadas entre 12 y 36 veces.

(c) Préstamos integrados: A diferencia del concepto anterior, este tipo de préstamos tiene ya un grado notable de integración social, determinado tanto por la alta frecuencia de menciones como por el desplazamiento de los sinónimos correspondientes de la lengua base. En este caso, los préstamos se extienden también a los hablantes monolingües de la misma comunidad, y los informantes no tienen conciencia de que utilizan una palabra propia de otra lengua. Este tipo de préstamos registra un porcentaje de aparición superior al 15%, y por tanto, cuenta con más de 36 menciones⁸⁶⁴.

(d) Cambio de código: Este caso se diferencia del primero en que se trata de términos gramaticales, y por tanto, con corrección lingüística, pero que no se corresponden con la lengua en la que se está completando la encuesta. El porcentaje de hablantes que los menciona es relativamente escaso, lo que prueba que carecen de integración social. Su presencia viene condicionada por el olvido momentáneo del término apropiado en la lengua referida, al menos en la situación que se está produciendo en el momento de la recogida de datos. Al igual que las interferencias, su porcentaje de aparición es inferior al 5%.

(e) Combinaciones híbridas: Con esta etiqueta agrupamos aquellos casos que suponen una amalgama entre las dos lenguas, en el sentido de que, generalmente en una combinación sintagmática, aparece un término en una lengua y otro en

⁸⁶³ Cabe indicar que algunas de las voces que aparecen como interferencias en el corpus catalán (como *avespa**, *garaig** o *llimpiar**), ya aparecen en los diccionarios dialectales del catalán de la región, en LNDLLC y en el DLL, cuyas características metodológicas ya advertimos (*vid.* 5.4.).

⁸⁶⁴ Téngase en cuenta que, como ya hemos indicado, las menciones coinciden con el número de informantes que actualiza el término en cuestión y que la muestra se compone de 240 sujetos.

otra, lo que hace imposible determinar la lengua a la que corresponde la unidad en su globalidad. Dentro de esta categoría distinguimos entre combinaciones sintagmáticas (con el núcleo en castellano, con el núcleo en catalán o con algún elemento híbrido) y las voces compuestas por dos lexemas de los cuales, generalmente, uno pertenece al castellano y el otro, al catalán⁸⁶⁵.

(f) Calco sintáctico: Esta denominación se corresponde con una lexía compuesta aunque, en este caso, lo que se transfiere de la otra lengua es el esquema de la combinación de los distintos elementos, ya que todos ellos pertenecen a la lengua base en la que se completa la encuesta.

(g) Calco semántico: A semejanza del caso anterior, tampoco en esta ocasión se transfiere un elemento de otra lengua, sino que una palabra adquiere un significado nuevo por el contacto con una lexía de la otra lengua con la que presenta una semejanza formal y semántica. Como las dos lenguas se hallan emparentadas genéticamente, todos los casos de nuestro corpus que se engloban bajo esta etiqueta comparten, además, un mismo étimo, generalmente latino.

Es de recibo comentar que no ha sido fácil delimitar este último tipo de calco en nuestros corpus. La gran similitud entre las dos lenguas y la ausencia de un contexto donde se inserten las unidades léxicas ha hecho que albergáramos muchas dudas sobre en qué ocasiones se añadía un significado a una forma ya existente en la lengua o bien se transfería una forma lingüística nueva. Cabe destacar que, en casi todos los ejemplos que hemos considerado como calcos semánticos, existe una relación entre el significado que comparten las dos formas y el que se transfiere de un idioma a otro, de modo que las realidades que se designan suelen presentar algún rasgo en común, aunque en unos casos la vinculación es más clara que en otros⁸⁶⁶. No obstante, no estamos completamente seguros de que no pudiera haber otro término en nuestros listados que cumpliera también estos requisitos. De cualquier modo, es realmente complicado establecer los límites entre préstamos y calcos semánticos, como otros autores, también

⁸⁶⁵ Aunque los ejemplos de compuestos son escasos en los dos corpus, la tipología que hallamos al analizarlos es amplia puesto que algunas veces los dos lexemas existen en cada una de las lenguas y otras se construyen a base de híbridos (v.gr. fregaplats*, lavavaixelles*, lavavajilles*...).

⁸⁶⁶ Como, por ejemplo, *palmera* en el corpus catalán, que adopta el significado de ‘pastel de hojaldre’, recogido en el DRAE y expresado en catalán normativo por *ulleres*. Es evidente, en este caso, la relación metafórica entre la realidad designada y el término que la designa en cada lengua.

dedicados al contacto castellano-catalán, ya han señalado con anterioridad (Hernández García 1998a: 218, Vila i Moreno 1996: 50)⁸⁶⁷.

Nuestra clasificación tiene algunas limitaciones. La principal se deriva del hecho de que aunamos criterios de dos tipos a la hora de clasificar los fenómenos: por una parte, siguiendo un criterio formal, distinguimos entre *calcos sintácticos*, *calcos semánticos* o entre distintos tipos de *combinaciones híbridas* y, por otra, teniendo en cuenta la integración social del término —basado en la frecuencia de mención o el porcentaje de aparición— diferenciamos entre *interferencias* o *cambios de código*, por un lado, y, por otro, entre *préstamos en vías de integración* y *préstamos integrados*. Por este motivo, entre las *combinaciones híbridas*, los *calcos sintácticos* y los *calcos semánticos*, pueden hallarse también ejemplos de *préstamos en vías de integración* o ya *integrados*, como analizamos en el último apartado (*vid.* 7.4.). La doble clasificación de estos fenómenos, un total de 11 entre las dos lenguas, se detalla en el análisis pormenorizado de cada centro de interés, que efectuamos a continuación, mediante una nota a pie de página⁸⁶⁸.

Seguidamente presentamos en 34 tablas los diferentes casos de contacto lingüístico detectados en el léxico disponible del alumnado de nuestra provincia del curso previo a la entrada en la universidad, según esta tipología que acabamos de definir. El orden que seguimos es el de los centros de interés recogidos en la encuesta. En cada categoría, ofrecemos una primera tabla con los datos del corpus castellano y una segunda con los datos del corpus catalán.

Somos conscientes del peligro de relacionar todos y cada uno de los casos que hemos detectado según la clasificación que les hemos dado. Más de uno puede ser

⁸⁶⁷ Vila i Moreno (1996: 50) hace hincapié en la dificultad que supone establecer los límites entre préstamos y calcos, especialmente cuando la lengua receptora suele tener un homófono de la palabra importada. Hernández García (1998a: 218) también afirma que la diferencia entre *calco formal* o *semántico* es fácil en teoría pero que conlleva muchos problemas en la práctica. De hecho, esta autora, cuando de forma clara cree que se da una coincidencia casual con formas de la otra lengua más que un desplazamiento de significado entre formas, lo considera un *calco formal*, y no un *calco semántico*. De otra parte, Galindo (2006: 170) comprueba, con ejemplos de su corpus, cómo entre lenguas parecidas estructuralmente es más frecuente la presencia de préstamos que la de calcos.

⁸⁶⁸ En el fondo de esta limitación subyace la doble perspectiva de análisis de los fenómenos de contacto lingüístico: como *procesos* en los cuales puedan determinarse distintas fases o como *resultados* en los que pueda analizarse la relación entre las formas interferidas y las estructuras interferentes. Hernández García (1998a: 175-177) explica y aborda idéntica problemática en su trabajo.

objeto de debate, como debate hemos tenido en algunos casos para clasificar el término o la combinación en una casilla o en otra. Pero hemos creído que esta información era de suma importancia para estudios posteriores que podamos acometer o que puedan acometer otros lingüistas interesados en este campo, y, por ello mismo, nos hemos arriesgado a presentarla.

Tabla 7.4. El cuerpo humano

Fenómenos de contacto lingüístico		
CORPUS CASTELLANO		
Interferencias		
avantbrazo*, mugrón*, olfacto*		
Cambio de código		
alimentació, anus, barbeta, braç, bufeta, cama, colze, cuixa, galta, geniva, gluti, malaltia, nas, oïda, òvul, penis, pïlor		
Préstamo en vías de integración		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Compuestos
Calcos semánticos		
Calcos sintácticos		

Tabla 7.5. El cos humà

Fenómenos de contacto lingüístico CORPUS CATALÁN		
Interferencias		
antebraç*, enfermetat*, meñic*, moflet*, pàrpada*, pessó*, testícul*		
Cambio de código		
ano, articulación, barbilla, brazo, cana, chocho, costilla, esófago, espinilla, esternocleidomastoideo, esternón, gemelo, glúteo, grasa, menisco, nalga, oído, pene, perilla, rabadilla, rabo, tendón, tobillo, vello		
Préstamo en vías de integración		
muñeca		
Préstamo integrado		
hombro		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Compuestos
Calcos semánticos		
Calcos sintácticos		

Tabla 7.6. Ropa y complementos

Fenómenos de contacto lingüístico		
CORPUS CASTELLANO		
Interferencias		
abrigo*		
Cambio de código		
brusa, calces, espardenya, faldilla, jaqueta, mitgeta, mitjó, mocador, passamuntanyes, sabatilla, (sabatilla) esportiva, samarreta, sostenidor, vel, xandall		
Préstamo en vías de integración		
armilla		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Con algún elemento híbrido
<i>samarreta</i> con manga larga, <i>samarreta</i> corta, <i>samarreta</i> sin mangas	camiseta <i>arrapada</i> , pantalón <i>de xandall</i> , <i>zapatilla esportiva</i>	zapato esportivo*
Calcos semánticos		
clip ‘pasador’		
Calcos sintácticos		

7.7. Roba i complements

Fenómenos de contacto lingüístico CORPUS CATALÁN		
Interferencias		
albornós*, camisó*, calentadors*, colgant*, xalec*		
Cambio de código		
albornoz, babero, batín, blusón, bolsa, botón, braga, calcetín, calzoncillo, camisón, chubasquero, chupa, cinturón, collarín, cordón, corpiño, corsé, delantal, faja, hombrera, leotardo, maletín, manguito, mariconera, media, minifalda, monedero, mono, mono de esquí, muñequera, nomeolvides, pantalón, pantalón corto, pelele, peluca, riñonera, sello, sombrero, sujetador, tabardo, tejano, vaquero, viso, (zapatilla) deportiva, zapato		
Préstamo en vías de integración		
blusa, chaleco, chándal ⁸⁶⁹ , pajarita, sostén, sudadera, traje		
Préstamo integrado		
bolso, (forro) (polar), gorro, sostens*		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Compuestos
caçadora <i>tejana</i>	<i>pantalón</i> curt, <i>traje</i> jaqueta	
Calcos semánticos		
falda ‘faldilla’		
Calcos sintácticos		

⁸⁶⁹ NDLLC incluye *txàndal*, lo que nos da una idea de la extensión social de esta palabra.

7.8. Partes de la casa (sin los muebles)

Fenómenos de contacto lingüístico		
CORPUS CASTELLANO		
Interferencias		
enván*, fonamento*, soterráneo*		
Cambio de código		
armari, bidet, claveguera, despatx, escala, façana, fumeral, jardí, menjador, parquet, passadís, pica, porxo, quarto de rentar, rajola, rebedor, safareig, saleta, sostre, teulada, xemeneia		
Préstamo en vías de integración		
(es)golfá, (es)golfas*, rebost		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Compuestos
Calcos semánticos		
Calcos sintácticos		
cielo abierto (>celobert, ‘patio de luces’), fuego a tierra (>foc a terra, ‘fuego, chimenea, hogar’)		

7.9. *Parts de la casa (sense els mobles)*

Fenómenos de contacto lingüístico		
CORPUS CATALÁN		
Interferencias		
garaig*, tabic*, sòtan*		
Cambio de código		
altillo, aseo, cañería, carpa, cemento, cocina, comedor, desván, enchufe, escalera, garaje, habitación de los invitados, habitación de los niños, habitación de matrimonio, lavadero, pared de contrapeso, pasillo, piso, porche, ropero, salita, salón, sótano, tabique, techo, terraza, trastero		
Préstamo en vías de integración		
buhardilla, despensa, recibidor		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Con algún elemento híbrido
		armari roper*
Calcos semánticos		
bodega ‘celler, rebost’ ⁸⁷⁰		
Calcos sintácticos		
quart de bany (> <i>cuarto de baño</i> , ‘bany, cambra de bany’), pati de llum (> <i>patio de luces</i> , ‘celobert’)		

⁸⁷⁰ También es un *préstamo en vías de integración*.

7.10. Muebles de la casa

Fenómenos de contacto lingüístico		
CORPUS CASTELLANO		
Interferencias		
calajera*, microonas*, prestaje*		
Cambio de código		
armari, armari de coberteria, bagul, bidet, bufet, cadira, calaixera, capçalera del llit, (em)postada, escriptori, estant, habitació, moble bar, pica, prestatge, rebedor, rentadora, safareig, tassa del vàter		
Préstamo en vías de integración		
escritorio		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Compuestos
<i>armari empotrado,</i> <i>prestatge de los libros</i>	<i>mesa de l'ordinador</i>	
Calcos semánticos		
Calcos sintácticos		
fuego a tierra (> <i>foc a terra</i> , 'fuego, chimenea, hogar')		

7.11. Els mobles de la casa

Fenómenos de contacto lingüístico			
CORPUS CATALÁN			
Interferencias			
arcó*, biom*, butacó*, escritori*, porticón*, taburet*			
Cambio de código			
alfombra, almohada, armario, armario de la cocina, bombilla, cama, [cama] nido, colchón, cuna, despensa, edredón, encimera, horno, lavadora, mármol, mecedora, mesa, mesita de centro, mostrador, ordenador, plegatín, recibidor, revistero, rinconera, secador, secadora, sillón, sofà cama, tumbona, zapatero			
Préstamo en vías de integración			
lámpara, mesita, silló*			
Préstamo integrado			
Combinaciones híbridas			
Núcleo en catalán	Núcleo en castellano	Compuestos	Con algún elemento híbrido
armari <i>del comedor</i> , taula <i>del ordenador</i>	armario de la cuina, botiquín de primers auxilis, mármol de la cuina, lámpara de peu, mesa de la tele(visió), mesita de nit	fregaplats*	armari empotrat* ⁸⁷¹ , armari roper*
Calcos semánticos			
hamaca ‘gandula, balancí’			
Calcos sintácticos			

⁸⁷¹ También es un *préstamo en vías de integración*.

7.12. Alimentos y bebidas

Fenómenos de contacto lingüístico		
CORPUS CASTELLANO		
Interferencias		
galletas*, mandonguillas*		
Cambio de código		
aigua, albergínia, anxova, bacallà, carxofa, cogombre, enciam, figa, isard, lluç, macarró, melmelada, mongeta, pa amb tomata, pastanaga, plàtan, poma, pruna, síndria, sípia, taronjada, vedella, vi		
Préstamo en vías de integración		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Con algún elemento híbrido
		pollo alás*
Calcos semánticos		
biquini ‘sándwich’, champú ‘cerveza con limonada’		
Calcos sintácticos		
jamón dulce (> <i>pernil dolç</i> , ‘jamón de york’), jamón salado (> <i>pernil salat</i> , ‘jamón serrano’)		

7.13. *Aliments i begudes*

Fenómenos de contacto lingüístico		
CORPUS CATALÁN		
Interferencias		
almejes*, bogavant*, chucheries*, chuletes*, empanadilles*, galletes*, golosines*, gominoles*, grasses*, guindilles*, melocotó*, ràban*, salchichó*, xorís*		
Cambio de código		
aguacate, alcachofa, arroz, atún, berberecho, bocata, bollería, boquerón, cacahuete, canela, carajillo, cebolla, chorizo, chucho, chupito, ensaladilla, flan, fresa, gazpacho, judía blanca, judía verde, lima, limón, limonada, lubina, lucio, melón, merluza, paraguay, pastel, pato, pavo, pechuga, pepino, pincho, plátano, ponche, puerro, pulpo, ron cola, salchicha, salmón, sandía, sifón, solomillo, tarta, ternasco, ternera, tocinillo, tocino, tortilla, zanahoria, zumo		
Préstamo en vías de integración		
gaseosa, mermelada, ron		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Compuestos
suc <i>de uva</i> , vodka <i>amb pomelo</i>		
Calcos semánticos		
palmera ‘ulleres’		
Calcos sintácticos		
ou fregit (> <i>huevo frito</i> , ‘ou ferrat’)		

7.14. *Objetos colocados en la mesa para la comida*

Fenómenos de contacto lingüístico			
CORPUS CASTELLANO			
Interferencias			
drapo*, ganivete*, horquilla*, sellón*			
Cambio de código			
aigua, càntir, estalvis, gerra, gerro de l'aigua, gerro del vi, porró, safata, setrill, setrilleres, tovalló			
Préstamo en vías de integración			
estovalles, plata			
Préstamo integrado			
Combinaciones híbridas			
Núcleo en catalán	Núcleo en castellano	Compuestos	Con algún elemento híbrido
<i>gerra</i> de agua		trespies*	ganivete* del pan
Calcos semánticos			
argolla 'servilletero', paella 'sartén', toallas 'mantel'			
Calcos sintácticos			
copa de vino negro (>[<i>copa de</i>]vi negre, 'vino tinto'), plato fondo (>plat fondo, 'plato hondo')			

7.15. *Objectes col·locats a la taula pel dinar*

Fenómenos de contacto lingüístico			
CORPUS CATALÁN			
Interferencias			
alinyadors*, bandejes*, paliller*, servilleter*, trap*			
Cambio de código			
adorno, bandeja, botella de gaseosa, candelabro, copa de agua, copa de vino, cubitera, cuchara, cuchara de postre, cucharón, cuchillo, cuchillo grande, encendedor, ensaladera, espumadera, fiambarrera, florero, gaseosa, mahonesa, mando, mando del televisor, mantelería, mostaza, palillo, pincho, plato, plato de carne, platô hondo, porrón, posavasos, puchero, rasera, salvamanteles, sartén, sifón, tapete, tarro, tenedor, tenedor de carne, tenedor de pescado, termo, vaso, vela			
Préstamo en vías de integración			
mantell*, servilleta			
Préstamo integrado			
mantel			
Combinaciones híbridas			
Núcleo en catalán	Núcleo en castellano	Compuestos	Con algún elemento híbrido
ampolla de <i>gaseosa</i> , got de <i>tubo</i> , plat <i>llano</i> , vas de <i>xupito</i>	<i>mando</i> de la tele(visió)	posagots*, salvaestovalles*, salvamantells*	mantell* individual
Calcos semánticos			
font ‘plata, plàtera’			
Calcos sintácticos			

7.16. La cocina y sus utensilios

Fenómenos de contacto lingüístico		
CORPUS CASTELLANO		
Interferencias		
agüera*, cazón*, drapo*, escombrarias*, escorredera*, ganivete*, graellas*, horquilla*		
Cambio de código		
aigüera, armari, batedora, cassola, davantal, destral, drap, escorredor, fogó, fruitera, fusta, gerro, gibrell, microones, setrill, torradora, tupí		
Préstamo en vías de integración		
pica, plata		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Compuestos	Con algún elemento híbrido
<i>gerra</i> del agua.	rentaplatos*	ganivete* de sierra
Calcos semánticos		
paella ‘sartén’ (paella grande, paella mediana, paella pequeña) ⁸⁷²		
Calcos sintácticos		

⁸⁷² También es un *préstamo integrado*. Hemos incluido en su número de menciones (77) las tres producidas por un solo informante relativas al tamaño del objeto (“paella pequeña”, “paella mediana” y “paella grande”), que dan fe de su uso como equivalente a una sartén cualquiera. En este caso, el número de menciones del término no coincidiría con el de informantes que lo actualizan, por tanto.

7.17. La cuina i els seus utensilis

Fenómenos de contacto lingüístico		
CORPUS CATALÁN		
Interferencias		
alinyadors*, assafrà*, devantal*, mantell*, microondes*, pany*, servilletes*, taburet*, trap*		
Cambio de código		
abridor, armario, bandeja, basura, butano, cacerola, calentador, cazo, cubitera, cubo de la basura, cuchara, cuchillo, delantal, despensa, encimera, escurridor, espumadera, estropajo, fiambra, fregadero, fregona, freidora, grifo, hacha, lavadora, manopla, mechero, mopa, palillo, papelera, parrilla, pincho, plato, rasera, rodillo, secadora, servilleta, servilletero, tenedor, termo, tostadora		
Préstamo en vías de integración		
batidora, cucharón, exprimidora, sandwichera, sartén		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Compuestos
rellotge grande	<i>exprimidora de suc</i>	fregaplats*, lavavaixelles*, lavavajilles*, salvamantells*, secaplats*
Calcos semánticos		
font ‘plata, plàtera’		
Calcos sintácticos		

7.18. La escuela: muebles y materiales

Fenómenos de contacto lingüístico		
CORPUS CASTELLANO		
Interferencias		
jaquetas*, prestaje*, prestajería*		
Cambio de código		
cadira, canó, cartabó, esborrador, escaire, estoig, gomet, mocador, motxilla, ordinador, paperera, penja-robres, penjador, porter, postada, prestatge, suro, trepadora		
Préstamo en vías de integración		
maquineta		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Compuestos
	mesa <i>de l'ordinador</i> , tablero de <i>suro</i>	cuelga-ropas*
Calcos semánticos		
Calcos sintácticos		

7.19. L'escola: mobles i materials

Fenómenos de contacto lingüístico		
CORPUS CATALÁN		
Interferencias		
canastes*, colchonetes*, enchuf*, escritori*, maniveles*, sillo* , taburet*		
Cambio de código		
atlas, basura, bolígrafo, canasta, cañón, casco, colgador, corcho, fichero, folio, fosforito, goma de borrar, lámpara, lápiz, monedero, mostrador, papelera recicladora, pared de corcho, pasillo, perchero, potro, taladradora, taladro, televisión, ténpera, tiza		
Préstamo en vías de integración		
carpessà*, pegament*, tissa*, carpesano, mochila, ordenador, papelera		
Préstamo integrado		
borrador		
Combinaciones híbridas		
Núcleo en catalán	Con algún elemento híbrido	Compuestos
	material deportiu*, tabló* d'anuncis, canastes* de bàsquet, aire acondicionat*	sacapuntes*
Calcos semánticos		
esquadra 'escaire' ⁸⁷³ , taquilla 'armariet, caseller' ⁸⁷⁴		
Calcos sintácticos		
tauló d'anuncis (>tablón de anuncios, 'tauler d'anuncis')		

⁸⁷³ También es un *préstamo en vías de integración*.

⁸⁷⁴ También es un *préstamo en vías de integración*.

7.20. Iluminación, calefacción y medios para airear un recinto

Fenómenos de contacto lingüístico		
CORPUS CASTELLANO		
Interferencias		
porticón*		
Cambio de código		
bombeta, calefacció de gasoil, corrent d'electricitat, díode, espelma, fluorescent, gas ciutat, llanterna, misto, vano ⁸⁷⁵ , xemeneia		
Préstamo en vías de integración		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Compuestos
	ojo <i>de bou</i>	
Calcos semánticos		
Calcos sintácticos		
aire condicionado > <i>aire condicionat</i> 'aire acondicionado', cielo abierto > <i>celobert</i> 'patio de luces', fuego a tierra > <i>foc a terra</i> 'fuego, chimenea, hogar' ⁸⁷⁶		

⁸⁷⁵ Aunque esta palabra también existe en castellano y, de hecho, aparece en algunos léxicos disponibles en este mismo centro de interés, como por ejemplo en Soria (Bartol 2004), aquí el hablante actualiza el término homógrafo del catalán (equivalente a "abanico"), tal como deducimos al comparar la frecuencia de mención de ambos términos en los dos corpus.

⁸⁷⁶ También es un *préstamo en vías de integración*.

7.21. *l*-luminació, calefacció i mitjans per airejar un recint

Fenómenos de contacto lingüístico			
CORPUS CATALÁN			
Interferencias			
alfombres*, canyeries*, lampareta*, llàmpara*, llinterna*, rejilles*, tuberies*			
Cambio de código			
abanico, acondicionador, aplique, butano, enchufe, esterilla, estufa de queroseno, hoguera, lámpara de gas, linterna, mechero, neón, respiradero, tubería, tubo			
Préstamo en vías de integración			
bombilla, calentador, vela			
Préstamo integrado			
foco, lámpara			
Combinaciones híbridas			
Núcleo en catalán	Núcleo en castellano	Compuestos	Con algún elemento híbrido
manta <i>calentadora</i>	<i>globo</i> de llum, <i>lámpara</i> de nit, <i>lámpara</i> de peu	portalámpares*	lampareta* de nit, (aparell d')aire acondicionat* ⁸⁷⁷
Calcos semánticos			
farola 'fanal'			
Calcos sintácticos			

⁸⁷⁷ También es un *préstamo integrado*.

7.22. La ciudad

Fenómenos de contacto lingüístico CORPUS CASTELLANO		
Interferencias		
carrejuela*, poliesportivo*, sabaterias*, vianante*		
Cambio de código		
ajuntament, avinguda, bomber, botiga, casal, cinema, claveguera, consell comarcal, església, jardi, mosso (d'esquadra), neó, parella, pastisseria, pavelló, restaurant, rogalia, semàfor, tiquet		
Préstamo en vías de integración		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Con algún elemento híbrido	Compuestos
<i>bufet</i> de abogados	paso de vianantes*, recinto esportivo*	gratacielo*
Calcos semánticos		
fanal 'farola'		
Calcos sintácticos		

7.23. La ciutat

Fenómenos de contacto lingüístico		
CORPUS CATALÁN		
Interferencias		
abogat*, agobi*, agobiant*, almacens*, alquitrà*, barull*, chucheries*, columpi*, escaleres*, lámpares*, palomes*, papeleres*, papeleries*, peluqueries*, polideportiu*, vivenda*		
Cambio de código		
andén, atasco, basura, bolera, botellón, ceda, chiringuito, club de alterne, gato, gimnasio, golondrina, iglesia, jardín, mendigo, mercadillo, monovolumen, museo, panadería, peluquería, perro, pijo, pito, playa, red eléctrica, semáforo, tenderete, valla, zona peatonal		
Préstamo en vías de integración		
acera, peató*		
Préstamo integrado		
Combinaciones híbridas		
Con algún ele híbrido	Núcleo en castellano	Compuestos
pas de peatons* ⁸⁷⁸		rascacels*
Calcos semánticos		
bodega ‘celler’, farola ‘fanal’ ⁸⁷⁹		
Calcos sintácticos		
medi de comunicació (> <i>medio de comunicación</i> ‘mitjà de comunicació)		

⁸⁷⁸ También es un *préstamo en vías de integración*.

⁸⁷⁹ También es un *préstamo integrado*.

7.24. *El campo*

Fenómenos de contacto lingüístico		
CORPUS CASTELLANO		
Interferencias		
embasamiento*, hivernaclo*, ramadero*, ramado*, tancas*		
Cambio de código		
adob, aigua, aixada, alfals, alzina, arada, blat, bolet, brutícia, carretó, celebració, collita, destral, fita, humitat, insecte, isard, mangrana, natura, ordi, oxigen, palet, palot, panís, papallona, pinso, plataner, porc fer, pou, ramaderia, regadiu, rierol, roure, solitud, terra, truja, trumfa, turó, xiprer		
Préstamo en vías de integración		
pagès		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Con un ele híbrido
		einas* de trabajo
Calcos semánticos		
cucharita ‘renacuajo’		
Calcos sintácticos		
fuego a tierra (> <i>foc a terra</i> , ‘fuego, chimenea, hogar’)		

7.25. *El camp*

Fenómenos de contacto lingüístico		
CORPUS CATALÁN		
Interferencias		
aislament*, avespa*, avispes*, cansanci*, carner*, culebres*, deport*, estanc*, gaviotes*, hortelà*, margarites*, matorrals*, rastrell*		
Cambio de código		
abonadora, abono, amapola, árbol, ardilla, bicho, borrego, césped, cisne, columpio, cortijo, corzo, cosechadora, crío, dominguero, gravilla, hachís, jabalí, laguna, libertad, manguera, margarita, mariquita, matadero, mono, mosquito, naturaleza, paloma, pantano, pato, poste de gas, pradera, rama, regadera, roble, saltamontes, sapo, sombra, tábano, tocino, topo, tormenta, trigo, valla, viñedo		
Préstamo en vías de integración		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Compuestos
	<i>cámara</i> de fruita, <i>poste</i> de llum, <i>poste</i> de telèfon	
Calcos semánticos		
Calcos sintácticos		

7.26. Medios de transporte

Fenómenos de contacto lingüístico		
CORPUS CASTELLANO		
Interferencias		
Cambio de código		
globus, monopatí, rai, raier		
Préstamo en vías de integración		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Compuestos
Calcos semánticos		
toro ‘carretilla elevadora’ ⁸⁸⁰		
Calcos sintácticos		

⁸⁸⁰ En el centro 13 aparece “torero”, que podría ser otro calco semántico a semejanza del presente. Sin embargo, finalmente lo hemos descartado por no estar seguros de si el único hablante que lo incluye se refiere al arte del toreo o al significado del mismo término en catalán, esto es, a la ‘persona que trabaja en el manejo de un toro o carretilla elevadora’, lo cual parece, en principio, más probable.

7.27. *Mitjans de transport*

Fenómenos de contacto lingüístico		
CORPUS CATALÁN		
Interferencias		
assafata*, crucer*, deportiu*, foraborda*, llançadera*, navier*, petroler*, planejador*, rancheres*, yat*		
Cambio de código		
a caballlito, aeronave, andando, arrastre, avión, buga, camión, catamarán, coche de San Fernando, cosechadora, embrague, globo, golondrina, helicóptero, monopatín, monovolumen, patín, patinete, remolque, tabla de snow, telearrastre, triciclo, trolebús, velero, yate, zepelín		
Préstamo en vías de integración		
barco, lancha, patera		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Con algún elemento híbrido
	<i>globo aerostàtic</i>	vehicle mixte*
Calcos semánticos		
Calcos sintácticos		
taula de snow > <i>tabla de snow</i> ‘planxa de snow’, taula de surf > <i>tabla de surf</i> ‘planxa de surf’		

7.28. Trabajos del campo y del jardín

Fenómenos de contacto lingüístico		
CORPUS CASTELLANO		
Interferencias		
conreo*, plantero*		
Cambio de código		
aclaridor, aclarir, boletaire, camperol, collar, collir, conreador, conrear, destriar, empeltar, ensinitrador, ensofrar, ensulfatador, esgotador, esllucar, esporgador, jardiner, llaurar, munyidor de vaques, munyir, passar la curra, plantació, ramader, ramaderia, rampilladora, rampillar, recol·lecció, recollida de fruita, regadiu, regadora, rotació dels conreus, tallar, tofonaire, trisellar, veremar, xollador		
Préstamo en vías de integración		
ensulfatar, esporgar, pagès, ramadero*		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Con algún elemento híbrido
<i>tallar</i> la hierba	recoger <i>préssecs</i> , transportar <i>palots</i>	cortar las tijas*, recoger pomas*
Calcos semánticos		
adobar(la tierra) ‘abonar’, adobo ‘abono’		
Calcos sintácticos		

7.29. *Treballs del camp i del jardí*

Fenómenos de contacto lingüístico		
CORPUS CATALÁN		
Interferencias		
carpinter*, ganader*, goteig*, hortelà*, llimpiar *, panader*, rastrell*		
Cambio de código		
abono, apisonar, arar, cambiar tubos, cazador, coger, cosecha, cosechador, descogollar, escarbar, escobar, injertar, labrar, matar bichos, ordeñar, pastar, piar, rastrillo, secar, tirar abono, vendimiari		
Préstamo en vías de integración		
cosechar		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Con algún elemento híbrido
<i>tallar</i> el césped		ingenier* de camins, ingenier* agrònom, lligar rames*
Calcos semánticos		
abonar ‘adobar’ ⁸⁸¹		
Calcos sintácticos		

⁸⁸¹ También es un *préstamo en vías de integración*.

7.30. *Animales*

Fenómenos de contacto lingüístico		
CORPUS CASTELLANO		
Interferencias		
ciérvol*, dofín*, ratolín*, vestruz*		
Cambio de código		
àliga, balena, bou, cadenera, carderola, cocodril, corder, cuc, esquiroi, euga, floquet de neu, gat, gos, guatlla, guineu, isard, llimac, musclo, mussol, ocell, oreneta, peix, petxina, pit-roig, porc, porc fer, puput, teranyina, voltor		
Préstamo en vías de integración		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Compuestos
Calcos semánticos		
helicóptero ‘libélula’		
Calcos sintácticos		
oso bruno (>ós <i>bru</i> , ‘oso pardo’)		

7.31. *Animals*

Fenómenos de contacto lingüístico		
CORPUS CATALÁN		
Interferencias		
avespa*, boqueró*, carner*, codorniu*, huró*, llagart*, lombriu*, murcièlag*, pelícan*, picantó*, pulgues*, sap*, terner*, tocí*, vestruç*		
Cambio de código		
alburno, alce, almeja, ardilla, atún, avestruz, avispa, berberecho, bermejuela, buey, búfalo, búho, buitre, camaleón, canario, cangrejo de mar, cangrejo de río, caracola, cardelina, cerda, chipirón, cigüeña, cisne, cocodrilo, corzo, cucaracha, culebra, faisán, flamenco, garrapata, gato, gaviota, golondrina, gorrión, guepardo, hipopótamo, jilguero, leopardo, lince, loro, lubina, lucio, mandarín, mantis (religiosa), mariquita, mejillón, mosquito, murciélago, ornitorrinco, oruga, oso, oso polar, pájaro, paloma, pato, pavo, pavo real, pelícano, perico, perro, pez araña, pez espada, pez globo, pez payaso, pingüino, pollo, potro, pulpo, rana, raya, reno, rinoceronte, saltamontes, serpiente, siluro, símio, ternero, tiburón, tiburón blanco, tiburón toro, topo, tórtola, tritón, trucha, vaquilla, víbora, yegua, zapatero, zorro		
Préstamo en vías de integración		
delfí*, ballena, jabalí, tiburó*, tocino		
Préstamo integrado		
mono		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Compuestos
	<i>copito</i> de neu	
Calcos semánticos		
Calcos sintácticos		

7.32. Juegos y distracciones

Fenómenos de contacto lingüístico		
CORPUS CASTELLANO		
Interferencias		
chafardear*, endevinanzas*, patacones*		
Cambio de código		
bitlles, cadira, cinema, coit, corda, corranda, cucut i amagar, escacs, jo mai, mocador, monopatí, ordinador, palet, pilota, pitxi, siset		
Préstamo en vías de integración		
botifarra, handbol		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Con algún elemento híbrido
<i>baldufa</i> saltarina	juego de <i>ordinador</i>	hacer cursas*
Calcos semánticos		
Calcos sintácticos		

7.33. *Jocs i distraccions*

Fenómenos de contacto lingüístico		
CORPUS CATALÁN		
Interferencias		
caniques*, crucigrames*, manualitats*, mentirós*, paracaidisme*		
Cambio de código		
ahorcado, ajedrez, atletismo, baloncesto, balonmano, barcos, bolera, bolos, bote bote, botellón, chiste, chivarse, comba, conejo de la suerte, corazones, corro de la patata, dados, frontón, fútbol playa, fútbolín, jabalina, laberinto, me gustas tú, natación, novia, palitos, pértiga, pito pito, polis y cacos, pozo, puenting, pulso, rabino, rellano, remigio, sexo, soldadito inglés, solitario, tazos, televisión, tiro con arco, tute, tragaperras, veo veo		
Préstamo en vías de integración		
escondite, ordenador, teto, uno		
Préstamo integrado		
pilla pilla, video(console)		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Compuestos
gallineta <i>ciega</i> , joc de la <i>escalera</i> , prendre <i>algo</i> , soldadet <i>inglés</i>	<i>soldadito</i> anglès	
Calcos semánticos		
Calcos sintácticos		
anar de tendes (> <i>ir de tiendas</i> ‘anar o sortir a comprar’)		

7.34. Profesiones y oficios

Fenómenos de contacto lingüístico		
CORPUS CASTELLANO		
Interferencias		
advocado*, cambrero*, encargado*, ingeniero*, ferrero*, fornero*, fustero*, libretero*, pastisero*, perruquero*, pescatero*, proveedores*, ramadero*		
Cambio de código		
banquer, botiguer, camioner, enginyer, escriptor, esportista, funcionari, fuster, hostessa, infermeria, informàtic, joier, mainadera, manobre, mantenidor de material, metge, ministre, pagès, polític, professor de català, ramader		
Préstamo en vías de integración		
infermera, mosso (d'esquadra)		
Préstamo integrado		
paleta		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Con algún elemento híbrido
venedor de fruta		ingeniero* agrónomo
Calcos semánticos		
lampista 'electricista' o 'fontanero'		
Calcos sintácticos		
mujer de hacer faenas (> <i>dona de fer feines</i> , 'señora de la limpieza')		

7.35. Professions i oficis

Fenómenos de contacto lingüístico		
CORPUS CATALÁN		
Interferencias		
assafata*, barrender*, basurer*, carpinter*, chatarrer*, cirugià*, cociner*, cristaler*, ganader*, panader*, recader*		
Cambio de código		
actriz, adivino, administrativo, albañil, aparcacoches, asistenta, asistente, asistente social, astrónomo, azafata, bailador, barrendero, basurero, biólogo, boina verde, bombero, butanero, cabo, camionero, canguro, carnicero, carpintero, catador, chacha, chapero, chulo, cocinero, cura, deportista, enfermería, enfermera, esteticien, fontanero, geo, guardia urbano, ingeniería, jefe, masajista, mercenario, modelo, monje, ordeñador, paro, peluquero, peón, pícaro, piloto, reponedor, segurata, soldado, vendedor		
Préstamo en vías de integración		
abogat*, camarer*, enfermer*, ingenier*, peluquer*		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Con algún elemento híbrido
		ingenier* aeronàutic, ingenier* agrònom, ingenier* industrial
Calcos semánticos		
fontaner ‘llauner, lampista’ ⁸⁸²		
Calcos sintácticos		

⁸⁸² También es un *préstamo integrado*.

7.36. Colores

Fenómenos de contacto lingüístico		
CORPUS CASTELLANO		
Interferencias		
carabaza*, daurado*, torrado*		
Cambio de código		
blanc, blau, blaugrana, carn, cru, groc, groc fluix, groc fort, groc pà·lid, merda d'oca, negre, roig, taronja, verd, vermell		
Préstamo en vías de integración		
Préstamo integrado		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Compuestos
Calcos semánticos		
Calcos sintácticos		
azul flojo, azul fuerte, marrón flojo, marrón fuerte, rosa fuerte, verde flojo, verde fuerte (> <i>nom color</i> + <i>fort, fluix</i> 'nombre color + oscuro, claro')		

7.37. Colors

Fenómenos de contacto lingüístico		
CORPUS CATALÁN		
Interferencias		
cobris*		
Cambio de código		
amarillo, amarillo claro, amarillo oscuro, ámbar, azul, azul claro, azul oscuro, berenjena, blanco, blanco lechoso, canela, esmeralda, fosforito, lila fosforito, lila melón, marrón claro, marrón oscuro, mate, naranja, negro, negro oscuro, pelirroja, plateado, porcelana, rojo, rojo claro, rosa fosforito, salmón, verde claro, verde fosforito, verde oscuro		
Préstamo en vías de integración		
marrón		
Préstamo integrado		
dorat*, granate		
Combinaciones híbridas		
Núcleo en catalán	Núcleo en castellano	Compuestos
blau <i>fosforito</i> , blau <i>pitufu</i> , groc <i>chillón</i> , groc <i>fosforito</i> , roig <i>terciopelo</i> , taronja <i>fosforito</i> , verd <i>esmeralda</i> , verd <i>fosforito</i> , verd <i>lima</i> .	verde poma	
Calcos semánticos		
Calcos sintácticos		

7.3. ANÁLISIS CUANTITATIVO

A continuación, pasamos a estudiar el peso cuantitativo que los fenómenos de contacto lingüístico adquieren sobre el conjunto total de los datos. Para ello, analizamos, en primer lugar, la cifra absoluta de transferencias halladas en cada lengua (tabla 7.38) y su distribución según la clasificación que hemos esbozado en el apartado anterior (gráficos 7.1 y 7.2):

7.38. Frecuencia absoluta y relativa de los fenómenos de contacto lingüístico en las dos lenguas⁸⁸³.

Tipos de Fenómenos	Presentes en el corpus castellano	%	Presentes en el corpus catalán	%
Interferencias	62	12.78%	140	14.68%
Cambios de Código	332	68.45%	640	67.09%
Préstamos en vías de integración	19	3.92%	61	6.39%
Préstamos Integrados	2	0.41%	17	1.78%
Combinaciones Híbridas	36	7.42%	75	7.86%
<i>Núcleo Castellano</i>	<i>10</i>	<i>27.8%</i>	<i>20</i>	<i>26.67%</i>
<i>Núcleo Catalán</i>	<i>11</i>	<i>30.6%</i>	<i>25</i>	<i>33.33%</i>
<i>Con algún elemento híbrido</i>	<i>11</i>	<i>30.6%</i>	<i>18</i>	<i>24%</i>
<i>Compuestos</i>	<i>4</i>	<i>11.1%</i>	<i>12</i>	<i>16%</i>
Calcos Semánticos	14	2.89%	13	1.36%
Calcos Sintácticos	20	4.12%	8	0.84%

⁸⁸³ Observará el atento lector que la suma de los distintos elementos no coincide con la que ofrecemos en la tabla 7.42, donde damos el total de transferencias en cifras absolutas en las dos lenguas. Esto es debido a que, como ya hemos explicado, hay algunas formas, *calcos semánticos* o *sintácticos* y *combinaciones híbridas*, que también son —por su porcentaje de aparición— *préstamos*, lo que provoca que estén sumadas en los dos grupos y, por tanto, aparezcan aquí dos veces. Esto sucede con 2 ejemplos del castellano y con 9 del catalán, de ahí que deban restarse estas cifras al global para obtener la cantidad exacta que señalamos en la tabla 7.42 (*infra*).

Gráfico 7.1. Distribución de los fenómenos de contacto lingüístico en el corpus castellano.

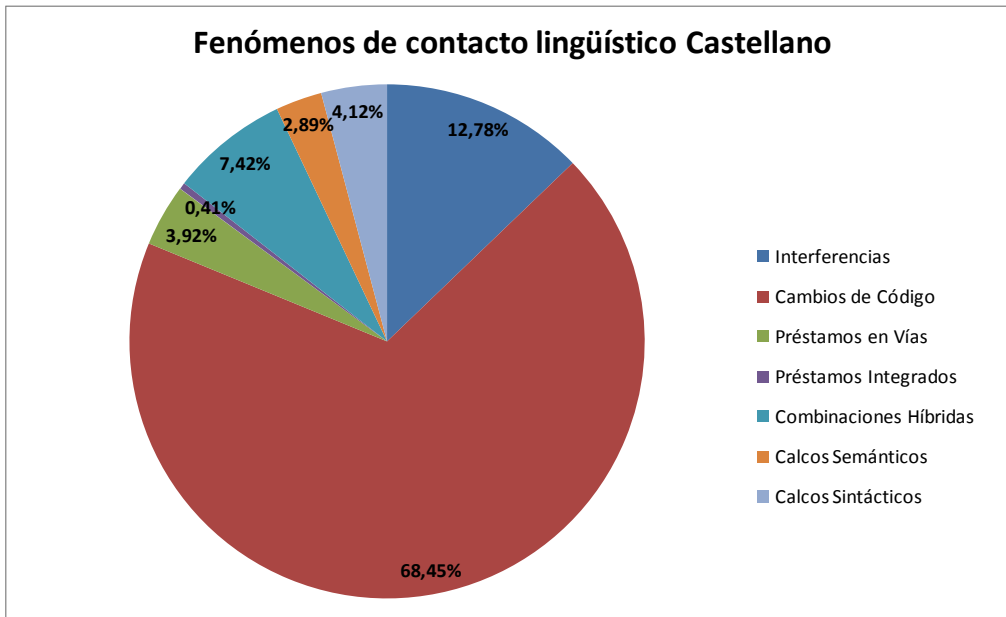
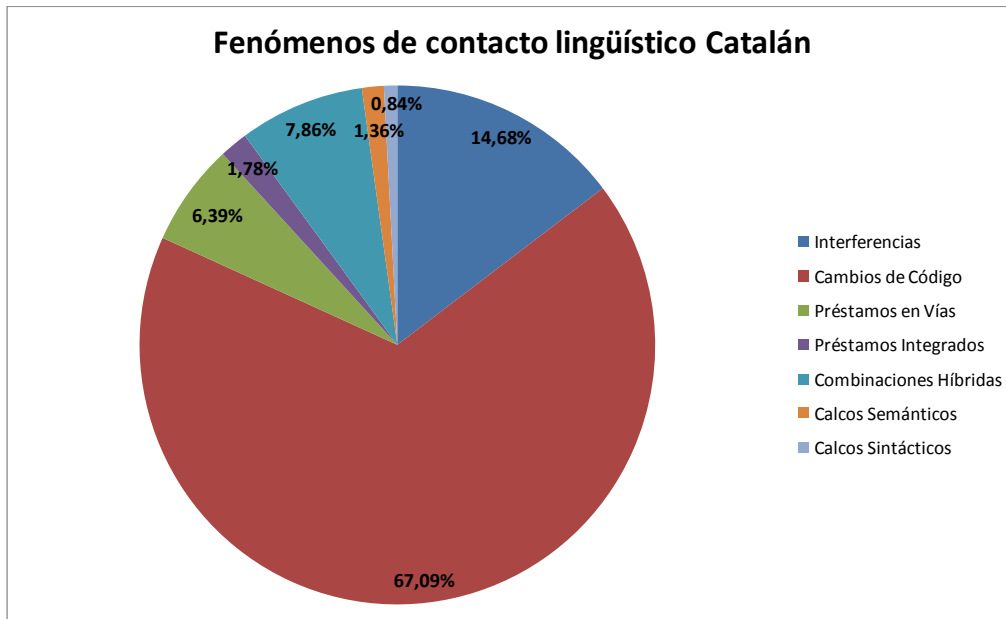


Gráfico 7.2. Distribución de los fenómenos de contacto lingüístico en el corpus catalán.



Tal como se observa en la tabla 7.38, el número de transferencias es superior en lengua catalana en prácticamente todas las categorías salvo en los *calcos sintácticos*, donde en castellano se recaba una cifra superior. De otra parte, en relación con la clasificación de los elementos, las dos lenguas presentan una distribución muy similar, donde los *cambios de código*, con una proporción prácticamente idéntica (68.45% en castellano y 67.09% en catalán), son la principal categoría. Las *interferencias* son el segundo tipo de fenómeno más frecuente en los dos *corpus* con mucha distancia, pero es

llamativo que también su porcentaje sea parecido en los dos idiomas (12.78% en castellano y 14.68% en catalán). En realidad, si sumamos el alcance de estos dos fenómenos (lo que representa un poco más del 80% del total), podemos afirmar que la mayoría de los elementos recogidos son actualizaciones individuales, dado que estos dos tipos de transferencias tienen una frecuencia de mención baja, de como máximo 11 menciones por fenómeno, como hemos indicado en el apartado anterior.

El hecho de que estos dos fenómenos sean los más frecuentes nos da una idea de la “separación” que existe entre los dos idiomas, toda vez que los elementos que gozan de una frecuencia de mención elevada y que, por ende, tienen un cierto grado de integración social (los *préstamos integrados* o *en vías de integración*) son escasos en ambas lenguas. Los dos idiomas, pese a la convivencia continuada que han tenido a lo largo de la historia (*vid.* 4.1. y 4.3.2.), mantienen su autonomía y no se aprecia que el léxico de una de las dos lenguas corra el peligro de ser sustituido por el de la otra, aunque hay que reconocer que, en el caso del catalán, la influencia del castellano es mucho mayor que la que se aprecia en sentido contrario, como a continuación comentaremos. No obstante, el fenómeno que más veces se constata en nuestros listados es la presencia de un término gramatical y correcto en la otra lengua estudiada causada —en nuestra opinión— por un olvido momentáneo del término adecuado en la lengua de la encuesta⁸⁸⁴ (*cambio de código*), seguida a gran distancia de la producción errónea de un término por generalización de las normas de derivación de una de las dos lenguas (*interferencia*). Esto es, imperan las producciones individuales esporádicas fruto del idiolecto, ya sean dentro de la norma (*cambios de código*) o fuera de ella (*interferencias*).

Las *combinaciones híbridas*, de otra parte, también presentan una proporción muy parecida en los dos idiomas (7.42% en castellano y 7.86% en catalán). Estas expresiones también podrían considerarse como un tipo de *interferencia*, dado que son producciones agramaticales donde se produce una amalgama de las dos lenguas, aunque esta se produzca fuera de los límites de la palabra simple, ya sea en una lexía compuesta o en una combinación sintagmática. Es curioso cómo, aunque con cifras superiores en

⁸⁸⁴ Hay que recordar que el informante está bajo unas condiciones concretas de realización de la prueba y es consciente de que cuenta con un tiempo límite (*vid.* 5.3). Como ya hemos indicado, sería necesario cotejar estos datos con otros recogidos mediante otra metodología para comprobar si el informante realmente incorpora estos *cambios de código* en un discurso real espontáneo.

catalán, se da también la misma distribución dentro de la tipología de fenómenos de este tipo (con el núcleo en castellano, en catalán, con un elemento híbrido o compuestos). La mayoría de los elementos de esta clase también son producciones individuales, toda vez que ningún ejemplo supera las 11 menciones, al igual que las *interferencias* y los *cambios de código*, con la excepción de 3 casos en el corpus catalán, todos ellos *combinaciones sintagmáticas con un elemento híbrido* —armari empotrat*, (aparell d')aire acondicionat* y pas de peatons*— que rebasan, algunos ampliamente, este límite.

La semejanza entre la distribución de las dos lenguas al comparar tanto los *préstamos* como los *calcos* cambia ostensiblemente. La proporción de *préstamos*, elementos que tienen ya una cierta integración social en la lengua receptora y superan las menciones que establecimos como límite, es claramente superior en el corpus catalán. El porcentaje que suponen los préstamos integrados en lengua catalana (1.78%) cuadruplica la proporción que estos representan en el corpus castellano, donde solo suponen el 0.41% de las transferencias. En esta lengua solo se hallan dos ejemplos con más de 36 menciones (*paella* y *paleta*) mientras que, en catalán, hay 17 palabras en estas circunstancias (*vid.* 7.4). En cuanto a los *préstamos en vías de integración*, que cuentan con un grado relativo de integración social y son susceptibles de convertirse en *préstamos integrados*, la proporción también difiere en las dos lenguas. En castellano hallamos 19 transferencias que cuentan con entre 11 y 36 menciones, es decir, el 3.92%; por el contrario, en catalán, los casos de este tipo suponen el 6.39% del total de fenómenos de contacto recogidos y la nómina de ejemplos se amplía hasta 61 casos. Estos dos datos nos muestran cuál es la dirección en la que la interferencia se manifiesta mayoritariamente en nuestra provincia, dado que no solo el número total de fenómenos registrados es claramente superior en catalán, como a continuación detallaremos, sino que dentro de esta lengua las transferencias cuentan con una mayor integración social, como veremos en el análisis concreto de los préstamos (*vid.* 7.4.).

Con respecto a los *calcos*, también se observan diferencias en su producción según el idioma que analicemos, aunque en este caso, es en castellano donde se observa una proporción superior. Si bien los ejemplos de *calcos semánticos* son prácticamente idénticos en número en los dos *corpus* (14 en castellano y 13 en catalán), el porcentaje que estos suponen sobre el total de los fenómenos de contacto hallados es más elevado

en castellano (2.89% frente a 1.36%) debido a que otras categorías, como los *préstamos*, son mucho menos abundantes en esta lengua. De cualquier modo, no es insignificante la presencia de *calcos semánticos* en el corpus castellano, si la comparamos con la de otros elementos como los *préstamos integrados* o *en vías de integración*. Los *calcos sintácticos* son más abundantes en el corpus castellano, tanto en relación con los restantes elementos como en cifras absolutas frente al corpus catalán. Los casos de este tipo en castellano suponen el 4.12% del total de transferencias y engloban 20 casos, mientras que en catalán solo representan el 0.84% y únicamente contamos con 8 ejemplos de este tipo.

Hemos intentado hallar una explicación a estas diferencias entre la proporción de *préstamos* y *calcos* en las dos lenguas. A nuestro modo de ver, la mayor presencia de los *calcos*, sobre todo los *sintácticos*, en el corpus castellano, se explica, una vez más (*cf.* capítulo 6), por los dos perfiles que tenemos en nuestra muestra: el alumnado cuya lengua materna es el castellano y el alumnado cuya lengua materna es el catalán. No podemos perder de vista que, a lo largo del presente capítulo, estamos manejando los datos de la muestra en su globalidad, y no desglosados por alguna variable social como la *lengua materna*. Por este motivo, la distinta distribución de transferencias en las dos lenguas y la mayor presencia de *calcos* en castellano se explica porque la mayoría de los informantes (65.42%) tiene el catalán como lengua materna y, además, todos ellos, con independencia de su lengua materna, han sido escolarizados en esta lengua. Su desconocimiento del término adecuado en la lengua de la encuesta (por ejemplo, de cómo decir *foc a terra* en castellano), y de los significados que puede tener un determinado sintagma en español, provoca que los encuestados busquen, uno a uno, la traducción de los elementos que integran la lexía compleja en catalán, dando lugar a un calco sintáctico en castellano. A pesar de que pocos de estos calcos cuentan con una integración social en el corpus (únicamente “fuego a tierra” es también un *préstamo en vías de integración*), sabemos por nuestra propia experiencia que algunos de ellos son corrientes en el castellano de nuestra provincia (como “azul fuerte/flojo”, “jamón dulce/salado” o “mujer de hacer faenas⁸⁸⁵”). El mismo desconocimiento de los límites semánticos de los conceptos en castellano se aprecia en la mayor proporción que este tipo de *calcos* (los *semánticos*) suponen en el repertorio léxico en castellano. Creemos

⁸⁸⁵ Wesch (2008: 57-58) cita este caso como ejemplo de lexía o construcción de origen catalán cuyo uso se limita a las tierras de habla catalana (*vid.* 4.2.).

que los informantes, en su mayoría de lengua materna catalana, tienen una mayor conciencia de los límites semánticos de los términos de su propia lengua materna y, por este motivo, es más frecuente que realicen *préstamos* y no *calcos* (*semánticos* o *sintácticos*) en lengua catalana. La importación de *préstamos*, donde se transfiere no solo un significado sino también una forma lingüística, se explica a su vez por los años de subordinación lingüística de la lengua catalana a la castellana y porque es todavía relativamente reciente la normalización lingüística llevada a cabo (*cf.* 4.1.). Muchos de estos préstamos, como veremos en el siguiente apartado (*vid.* 7.4), presentan un grado notable de integración social y han desplazado a los sinónimos correspondientes de la lengua catalana. Este hecho nos da una idea de la importancia que adquiere la transferencia entre las dos lenguas, sobre todo en dirección castellano>catalán, en nuestra comunidad.

En segundo lugar, analizamos el peso cuantitativo global de los fenómenos de contacto sobre el total de palabras, basándonos para ello en el número de respuestas que contienen un fenómeno de este tipo. Cabe destacar que, tradicionalmente, la mayoría de estudios sobre disponibilidad léxica realizados en territorios bilingües (*vid.* 3.2 y 3.3) ha analizado el porcentaje que suponen sobre el conjunto léxico total las transferencias, sirviéndose para ello de la cifra absoluta de fenómenos de este tipo recabados en cada centro de interés y relacionando este dato, por ende, con el total de vocablos de cada área. Si bien esta información nos da una idea del peso que suponen las transferencias en el recuento léxico general, nos ha parecido también interesante conocer el alcance que tienen estos elementos sobre el total de palabras, con lo que creemos que se observa mejor el valor que estos fenómenos adquieren en el habla “real” de los bilingües, siempre teniendo en cuenta las salvedades que ya hemos comentado sobre la artificialidad de este tipo de pruebas y su distancia con respecto a un discurso real espontáneo. Como ya dijimos en el capítulo anterior (*vid.* 6.1.4.), los vocablos se encuentran más condicionados que las palabras por la presencia de informantes “atípicos” en el aula, por ejemplo, un estudiante que escriba casi todas las palabras de un centro en una lengua distinta a la de la encuesta. Por este motivo, creemos que el cálculo del porcentaje de respuestas que contienen una transferencia sobre el conjunto total de palabras confiere una visión más aproximada de la extensión social de los fenómenos de contacto lingüístico en una comunidad de habla. Conseguir esta información es, sin duda alguna, tarea mucho más laboriosa a la del recuento general de

fenómenos, dado que el programa informático no la ofrece⁸⁸⁶. A continuación analizamos, pues, este dato, y lo relacionamos con algunos de los índices generales que ofrecimos en el capítulo anterior (*vid.* 6.1.), como el total de palabras, el promedio de palabras por centro de interés y el promedio de palabras por informante y centro de interés:

Tabla 7.39. Respuestas con algún fenómeno de contacto lingüístico sobre el total de palabras en ambas lenguas.

Total de palabras		Total de menciones de transferencias		Porcentaje de transferencias (menciones) sobre el total de palabras		Promedio de transferencias por informante	
Castellano	Catalán	Castellano	Catalán	Castellano	Catalán	Castellano	Catalán
73038	76061	1509	4141	2.07	5.44	6.29	17.25
Promedio de palabras por centro de interés		Promedio de transferencias por centro de interés		Promedio de palabras por informante y centro de interés		Promedio de transferencias por informante y centro de interés	
Castellano	Catalán	Castellano	Catalán	Castellano	Catalán	Castellano	Catalán
4296.35	4474.18	88.76	243.5	17.90	18.64	0.37	1.01

Tal como se observa en la tabla 7.39, la proporción de transferencias es siempre superior en lengua catalana en relación con el número total de palabras recabadas. De hecho, las respuestas con algún fenómeno de contacto lingüístico en catalán superan ampliamente las que se dan en castellano, con 4141 menciones de transferencias frente a 1509. No solo esta cifra absoluta es superior sino, más importante aún, el porcentaje que esta cantidad supone sobre el total de palabras, que es también superior en catalán (*vid.* 6.1.), es más del doble en la lengua vernácula de la provincia (5.44% frente a 2.07%). Por esta razón, todos los restantes índices (el promedio de transferencias por informante, por centro de interés y por informante y centro de interés), que derivan matemáticamente del total de respuestas con algún elemento transferido, siempre son superiores en lengua catalana. De todos estos datos, ha llamado nuestra atención el hecho de que cada informante produzca, de media, una transferencia en la encuesta en

⁸⁸⁶ Para ello, hemos aplicado una sencilla regla de tres al porcentaje de aparición de cada término (tercera columna en los diccionarios finales, *vid.* capítulos 10 y 11) que nos ha servido para extraer la cantidad de menciones de cada término transferido. A continuación, hemos procedido, en una tabla de Excel, a sumar las menciones de cada transferencia (en total, por centros de interés, etc.).

lengua catalana (1.01) y, en cambio, esa cantidad no llegue ni a la media respuesta en el caso del castellano (0.37)⁸⁸⁷.

Tabla 7.40. Porcentaje de respuestas por centro de interés con algún fenómeno de contacto lingüístico en castellano (frecuencia absoluta y relativa).

Nº	Centro de Interés	Total de palabras	Transferencias (menciones)	%
1	El cuerpo humano	6129	32	0.52
2	Ropa y complementos	5189	76	1.46
3	Partes de la casa (sin los muebles)	3400	92	2.71
4	Muebles de la casa	2975	72	2.42
5	Comidas y bebidas	5811	62	1.07
6	Objetos colocados en la mesa para la comida	3169	107	3.38
7	La cocina y sus utensilios	3691	183	4.96
8	La escuela (muebles y materiales)	5080	82	1.61
9	Calefacción, iluminación y medios de airear un recinto	2660	65	2.44
10	La ciudad	5124	59	1.15
11	El campo	4369	79	1.81
12	Medios de transporte	4043	10	0.25
13	Trabajos del campo y del jardín	2125	170	8.00
14	Animales	6245	65	1.04
15	Juegos y distracciones	3914	84	2.15
16	Profesiones y oficios	4709	226	4.80
17	Colores	4405	45	1.02

Si observamos, ahora, el porcentaje que suponen las respuestas con un elemento de contacto lingüístico sobre cada centro de interés en castellano (tabla 7.40), podemos apreciar con claridad cómo destaca, por encima de todas las áreas, *Trabajos del campo y del jardín*, con un 8% de palabras evocadas correspondientes a fenómenos de contacto lingüístico con el catalán. A continuación, y a una cierta distancia, hallamos *La cocina y sus utensilios*, con el 4.96% de transferencias, *Profesiones y oficios*, con un 4.80% y *Objetos colocados en la casa para la comida*, con un 3.38%. Además de estos centros, también *Partes de la casa*, *Calefacción, iluminación y formas de airear un recinto*, *Muebles de la casa* y *Juegos y distracciones* quedan por encima de la media general en lengua castellana. Entre estos ocho centros, que presentan un porcentaje superior al

⁸⁸⁷ Esto nos ha llevado a reflexionar sobre los datos que obtuvimos en el capítulo anterior, es decir, sobre la propia relevancia de las conclusiones de este estudio: ¿Realmente la disponibilidad léxica en lengua catalana es superior a la del castellano si tenemos en cuenta que, de media, hay una palabra por centro de interés en la encuesta en catalán que es, a su vez, una transferencia del castellano? Como vimos, la diferencia entre el promedio general de las dos lenguas es de 0.74 unidades (*cf.* 6.1.). El hecho de que el informante incorpore estas palabras en la encuesta, a nuestro modo de ver, no descarta la validez de nuestras conclusiones porque es un reflejo de la propia competencia léxica del individuo, y también al fin y al cabo, de la integración social de algunos préstamos del castellano en la lengua catalana (*vid.* 7.4.).

2.07% de menciones de transferencias, destacan, a nuestro modo de ver, aquellos relacionados con el ámbito estricto del hogar, y donde, por tanto, mejor se vislumbra el peso de la lengua materna, como son *La cocina y sus utensilios*, *Objetos colocados en la mesa para la comida*, *Partes de la casa* y *Muebles de la casa*. Recordemos, una vez más, que la mayoría de los informantes se declaran catalanohablantes y, por consiguiente, es de esperar que al completar la encuesta en castellano produzcan transferencias en estas áreas directamente relacionadas con la lengua empleada en la familia⁸⁸⁸.

De otra parte, es más que destacable la cantidad de elementos de contacto detectados en *Trabajos del campo y del jardín*⁸⁸⁹. Como ya se ha constatado en ocasiones anteriores (Bartol 2011), esta área temática es una de las que presenta más préstamos de la lengua vernácula, así como términos dialectales⁸⁹⁰. Como ya explicamos (*vid.* 4.3.2.), el catalán es la lengua predominante en los pequeños núcleos rurales de nuestra provincia, además de que hay una relación implícita entre este ámbito y la experiencia vital familiar. Otra área que presenta un gran número de transferencias de la lengua catalana es *Profesiones y oficios*. El hecho de que se trabaje el vocabulario de esta área en la escuela, a la vez de que se trate de un léxico aprendido tempranamente puede explicar, tal vez, esta circunstancia, que reforzaría claramente el vocabulario de la lengua catalana.

Por la parte final de la clasificación, encontramos centros con un alto grado de cohesión interna: *Medios de transporte*, *El cuerpo humano*, *Los colores*, *Animales*... Todos ellos presentan porcentajes bajos de elementos de contacto lingüístico sobre el total de palabras.

⁸⁸⁸ Como ya hemos comentado, sería muy interesante analizar estos mismos datos según la variable lengua materna para comprobar si entonces se producen más coincidencias entre los centros que contienen más transferencias de la otra lengua de la comunidad.

⁸⁸⁹ Un centro similar como *El campo* se sitúa, por el contrario, por debajo de la media.

⁸⁹⁰ Bartol (2011) llega a esta conclusión en su estudio comparativo donde tiene en cuenta los datos de Galicia, Valencia, Alicante, Vizcaya, Álava, Navarra y nuestros propios datos (Serrano Zapata 2003b, 2006), correspondientes a seis centros de interés.

Tabla 7.41. Porcentaje de respuestas por centro de interés con algún fenómeno de contacto lingüístico sobre el total de palabras en catalán.

Nº	Centro de Interés	Catalán	Transferencias	%
1	El cuerpo humano	6924	121	1.75
2	Ropa y complementos	5566	507	9.11
3	Partes de la casa (sin los muebles)	3503	170	4.85
4	Muebles de la casa	3072	173	5.63
5	Comidas y bebidas	6357	229	3.60
6	Objetos colocados en la mesa para la comida	3325	201	6.05
7	La cocina y sus utensilios	3954	289	7.31
8	La escuela (muebles y materiales)	5310	318	5.99
9	Calefacción, iluminación y medios de airear un recinto	2652	422	15.91
10	La ciudad	5100	165	3.24
11	El campo	4414	102	2.31
12	Medios de transporte	4102	149	3.63
13	Trabajos del campo y del jardín	2332	106	4.55
14	Animales	6317	402	6.36
15	Juegos y distracciones	3771	272	7.21
16	Profesiones y oficios	4888	284	5.81
17	Colores	4474	230	5.14

Con respecto al catalán, en esta lengua, es precisamente *Calefacción, iluminación y formas de airear un recinto*, con casi un 16%, el área donde obtienen un peso mayor los fenómenos de contacto⁸⁹¹. A pesar de que no es el centro que más respuestas de este tipo contiene en cifras absolutas —lugar que corresponde a *Animales*, como veremos a continuación (*infra* tabla 7.45)— ostenta el primer lugar en esta clasificación porque cuenta con algunos préstamos que tienen una frecuencia alta de aparición (“bombilla”, “lámpara”, “foco”...), entre los que destaca (aparell d’) aire acondicionat*, que menciona casi el 60% de los encuestados. De modo similar a como veíamos en castellano, también aquí áreas relacionadas con el entorno familiar ocupan las primeras posiciones, como *La cocina y sus utensilios* (7.31%) u *Objetos colocados en la mesa para la comida*⁸⁹² (6.05%), donde es de esperar que, en este caso, los informantes castellanohablantes incorporen vocabulario de su propia lengua materna. Un caso similar es el de *Ropa y complementos*, en la segunda posición del ranquin con un 9.11%, aunque en esta ocasión creemos que, a la posible vinculación con el entorno familiar, se suma la mayor presencia del castellano en los mercados de consumo, reflejo final de los años de supremacía lingüística de esta lengua frente a la vernácula (*vid.* 4.1.), que influiría en el léxico catalán de todos los informantes.

⁸⁹¹ Nótese que es el doble de la proporción que supone el centro de interés con mayor transferencia en el corpus castellano, esto es, *Trabajos del campo y del jardín*.

⁸⁹² También *Muebles de la casa* queda por encima de la media, situada en 5.44 (*vid.* tabla 7.39).

En cuarta posición y con un 7.24% de respuestas con algún elemento de contacto, hallamos *Juegos y distracciones*, uno de los pocos centros donde, como vimos (*vid.* 6.1.3.), los informantes aportan un promedio superior de palabras en castellano que en catalán. Como en aquella ocasión afirmamos, el léxico suscitado bajo esta etiqueta, donde abundan distintas formas de ocio como los deportes, la música o las nuevas tecnologías, parece ser propicio a la influencia de la lengua castellana. A *Juegos y distracciones* le sigue *Animales*, que ya se ha comprobado que es una de las áreas que más transferencias suscita (Bartol 2011). A nuestro modo de ver, quizá pueda también explicarse por la presencia mayoritaria de voces en este campo aprendidas tempranamente (como demuestra Hernández Muñoz 2006) que haría que el alumnado castellanohablante incorporase palabras de su propia lengua materna. No obstante, llama la atención este resultado cuando sabemos que este alumnado, minoría frente al catalanohablante en la muestra, ha sido escolarizado íntegramente en lengua catalana, y dado que en la escuela se trabaja el léxico de este campo semántico. De otra parte, precisamente en castellano no sucede lo mismo, ya que este campo queda por debajo de la media y en decimocuarta posición en cuanto a las respuestas con transferencias del vernáculo (*infra* tabla 7.42), lo que parece ilógico dado que la mayoría de los informantes tiene el catalán como lengua materna. Por este motivo, creemos que aquí se evidencia, además de la posibilidad que hemos mencionado, la subordinación lingüística histórica del catalán al castellano, que provoca que incluso el alumnado catalanohablante evoque algunos nombres de animales en lengua castellana, posiblemente porque ya los aprendió en esta lengua⁸⁹³.

También *La escuela* se sitúa por encima de la media; en este ámbito, el catalán es la única lengua vehicular y, por este motivo, no sería esperable un porcentaje alto de transferencia del castellano, a pesar de que las conclusiones de algunos estudios recientes demuestran lo contrario (Galindo 2006)⁸⁹⁴. Sin embargo, hallamos en este campo, nuevamente, algunos préstamos concretos caracterizados por una elevada frecuencia de mención, como “borrador”, “carpesano”, “mochila”, “papelera”... Con

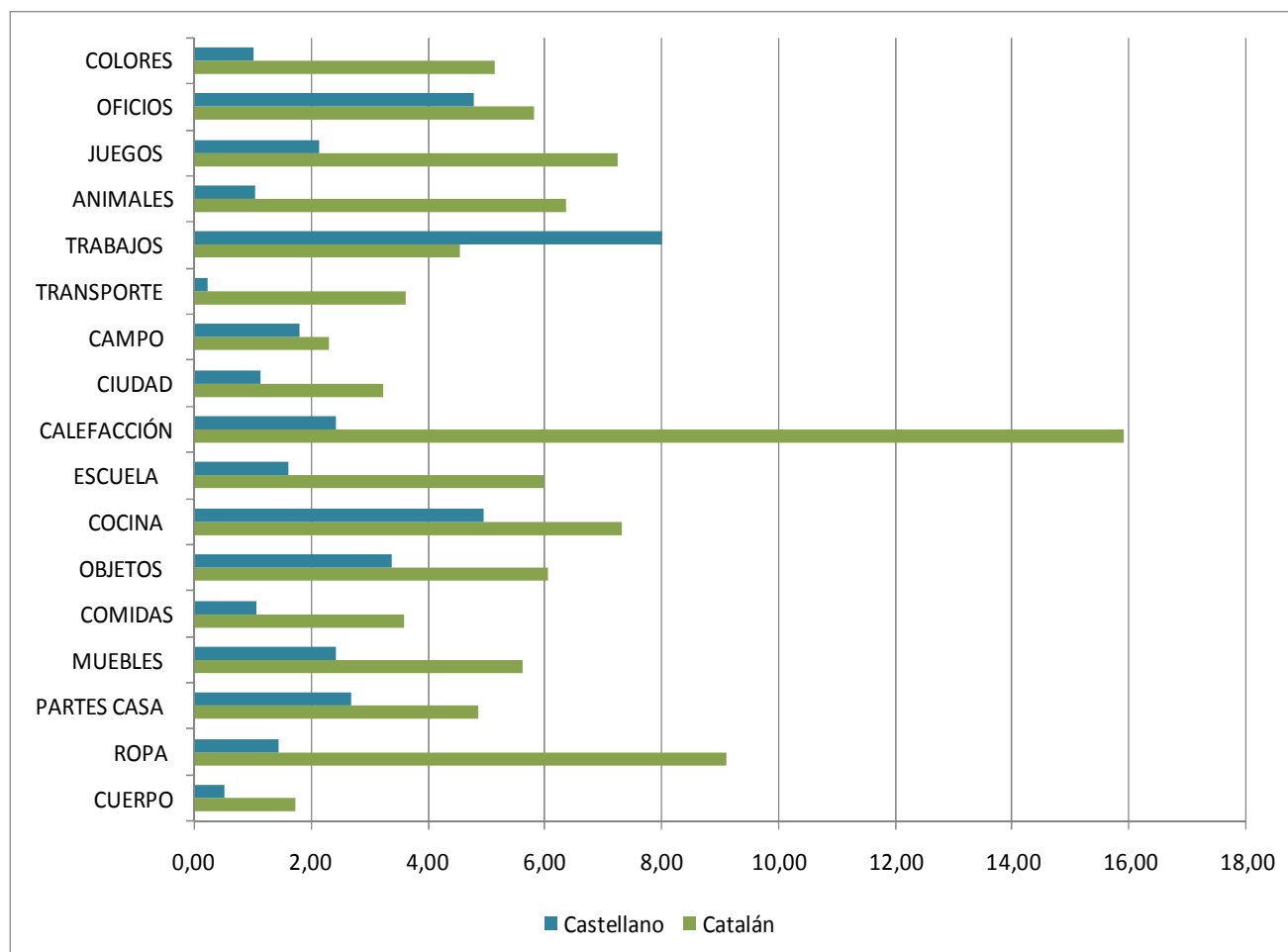
⁸⁹³ Hasta que no analicemos los datos en función de la variable *lengua materna* no podremos corroborar esta hipótesis, pero parece difícil que las 402 respuestas con elementos transferidos que se dan en este centro de interés en catalán pertenezcan únicamente a los 66 informantes cuya lengua materna es el castellano. Tampoco hay que despreciar la posible influencia de la televisión y el cine en nuestros informantes, dado que muchas películas o series infantiles las protagonizan animales.

⁸⁹⁴ La autora concluye que en el ámbito concreto de la hora del patio se usa más el castellano que el catalán. Recordemos que su estudio se basa en los usos lingüísticos de los escolares de primaria (*vid.* 4.1).

respecto a los centros que quedan últimos en la clasificación, cabe destacar el bajo porcentaje de transferencias que se recoge en torno a *El cuerpo humano*, que, como ya habían concluido otros estudios, no es un centro que se preste a la transferencia de palabras de otras lenguas (Bartol 2011: 160-162).

De otro lado, si comparamos el porcentaje que alcanza la transferencia en cada centro de interés en las dos lenguas, se constata en seguida cómo la proporción de respuestas con un elemento procedente de la otra lengua cooficial es siempre mayor en lengua catalana. Esta información, fruto de la comparación de las tablas anteriores (7.40 y 7.41), se plasma con claridad en el gráfico 7.3. La única excepción al predominio del castellano sobre el vernáculo se da en *Trabajos del campo y del jardín*. En esta área tanto el número total de respuestas con algún elemento de contacto lingüístico (170 frente a 106) como el porcentaje que estas representan sobre el total de palabras (8% frente a 4.55%) es superior en castellano. Se advierte aquí, por tanto, la dificultad de la mayoría de la muestra, de lengua materna catalana, de enunciar tareas agrícolas en castellano. Por otra parte, las disimilitudes entre las menciones de transferencias en las dos lenguas son de alcance considerable en algunas áreas, como en *Calefacción, iluminación y formas de airear un recinto* con una distancia de más del 10% (un 13.47% concretamente), o en *Ropa y complementos*, con un 7.65% de diferencia. También en otros centros, de los que ya hemos hablado, como *Juegos y distracciones* y *Animales*, la distancia entre las dos lenguas es destacable y sobrepasa el 5%, siempre a favor de la transferencia en lengua catalana. En otros campos semánticos, en cambio, la distancia entre los dos idiomas se reduce, como sucede en *El campo* (1.81% en castellano y 2.31% en catalán), *Profesiones y oficios* (4.8% en castellano y 5.81% en catalán) o *El cuerpo humano* (0.52% en castellano y 1.75% en catalán).

Gráfico 7.3. Porcentaje de respuestas con algún fenómeno de contacto lingüístico sobre el total de palabras por centro de interés en las dos lenguas.



En la tabla 7.42 comparamos la clasificación de los centros de interés en las dos lenguas según el porcentaje que suponen las menciones de las transferencias sobre el total de palabras. Es fácil concluir que las dos lenguas siguen patrones distintos en cuanto a los centros que presentan mayor y menor proporción de respuestas con un elemento de contacto lingüístico; es más, ningún centro ostenta el mismo rango. Sin embargo, sí es posible reconocer algunas tendencias comunes, como por ejemplo, que *La cocina y sus utensilios*, *Objetos colocados en la mesa para la comida* y *Muebles de la casa* —áreas como ya hemos indicado relacionadas con la lengua usada en el hogar— quedan, en ambos idiomas, por encima de la media y también otras áreas como *Profesiones y oficios*, *Calefacción, iluminación y formas de airear un recinto* y *Juegos y distracciones*. Asimismo, hay coincidencias con respecto a algunos centros que se sitúan por debajo de la media en ambos casos: algunos abiertos o difusos como *El campo* o *La ciudad*, pero también otros con un alto grado de cohesión, como *Colores*, *Medios de transporte*, *El cuerpo humano* y, en menor medida, *Comidas y bebidas* (vid.

6.1.5.). Así pues, el grado de cohesión interna de cada categoría semántica, como ya dijimos en un trabajo anterior (Serrano Zapata 2003b: 80), no parece guardar ninguna relación con una mayor o menor aparición de elementos de contacto lingüístico.

No obstante, las diferencias entre las posiciones de los centros en las dos lenguas son más que notables, sobre todo con respecto a algunas áreas concretas, como *Trabajos del campo y del jardín*, que ocupa el primer puesto en castellano y, en cambio, es duodécimo en catalán; al igual que *Partes de la casa*, que también ocupa una posición más aventajada en castellano que en catalán. En lengua catalana, tres áreas (*Ropa y complementos*, *Animales* y *La escuela*) quedan por encima de la media y, en cambio, se sitúan por debajo en castellano. Otro caso relevante es el de *Animales*, que pasa de la quinta posición en catalán a la decimocuarta en castellano.

Tabla 7.42. Rango de los centros de interés en cada una de las lenguas según el porcentaje de respuestas con algún fenómeno de contacto lingüístico sobre el total de palabras⁸⁹⁵.

Rango	Castellano	%	Catalán	%
1	TRABAJOS	8.00	CALEFACCIÓN	15.91
2	COCINA	4.96	ROPA	9.11
3	PROFESIONES	4.80	COCINA	7.31
4	OBJETOS	3.38	JUEGOS	7.24
5	CASA	2.71	ANIMALES	6.36
6	CALEFACCIÓN	2.44	OBJETOS	6.05
7	MUEBLES	2.42	ESCUELA	5.99
8	JUEGOS	2.15	PROFESIONES	5.81
9	CAMPO	1.81	MUEBLES	5.63
10	ESCUELA	1.61	COLORES	5.14
11	ROPA	1.46	CASA	4.85
12	CIUDAD	1.15	TRABAJOS	4.55
13	COMIDAS	1.07	TRANSPORTE	3.63
14	ANIMALES	1.04	COMIDAS	3.60
15	COLORES	1.02	CIUDAD	3.24
16	CUERPO	0.52	CAMPO	2.31
17	TRANSPORTE	0.25	CUERPO	1.75

⁸⁹⁵ Con una línea más gruesa se indica donde se sitúa la media respectiva.

A continuación, analizamos la cifra absoluta de fenómenos de contacto lingüístico sobre el total de vocablos, donde todas las transferencias, con independencia de las menciones que hayan recibido, adquieren el mismo valor y, por ende, donde se da una mayor cabida a los fenómenos de tipo individual.

Tabla 7.43. Fenómenos de contacto lingüístico sobre el total de vocablos en las dos lenguas.

Total de vocablos		Total de transferencias (valor absoluto)		Promedio de transferencias por centro de interés		Porcentaje de transferencias sobre el total de vocablos	
Castellano	Catalán	Castellano	Catalán	Castellano	Catalán	Castellano	Catalán
6824	7064	483	945	28.41	55.59	7.08	13.38

En la tabla 7.43, observamos la cifra absoluta de transferencias recogidas en las dos lenguas (483 en castellano y 945 en catalán) y, también, el promedio de estas por centro de interés —resultado de dividir estos datos entre 17, el número de campos semánticos estudiados—. Ambas magnitudes, al igual que el porcentaje sobre el total de vocablos (7.08 en castellano y 13.38 en catalán) plasman con claridad que los fenómenos de contacto lingüístico se producen en mayor medida en lengua catalana que en lengua castellana⁸⁹⁶.

Tabla 7.44. Porcentaje de fenómenos de contacto lingüístico sobre el total de vocablos en castellano por centro de interés.

C.I.	Centro de Interés	Castellano	Transferencias	%
1	El cuerpo humano	386	20	5.18
2	Ropa y complementos	291	25	8.59
3	Partes de la casa	238	29	12.18
4	Muebles de la casa	263	27	10.27
5	Comidas y bebidas	476	30	6.30
6	Objetos colocados en la mesa para la comida	238	25	10.50
7	La cocina y sus utensilios	347	31	8.93
8	La escuela	420	25	5.95
9	Calefacción, iluminación y medios de airear un recinto	297	16	5.39
10	La ciudad	678	28	4.13
11	El campo	727	48	6.60
12	Medios de transporte	276	5	1.81
13	Trabajos del campo y del jardín	434	49	11.29
14	Animales	347	35	10.09
15	Juegos y distracciones	576	24	4.17

⁸⁹⁶ Nótese que, aunque sobre el porcentaje no hay tanta diferencia, el total de elementos recogidos de este tipo en cifras absolutas es casi el doble en catalán que en castellano.

16	Profesiones y oficios	591	41	6.94
17	Colores	239	25	10.46

En castellano, al analizar las áreas según el porcentaje que suponen los elementos de contacto lingüístico en cifras absolutas sobre el total de vocablos, destaca en primer lugar *Partes de la casa*, con un 12.18% del conjunto léxico integrado por transferencias del catalán. A corta distancia, hallamos *Trabajos del campo y del jardín*, con un 11.29% y *Objetos colocados en la mesa para la comida*, con un 10.50% de transferencia. Por encima del 10% del total se sitúan asimismo *Colores*, *Muebles de la casa* y *Animales* y, por encima de la media, además de los ya citados, *La cocina y sus utensilios* y *Ropa y complementos*. En todos estos casos, creemos que se constatan tres tendencias distintas. En primer lugar, la influencia de la lengua usada en el contexto familiar, en este caso el catalán, sobre el léxico de la segunda lengua (tal como se aprecia en las posiciones destacadas de *Ropa y complementos*, *Partes de la casa*, *Muebles de la casa*, *Objetos colocados en la mesa para la comida* y *La cocina y sus utensilios*); en segundo, la relación existente entre la lengua vernácula y la evocación de tareas agrícolas (*Trabajos del campo y del jardín*) y, en último lugar, la posible vinculación entre la lengua materna y el vocabulario tempranamente aprendido, visible en *Animales* y *Colores*. Por el contrario, se relegan ahora a las últimas posiciones de la clasificación *Calefacción, iluminación y formas de airear un recinto* y *El cuerpo humano*, con algo más del 5% de vocabulario del catalán, *Juegos y distracciones*, con un 4.17%, *La ciudad*, con 4.13% y, *Medios de transporte*, donde los fenómenos de contacto no llegan al 2%. Tiene cierta lógica que algunas de estas áreas, como *Calefacción, iluminación y formas de airear un recinto*, *La ciudad* y *Juegos y distracciones* se muestren poco propicias a la transferencia del catalán sobre el castellano, ya que el predominio de la lengua castellana ya se advertía al analizar el total de palabras y vocablos. Recordemos que estas son unas de las pocas categorías donde la productividad y la riqueza léxica son superiores en castellano (*vid.* 6.1.3. y 6.1.4.)⁸⁹⁷.

Antes de pasar a analizar los datos de la lengua catalana, cabe señalar que esta clasificación difiere considerablemente de la obtenida al estudiar los mismos datos según el porcentaje de menciones de transferencias sobre el total de respuestas (*cf.*

⁸⁹⁷ En 6.1.3 y 6.1.4. propusimos algunas razones para que esto fuera así, de un lado, la relación del centro 15 con las formas de ocio juvenil donde el castellano está presente (ya sea a través de los medios de comunicación, los mercados de consumo, etc.), y de otro, la mayor presencia del castellano en las ciudades, que se contrapone con la vinculación del catalán a las áreas rurales (*vid.* 4.3.2.).

tabla 7.40). Algunos centros que allí se sitúan en las primeras posiciones, como *Profesiones y oficios*, *Calefacción, iluminación y formas de airear un recinto* y *Juegos y distracciones*, se ubican en este ranquin por debajo de la media. De forma inversa, *Ropa y complementos* y, sobre todo, *Animales* y *Colores* ascienden en la clasificación, al tener ahora igual consideración todos los elementos de contacto lingüístico, y ser estos centros donde predominan los fenómenos individuales; principalmente, los cambios de código.

Tabla 7.45. Porcentaje de fenómenos de contacto lingüístico sobre el total de vocablos en catalán por centro de interés.

Rango	Centro de Interés	Catalán	Transferencias	%
1	El cuerpo humano	364	33	9.07
2	Ropa y complementos	313	65	20.77
3	Partes de la casa	258	37	14.34
4	Muebles de la casa	274	51	18.61
5	Comidas y bebidas	523	74	14.15
6	Objetos colocados en la mesa para la comida	282	61	21.63
7	La cocina y sus utensilios	373	63	16.89
8	La escuela	444	49	11.04
9	Calefacción, iluminación y medios de airear un recinto	286	35	12.24
10	La ciudad	658	51	7.75
11	El campo	748	61	8.16
12	Medios de transporte	303	43	14.19
13	Trabajos del campo y del jardín	418	34	8.13
14	Animales	434	111	25.58
15	Juegos y distracciones	560	61	10.89
16	Profesiones y oficios	577	71	12.31
17	Colores	249	45	18.07

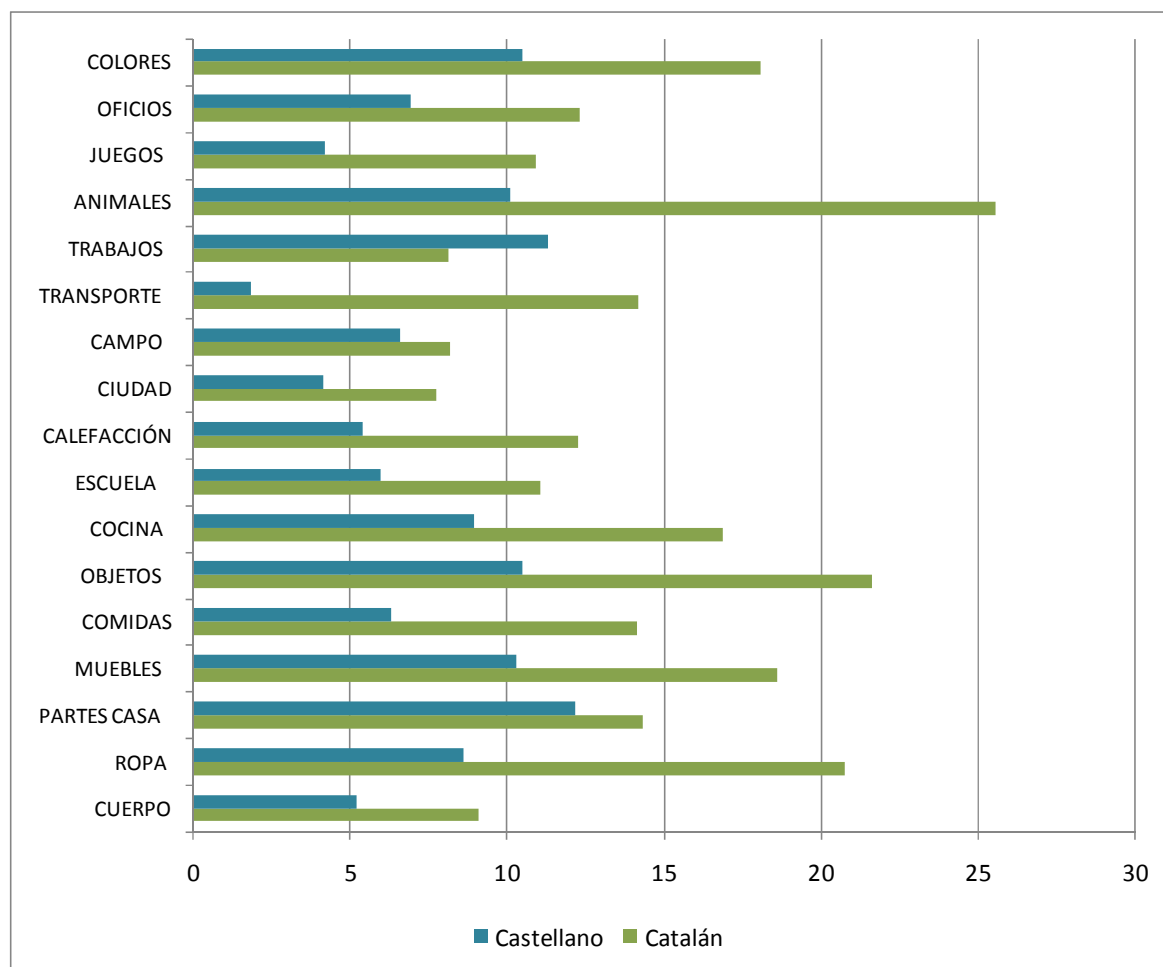
En los datos relativos a la lengua catalana, donde las transferencias suponen una proporción mayor sobre el conjunto léxico que en castellano como hemos visto, destaca en primera posición con un 25.58% *Animales*, seguido de *Objetos colocados en la mesa para la comida* y de *Ropa y complementos*, áreas donde la transferencia supera en ambos casos el 20% del vocabulario recabado. A corta distancia le siguen *Muebles de la casa* y *Colores*, con cifras que rebasan el 18%. *La cocina y sus utensilios*, *Partes de la casa*, *Medios de transporte* y *Comidas y bebidas* completan la nómina de centros que se sitúan por encima de la media (13.38). El panorama, por tanto, de áreas donde se aprecia en mayor medida la transferencia del castellano parece responder, al menos, a dos de las tres tendencias vistas en el castellano: la vinculación con la lengua del hogar

(*Objetos colocados en la mesa para la comida, Ropa y complementos, Muebles de la casa, La cocina y sus utensilios, Partes de la casa y Comidas y bebidas*) y la presencia de áreas donde predomina el léxico adquirido tempranamente como *Animales, Colores* o *Medios de transporte*. No obstante, como ya hemos comentado, parece difícil que la notable presencia en estos centros de elementos de contacto lingüístico del castellano se deba en su totalidad a la influencia de la lengua materna de los informantes castellanohablantes, que son minoría en la muestra. De hecho, en todas estas áreas se aprecia el predominio social que ha tenido el castellano sobre la lengua vernácula, favorecido por las circunstancias sociopolíticas de buena parte del siglo XX. Por la información que manejamos hasta el momento, todo apunta a que los informantes catalanohablantes ya aprenden las transferencias en su propia lengua materna, esto es, que la lengua catalana contiene un porcentaje nada despreciable de términos del castellano que se transmiten así de generación en generación, como consecuencia de que las generaciones mayores no han sido escolarizadas en catalán. De otro modo, no se explica que categorías donde predomina el vocabulario tempranamente aprendido ocupen las primeras posiciones en cuanto a transferencias en lengua catalana, siendo que la mayoría de la muestra es catalanohablante.

Por la parte inferior de la clasificación, como centros menos propicios a la incorporación de fenómenos de contacto lingüístico, hallamos *Juegos y distracciones* (10.89%), *El cuerpo humano* (9.07%), *El campo* (8.16%), *Trabajos del campo y del jardín* (8.13%) y *La ciudad* (7.75%). Es notable que la menor proporción de transferencia se dé en dos centros estrechamente relacionados con el entorno rural, como *Trabajos del campo y del jardín* y *El campo*, donde ya hemos dicho que predomina de forma clara el vernáculo. Por otro lado, al igual que hemos comentado con respecto al castellano, no hay relación entre esta clasificación de los centros de interés y la que hemos realizado anteriormente con respecto al porcentaje de menciones de transferencias sobre el total de palabras; baste comparar las posiciones que ocupan en ambas clasificaciones *Calefacción, iluminación y formas de airear un recinto* o *Juegos y distracciones* para comprobarlo⁸⁹⁸.

⁸⁹⁸*Calefacción, iluminación y formas de airear un recinto* es primera, según aquella clasificación y, en cambio, aquí desciende a la undécima posición; por su lado, *Juegos y distracciones*, desciende de la cuarta a la decimotercera. Como ya dijimos, esto es debido a la presencia de algunos préstamos con una alta frecuencia de mención que provocan que estos campos léxicos asciendan al analizar los datos según

Gráfico 7.4. Porcentaje de fenómenos de contacto lingüístico sobre el total de vocablos por centro de interés en las dos lenguas.



Como se puede observar en el gráfico 7.4, el porcentaje de transferencias sobre el total de vocablos es superior en catalán en todas las áreas, salvo en *Trabajos del campo y del jardín*, donde en castellano se da una proporción más elevada de fenómenos de contacto lingüístico (11.29% frente a 8.13%)⁸⁹⁹. La clara vinculación entre la lengua vernácula y los núcleos rurales ubicados fuera de la capital del *Segrià*, justifica, como ya hemos comentado, que el alumnado catalanohablante —mayoritario en la muestra— realice transferencias de su lengua materna al enunciar acciones relacionadas con el contexto rural, al no usar en este ámbito nunca, o casi nunca, la lengua castellana. Por el contrario, el alumnado castellano hablante que se halla en la

el porcentaje de menciones de transferencias. Otros campos como *Colores*, *Partes de la casa*, *Medios de transporte*, *Comidas y bebidas* o *La escuela*, también varían su posición considerablemente.

⁸⁹⁹ Recordemos que el porcentaje también era superior al analizar las menciones de transferencias sobre el total de palabras (*vid.* tabla 7.42) y que, en el análisis cuantitativo según el colectivo lingüístico, los catalanohablantes mostraban de forma firme su superioridad léxica al completar esta área en su propia lengua materna (*vid.* 6.2.5.).

situación inversa presenta poca transferencia de su lengua materna al completar la encuesta en catalán, en nuestra opinión, porque en este caso el léxico que se evoca se aprende a una edad más tardía —al contrario de lo que sucede con *Animales* y *Colores*—, y aquí sí que el vocabulario aprendido en la escuela palió los efectos de que la lengua materna sea distinta.

A pesar de que los datos siempre son superiores en catalán, hay algunas áreas (*El campo*, *Partes de la casa*, *La ciudad* y *El cuerpo humano*, por este orden) donde la diferencia entre las dos lenguas se reduce. Por el contrario, en muchos más, la proporción de transferencias en lengua catalana dobla al porcentaje del castellano (*Ropa y complementos*, *Comidas y bebidas*, *Objetos colocados en la mesa para la comida*, *Calefacción*, *iluminación y formas de airear un recinto*, *Animales* y *Juegos y distracciones*) y, en otros (*Muebles de la casa*, *La cocina y sus utensilios*, *La escuela*, *Profesiones y oficios* y *Colores*), la distancia no llega a ser del doble por pocas décimas. La divergencia más acusada entre las distintas áreas se constata en *Animales* (con un 25.58% de transferencias en catalán y solo un 10.09% en castellano), en *Medios de transporte* (con 14.19% frente a 1.81%), y en *La ropa* (20.77% frente a 8.59%). Es de justicia reconocer que la influencia del castellano siempre se evidencia en mayor medida en el corpus catalán que en sentido inverso; de hecho, el centro que ocupa la última posición en catalán, *La ciudad*, presenta un porcentaje de transferencia de léxico castellano del 7.75%, esto es, más de la media del corpus castellano.

Tabla 7.46. Rango de los centros de interés en cada una de las lenguas según el porcentaje de fenómenos de contacto lingüístico sobre el total de vocablos (de mayor a menor)⁹⁰⁰.

Rango	Castellano	%	Catalán	%
1	CASA	12.18	ANIMALES	25.58
2	TRABAJOS	11.29	OBJETOS	21.63
3	OBJETOS	10.50	ROPA	20.77
4	COLORES	10.46	MUEBLES	18.61
5	MUEBLES	10.27	COLORES	18.07
6	ANIMALES	10.09	COCINA	16.89
7	COCINA	8.93	CASA	14.34
8	ROPA	8.59	TRANSPORTE	14.19
9	PROFESIONES	6.94	COMIDAS	14.15

⁹⁰⁰ Los centros que alcanzan el mismo rango se señalan de forma coloreada en la tabla; la media global, situada en 7.08 en castellano y en 13.38 en catalán, se destaca con una doble raya.

10	CAMPO	6.60	PROFESIONES	12.31
11	COMIDAS	6.30	CALEFACCIÓN	12.24
12	ESCUELA	5.95	ESCUELA	11.04
13	CALEFACCIÓN	5.39	JUEGOS	10.89
14	CUERPO	5.18	CUERPO	9.07
15	JUEGOS	4.17	CAMPO	8.16
16	CIUDAD	4.13	TRABAJOS	8.13
17	TRANSPORTE	1.81	CIUDAD	7.75

En la tabla 7.46 se evidencia que hay una mayor coincidencia entre las dos lenguas que la que observamos al analizar los mismos datos respecto a la proporción de menciones de transferencias sobre el total de palabras (*vid.* tabla 7.42). En concreto, hay dos centros que ostentan el mismo rango: *La escuela*, en duodécimo puesto y *El cuerpo humano*, decimocuarto en los dos idiomas. Muchos más difieren en solo una posición (*Muebles de la casa*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios*, *La ciudad*, *Profesiones y oficios* y *Colores*) y algunos en dos (*Calefacción*, *iluminación y formas de airear un recinto*, *Juegos y distracciones* y *Comidas y bebidas*). Las coincidencias son incluso mayores si analizamos las áreas que se sitúan por encima y por debajo de la media respectiva en ambas lenguas. Como ya hemos indicado, en los primeros lugares se detecta la presencia de áreas estrechamente relacionadas con el hogar y, por tanto, con la lengua materna del informante (*Ropa y complementos*, *Partes de la casa*, *Muebles de la casa*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios* y, solo en catalán, *Comidas y bebidas*). Es de esperar que en estas áreas, el alumnado que completa la encuesta en su segunda lengua, es decir, el castellanohablante en la encuesta en catalán y el catalanohablante en la del castellano, tenga dificultad para evocar términos relacionados con estos enunciados y de ahí que realice transferencias, del tipo que sea, para nombrar las realidades que se incluyen bajo estos centros de interés. Cabe recordar en este punto, que, en el análisis cuantitativo, vimos cómo tres de estos centros (*Partes de la casa*, *Objetos colocados en la mesa para la comida* y en *La cocina y sus utensilios*) presentaban una covariación estadísticamente significativa con la variable *lengua materna*, siempre a favor del alumnado que completa la encuesta en su primera lengua (*vid.* 6.2.5.).

De otra parte, también se sitúan en las primeras posiciones en los dos idiomas *Animales*, sexto en castellano y primero en catalán y *Colores*, cuarto en castellano y

quinto en catalán. Como también hemos indicado, creemos que la edad de adquisición de este vocabulario incide, sobre todo en castellano, en el número de transferencias que realiza un informante al tratar de completar estas áreas en su segunda lengua. En el análisis cuantitativo, también constatamos cómo estos dos centros de interés presentaban una covariación significativa estadísticamente con respecto a la variable *lengua materna* (*Animales* en las dos *corpus* y *Colores* solo en castellano) y en el análisis de la producción en las dos lenguas según la adscripción lingüística, cada colectivo mostraba superioridad léxica en su propia lengua materna⁹⁰¹. Asimismo, en el análisis del grado de cohesión de cada área temática, también constatamos que este centro era más cerrado o compacto en castellano que en catalán. Aunque en aquel momento expusimos otra hipótesis posible (*vid.* 6.1.5.), ahora podemos argumentar que esto se debe, con mucha más probabilidad, a la gran cantidad de transferencias que se producen en torno a este centro de interés en lengua catalana, que provoca el aumento del número de vocablos y, por tanto, reduce la posibilidad de que estos coincidan entre ellos.

También podemos hablar de tendencias comunes al analizar los campos semánticos que ocupan las últimas posiciones. Por ejemplo, *Juegos y distracciones* o *La ciudad* parecen propiciar poco la aparición de elementos de una segunda lengua, dado que comparten las últimas posiciones en las dos lenguas. También es el caso de *El cuerpo humano*, decimocuarto en los dos idiomas y, en menor medida, *Calefacción, iluminación y formas de airear un recinto*, decimotercero en castellano y undécimo en catalán. Un caso peculiar es el de *La escuela*, que, a pesar de ostentar el mismo rango en las dos lenguas, contiene una proporción superior de fenómenos de contacto en catalán, cuando no sería esperable dado su posición privilegiada como única lengua vehicular de la enseñanza, como ya hemos comentado.

Los ejemplos en los que se aprecian tendencias contrarias en las dos lenguas son dos. En primer lugar, *Trabajos del campo y del jardín* ocupa el segundo lugar en castellano, debido al alto porcentaje de transferencias del catalán que presenta y, en cambio, se relega a la última posición en lengua catalana. En segundo lugar y de forma contraria, *Medios de transporte* es último en castellano y octavo en catalán. Mientras

⁹⁰¹ En concreto, en *Animales*, por ejemplo, los castellanohablantes actualizaban en castellano hasta 3.14 respuestas de media más que en catalán (*vid.* 6.2.5.).

que, en lengua castellana, este centro apenas presenta fenómenos de contacto procedentes del catalán, en catalán presenta una proporción considerable (1.81% frente al 14.19%), con lo que concluimos que el léxico de este campo semántico está bastante influido por el castellano, quizá por su relación estrecha con los mercados de consumo⁹⁰².

En cualquier caso, lo que queda evidenciado es que el peso de los fenómenos de contacto lingüístico es notablemente superior en catalán que en castellano, lo que no deja de llamar la atención cuando sabemos que la mayoría de la muestra tiene como lengua materna el catalán y ha sido escolarizada por completo en esta lengua. El análisis de los datos nos lleva a concluir que la influencia social que ha tenido el castellano sobre el catalán a lo largo del siglo XX, por las circunstancias ya explicadas con anterioridad (*vid.* 4.1.)⁹⁰³, sigue plasmándose en la lengua con un mayor peso de la transferencia del castellano en la lengua catalana que en sentido inverso. A pesar del éxito del proceso de normalización lingüística del catalán, la influencia del castellano sigue estando presente sobre la lengua vernácula y, excepto en un área relacionada estrictamente con el léxico del trabajo rural, el alcance de los fenómenos de contacto lingüístico es siempre claramente superior en catalán que en castellano.

No obstante, la influencia de la lengua catalana sobre el castellano no es precisamente escasa. Si comparamos nuestros datos con los de otras sintopías donde el castellano convive con otra lengua dentro de la Península Ibérica, nos percatamos en seguida de que en ningún otro lugar, el castellano recibe una cantidad ni tan siquiera similar de transferencias de la lengua vernácula. Los datos de Bartol (2011: 173), que comentamos a continuación⁹⁰⁴, comparan el porcentaje que suponen los elementos de contacto lingüístico sobre el total de vocablos en distintos estudios sobre léxico disponible realizados con condiciones metodológicas idénticas al nuestro. La presencia

⁹⁰² Recordemos que en el análisis cuantitativo de este centro en función de la lengua materna, se observa, aunque por poca diferencia, una productividad superior del colectivo catalanohablante, tanto en castellano como en catalán, lo que también sucede con *La ciudad* (*vid.* 6.2.5.).

⁹⁰³ Principalmente, un largo periodo de subordinación política y la llegada masiva de inmigrantes monolingües en un entorno en que los hablantes o ya eran bilingües o acabaron siéndolo (Boix 2004). Además, como explicamos en el capítulo cuarto, cabe recordar que los hábitos lingüísticos de la población se han caracterizado durante muchos años por una convergencia hacia el español, aunque la situación esté cambiando en la actualidad (*vid.* 4.1., nota 367).

⁹⁰⁴ Excepto los datos de Castellón, extraídos de Casanova (2013: 534), los restantes porcentajes que enumeramos proceden todos de este autor.

mayor de palabras de la otra lengua cooficial —sin tener en cuenta nuestros datos provisionales de 2003 (Serrano Zapata 2003b, 2006)— se da en Galicia, con casi un 3% de vocablos gallegos (concretamente un 2.88%), seguida de Vizcaya (con un 2.35%) y de Valencia (con un 2.3%). Algo menor es la proporción de catalanismos en Castellón (Casanova 2013: 534), que suponen el 1.56% del total de vocablos, y aún es más escaso el porcentaje de entrada de elementos de la lengua autóctona en el castellano de Álava (0.83%), Alicante (0.57%) y Navarra (0.55%). Como es evidente, todas estas cifras se sitúan muy por debajo del porcentaje de transferencias que hemos hallado en nuestro corpus castellano, situado en un 7.08%. En estas diferencias es posible apreciar las distintas situaciones sociolingüísticas que se dan en la Península, determinadas por el prestigio y la extensión social del vernáculo en cada región. Como ya dijimos (*vid.* 6.1.7), al comparar el promedio de palabras en distintos territorios donde convive más de una lengua, en ninguna de las sintopías analizadas la lengua vernácula, u otra cualquiera distinta al español, tiene una extensión social ni siquiera similar a la del catalán en Lleida. Asimismo, el nuestro es el único territorio donde, con independencia de su lengua materna, no hay ningún informante escolarizado en lengua castellana. Estas dos circunstancias explican, en nuestra opinión, que el porcentaje de transferencia sea mayor en nuestra provincia que en los restantes territorios españoles analizados.

Con respecto al catalán, la situación es distinta si cotejamos nuestros datos con los de Llopis (2009: 321-323), los únicos existentes hasta el momento sobre la lengua catalana. La media del porcentaje de fenómenos de contacto es, en su caso, de 28.38%⁹⁰⁵, cifra bastante superior a la que se registra en nuestro corpus en catalán, situada en 13.38%, como hemos visto. La situación sociolingüística del catalán y del castellano en las dos provincias, Valencia y Lleida, explica de nuevo estas diferencias con respecto al volumen de fenómenos de contacto lingüístico. Aunque el promedio de respuestas por centro de interés es bastante similar en las dos comunidades (*vid.* 6.1.7.), el alumnado valencianohablante parece desconocer, en mayor medida, el léxico catalán normativo, al menos en las condiciones artificiales de la encuesta de disponibilidad.

⁹⁰⁵ Media calculada por nosotros a partir de los datos que da el autor de cada uno de los centros de interés en las páginas citadas. Recordemos que el estudio de Llopis (2009) solo engloba las comarcas catalanohablantes de la provincia de Valencia y entre estos fenómenos de contacto el autor incluye aquellos procedentes de otras lenguas distintas al castellano, clasificados como *cambios de código*. Estas voces son claramente minoritarias en su corpus, frente a las que proceden del castellano.

Este hecho provoca una sustitución mayor del vocabulario de esta lengua por el de la lengua castellana.

7.4. ESTUDIO DE LOS PRÉSTAMOS EN LAS DOS LENGUAS

En el presente apartado, abordamos el análisis de todos los préstamos hallados en las dos lenguas. Recordemos que para su clasificación hemos usado el porcentaje de aparición o la frecuencia de mención de cada término, diferenciando los préstamos en vías de integración de los ya integrados. Conocer el alcance que estos elementos cobran en la lengua no es fácil a partir de los datos con los que contamos en este estudio. Nos es imposible evaluar si el préstamo se produce en distintos niveles de la lengua o si se combina adecuadamente con los formantes de la lengua receptora. No obstante, una de las posibilidades más interesantes que sí tenemos a partir de las encuestas de disponibilidad es comprobar el desplazamiento de los sinónimos equivalentes en cada listado respectivo, posibilidad que ya usó Mackey (1971a, 1976) en su estudio pionero sobre el francés de Canadá. Además del desplazamiento de los equivalentes respectivos en la lengua de la encuesta, también es posible comparar dos índices matemáticos que cobran un especial interés aquí: el porcentaje de aparición, esto es la proporción de informantes que actualiza un determinado término, y el índice de disponibilidad, que combina la frecuencia de una unidad léxica con la posición que esta ocupa en cada listado.

A continuación, presentamos dos tablas, relativas a los datos del corpus del castellano, con los índices mencionados que obtiene tanto cada préstamo como su sinónimo respectivo. En cada tabla es posible observar el centro de interés en el que aparece cada préstamo (CI), el rango o posición que obtiene en el diccionario final (P), el porcentaje de aparición (% A) y el índice de disponibilidad (ID). En primer lugar, analizamos los *préstamos en vías de integración* (tabla 7.47) y seguidamente, los *préstamos integrados* (tabla 7.48):

Tabla 7.47. Préstamos en vías de integración en el corpus castellano.

CI	Préstamo	P	% A	ID	Equivalente	P	% A	ID
2	armilla	64	5.000 %	0.02566	chaleco	33	20.833 %	0.10266
3	(es)golfa	51	5.000 %	0.02226	desván, buhardilla	28 22	12.917 % 18.750 %	0.06570 0.09363

3	(es)golfas*	44	6.667 %	0.02985	desván, buhardilla	28 22	12.917 % 18.750 %	0.06570 0.09363
3	rebot	54	5.000 %	0.02092	despensa	21	21.667 %	0.10059
4	escriptorio*	27	5.000 %	0.03548	escritorio	7	43.333 %	0.24191
6	estovalles	32	7.500 %	0.04133	mantel	7	58.333 %	0.32632
6	plata	25	10.000 %	0.04913	fuelle bandeja	80 17	3.750 % 15.833 %	0.01328 0.06752
7	pica	35	10.000 %	0.05031	pila fregadero	142 19	1.250 % 21.667 %	0.00561 0.13296
7	plata	57	5.833 %	0.02812	fuelle bandeja	160 32	0.833 % 12.083 %	0.00438 0.06629
8	maquineta	34	12.917 %	0.05826	sacapuntas	44	10.417 %	0.04589
9	fuego a tierra	28	7.917 %	0.04004	chimenea ⁹⁰⁶	41	5.417 %	0.02705
11	pagès	81	5.417 %	0.02499	payés	37	11.250 %	0.06311
13	ensulfatar	26	8.333 %	0.04481	sulfatar	20	9.167 %	0.05286
13	esporgar	19	8.750 %	0.05323	podar	5	37.500 %	0.26375
13	pagès	18	8.750 %	0.07147	payés	8	15.000 %	0.12099
13	ramadero*	17	11.250 %	0.08354	ganadero	43	4.167 %	0.02751
15	botifarra	28	10.833 %	0.06664	_____			
15	handbol	56	6.250 %	0.03185	balonmano	44	7.917 %	0.04079
16	infermera	48	10.000 %	0.06396	enfermera	17	23.333 %	0.14063
16	mosso (d'esquadra)	72	8.333 %	0.03702	mozo de escuadra	233	0.833 %	0.00711

Tabla 7.48. Préstamos integrados en el corpus castellano.

CI	Préstamo	P	%A	ID	Equivalente	P	%A	ID
7	paella	11	32.083 %	0.22549	sartén	5	54.167 %	0.38268
16	paleta	4	35.417 %	0.21888	albañil	32	16.250 %	0.09072

Tal como se plasma en las dos tablas anteriores, la mayoría de los préstamos del catalán no desplaza el término equivalente en la lengua castellana. Del total de 21

⁹⁰⁶ Otros sinónimos posibles para este calco serían “fuego” u “hogar”, que brillan por su ausencia en los listados.

préstamos analizados⁹⁰⁷, solo en 6 (*maquineta*, “fuego a tierra”, *ramadero**, *botifarra*, *mosso d’esquadra* y *paleta*) se da esta circunstancia. En estos casos, el préstamo alcanza un índice de disponibilidad mayor que su sinónimo respectivo y su porcentaje de aparición es también superior. Entre estos fenómenos de contacto, hallamos dos préstamos culturales, *botifarra* y *mosso d’esquadra*, que designan realidades particulares de la cultura catalana, razón por la cual el equivalente castellano no es frecuente o ni siquiera existe. Hay que señalar que hemos dudado sobre la inclusión en estos 6 casos de dos ejemplos más: *pica* y *plata*. En ambos casos, los préstamos superan un posible sinónimo (como es “fuente” para el caso de *plata* y “pila” para el caso de *pica*), pero la presencia de otros sinónimos que sí alcanzan mayor disponibilidad que los préstamos en los listados (“bandeja” y “fregadero”) ha hecho que, finalmente, los hayamos excluido.

Según la integración o no del préstamo, la mitad de los escasos ejemplos de préstamos integrados de nuestro corpus (únicamente contamos con *paella* y *paleta*) desplaza un sinónimo en la lengua base (lo que ocurre con *paleta* respecto a “albañil”), y, en cambio, solo el 26.32% de los *préstamos en vías de integración* (esto es, 5 sobre 19) lo hace. Finalmente, en cuanto a la clasificación formal de todos los préstamos, hemos de decir que, de un total de 21 unidades léxicas, el 76.2% son formas correctas en lengua catalana (16 casos), un 14.29% son unidades híbridas, inexistentes en las dos lenguas —señaladas siempre con un asterisco ((es)golfas*, escritorio* y ramadero*)—, un 4.76% corresponde a calcos semánticos (únicamente, *paella*) y el mismo porcentaje representan los casos de calcos sintácticos (*fuego a tierra*). La misma información, con respecto al corpus catalán se puede observar en las siguientes tablas:

Tabla 7.49. Préstamos en vías de integración en el corpus catalán.

CI	Préstamo	P	%A	ID	Equivalente	P	%A	ID
1	muñeca	71	7.500 %	0.02974	canell	35	25.000 %	0.10994
2	blusa	49	10.833 %	0.05947	brusa	55	6.667 %	0.04250
2	chaleco	56	9.167 %	0.03724	armilla	53	10.417 %	0.04375
2	chándal	57	8.333 %	0.03622	xandall	34	23.333 %	0.10724

⁹⁰⁷ En realidad, si sumamos los ejemplos de las tablas 7.46 y 7.47, obtenemos 22 préstamos pero hemos suprimido uno dado que *plata* se repite en los centros 6 y 7.

CI	Préstamo	P	%A	ID	Equivalente	P	%A	ID
2	pajarita	73	6.667 %	0.02709	corbati ⁹⁰⁸	161	0.833 %	0.00535
2	sostén	66	5.833 %	0.02969	sostenidor	39	17.500 %	0.09170
2	sudadera	50	10.833 %	0.05700	dessuadora	—		
2	traje	71	6.667 %	0.02827	vestit ⁹⁰⁹	25	29.167 %	0.15688
3	buhardilla	45	5.417 %	0.02648	(es)golfa	16	37.500 %	0.17972
3	despensa	25	10.833 %	0.05370	rebost	15	36.250 %	0.18702
3	recibidor	40	5.000 %	0.03321	rebedor	4	65.833 %	0.45803
4	armari empotrat*	38	6.667 %	0.03307	armari encastat	—		
4	lámpara	22	11.667 %	0.06239	làmpada llum	69 109	2.500 % 1.667 %	0.01436 0.00669
4	mesita	26	7.917 %	0.04961	tauleta	10	21.667 %	0.14213
4	silló*	36	5.417 %	0.03398	butaca ⁹¹⁰	6	52.500 %	0.31209
5	gaseosa	40	14.167 %	0.08821	gasosa	132	3.750 %	0.02130
5	mermelada	158	5.000 %	0.01527	melmelada	503	0.417 %	0.00125
5	ron	94	7.500 %	0.03853	rom	265	0.833 %	0.00590
6	mantell*	26	11.667 %	0.06637	(es)tovalles	7	47.917 %	0.29394
6	servilleta	37	5.833 %	0.03986	tovalló	5	77.917 %	0.49780
7	batidora	27	14.583 %	0.07752	batedora	40	10.000 %	0.04982
7	cucharón	56	5.833 %	0.02914	cullerot	13	23.333 %	0.16663
7	exprimidora	41	10.833 %	0.04881	espremedora	221	1.250 %	0.00327
7	sandwichera	48	8.750 %	0.03731	sandvitxera	122	2.500 %	0.00884
7	sartén	33	9.167 %	0.06294	paella	1	65.833 %	0.46439
8	carpessà*	57	10.000 %	0.03669	carpeta d'anelles	—		
8	carpesano	85	5.833 %	0.01870	carpeta d'anelles	—		

⁹⁰⁸ Otra expresión sinónima como *corbata de llacet* no aparece.

⁹⁰⁹ Más que a un *vestit*, intuimos que los informantes, cuando incluyen esta transferencia del castellano, se refieren al conjunto de pantalón, chaqueta y chaleco, que en catalán normativo sería *tern*, que ningún informante aporta.

⁹¹⁰ Otro término sinónimo, *cadira de braços*, no aparece.

CI	Préstamo	P	%A	ID	Equivalente	P	%A	ID
8	esquadra	63	7.500 %	0.03324	escaire	53	8.750 %	0.04343
8	mochila	59	8.750 %	0.03563	motxilla	23	23.750 %	0.10621
8	ordenador	40	12.083 %	0.06356	ordinador	25	23.333 %	0.09865
8	pegament*	74	6.250 %	0.02528	pega ⁹¹¹	115	2.500 %	0.01014
8	tissa*	61	7.083 %	0.03458	guix	9	56.667 %	0.32722
8	papelera	65	7.917 %	0.03199	paperera	29	20.000 %	0.08904
8	taquilla	54	8.333 %	0.04107	caseller, armariet	192 331	0.833 %, 0.417 %	0.00425 0.00173
9	bombilla	17	12.083 %	0.08489	bombeta	3	61.667 %	0.45370
9	calentador	31	5.000 %	0.03192	escalfador	71	1.250 %	0.00965
9	vela	32	6.250 %	0.03152	espelma	18	15.417 %	0.07945
10	acera	52	7.500 %	0.05420	vorera	36	14.167 %	0.07914
10	pas de peatons*	60	6.667 %	0.04103	pas de vianants	90	3.333 %	0.02307
10	peató*	63	6.667 %	0.04016	vianant	65	7.500 %	0.03862
12	barco	32	7.083 %	0.04183	vaixell	7	77.500 %	0.44355
12	lancha	40	7.083 %	0.02625	llanxa	52	4.167 %	0.01851
12	patera	48	5.417 %	0.02300	pastera	—		
13	abonar	16	14.583 %	0.09032	adobar (la terra)	18	14.167 %	0.08837
13	cosechar	26	7.500 %	0.04663	collir recol·lectar	6 25	37.500 % 8.750 %	0.24293 0.04952
14	ballena	64	12.083 %	0.05193	balena	26	33.333 %	0.13752
14	delfi*	91	6.667 %	0.02710	dofi	22	34.583 %	0.15820
14	jabalí	87	5.833 %	0.02877	porc senglar porc fer	5 223	15.000 % 0.833 %	0.06330 0.00508
14	tiburó*	72	9.167 %	0.04184	tauró	28	34.583 %	0.13645
14	tocino	94	5.000 %	0.02581	porc	8	50.833 %	0.29296
15	escondite	28	13.333 %	0.07361	amagatall	94	2.917 %	0.01798

⁹¹¹ Otra expresión equivalente, *goma d'enganxar*, no tiene cabida en nuestros listados.

CI	Préstamo	P	%A	ID	Equivalente	P	%A	ID
15	ordenador	33	8.750 %	0.05546	ordinador	11	23.333 %	0.14173
15	teto	44	5.000 %	0.04227	_____			
15	uno	51	5.417 %	0.03493	_____			
16	abogat*	47	10.833 %	0.05756	advocat	9	29.167 %	0.18403
16	camarar*	75	5.833 %	0.03735	cambrer	15	26.250 %	0.14843
16	enfermer*	48	9.583 %	0.05744	infermera	14	26.667 %	0.14858
16	ingenier*	84	5.833 %	0.03119	enginyer	29	17.500 %	0.09723
16	peluquer*	81	6.250 %	0.03230	perruquer	40	12.500 %	0.07393
17	marrón	44	5.000 %	0.02877	marró	11	82.083 %	0.44776

Tabla 7.50. Préstamos integrados en el corpus catalán.

CI	Préstamo	P	FA	ID	Equivalente	P	FA	ID
1	hombro	43	18.750 %	0.08407	espatlla	34	27.083 %	0.11523
2	bolso	21	32.083 %	0.17374	bossa de mà ⁹¹²	175	0.833 %	0.00417
2	(forro) (polar)	44	17.500 %	0.07831	folre polar	_____		
2	gorro	35	23.333 %	0.10473	gorra ⁹¹³	24	32.917 %	0.16132
2	sostens*	29	28.333 %	0.13631	sostenidor	39	17.500 %	0.09170
6	mantel	13	18.750 %	0.11543	(es)tovalles	7	47.917 %	0.29394
8	borrador	17	31.667 %	0.15949	esborrador	44	10.833 %	0.05734
9	(aparell d')aire acondicionat*	5	57.917 %	0.40986	(aparell d')aire condicionat	11	17.917 %	0.13007
9	foco	15	18.750 %	0.11067	focus	_____		
9	lámpara	8	40.833 %	0.27161	llum, làmpada	13 23	15.833 % 8.750 %	0.11672 0.05869
10	farola	23	18.333 %	0.11640	fanal	117	2.500 %	0.01464
14	mono	15	33.750 %	0.16687	mona, simi,	122	2.500 %	0.01654

⁹¹² El equivalente exacto es este, raramente usado tal como se ve. La palabra *bossa* sin especificación aparece en la posición 54, también por debajo de “bolso”, aunque puede referirse a otros referentes, muy alejados de “bolso”.

⁹¹³ Otros sinónimos como *casquet*, *capell* o *capçó* no aparecen.

CI	Préstamo	P	FA	ID	Equivalente	P	FA	ID
					mico	171 53	1.667 % 13.333%	0.00822 0.06306
15	pilla pilla	17	17.500 %	0.11002	tocar i parar	80	3.750 %	0.02192
15	(vídeo)conso- la	21	15.833 %	0.09422	consola de joc	—		
16	fontaner	31	17.083 %	0.09438	llauner, lampista	123 , 97	3.333 % 4.583 %	0.01865 0.02443
17	dorat*	23	27.917 %	0.11578	daurat	41	10.000 %	0.04415
17	granate	17	37.500 %	0.18105	granat	129	0.833 %	0.00317

El alcance de los préstamos en el corpus catalán es bastante diferente al que acabamos de ver en castellano. Se recoge un total de 76 préstamos distintos, de los cuales 36 desplazan el sinónimo correspondiente y obtienen un índice de disponibilidad mayor y una frecuencia de mención también más elevada que el término apropiado de la lengua catalana⁹¹⁴. Esto supone que casi la mitad de los préstamos recogidos, un 47.37%, sustituya el correlato catalán, circunstancia que en el corpus castellano solo ocurre, en cambio, con el 28.57% del total de préstamos. Si atendemos a la distribución de los préstamos en función de su integración, constatamos que, de los *préstamos en vías de integración*, un 38.33% supera el término equivalente en lengua catalana (23 sobre 60); esto se da, por el contrario, en el 82.35% de los préstamos ya integrados (14 sobre 17).

De otra parte, los préstamos recabados en las encuestas en catalán son, mayoritariamente, términos normativos del castellano (52 ejemplos, el 68.42%). Al igual que en el corpus castellano, es menor la presencia de calcos semánticos (6 casos, el 7.89% del total), y aún más ínfima la de combinaciones híbridas (3 ejemplos, el 3.95%). Los préstamos agramaticales, es decir, basados en interferencias que se han difundido, tienen ya una mayor presencia (15 préstamos, el 19.74%). Asimismo, consideramos bastante significativo el hecho de que en 9 de los 36 préstamos que desplazan el léxico básico catalán —esto es, en el 25% de casos— el término apropiado en la lengua de la encuesta ni tan siquiera aparezca en nuestros listados, como ocurre con *dessuadora*, *armari encastat*, *carpeta d'anelles*, *focus*, *consola de joc*, *folre polar*...

⁹¹⁴ En realidad, si se suman las dos tablas se observará que la cantidad exacta de préstamos es 78 y la de préstamos que superan su correspondiente sinónimo, 37. Las cifras que aparecen en el texto son resultado de restar a estas cantidades los ejemplos repetidos, a saber, “lámpara” y “ordenador”.

En dos casos más, “teto” y “uno”, el equivalente no aparece dado que se trata de expresiones de mayor extensión totalmente hechas en castellano o de juegos que no cuentan con un correlato en lengua vernácula.

En conclusión, la hipótesis de la que partimos, que era esperable que los fenómenos de contacto lingüístico se manifestaran en ambas direcciones en nuestro estudio, dada la situación de bilingüismo extenso, estable y no diglósico en la que nos hallamos, se ha visto sobradamente corroborada. Ahora bien, si en el capítulo 4 (*vid.* 4.1.) vimos detalladamente que existen en la actualidad razones para que la interferencia se manifieste tanto en un sentido como en otro, es momento ahora de reconocer que la influencia del castellano sobre el catalán es superior a la que se registra en sentido contrario, tal como también constatan los pocos estudios dedicados al análisis de los fenómenos de contacto lingüístico de forma simultánea en las dos lenguas (Galindo 2006: 78). Como acabamos de ver en este capítulo, en nuestro corpus, la proporción de transferencias es siempre mayor en catalán, salvo en un centro de interés relacionado estrictamente con las tareas agrícolas. Además, no solo el porcentaje de elementos de este tipo sobre el total de vocablos es superior en catalán (con un 13.38 frente a 7.08), sino que además, en el análisis concreto de los préstamos que acabamos de realizar, se corrobora que en muchas más ocasiones en lengua catalana (47.37% frente a 28.57%) estos términos, algunos ya integrados como demuestra su frecuencia de mención, sustituyen al vocabulario básico de la lengua de la encuesta. Por tanto, no podemos más que concluir que, a pesar del esfuerzo realizado por las distintas instituciones por normalizar la lengua catalana, y aunque el catalán es lengua de prestigio en la actualidad en Cataluña (*vid.* 4.1.), el peso de los más de 40 años en que su uso estuvo relegado al ámbito familiar durante el pasado siglo, sigue evidenciándose en una proporción nada desdeñable de transferencias del castellano en el léxico disponible catalán.

8. CONCLUSIONES

En el presente capítulo realizamos una valoración general de los objetivos que nos propusimos al emprender este trabajo, así como una revisión del cumplimiento o no de las hipótesis que manejamos en la introducción. A pesar de que a medida que ahondábamos en el análisis hemos ido ofreciendo algunas conclusiones parciales, consideramos que el volumen de los datos manejados también hace necesaria esta recapitulación. En primer lugar, es de recibo señalar que la disponibilidad léxica se ha mostrado, una vez más, como una herramienta eficiente para describir la estratificación sociolectal de una comunidad lingüística, y, en concreto, se ha revelado como método de análisis adecuado para medir el contacto lingüístico entre dos lenguas, tal como ya habían notado con anterioridad otros autores (Mackey 1971a).

Con respecto a los objetivos que nos propusimos al emprender este trabajo, el primero de ellos, establecer el léxico disponible en lengua castellana y catalana, se ha visto cumplido con la elaboración de los diccionarios de disponibilidad que ocupan los capítulos 10 y 11 del presente trabajo.

El segundo de nuestros objetivos era mostrar la variación de cada uno de los dos corpus en relación con las distintas variables lingüísticas —los centros de interés— y las cinco variables sociales. A continuación, reseñamos las principales conclusiones a las que hemos llegado tras la comparación cuantitativa de los datos recabados en esta investigación.

En relación con los índices totales de palabras y de vocablos se advierte que, tanto en un parámetro como en otro, el catalán supera ampliamente al castellano, toda vez que tanto la productividad léxica (la cantidad de respuestas bajo cada estímulo) como la riqueza de los centros de interés (las palabras diferentes actualizadas en cada área) es mayor en lengua catalana. La hipótesis que manejamos en la introducción sobre el predominio del catalán en nuestra área de estudio se ve, por lo tanto, refrendada con los datos obtenidos. Tal como hemos ido defendiendo a lo largo de este trabajo, los altos índices de uso de la lengua catalana en nuestra provincia, así como el elevado porcentaje de población autóctona que reside en ella, justifican este resultado. La muestra, reflejo final del universo estudiado, avala esta distribución, dado que la

mayoría de informantes tiene como lengua materna el catalán. En este sentido, los resultados evidencian la preferencia de los encuestados por su propia lengua materna y la influencia de la escolarización que se realiza también íntegramente en esta lengua.

Concretamente, sobre el total de palabras, en lengua catalana se recogen 3023 respuestas más (73038 en castellano frente a 76061 en catalán), lo que supone 177.83 en el promedio de palabras por cada centro de interés. Las diferencias se reducen al evaluar el comportamiento individual de los informantes y son de 12.6 por informante a lo largo de toda la encuesta y de 0.74 por informante y centro de interés. En cuanto a los vocablos, en total se actualizan 240 más en catalán (6824 en castellano frente a 7064 en catalán), lo que supone una diferencia de 14.12 términos por centro de interés.

Según el número de palabras —magnitud de respuestas— ya sea el total o el promedio, los centros se ordenan del siguiente modo:

- En castellano: *Animales, El cuerpo humano, Comidas y bebidas, Ropa y complementos, La ciudad, La escuela, Profesiones y oficios, Colores, El campo, Medios de transporte, Juegos y distracciones, La cocina y sus utensilios, Partes de la casa, Objetos colocados en la mesa para la comida, Muebles de la casa, Calefacción, iluminación y formas de airear un recinto y Trabajos del campo y del jardín.*
- En catalán: *El cuerpo humano, Comidas y bebidas, Animales, Ropa y complementos, La escuela, La ciudad, Profesiones y oficios, Colores, El campo, Medios de transporte, La cocina y sus utensilios, Juegos y distracciones, Partes de la casa, Objetos colocados en la mesa para la comida, Muebles de la casa, Calefacción, iluminación y formas de airear un recinto y Trabajos del campo y del jardín.*

Según el número de vocablos, los centros siguen el siguiente orden:

- En castellano: *El campo, La ciudad, Profesiones y oficios, Juegos y distracciones, Comidas y bebidas, Trabajos del campo y del jardín, La escuela, El cuerpo humano, La cocina y sus utensilios, Animales, Calefacción, iluminación y formas de airear un recinto, Ropa y complementos, Medios de transporte, Muebles de la casa y Colores.*
- En catalán: *El campo, La ciudad, Profesiones y oficios, Juegos y distracciones, Comidas y bebidas, La escuela, Animales, Trabajos del campo y del jardín, La*

cocina y sus utensilios, El cuerpo humano, Ropa y complementos, Medios de transporte, Calefacción, iluminación y formas de airear un recinto, Objetos colocados para la comida y Muebles de la casa.

Tal como se observa en las clasificaciones obtenidas, y al igual que han señalado todas las investigaciones que nos preceden, no se observa relación alguna en la clasificación de las áreas en función de estos dos índices. Del mismo modo, hay una mayor coincidencia entre los rangos que alcanza cada centro de interés en las dos lenguas estudiadas según el número de palabras –donde diez áreas obtienen el mismo rango– que según el número de vocablos –donde solo siete coinciden–, tal como preveíamos en la introducción. Los resultados realizados en distintos lugares hispánicos ya habían confirmado esta hipótesis, que ahora corroboramos en la comparación de dos lenguas distintas.

Otro de los objetivos básicos de nuestro trabajo era comparar la producción en las dos lenguas. Para ello nos servimos tanto de los totales y de los promedios de palabras como del número de vocablos. Sin embargo, hay que aclarar que la información que nos ofrecen los promedios de palabras es más fiable por dos motivos: en primer lugar, porque hemos sometido nuestros datos a un análisis estadístico que ha avalado estos resultados y, en segundo lugar, debido a que los totales de vocablos por centro de interés son más alterables que los promedios, que presentan una mayor estabilidad frente a posibles informantes cuyo comportamiento se aleje de los patrones habituales.

Mediante la comparación de los promedios por centro de interés de cada área se observa cómo, de nuevo, el catalán obtiene siempre el mejor resultado salvo en tres áreas: *Calefacción, iluminación y formas de airear un recinto, La ciudad y Juegos y distracciones*, donde el resultado es favorable al castellano. Salvo la divergencia observada en *Juegos y distracciones* (de 0.6 en el promedio y de 143 palabras en el total), las restantes diferencias a favor del castellano son escasas. El predominio del castellano en esta última área se justifica por la influencia de esta lengua en los distintos modos de ocio juvenil, principalmente los relacionados con los medios de comunicación y los mercados de consumo.

En las restantes categorías, los promedios y los totales son superiores siempre en catalán. Las mayores divergencias se dan en *El cuerpo humano*, donde se superan las tres respuestas de media, en *Comidas y bebidas*, con 2.28 de diferencia, y en *Ropa y complementos*, con 1.57. Solo en un centro más, *La cocina y sus utensilios*, las diferencias superan la unidad. Tras aplicar el análisis T de Student para muestras emparejadas, se constata cómo la diferencia entre el promedio general en las dos lenguas (17.90 en castellano *versus* 18.64 en catalán) es significativa, así como también lo es la observada en ocho centros de interés, donde el resultado es favorable al catalán en función de la lengua en que se completa la encuesta: *El cuerpo humano*, *Ropa y complementos*, *Comidas y bebidas*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios*, *La escuela*, *Trabajos del campo y del jardín* y *Profesiones y oficios*. En ninguno de los centros en los que el castellano aventaja al catalán, dicha diferencia resulta significativa estadísticamente.

Por tanto, cabe señalar que los resultados acerca de la productividad léxica apuntan claramente a que el caudal léxico de los informantes, analizados en general, es mejor en lengua catalana, toda vez que, en los únicos centros donde el castellano supera al catalán, esta diferencia no tiene validez estadística. Es importante recordar en este punto que, como bien argumenta Hernández Muñoz (2006: 320), no siempre un caudal léxico mayor conlleva un conocimiento superior de vocabulario, sino que las diferencias que evalúa la prueba de disponibilidad léxica pueden radicar simplemente en una mayor habilidad para evocar palabras.

El predominio del catalán, en casi todas las áreas temáticas, puede explicarse, de un lado, por el peso de los informantes cuya lengua materna es el catalán, que suponen el 65.42% de la muestra. Este hecho adquiere más relevancia cuando sabemos, desde la investigación de Hernández Muñoz (2006), que la disponibilidad se asocia a factores cognitivos como la edad de adquisición o la familiaridad, estrechamente ligados a la lengua materna del individuo. En segundo lugar, el peso de la escolarización, realizada íntegramente en lengua catalana, acentúa las diferencias entre las dos lenguas y las extiende a las áreas temáticas que se trabajan en la escuela. En último lugar, la influencia del entorno (urbano o rural) también puede percibirse en las diferencias que se dan en algunos centros de interés en concreto. Como es sabido, la inmigración castellano hablante llegada en la posguerra se instaló principalmente en las ciudades,

hecho que hace que el catalán se asocie más a los pequeños núcleos poblacionales y el castellano con las grandes ciudades. En general, creemos que las razones que llevan al predominio del catalán en las áreas que se revelan como estadísticamente significativas son distintas en cada caso (*vid.* 6.1.3.), aunque en todas ellas se observan los factores que acabamos de mencionar, así como la vinculación estrecha de algunas áreas con la lengua utilizada en el hogar.

En cuanto a las diferencias entre las dos lenguas en función del número de vocablos o palabras distintas, el catalán también predomina en la mayoría de centros, concretamente en once de las diecisiete áreas se recogen más vocablos en catalán. No parece ser casual que en las mismas áreas en las que el castellano tiene una mayor productividad (*Calefacción, iluminación y formas de airear un recinto, La ciudad y Juegos y distracciones*) también se detecte un mayor número de vocablos en esta lengua, lo que puede indicar una cierta preferencia por el uso del castellano en estos ámbitos. En tres centros de interés más, *El cuerpo humano, Trabajos del campo y del jardín y Profesiones y oficios*, se alcanza un mayor número de vocablos en castellano, aunque son áreas donde la productividad en catalán es mayor. En general, las diferencias numéricas son superiores en los centros que presentan un mayor número de vocablos en catalán que en los que presentan un resultado favorable al castellano. La mayor disimilitud detectada es la de *Animales*, con 87 vocablos más en catalán, *Comidas y bebidas*, con 47, y *Objetos colocados en la mesa para la comida*, con 44. Al igual que sucedía con el total de palabras, en general, se percibe en estos resultados la influencia de la lengua materna de la mayoría de la muestra así como el peso de la escolarización realizada íntegramente en lengua catalana.

Con respecto a los índices que examinan el grado de convergencia entre las respuestas, a pesar de que ofrecemos también el de la densidad léxica (*vid.* 6.1.5), nos basamos en los datos obtenidos mediante el índice de cohesión. Según este parámetro se obtiene una clasificación muy similar de las áreas temáticas en las dos lenguas: en concreto, nueve áreas coinciden exactamente en el mismo valor. En la clasificación, las coincidencias también son notables, sobre todo en los centros más abiertos o difusos. Según este índice, las áreas temáticas se clasifican según el modo que sigue en los dos idiomas analizados:

(a) centros de cohesión alta, es decir, muy compactos o cerrados, cuyo índice supera el 0.6: *El cuerpo humano, Ropa y complementos, Animales, Colores* y, en lengua castellana, *Medios de transporte*;

(b) centros de cohesión media, cuyo índice oscila entre 0.4 y 0.6: *Partes de la casa, Muebles de la casa, Comidas y bebidas, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, La escuela*, y, en catalán, *Medios de transporte*; y, finalmente,

(c) centros dispersos o difusos, cuyo índice es inferior al 0.4: *Calefacción, iluminación y formas de airear un recinto, La ciudad, El campo, Trabajos del campo y del jardín, Juegos y distracciones y Profesiones y oficios*.

Las similitudes que acabamos de observar en los dos idiomas nos informan de que la caracterización de los centros, según esta magnitud, es constante por encima de la lengua en que se completa la encuesta, de una forma similar a como se veía al analizar los promedios de respuesta por cada centro. Esto es lógico puesto que, como ya se ha señalado (Hernández Muñoz 2006: 155), este valor se relaciona claramente con la estructura interna de cada centro de interés, que permanece inalterada al cambiar de idioma. Las diferencias mayores en este sentido se dan en dos áreas temáticas: *El cuerpo humano*, claramente más cohesionado en catalán, y *Animales*, más compacto en castellano. La notable mayor productividad en lengua catalana explica la diferencia en el primero de los centros de interés, mientras que la mayor riqueza léxica en lengua vernácula de *Animales* —probablemente influida por el gran número de fenómenos de contacto lingüístico que contiene esta área (*vid.* 7.3)— lleva a que las respuestas sean más variadas en lengua catalana, y, por tanto, el índice de cohesión sea menor. En nuestros datos también se observa, de forma muy clara, la relación inversamente proporcional ya constatada (Gómez Molina y Gómez Devís 2004: 87, Hernández Muñoz 2004: 61-62 y Fernández Smith *et al.* 2008: 97-98) entre el índice de cohesión y el número de vocablos. De este modo, las áreas que tienen más vocablos en castellano tienen un índice de cohesión mayor en catalán, y, de forma contraria, los estímulos con más vocablos en catalán tienen un índice de cohesión generalmente igual o mayor en

castellano⁹¹⁵. No obstante, las coincidencias entre el valor que alcanza cada área en cada una de las dos lenguas son numerosas, como ya hemos comentado.

Un punto muy interesante de nuestro estudio es la comparación de nuestros datos con los recogidos, en idénticas condiciones metodológicas en otros territorios. A pesar de que estas comparaciones presentan algunos inconvenientes (*vid.* 6.1.7), nos dan una idea del caudal de léxico disponible existente en distintos lugares. Si comparamos la media que obtienen nuestros informantes en lengua castellana (17.90) con la de otros lugares, se constata claramente cómo nuestros datos están por debajo de la media de la mayoría de territorios, tanto en relación con otras sintopías hispánicas monolingües como en relación con otras provincias, ciudades o comunidades autónomas donde se emplean dos códigos lingüísticos. En nuestra opinión, estos bajos resultados se explican porque, en nuestra muestra, reflejo último del universo estudiado, la mayoría de los informantes tiene una lengua materna distinta al castellano, hecho que no se da en ninguna de las restantes investigaciones, con la excepción de Bilbao⁹¹⁶. Además, en ninguna de las sintopías analizadas la segunda lengua tiene la extensión social que alcanza el catalán en nuestra provincia, donde además, todos los informantes han sido escolarizados en esta lengua. En la comparación de nuestros datos sobre el catalán con los recabados por Llopis (2009) en Valencia, en cambio, se observa una gran concomitancia; de hecho el promedio general es prácticamente idéntico en las dos provincias (18.60 en Lleida y 18.30 en Valencia).

A continuación ofrecemos las principales conclusiones a las que hemos llegado tras el análisis cuantitativo de la estratificación lingüística que ejercen las cinco variables sociales analizadas:

⁹¹⁵ De igual forma, en el análisis de la cohesión según los distintos subgrupos sociales también se aprecia, de forma más evidente si cabe, cómo el tamaño de la muestra influye en el valor de este índice. Esta es la causa por la que siempre el grupo con un menor número de informantes, que recaba también un número inferior de vocablos, alcanza siempre el índice de cohesión más alto en comparación con los restantes subgrupos sociales (Fernández Smith *et al.* 2008: 93).

⁹¹⁶ De hecho, nuestros resultados mejoran, en ambas lenguas, si se toma en consideración únicamente los datos de los informantes que completan la lengua en su propia lengua materna, esto es, los castellanohablantes en castellano y los catalanohablantes en catalán, respectivamente. Sin embargo, la media en lengua castellana continúa estando por debajo de la mayor parte de regiones hispánicas (*vid.* 6.2.5.).

- (a) Variable social *sexo*: esta una de las variables con mayor valor discriminante de las cinco analizadas, dado que la diferencia entre el promedio global que alcanzan varones y mujeres es significativa estadísticamente tanto en castellano como en catalán. También es posible extender al universo estudiado la distancia observada entre la productividad de los dos sexos en ocho áreas temáticas en castellano (*Ropa y complementos, Muebles de la casa, Comidas y bebidas, La cocina y sus utensilios, La escuela, La ciudad, Animales y Colores*) y en trece en catalán (*Ropa y complementos, Muebles de la casa, Comidas y bebidas, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, La escuela, Calefacción, iluminación y formas de airear un recinto, La ciudad, El campo, Animales, Juegos y distracciones, Profesiones y oficios y Colores*), donde el colectivo femenino siempre obtiene el mejor resultado.

Cabe destacar que, aunque los datos siempre favorecen al colectivo femenino, la diferencia entre las medias es mayor en catalán (2.09 respuestas) que en castellano (1.05). El mismo hecho se observa en el análisis de los centros de interés: en castellano, los varones obtienen promedios superiores en cinco áreas, mientras que, en catalán, únicamente en una. En el análisis detallado de las categorías, se puede observar aún cierta preferencia por determinados aspectos de la realidad o la pervivencia del rol tradicionalmente desempeñado por cada uno de los dos sexos. De este modo, las mujeres obtienen mejores promedios en las dos lenguas en *Ropa y complementos, Comidas y bebidas, La cocina y sus utensilios* o *Colores*, mientras que los varones aventajan en la producción en *Medios de transporte* –sólo en castellano– y en *Trabajos del campo y del jardín* –en las dos lenguas⁹¹⁷–. Sin embargo, las áreas donde se observa esta superioridad masculina no resultan significativas estadísticamente, y, de otro lado, las diferencias a favor de las mujeres se extienden a áreas aparentemente neutras, como *La escuela, La ciudad* o *Animales*, lo que nos lleva a relacionar nuestros resultados con las teorías que otorgan una mayor competencia léxica a

⁹¹⁷ Estas preferencias por ciertos aspectos de la realidad también se observan en los vocablos, a pesar de ser menos, los varones actualizan más palabras diferentes en torno a *Medios de transporte* o *Trabajos del campo y del jardín*.

las mujeres con respecto al aprendizaje de primeras o segundas lenguas (Lozano 1995, Calero 2007)⁹¹⁸.

- (b) En cuanto a la variable social *tipo de centro*, este factor no ejerce un efecto estadístico ni en el promedio global ni en la mayor parte de las áreas asociativas de las dos lenguas. No obstante, es posible refutar la hipótesis nula en algunos centros de forma aislada como en *Partes de la casa* (en las dos lenguas), en *Ropa y complementos* y *Muebles de la casa*, en castellano y en *Calefacción, iluminación y formas de airear un recinto*, en catalán. En estos centros de interés, y en el promedio general, se observa una tendencia contraria según el corpus que analicemos. Con una diferencia casi idéntica (de 0.53 y de 0.52 respectivamente), los centros privados obtienen un promedio mejor en castellano y, en cambio, los públicos muestran un resultado más favorable al catalán. Entre los colegios concertados, ubicados todos ellos en la capital del *Segrià*, el resultado es superior en castellano en doce centros de interés; los centros públicos se decantan también en doce áreas por el catalán.
- (c) La variable *zona geográfica* tampoco ejerce un valor discriminatorio destacable en los resultados. Tras realizar el análisis estadístico pertinente, se observa que la diferencia no es significativa ni en los promedios globales ni en la mayoría de las áreas de interés. No obstante, al igual que sucedía con la variable *tipo de centro*, en algunas categorías semánticas aisladas la probabilidad se sitúa por debajo del 0.05. Estas son *Ropa y complementos*, *Objetos colocados en la mesa para la comida*, *La ciudad* y *Trabajos del campo y del jardín*, en castellano, y *Partes de la casa*, *Calefacción, iluminación y formas de airear un recinto*, *Trabajos del campo y del jardín* y *Animales*, en catalán. De forma muy similar a la que se veía al observar la estratificación del léxico según la variable social *tipo de centro*, también se aprecian tendencias distintas según el corpus que analicemos. En lengua castellana, los informantes procedentes de centros urbanos consiguen mejores promedios, tanto en el global como en doce áreas temáticas. En catalán, en cambio, el alumnado de los institutos situados fuera de la capital consigue medias más favorables, concretamente, su resultado es superior en global y en quince centros de interés. La distancia en ambos casos,

⁹¹⁸ Cabe recordar que la noción de léxico disponible no solo se asocia a “poseer más palabras” sino sobre todo a tener más recursos para evocarlas. De otro lado, sería necesario saber cómo esas palabras se integran realmente en el discurso de los hablantes para sacar conclusiones rigurosas sobre la competencia léxica a partir de estos datos.

entre los promedios generales, es bastante similar (0.46 frente a 0.6). En el análisis de esta variable llama la atención cómo el entorno inmediato puede influir en el caudal de léxico disponible, dado que no parece un hecho casual que los informantes urbanos obtengan, independientemente de la lengua, un resultado superior en torno a *La ciudad* y, de forma inversa, los sujetos rurales ostenten siempre un promedio superior en *Trabajos del campo y del jardín*, en cualquiera de las dos lenguas analizadas.

(d) La muestra se distribuye, según la variable *nivel sociocultural*, en tres niveles (alto, bajo y medio) teniendo en cuenta no solo el nivel de estudios y la profesión de los progenitores, sino también un parámetro no habitual, al que denominamos “nivel cultural de la familia”, que obtuvimos tras incorporar en la encuesta una serie de preguntas sobre la cantidad de libros que leía el informante o la asiduidad con la que sus progenitores asistían a ciertos eventos culturales, entre otras magnitudes. Tras analizar los resultados obtenidos, se observa cómo en las dos lenguas se crea el mismo patrón de estratificación considerando los promedios globales, y es el nivel medio el que lidera las dos clasificaciones, seguido del alto y del bajo, en este orden. La distancia entre el resultado mejor y peor (esto es, entre los niveles bajo y medio, como acaba de explicarse) es similar en las dos lenguas (1.35 en castellano y 1.29 en catalán). En el análisis por centros de interés, se puede ver cómo el nivel medio predomina en diez centros de interés en castellano, mientras que en los siete restantes es el nivel alto el que obtiene un resultado mejor. En catalán, el nivel medio recaba los mejores promedios en catorce áreas, en tres los ostenta el nivel alto y en una coinciden ambos sociolectos en el resultado.

Al realizar el análisis mediante ANOVA de la diferencia entre los promedios y tras llevar a cabo la prueba *post-hoc* pertinente, se constata cómo los niveles implicados en la significación son el medio y el bajo. La posibilidad de refutar la hipótesis nula se extiende a cinco áreas en castellano (*El cuerpo humano*, *Comidas y bebidas*, *Objetos colocados en la mesa para la comida*, *El campo* y *Medios de transporte*) y a seis en catalán (*El cuerpo humano*, *Muebles de la casa*, *Objetos colocados en la mesa para la comida*, *El campo* y *Medios de transporte*). Estos resultados están en consonancia con el análisis descriptivo, ya que, al comparar los promedios de cada centro de interés, se constata cómo,

tanto en castellano como en catalán, el mayor contraste se da entre los niveles medio y bajo. De hecho, en las dos lenguas, el estrato medio supera en todas las categorías semánticas al bajo, salvo en un centro de interés en castellano donde ambos niveles obtienen exactamente la misma media. En la relación entre el nivel alto y el medio se advierte cómo en catalán la actuación lingüística difiere más que en castellano, donde ambos estratos obtienen cifras más similares. Por el contrario, en lengua española se observa una mayor distancia en la media global entre los niveles alto y bajo; en catalán, en cambio, la diferencia general es menor y diverge en función del área concreta analizada. Finalmente, hay que considerar que nuestros datos están en consonancia con otras investigaciones recientes de disponibilidad, donde también el estrato alto se sitúa por debajo del medio (*cf.* Martínez Olmos 2008 y Fernández Smith *et al.* 2008), en la línea de los primeros estudios de sociolingüística variacionista⁹¹⁹. Las investigaciones futuras corroborarán o no esta tendencia, aunque también sería conveniente profundizar en las diferencias entre estratos sociales, a la luz de los conocimientos sociológicos actuales⁹²⁰.

- (e) La última de las variables analizadas, la *lengua materna*, divide a los informantes en cuatro grupos de tamaño muy desigual. Mientras la mayoría de la muestra se adscribe al castellano o al catalán, también existen dos subgrupos muy reducidos: los que declaran tener ambas lenguas (castellano y catalán) como maternas y los hablantes de aranés. En la comparación general de los cuatro grupos lingüísticos, se observa cómo es, precisamente, este último colectivo —el de *la Vall d’Aran*— el que obtiene los mejores resultados tanto en castellano (con una media de 20.59) como en catalán (con una media de 19.21). Al resultado tan favorable de este grupo le sigue, en castellano, el del grupo que tiene ambas lenguas como maternas y, en catalán, el de los catalanohablantes, que superan así a los informantes de “ambas lenguas”. Los peores resultados los ostentan, respectivamente, los catalanohablantes en el corpus castellano y los castellanohablantes en el corpus catalán.

⁹¹⁹ Labov (1972) demostró que la clase social media es la que más se aproxima a la norma, lo que podría traducirse aquí por un mayor interés de este estrato social en la ampliación del vocabulario de sus hijos y en el uso del dominio de la lengua como modo de promoción social.

⁹²⁰ Tampoco hay que olvidar que se debe tener en cuenta las características concretas de la población estudiada para obtener un resultado relevante en la comparación entre estratos sociales, tal como se colige del trabajo de Hernández Muñoz (2006). De otra parte, también es fundamental contar con un número adecuado de miembros que sean suficientemente representativos de cada estrato social.

Con la finalidad de profundizar en las diferencias entre los dos colectivos mayoritarios, hemos prescindido de los datos de los dos grupos lingüísticos menores (araneses y “ambas lenguas”) y hemos comparado únicamente los promedios de los informantes cuya lengua materna es el castellano o el catalán, que suponen, en total, el 92.92% de la muestra⁹²¹. En el cotejo de estos dos grupos se observa, con total claridad, cómo el colectivo que completa la encuesta en su propia lengua materna obtiene, tanto en el global como en la mayoría de centros de interés, el resultado más favorable, lo que nos indica que el léxico que se evoca a través de los diecisiete centros de interés está antes ligado al entorno básico y a la familia que al currículum escolar⁹²².

En concreto, en el corpus castellano, los castellanohablantes tienen mejores promedios en todas las áreas, salvo en dos, *La ciudad* y *Medios de transporte*, mientras que, en el corpus catalán, la preponderancia de los catalanohablantes es constante, excepto en *Muebles de la casa*. Del análisis estadístico —realizado también solo con los datos de los dos grupos lingüísticos mayoritarios— se infiere que la variable *lengua materna* provoca diferencias estadísticamente significativas, dado que el valor de la probabilidad es inferior a 0.05 tanto en castellano como en catalán. Además, es posible extender esta diferencia, en el caso del castellano, a siete áreas (*Ropa y complementos*, *Partes de la casa*, *Muebles de la casa*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios*, *Animales y Colores*)⁹²³ y, en el caso del catalán, a diez (*El cuerpo humano*, *Partes de la casa*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios*, *Calefacción, iluminación y formas de airear un recinto*, *La ciudad*, *El campo*, *Trabajos del campo y del jardín*, *Animales* y *Profesiones y oficios*).

⁹²¹ Varias razones nos han llevado a tomar esta decisión, de un lado, el exiguo número de representantes de los otros dos colectivos, y de otro, el hecho de que las características sociológicas concretas de estos informantes (por ejemplo, todos los informantes con lengua materna aranesa son mujeres) podían interferir en los resultados.

⁹²² Si no fuera así, sería difícil que se detectaran diferencias en función de la lengua materna, dado que todos se escolarizan en la misma lengua.

⁹²³ Cabe destacar que el centro 8, *La escuela*, se sitúa justo en el límite de la significación estadística (0.050).

Mientras que la actuación lingüística en el corpus castellano es más similar en función de la lengua materna (los castellanohablantes tienen un promedio de 18.52 y los catalanohablantes de 17.46), en el corpus catalán se detectan más diferencias a favor de la lengua materna de los informantes (17.53 y 19.08 respectivamente), por tanto, la distancia que existe en el corpus castellano entre la media global de los dos grupos lingüísticos (1.06) es inferior a la que se detecta en el corpus catalán, situada en 1.55 respuestas. En nuestra opinión, esta disimilitud se debe a que los dos perfiles de alumnado (el castellanohablante y el catalanohablante) tienen una relación distinta con las dos lenguas cooficiales (*vid.* 6.1.3 y 6.2.5.)⁹²⁴ y, por esta razón, las diferencias son mayores en el corpus catalán. En concreto, la cifra que alcanzan los catalanohablantes en su propia lengua materna es superior a la de los castellanohablantes en la suya (19.08 frente a 18.52), dado que estos informantes refuerzan el léxico de su lengua materna a través de la escuela, donde se trabaja el vocabulario de varias de las áreas de la encuesta. Por el contrario, el límite inferior es prácticamente idéntico en los dos corpus (17.46 los catalanohablantes en el corpus castellano y 17.53 los castellanohablantes en el corpus catalán), lo que nos da una idea de la coexistencia de las dos sistemas lingüísticos en la sociedad catalana⁹²⁵.

También las divergencias entre los dos perfiles de alumnado pueden explicar la diferencia entre los centros que resultan estadísticamente significativos en cada lengua. No obstante, en algunas áreas donde el factor lengua materna tiene relevancia estadística, como *Partes de la casa*, *Objetos colocados en la mesa para la comida* y *La cocina y sus utensilios*, se aprecia con total claridad la relación del léxico evocado en ellas con la lengua usada en el hogar. En otras, como en *Animales*, es llamativo que se sigan produciendo diferencias en función de la lengua materna en los dos corpus, dado que es una

⁹²⁴ A grandes rasgos, mientras un colectivo, el castellanohablante, se escolariza íntegramente en catalán, el otro, el catalanohablante, inicia el contacto con el castellano en la educación primaria, a los 6 años, con una asignatura de tres horas por semana. Sin embargo, como hemos venido repitiendo (*vid.* 6.1.3 y 6.2.5.), en una situación como la nuestra de bilingüismo social extenso, el contacto de cada individuo con las dos lenguas depende de sus particulares circunstancias vitales, aunque hay que reconocer el gran predominio social del catalán en nuestra provincia, y, sobre todo, en los núcleos rurales, a pesar de que el castellano impera en algunos ámbitos concretos (*vid.* 4.3.2.).

⁹²⁵ Parece difícil que se pudiera alcanzar este resultado tan similar si el contacto de los catalanohablantes con el castellano se limitara a las tres horas semanales de enseñanza del castellano en la escuela.

categoría ampliamente trabajada en la escuela, lo que podría determinar *a priori* diferencias únicamente en el corpus castellano. En este caso, tal como demuestra Hernández Muñoz (2006), las divergencias se explican porque esta categoría comprende palabras aprendidas antes que las pertenecientes a otras categorías y, como sabemos, la edad de adquisición, al igual que la familiaridad y la tipicidad, son los factores cognitivos que explican la producción del léxico disponible como L1 (Hernández Muñoz 2006).

A la vista de la influencia que ejerce la *lengua materna* en los resultados, es fácil explicar ahora la causa de que se observaran dos tendencias opuestas al analizar los datos de las variables *tipo de centro* y *zona geográfica*. Como ya reparamos al analizar estas dos variables (*vid.* 6.2.2 y 6.2.3.), el mejor resultado en castellano de los informantes de centros privados y urbanos y, de forma inversa, la preferencia clara de los sujetos de institutos públicos y situados fuera de la capital se relaciona con la propia distribución de la muestra y con la influencia que ejerce la variable *lengua materna*, toda vez que hay un mayor número de alumnado cuya lengua materna es el castellano en los centros privados y en los urbanos y un porcentaje mayor de estudiantes cuya lengua materna es el catalán en los centros públicos y rurales⁹²⁶. El mismo comportamiento se observará, como veremos a continuación, al analizar las diferencias de cada colectivo social según la lengua de la encuesta.

Tal como supusimos en la introducción, a la vista de los resultados obtenidos, podemos concluir que la estratificación que se observa en función de las variables sociales es idéntica en los dos corpus y otorga únicamente significación estadística a la lengua materna, al sexo y al nivel sociocultural. De otro lado, también es interesante señalar que en lengua catalana se constata una mayor variación tanto en la producción de palabras en torno a cada área léxica como con respecto al número de vocablos por centro de interés, en el sentido de que, si comparamos la diferencia entre los centros que alcanzan un mayor y un menor número de palabras o vocablos, la distancia siempre es

⁹²⁶ Como ya dijimos en el análisis (*vid.* 6.2.2), desconocemos si en los centros privados se concentra mayor número de castellanohablantes o este es un hecho casual de nuestra muestra. Por el contrario, la distribución de la población en la provincia de Lleida sí responde al mismo patrón de nuestros informantes: en la capital, la población de origen castellanohablante tiene un peso mayor que en el resto de la provincia, dado que la mano de obra y los miembros de la administración pública llegados durante la posguerra tendieron a concentrarse en las ciudades (*vid.* 4.3.2).

mayor en catalán. Lo mismo sucede al analizar la incidencia de las variables sociales en cada corpus por separado: las diferencias son mayores en catalán que en castellano.

En el análisis de las diferencias en la producción cruzando la lengua de la encuesta y cada grupo social, hemos de aclarar que, como ya indicamos (*vid.* 6.2), no hemos sometido estos datos a ningún programa estadístico (con la excepción de la variable *lengua materna*), lo que esperamos hacer en un futuro próximo, por lo que lo único que podemos describir son posibles tendencias. Presentamos, seguidamente, las conclusiones parciales del análisis de ese cruce de datos.

- (1) En relación con la variable social *sexo*, el colectivo masculino muestra un comportamiento bastante más equilibrado en la producción de las dos lenguas que el femenino, que muestra predilección por el catalán. De hecho, el promedio entre los varones solo dista en una décima (17.26 en castellano frente a 17.36 en catalán) y al analizar los resultados según el centro de interés también se observa una gran equivalencia (en nueve centros el resultado es mejor en castellano y en ocho, en catalán). Por el contrario, entre las mujeres la diferencia en la media global a favor del catalán es de 1.14 respuestas (18.31 frente a 19.45) y el predominio de la lengua vernácula se observa de forma constante en todas las áreas (salvo en *Juegos y distracciones*, por las razones que ya hemos comentado). Al analizar la riqueza léxica mediante el número de vocablos, se aprecia la misma tendencia en los dos sexos: a pesar de que el resultado global en ambos casos es mejor en catalán, en las mujeres se observa de forma más acusada la preferencia hacia la lengua materna de la mayoría de la muestra tanto en la diferencia total como en el número de áreas donde el promedio es mejor en catalán⁹²⁷.
- (2) En la variable *tipo de centro* se observan las mismas tendencias ya comentadas en la estratificación de los dos corpus según esta variable social. De este modo, el alumnado de los centros concertados ostenta un promedio favorable al castellano en diez centros de interés y en el promedio global, aunque por una diferencia insignificante de 0.05 centésimas (18.30 frente a 18.25). Con una distancia mayor —de una respuesta— los informantes de los centros públicos se

⁹²⁷ Las mujeres producen 256 vocablos más en lengua vernácula y los varones 63. Téngase en cuenta, sin embargo, que las mujeres son mayoría en la muestra (61.25% frente a 38.75%).

decantan por el catalán (17.77 frente a 18.77 en la media global), también su resultado es superior en esta lengua en quince de las diecisiete categorías semánticas analizadas. En el número de vocablos se detecta el mismo patrón: el alumnado de centros concertados actualiza más palabras distintas en castellano, mientras que el de los centros públicos tiene mayor riqueza léxica en catalán. Como ya hemos comentado, la causa de este comportamiento radica en la concentración de un mayor número de alumnos cuya lengua materna es la catalana en los institutos públicos que en los concertados, donde se aprecia una distribución más equitativa del alumnado en función de esta variable social.

- (3) En cuanto a la variable social *zona geográfica*, ambos colectivos alcanzan promedios favorables al catalán, aunque, de forma similar a como se observaba en la variable anterior, los informantes de los centros ubicados en la capital ostentan una media muy similar en ambas lenguas, con una diferencia de solo 0.21 (18.13 frente a 18.34). Concretamente, en nueve centros de interés su resultado es mejor en castellano, y en ocho, por el contrario, en catalán. El alumnado de los centros rurales, situados en distintos puntos de la provincia, muestra una preferencia mayor por el catalán, con una diferencia de 1.27 respuestas en esta lengua (17.67 frente a 18.94). En concreto, los datos que ofrecen son mejores en todas las áreas en lengua vernácula, salvo en *La ciudad y Juegos y distracciones*, donde se observa la influencia de la lengua castellana. En la riqueza léxica se constata la misma tendencia: ambos grupos exhiben un mayor número de vocablos en catalán, pero el colectivo rural con una mayor diferencia (246 frente a 58)⁹²⁸. Al igual que sucedía en la variable anterior, las diferencias entre el comportamiento lingüístico de estos dos colectivos se explican por la propia distribución de la muestra y por el peso de la variable *lengua materna*, dado que, proporcionalmente, hay un mayor número de alumnado castellano hablante en los centros urbanos que en los rurales, como también se observa en el conjunto de la población del territorio.
- (4) La variable *nivel sociocultural* determina la existencia de tres sociolectos: el alto, el medio y el bajo, cuya relación con las dos lenguas cooficiales presenta distintos patrones. El nivel bajo es el que más claramente se decanta hacia el catalán, tanto en el promedio general (16.99 en castellano frente a 17.84) como

⁹²⁸ Recordemos que los dos grupos de informantes (urbanos y rurales) se componen de un número idéntico de miembros (120).

en la cantidad de áreas en que el resultado en la lengua vernácula es superior (todas salvo *Juegos y distracciones*). En el nivel medio, también se observa esta preferencia por el catalán, aunque con una distancia un poco menor entre los promedios (18.34 en castellano, frente a 19.13 en catalán). En catorce áreas asociativas el resultado también es favorable a la lengua catalana. Por el contrario, en el análisis del nivel alto destaca la cercanía entre los resultados obtenidos en cada lengua, únicamente separadas por 0.32 décimas en el promedio global (18.02 en castellano frente a 18.34 en catalán). La misma tendencia se observa en el desglose por centros de interés: en diez, el catalán obtiene medias más altas y, en siete, el castellano. Idéntico patrón se detecta en el análisis de la riqueza léxica de cada sociolecto: los estratos bajo y medio muestran un mayor número de vocablos en catalán, mientras que el estrato alto recaba una cantidad superior en castellano. Concluimos que sería necesario profundizar en las conclusiones alcanzadas según esta variable dado que, en general, el uso del castellano se asocia en Cataluña a las clases más populares, lo que contrasta con el resultado obtenido⁹²⁹.

- (5) Con respecto a la *lengua materna*, los dos colectivos minoritarios, esto es, el aranés y el de ambas lenguas como maternas, se decantan por el castellano, aunque en el caso de los araneses de forma mucho más contundente con un promedio de 20.59 en castellano y de 19.21 en catalán. En cambio, la actuación lingüística de los informantes que componen el grupo que tiene las dos lenguas oficiales como maternas es muy similar en los dos idiomas (18.85 en castellano y 18.68 en catalán). En el análisis por áreas temáticas, el grupo aranés consigue mejores resultados en quince áreas en castellano, mientras que el colectivo de ambas lenguas lo hace en doce centros de interés. La misma tendencia se observa en el número de vocablos: los araneses muestran, de forma clarísima, una mayor riqueza léxica en castellano, por el contrario, el colectivo de ambas lenguas, aunque obtiene un mayor número de palabras distintas en catalán, presenta unos resultados mucho más aproximados⁹³⁰. A pesar del tamaño tan reducido de estos dos grupos, vale la pena destacar que el grupo de ambas

⁹²⁹ Sin embargo, como ya comentamos (*vid.* 4.3.2), también hay, en la capital del *Segrià*, una parte importante de funcionariado procedente de territorios castellanohablantes.

⁹³⁰ Concretamente, los araneses producen 141 vocablos más en castellano, mientras que en el colectivo de ambas lenguas, la diferencia es de 45 palabras distintas más en catalán. Téngase en cuenta, para evaluar estas disimilitudes, que el primer grupo lo integran 7 informantes y el segundo, 10.

lenguas, tal como era esperable, exhibe el resultado más cercano de los cuatro grupos lingüísticos en las dos lenguas, mientras que el colectivo aranés muestra una preferencia clara por el castellano, lo que también está en consonancia con el papel de esta lengua en esta parte de la provincia⁹³¹.

En relación con los dos colectivos mayoritarios, el grupo castellanohablante, donde se engloban los informantes que eligen el castellano como lengua materna, muestra una preferencia clara por el castellano, con un promedio de 18.52 en su lengua materna y de 17.53 en su segunda lengua. Tras aplicar la prueba T de Student —para lo cual aislamos únicamente a los informantes cuya lengua materna era el castellano— se constata cómo esta diferencia es significativa estadísticamente, lo que nos permite extrapolar este resultado al conjunto de población estudiada. En el análisis por centros de interés, se observa también cómo los resultados son superiores en lengua castellana en todos los centros, salvo en *El cuerpo humano* y *Comidas y bebidas*. De modo inverso, el colectivo integrado por los informantes cuya lengua materna es la catalana, también presenta un promedio superior en esta lengua (19.08 frente a 17.46); su predominio es constante en todas las áreas, excepto en *Juegos y distracciones*, donde, por pocas décimas, el promedio es mejor en castellano.

Tras comprobar que también en el colectivo catalanohablante la diferencia entre las medias globales es significativa, es digno de mención notar cómo la distancia entre los promedios de cada grupo, a favor de su propia lengua materna, es superior entre los informantes que tienen el catalán como primera lengua que entre los que tienen el castellano. Entre los catalanohablantes la diferencia se extiende a 1.62 respuestas, mientras que en los castellanohablantes esta cifra se sitúa en torno a la unidad (0.99). Una vez más, consideramos que la causa de que la actuación lingüística de estos dos colectivos difiera radica en que el grupo catalán ha sido escolarizado en su propia lengua materna. En el análisis por categorías, es interesante notar la relación de algunos centros de interés con la lengua utilizada en el hogar, como sucede en *Objetos colocados en la mesa para la comida*, donde se superan las dos respuestas de diferencia tanto entre los informantes cuya lengua materna es el castellano como entre los que

⁹³¹ En *la Vall d'Aran* la situación lingüística es compleja; la última encuesta de usos lingüísticos sitúa el castellano y el aranés como las lenguas preferidas por la población en los parámetros de *lengua inicial*, *de identificación* y de *uso habitual*, mientras que el catalán queda relegado a una tercera o cuarta posición, dada la importancia de otras lenguas en el valle pirenaico (*vid.* 4.3.2, especialmente el gráfico 4.28).

optan por el catalán. Cabe destacar también que entre los castellanohablantes las mayores diferencias se dan en *Animales* y en *El campo*, mientras que el grupo catalanohablante presenta las divergencias más acusadas en *El cuerpo humano* y en *Comidas y bebidas*, donde el resultado en su propia lengua materna es claramente superior. La misma tendencia constatada con respecto a la productividad se ve en la riqueza léxica de cada grupo lingüístico. De este modo, los castellanohablantes tienen más vocablos en castellano, con una diferencia de 203 palabras diferentes y los catalanohablantes, por su lado, también obtienen una cifra más alta de palabras distintas en su lengua materna, concretamente, 392 más⁹³².

Como acabamos de ver, hay una relación clara entre la productividad y la riqueza léxica, en torno a las lenguas de la encuesta. Es decir, en general, los informantes agrupados según cada variable social que presentan una superioridad léxica en la producción de palabras en un idioma, también actualizan más vocablos, o palabras distintas, en esa misma lengua. Esto sucede en la casi todos los subgrupos y, concretamente, los varones, las mujeres, los informantes de centros públicos, de centros urbanos, de centros rurales, de nivel bajo y medio y, finalmente, los catalanohablantes evocan más palabras y más vocablos, en un mismo tiempo de reacción, en lengua catalana que en lengua castellana, siguiendo la tendencia general a favor de la lengua vernácula que hemos visto a lo largo de todo este trabajo.

De cualquier modo, en general, las diferencias que se aprecian en la producción analizando la variación, de un lado, en función de las variables lingüísticas y, de otro, según los factores sociales, siempre son mayores en el primero de los casos, es decir, en función del centro de interés. Las semejanzas en este sentido entre las dos lenguas son más que notables, lo que nos da una idea de la universalidad de una prueba de evocación de léxico disponible.

Dentro ya del análisis cualitativo, y en relación con el tercer objetivo que nos propusimos —el de analizar en qué medida el léxico de cada una de las lenguas en contacto se ve influido por el de la otra lengua— hemos clasificado todos los elementos

⁹³² Recuérdese que el número de informantes influye sobre el total de vocablos recogido en cada área y que el grupo catalanohablante se compone de 157 informantes, mientras que el castellanohablante tiene 66.

de contacto lingüístico castellano-catalán, recogidos en los dos corpus. Para la clasificación de estas unidades nos hemos servido principalmente de dos criterios: el grado de integración social de cada término, medido a través de las menciones recibidas, y la noción de gramaticalidad/agramaticalidad. De este modo hemos obtenido la siguiente tipología: 1) *interferencias*, 2) *préstamos en vías de integración*, 3) *préstamos integrados*, 4) *cambios de código*, 5) *combinaciones híbridas*, 6) *calcos sintácticos* y 7) *calcos semánticos*, y los hemos agrupado según el centro de interés en que han sido encontrados. En el análisis de estos elementos, hemos constatado cómo la mayoría son producciones individuales, ya sean dentro de la norma (*cambios de código*) como fuera de ella (*interferencias*); estos dos fenómenos, entre los que predominan claramente los primeros, suponen, concretamente, el 80% del total de trasvases lingüísticos. Cabe señalar que el número absoluto de transferencias halladas es superior en lengua catalana en todas las categorías que acabamos de enumerar, salvo en el caso de los *calcos sintácticos*, donde en castellano se obtiene una cifra superior (20 frente a 8 casos).

En el análisis cuantitativo de este tipo de elementos, hemos analizado el peso que estos alcanzan, en los dos corpus, no solo sobre el total de vocablos, tal como se hace en muchos estudios actuales sobre disponibilidad (*vid.* 3.2 y 3.3.) sino también sobre el total de palabras. Es decir, hemos calculado el porcentaje de respuestas que contiene alguna transferencia en las dos lenguas, parámetro con el que consideramos que se observa mejor el valor real que adquieren los fenómenos de contacto en el uso. En este sentido, en castellano se dan 1509 respuestas con algún elemento transferido, que suponen el 2.07% del total; en catalán, en cambio, se registran 4141 respuestas de este tipo, lo que equivale al 5.44% del total de palabras. La misma tendencia se aprecia al analizar el porcentaje que suponen las transferencias (sin tener en cuenta las menciones que recibe cada una) sobre el conjunto léxico total: tanto su valor absoluto como su proporción en relación con el total de vocablos, son siempre superiores en el corpus catalán. En concreto, en esta lengua se recogen 945 elementos, el 13.38% del total de palabras distintas y, en castellano, estas mismas cifras se reducen prácticamente a la mitad (483 elementos, el 7.08%).

Si comparamos las dos lenguas y atendemos a la clasificación por centros de interés, constatamos cómo en los dos parámetros analizados la proporción y las cifras

absolutas de fenómenos de contacto lingüístico alcanzan índices más altos en lengua catalana. El único campo semántico que contiene más transferencias en castellano que en catalán es *Trabajos del campo y del jardín*, donde se advierte la dificultad de la mayoría de la muestra, de lengua materna catalana, de enunciar las distintas tareas agrícolas en castellano, hecho que refleja, en última instancia, la clara vinculación entre la lengua vernácula y los núcleos rurales. En concreto, si comparamos las divergencias entre el peso que supone en las dos lenguas el porcentaje de transferencias sobre el total de vocablos, las distancias más acusadas se dan en *Animales* (con un 25.58% de transferencias en catalán y solo un 10.09% en castellano), en *Medios de transporte* (con un 14.19% frente a un 1.81%), y en *La ropa* (20.77% frente a 8.59%).

Sobre el total de vocablos, los campos semánticos que presentan un mayor porcentaje de trasvases lingüísticos en castellano son *Partes de la casa* (12.18%), *Trabajos del campo y del jardín* (11.29%) y *Objetos colocados en la mesa para la comida* (10.50%). En lengua catalana, *Animales* (con un 25.58%), *Objetos colocados en la mesa para la comida* (con un 21.63%) y *Ropa y complementos* (20.77%) ocupan estas posiciones, con cifras notablemente superiores. En general, la clasificación de los centros en función de este parámetro es distinta en las dos lenguas, aunque se pueden reconocer dos tendencias comunes que propician la aparición de elementos de la otra lengua oficial. En primer lugar, la relación de ciertas áreas con la lengua materna del informante, que hacen que seguramente el alumnado que completa la encuesta en su segunda lengua tenga dificultad para evocar ciertos términos⁹³³ (tal como se ve en las posiciones adelantadas que alcanzan en ambas lenguas *Objetos colocados en la mesa para la comida* o *La cocina y sus utensilios*) y, en segundo lugar, la presencia abundante de transferencias en ciertas categorías caracterizadas por predominar en ellas un léxico tempranamente aprendido (como *Animales* y *Colores*). Otras áreas, como *Juegos y distracciones*, *La ciudad* o *El cuerpo humano*; por el contrario, son poco permeables a la presencia de fenómenos de contacto, tanto en castellano como en catalán. De cualquier modo, la hipótesis que manejamos en la introducción, relativa a que la presencia de fenómenos de contacto lingüístico variaría en función de los centros de interés, se ha visto sobradamente corroborada.

⁹³³ Hasta que no analicemos los datos en función de la *lengua materna* no podremos corroborar esta hipótesis, puesto que no conocemos cuál es el perfil social que evoca un mayor número de transferencias.

Finalmente, hemos llevado a cabo un estudio de los préstamos, ya estén en vías de integración o integrados, comparándolos, en cada corpus respectivo, con el término equivalente en la lengua de la encuesta y cotejando tanto el porcentaje de aparición como el índice de disponibilidad de los dos términos. En catalán, de los 76 préstamos detectados, 36 desplazan el término equivalente en la lengua catalana (esto es, prácticamente la mitad, el 47.37%); en castellano, en cambio, esta circunstancia solo se da en el 28.57% de los casos. La cifra total de préstamos (21) es también inferior en el corpus español.

Ante los resultados obtenidos en el apartado cualitativo, no podemos más que admitir que, aunque en la introducción nos parecía que la influencia de las dos lenguas sería similar, la realidad es otra y la influencia del castellano sobre el catalán es, a todas luces, mayor que en sentido contrario. No solo el porcentaje de transferencias es superior en lengua catalana, sino que, además, en el análisis se comprueba que en muchas más ocasiones en el corpus catalán estos términos sustituyen el vocabulario básico de la lengua de la encuesta. El predominio social que ha tenido el castellano en Cataluña sobre la lengua vernácula, favorecido por las circunstancias sociopolíticas de buena parte del siglo XX, explica esta importante presencia de elementos procedentes del castellano en el catalán, algunos de los cuales, tal como se ve en la frecuencia de mención, es muy probable que se transmitan ya así en lengua catalana. De cualquier modo, la influencia del catalán sobre el castellano tampoco es irrelevante: el porcentaje que suponen las transferencias sobre el total de vocablos (7.08%) es el mayor que se detecta en comparación con los estudios de disponibilidad llevados a cabo en otros lugares de la Península donde conviven dos lenguas, como Galicia, Vizcaya o Valencia (Bartol 2011).

Es lugar común en las conclusiones de las tesis doctorales, abordar las líneas de estudio que se abren a partir del presente trabajo. En nuestro caso, dada la ingente cantidad de datos manejada, son muchas las posibilidades que se ofrecen a partir de la explotación de los dos corpus recogidos, desde el análisis de las interferencias ortográficas o de los dialectalismos del catalán, estudios que podrían abordarse con los datos en bruto que conservamos, hasta el análisis de alguna variable más, como la *lengua habitual*, que en esta ocasión no hemos contemplado (véase el primer apéndice). De modo similar, y tal como hemos venido repitiendo, podríamos también someter los

datos obtenidos respecto a la variación del léxico disponible en relación con la lengua de la encuesta a un programa estadístico que determinara si las tendencias que hemos detectado hasta el momento son o no significativas.

Sin lugar a dudas, otra de las posibilidades más interesantes que se abren es la comparación interlingüística entre las lexías halladas en los dos corpus, para dilucidar si coinciden o no los conceptos que se evocan en las dos lenguas, o el análisis cualitativo de la convergencia entre las distintas variables sociales en cada lengua, basado en las 10, 20 ó 50 primeras palabras. No obstante, si tuviéramos que continuar ahora mismo con la explotación de nuestros datos, creemos que en primer lugar acometeríamos el estudio de los elementos de contacto lingüístico según las distintas variables sociales, principalmente la *lengua materna*, con lo que obtendríamos el perfil social que propicia una mayor aparición de este tipo de fenómenos y veríamos si este coincide o no en las dos lenguas. Son, todos ellos, caminos que se nos abren para seguir investigando y avanzando en el conocimiento.

9. BIBLIOGRAFÍA

- AGUSTÍ, Alfred. 1998. *La llengua catalana a la ciutat de Lleida: estudi sociolingüístic per barris*. Lleida: Universitat de Lleida.
- AHUMADA LARA, Ignacio. 2006. *El léxico disponible de los estudiantes preuniversitarios de la provincia de Jaén*. Jaén: Servicio de Publicaciones de la Universidad de Jaén.
- ALBA, Orlando. 1995a. *Léxico disponible de la República Dominicana*. Santiago de los Caballeros: Pontificia Universidad Católica Madre y Maestra.
- ALBA, Orlando. 1995b. "Anglicismos léxicos en el español dominicano: análisis cuantitativo" en *El español dominicano dentro del contexto americano*. Santo Domingo: Librería La Trinitaria, 11-38.
- ALBA, Orlando. 1996. "Disponibilidad léxica en el español dominicano: aspectos sociolingüísticos" en ARJONA IGLESIAS, Marina *et al.* (eds.). *Actas del X Congreso Internacional de la ALFAL*. México: UNAM, 742-749.
- ALBA, Orlando. 1998. "Variable léxica y dialectología hispánica", *La Torre. Revista de la Universidad de Puerto Rico* 7-8, 299-316.
- ALBA, Orlando. 1999. "Densidad de anglicismos en el léxico disponible de la República Dominicana" en SAMPER PADILLA, José Antonio *et al.* (eds.). *Actas del XI Congreso Internacional de ALFAL*. Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria y Librería Nogal, tomo II, 853-865.
- ALMEIDA, Manuel. 1999. *Sociolingüística*. Santa Cruz de Tenerife: Universidad de la Laguna.
- ÁLVAREZ DE LA GRANJA, María. 2011. "La presencia del gallego en el léxico disponible del español de Galicia. Análisis formal y funcional" en LÓPEZ MEIRAMA, Belén (ed.). *Estudios sobre disponibilidad léxica en el español de Galicia*. Santiago de Compostela: Universidade de Santiago de Compostela, 17-102.
- APPEL, René y Pieter MUYSKEN. 1996. *Bilingüismo y lenguas en contacto*. Barcelona: Ariel.
- ARACIL, Lluís. 1986. *Papers de sociolingüística*. Barcelona: La Magrana. [1ª ed. 1982]

- ARETA LARA, María. 2009. *El léxico disponible de los estudiantes preuniversitarios navarros. Estudio sociolingüístico*. Pamplona: Universidad de Navarra. Tesis doctoral inédita.
- ARMENGOL, R. *et al.* 2007. *Llengua catalana i literatura I*. Barcelona: Castellnou.
- ARNAL PURROY, M^a Luisa (coord.) *et al.* 2004. *Léxico disponible de Aragón*. Zaragoza: Libros Pórtico.
- ARNAL PURROY, M^a Luisa (ed.). 2008. *Estudios sobre disponibilidad léxica en los jóvenes aragoneses*. Zaragoza: Institución “Fernando el Católico”.
- ARNAL PURROY, M^a Luisa. 2008a. “Los dialectalismos en el léxico disponible de los jóvenes aragoneses” en ARNAL, M^a Luisa (ed.). *Estudios sobre disponibilidad léxica en los jóvenes aragoneses*. Zaragoza: Institución “Fernando el Católico”, 17-49.
- ARNAL PURROY, M^a Luisa. 2008b. “Niveles socioculturales y léxico dialectal en el vocabulario disponible de Aragón” en BLAS ARROYO, José Luis (ed.). *Discurso y sociedad II. Nuevas contribuciones al estudio de la lengua en contexto social*. Castellón de la Plana: Universitat Jaume I, 457-469.
- ARNAL PURROY, M^a Luisa. 2009. “Áreas lingüísticas y dialectalismos en los jóvenes aragoneses (materiales de disponibilidad léxica de Aragón)” en LAGÜENS, Vicente (ed.). *Baxar para subir. Colectánea de estudios en memoria de Tomás Buesa Oliver*. Zaragoza: Institución “Fernando el Católico”, 265-294.
- ATIENZA, Encarna *et al.* 1997. “Una tipología de interferencias catalán-castellano a partir de las producciones escritas de los estudiantes universitarios” en CANTERO, Francisco José *et al.* (eds.). *Didáctica de la lengua y la literatura para una sociedad plurilingüe del Siglo XXI. Actas del IV Congreso Internacional de la Sociedad Española de Didáctica de la Lengua y la Literatura; 27-29 de noviembre de 1996*. Barcelona: Universitat de Barcelona, 577-582.
- ÁVILA MUÑOZ, Antonio Manuel. 2006. *Léxico Disponible de los estudiantes preuniversitarios de Málaga*. Málaga: Universidad de Málaga.
- AYORA ESTEBAN, Carmen. 2002. “La disponibilidad léxica en una situación de contacto de lenguas” en DOLORES MUÑOZ, M^a T. *et al.* (eds.). *IV Congreso de*

- Lingüística General*. Cádiz: Universidad de Cádiz y Universidad de Alcalá, vol. II, 159-170.
- AYORA ESTEBAN, Carmen. 2004. “Deficiencias en la adquisición del léxico como problema específico de las lenguas en contacto” en VILLEYANDRE LLAMAZARES, Milka (ed.). *Actas del V Congreso de Lingüística General*. Madrid: Arco Libros, vol I, 281-293.
- AYORA ESTEBAN, Carmen. 2006. *Disponibilidad léxica en Ceuta: aspectos sociolingüísticos*. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.
- AYORA ESTEBAN, Carmen. 2007. “Los estudios de disponibilidad léxica: ámbitos de aplicación” en CANO LÓPEZ, Pablo (coord.). *Actas del VI Congreso de Lingüística General*. Madrid: Arco Libros, vol. II, tomo 2, 2367-2378.
- AZURMENDI, María José. 1983. *Elaboración de un modelo para la descripción sociolingüística del bilingüismo y su aplicación parcial a la comarca de San Sebastián*. San Sebastián: Caja de Ahorros Provincial de Guipúzcoa.
- AZURMENDI, María José. s.d. “Algunos aspectos del euskara utilizado hoy”, disponible en: <http://www.euskaltzaindia.net/dok/ikerbilduma/11116.pdf>
- BADIA CAPDEVILA, Ignasi. 2002. *Diccionari de les llengües d'Europa*. Barcelona: Enciclopèdia Catalana.
- BADIA i MARGARIT, Antoni Maria. 1979. *Llengua i cultura als Països Catalans*. Barcelona: Edicions 62.
- BAETENS, Hugo. 1989. *Principis bàsics del bilingüisme*. Barcelona: Edicions de la Magrana. [1ª ed. 1986]
- BÁEZ DE AGUILAR GONZÁLEZ, Francisco. 1997. *El conflicto lingüístico de los inmigrantes castellanohablantes en Barcelona*. Málaga: Universidad de Málaga.
- BALDAQUÍ ESCANDELL, Josep M^a. 2002. “El model lingüístic dels escolars: interferència i seguretat lingüístiques” en CANO, M^a Antonia *et al.* (eds.). *Les claus del canvi lingüístic*. Alacant: Institut interuniversitari de Filologia Valenciana. Ajuntament de La Núcia. Caja de ahorros del Mediterráneo.
- BAILEY VICTERY, John. 1971. *A study of lexical availability among monolingual-bilingual speakers of spanish and english*. Houston: Rice University. Tesis inédita.

- BARTOL HERNÁNDEZ, José Antonio. 1998. "Anglicismos en el español de finales de siglo. Anglicismos en el léxico disponible de dos regiones hispanas" en KENT, C. y M. D. DE LA CALLE (eds.). *Visiones salmantinas*. Salamanca: Universidad de Salamanca y Ohio State University, 97-124.
- BARTOL HERNÁNDEZ, José Antonio. 2001. "Reflexiones sobre la disponibilidad léxica" en BARTOL, José Antonio *et al.* (eds.). *Nuevas aportaciones al estudio de la lengua española. Investigaciones filológicas*. Salamanca: Luso-Española Ediciones, 221-235.
- BARTOL HERNÁNDEZ, José Antonio. 2003. "Léxico Disponible y norma lingüística" en MORENO FERNÁNDEZ, F. *et al.* (coords.). *Lengua, variación y contexto. Estudios dedicados a Humberto López Morales*. Madrid: Arco Libros, vol. I, 127-144.
- BARTOL HERNÁNDEZ, José Antonio. 2004. *Léxico disponible de Soria*. Burgos: Instituto Castellano y Leonés de la Lengua.
- BARTOL HERNÁNDEZ, José Antonio. 2006. "La disponibilidad léxica", *Revista Española de Lingüística (RSEL)* 36, 379-384.
- BARTOL HERNÁNDEZ, José Antonio. 2008. "Variación léxica del español: los léxicos disponibles de Aragón y Soria" en ARNAL, M^a Luisa (ed.). *Estudios sobre disponibilidad léxica en los jóvenes aragoneses*. Zaragoza: Institución "Fernando el Católico", 207-226.
- BARTOL HERNÁNDEZ, José Antonio. 2011. "Léxicos disponibles de zonas bilingües: interferencias sobre el español" en LÓPEZ MEIRAMA, Belén (ed.). *Estudios sobre disponibilidad léxica en el español de Galicia*. Santiago de Compostela: Universidade de Santiago de Compostela, 157-188.
- BARTOL HERNÁNDEZ, José Antonio y Natividad HERNÁNDEZ MUÑOZ. 2006. "Proyecto del léxico disponible de España" en BLAS ARROYO, José Luis *et al.* (eds.). *Discurso y sociedad. Contribuciones al estudio de la lengua en su contexto social*. Castellón de la Plana: Universitat Jaume I, 725-736.
- BASTARDAS, Albert. 1986. *Llengua i immigració. La segona generació immigrant a la Catalunya no-metropolitana*. Barcelona: La Magrana.

- BASTARDAS, Albert. 2004. "Les llengües a Catalunya: diversitat sociolingüística i perspectives de futur" en PAYRATÓ, Lluís y Francesc Xavier VILA (dirs.). *Les llengües a Catalunya*. Cicle Joan Coromines III. Sabadell: Fundació Caixa de Sabadell, 175-183.
- BELLÓN, Juan José. 2011. *Léxico disponible de la provincia de Córdoba*. Las Palmas de Gran Canaria-Córdoba: Universidad de las Palmas de Gran Canaria-Universidad de Córdoba.
- BERNAL, Jaana. 2003. *Análisis del bilingüismo y de la disponibilidad léxica en finés y español de un niño bilingüe*. Universidad de Turku. Departamento de Lengua Española. Trabajo de investigación inédito.
- BERNAL, Jaana. 2004-2005. "Aspectos cuantitativos del léxico disponible de bilingües finés-español", *Documentos de Español Actual*, 6-7, 7-34.
- BENÍTEZ PÉREZ, Pedro. 1991. "Reseña a Azurmendi 1983", *REL* 21, 2, 355-357.
- BENÍTEZ PÉREZ, Pedro. 1992a. "Listas abiertas y listas cerradas de palabras en disponibilidad léxica" en *Actas del VIII Congreso Nacional de la Asociación Española de Lingüística Aplicada*. Vigo: AESLA, 117-127.
- BENÍTEZ PÉREZ, Pedro 1992b. "Disponibilidad léxica en la zona metropolitana de Madrid", *Boletín de la Academia Puertorriqueña de la Lengua Española*, 1, 1, 71-102.
- BENÍTEZ PÉREZ, Pedro. 1994. "Convergencia y divergencias en el léxico de alumnos de COU", *Revista de estudios de adquisición de lengua española (REALE)* 2, 39-45.
- BENÍTEZ PÉREZ, Pedro, Clara Eugenia HERNÁNDEZ y José Antonio SAMPER. 1995. "Léxicos básicos de España (LEBAES) y de Canarias (LEBAICan). Proyectos de Investigación", *Revista de estudios de adquisición de lengua española (REALE)* 3, 9-17.
- BERNSTEIN, Basil. 1989. *Clases, códigos y control*. Madrid: Akal. [1ª ed. 1986]
- BLANCO, Marta. 2011. "La ortografía en el léxico disponible del español del Galicia" en LÓPEZ MEIRAMA Belén (ed.). *Estudios sobre disponibilidad léxica en el español de Galicia*. Santiago de Compostela: Universidade de Santiago de Compostela, 189-216.

- BLAS ARROYO, José Luis. 1993a. *La interferencia lingüística en valencia (dirección catalán--cast): estudio sociolingüístico*. Castellón: Universitat Jaume I.
- BLAS ARROYO, José Luis. 1993b. “Perspectiva sociofuncional del cambio de código. Estado de la cuestión y aplicaciones a diversos casos del bilingüismo peninsular”, *Contextos* XI/21-22, 221-263.
- BLAS ARROYO, José Luis. 1999. *Lenguas en contacto. Consecuencias lingüísticas del bilingüismo social en algunas comunidades del este peninsular*. Frankfurt am Main/Madrid: Iberoamericana/Vervuert.
- BLAS ARROYO, José Luis. 2004. “El español actual en las comunidades del ámbito lingüístico catalán” en CANO AGUILAR, Rafael (coord.). *Historia de la lengua española*. Barcelona: Ariel, 1065-1086.
- BLAS ARROYO, José Luis. 2005. *Sociolingüística del español. Desarrollos y perspectivas en el estudio de la lengua española en contexto social*. Madrid: Cátedra.
- BLAS ARROYO, José Luis y Manuela CASANOVA ÁVALOS. 2001-2003. “Interacción de factores sociales y de adscripción lingüística en el léxico disponible de una comunidad escolar española”, *Lenguas Modernas* 28-29, 165-189.
- BLAS ARROYO, José Luis y Manuela CASANOVA ÁVALOS. 2003a. “La influencia de la red educativa y del entorno sociocultural en la disponibilidad léxica: estudio de las comunidades de habla castellanenses” en SÁNCHEZ MIRET, Fernando (ed.). *Actas del XXIII Congreso Internacional de Lingüística y Filología Románica*, Salamanca-Tübingen: Niemeyer, 17-33.
- BLAS ARROYO, José Luis y Manuela CASANOVA ÁVALOS. 2003b. “La influencia de la lengua materna en algunos valores de la disponibilidad léxica. Estudio de las comunidades de habla castellanenes” en DOVAL, Irene y M^a Rosa PÉREZ RODRÍGUEZ (eds.). *Adquisición, enseñanza y contraste de lenguas, bilingüismo y traducción*. Vigo: Universidade de Vigo, 49-61.
- BLOOMFIELD, Leonard. 1933. *Language*. Nueva York: Holt, Rinehart and Winston.
- BOIX, Emili. 1993. *Triar no és trair. Identitat i llengua en els joves de Barcelona*. Barcelona: Edicions 62.

- BOIX, Emili. 2004. "El camí fins a la situació contemporània del català i del castellà a Catalunya: entre la complexitat i la fragilitat" en PAYRATÓ, Lluís y Francesc Xavier VILA (dirs.). *Les llengües a Catalunya*. Cicle Joan Coromines III. Sabadell: Fundació Caixa de Sabadell, 13-27.
- BOIX, Emili. 2008. "Un panorama sociolingüístic dels territoris de llengua catalana" en BOIX, E. (ed.). *Els futurs del català*. Barcelona: Edicions Universitat de Barcelona, 11-25.
- BOIX, Emili (ed.). 2008. *Els futurs del català*. Barcelona: Edicions Universitat de Barcelona.
- BOIX, Emili y Francesc Xavier VILA. 1998. *Sociolingüística de la llengua catalana*. Barcelona: Ariel.
- BORREGO, Julio. 2002. "Estudiantes universitarios y jubilados, ¿dos culturas léxicas diferentes?", comunicación presentada en las *Terceras Jornadas de Reflexión Filológica*. Salamanca: Universidad de Salamanca.
- BORREGO Julio y Carmen FERNÁNDEZ JUNCAL. 2002. "Léxico disponible: aplicaciones a los estudios dialectales" en DOLORES MUÑOZ, M^a, T. *et al.* (eds.). *IV Congreso de Lingüística General*. Cádiz: Universidad de Cádiz y Universidad de Alcalá, vol II, 297-306.
- BORREGO, Julio y Carmen FERNÁNDEZ JUNCAL. 2003. "¿En qué cambia la universidad la disponibilidad léxica de los preuniversitarios?" en MORENO FERNÁNDEZ *et al.* (coord.). *Lengua, variación y contexto. Estudios dedicados a Humberto López Morales*. Madrid: Arco Libros, vol. I, 167-178.
- BUTRÓN, Gloria. 1991. "Nuevos índices de disponibilidad léxica" en LÓPEZ MORALES, Humberto (ed.). *La enseñanza del español como lengua materna*. Río Piedras: Universidad de Puerto Rico, 79-89.
- CALERO FERNÁNDEZ, M^a Ángeles. 2003. Estudio sociolingüístico del habla de Toledo. Lleida: Pagès editors.
- CALERO FERNÁNDEZ, M^a Ángeles. 2007. *Percepción social de los sexolectos*. Cádiz: Universidad de Cádiz.

- CAÑIZAL ARÉVALO, Valentina. 1987. *Disponibilidad léxica en escolares de primaria terminada. Análisis de seis centros de interés*. México, Universidad Nacional Autónoma de México. Memoria de licenciatura inédita.
- CAÑIZAL ARÉVALO, Alva. 1991. "Redes semánticas y disponibilidad léxica en el español de escolares mexicanos" en HERNÁNDEZ, César *et al.* (eds.). *El español de América*. Valladolid: Junta de Castilla y León, 631-641.
- CARCEDO GONZÁLEZ, Alberto. 1998a. "Tradición y novedad en las aportaciones hispánicas a los estudios de disponibilidad léxica", *Lingüística* 10, 5-68.
- CARCEDO GONZÁLEZ, Alberto. 1998b. "Sobre las pruebas de disponibilidad léxica para estudiantes de español/LE", *RILCE* 14.2, 205-224.
- CARCEDO GONZÁLEZ, Alberto. 1999a. "Desarrollo de la competencia léxica en español LE: análisis de cuatro fases de disponibilidad", *Pragmalingüística* 5-6, 75-94.
- CARCEDO GONZÁLEZ, Alberto. 1999b. "Estudio comparativo del vocabulario español (LE) disponible de estudiantes finlandeses y el de la sintopía madrileña: propuestas didácticas", *Documentos de Español Actual* 1, 73-87.
- CARCEDO GONZÁLEZ, Alberto. 1999c. "Análisis de errores léxicos del español en la interlengua de los finlandeses" en JIMÉNEZ, T. *et al.* (eds.). *Español como lengua extranjera: enfoque comunicativo y gramática*. Universidad de Santiago de Compostela, 465-472.
- CARCEDO GONZÁLEZ, Alberto. 2000a. *Disponibilidad léxica en español como lengua extranjera: el caso finlandés (estudio del nivel preuniversitario y cotejo con tres fases de adquisición)*. Turku: Turun Yliopisto.
- CARCEDO GONZÁLEZ, Alberto. 2000b. "Índices léxico-estadísticos y graduación del vocabulario en la enseñanza de E/LE (aspectos culturales)", en FRANCO FIGUEROA, Mariano *et al.* (eds.). *Nuevas perspectivas en la enseñanza del español como lengua extranjera*. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz, 175-183.
- CARCEDO GONZÁLEZ, Alberto. 2000c. "La lengua como manifestación de otredad cultural (o convergencia intercultural)", en AGUIRRE ROMERO, Joaquín (ed.), *Espéculo monográfico Cultura e Intercultura en la enseñanza del español como*

- lengua extranjera*. Universidad Complutense de Madrid.
<http://www.ucm.es/info/especulo/ele/carcedo.html>
- CARCEDO GONZÁLEZ, Alberto. 2001. *Léxico disponible de Asturias*. Turku: Universidad de Turku, Publicaciones del Departamento de Lengua Española.
- CARCEDO GONZÁLEZ, Alberto. 2003. “Unidad y variedad diatópica de la disponibilidad léxica del español: comparación de los inventarios de Puerto Rico, Cádiz y Asturias” en MORENO FERNANDEZ, F. *et al.* (coords.). *Lengua, variación y contexto. Estudios dedicados a Humberto López Morales*, Madrid: Arco Libros, vol. I, 199-226.
- CARCEDO GONZÁLEZ, Alberto. 2004. “La variable léxica disponible en la comparación interdialectal: compatibilidad de la norma asturiana con otras sintopías hispanohablantes” en SÁNCHEZ CORRALES, Víctor Manuel (ed.), *Actas del XIII Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina*. San José de Costa Rica: Universidad de Costa Rica, 661-679.
- CARRERA BAIGET, Aitor. 2003. *Una aproximació a les varietats geogràfiques de l'occità de la Vall d'Aran*. Barcelona: Publicacions de l'Abadia de Montserrat.
- CARRERA BAIGET, Aitor. 2007. *Gramàtica aranesa*. Lleida: Pagès editors.
- CARRERA DE LA RED, Fátima y Wini E. BRADLEY. 2004. “Extranjerismos en el léxico disponible de Cantabria” en VILLAYANDRE LLAMAZARES, Milka (ed.). *Actas del V Congreso de Lingüística General*. Madrid: Arco Libros, vol. I, 535-545.
- CASANOVA ÁVALOS, Manuela. 2006. “La disponibilidad léxica en la Comunidad Valenciana” en BLAS ARROYO, José Luis *et al.* (ed.). *Discurso y sociedad. Contribuciones al estudio de la lengua en su contexto social*. Castellón de la Plana: Universitat Jaume I, 737-751.
- CASANOVA ÁVALOS, Manuela. 2008. “Anglicismos en el léxico disponible de Castellón” en BLAS ARROYO, José Luis *et al.* (ed.). *Discurso y sociedad II. Nuevas contribuciones al estudio de la lengua en contexto social*. Castellón de la Plana: Universitat Jaume I, 471-499.

- CASANOVA ÁVALOS, Manuela. 2013. "Valencianismos en el léxico disponible de Castellón" en CASANOVA, Emili y Cesáreo CALVO (eds.). *Actas del XXVI Congreso internacional de Lingüística y Filología Románicas (Valencia 6-11 de septiembre de 2010)*, De Gruyter, 531-543.
- CASANOVAS CATALÀ, Montse. 1995. "La interferencia fonética en el español de Lleida: algunos apuntes para su estudio", *Sintagma* 7, 53-59.
- CASANOVAS CATALÀ, Montse. 1996a. "Consecuencias de la interferencia lingüística en la morfosintaxis del español hablado en Lleida", *Verba* 23, 405-415.
- CASANOVAS CATALÀ, Montse. 1996b. "Algunos rasgos propios del español en las comunidades de habla catalanas: fonética, morfosintaxis y léxico", *Analecta Malacitana* XIX, 1, 149-160.
- CASANOVAS CATALÀ, Montse. 1996c. "El contacto lingüístico en Lleida: algunas consecuencias en el léxico español de los catalanohablantes", *Sintagma* 8, 57-63.
- CASANOVAS CATALÀ, Montse. 1996d. *La variación sintáctica en la ciudad de Lleida: la voz verbal*. Lleida: Universitat de Lleida. Tesis de licenciatura inédita.
- CASANOVAS CATALÀ, Montse. 1998. "Aproximación al estudio de la interferencia lingüística en la lengua escrita", *Revista de Filología de la Universidad de La Laguna* 16, 9-17.
- CASANOVAS CATALÀ, Montse. 2000a. *Análisis cualitativo y cuantitativo de la morfosintaxis de una segunda lengua: el caso del español en contacto con el catalán*. Lleida: Universitat de Lleida. Tesis doctoral.
- CASANOVAS CATALÀ, Montse. 2000b. "No cale que vengas, porque hoy plegaré tarde: Mecanismos de adaptación léxica en el español de los catalanohablantes", *Analecta Malacitana* XXIII 2, 697-709.
- CASANOVAS CATALÀ, Montse. 2002. "Modelos de incorporación léxica en un caso de contacto de lenguas: cuando el español es segunda lengua". *Verba* 29, 261-289.
- CASANOVAS CATALÀ, Montse. 2004. "Sobre la interferencia léxica: patrones léxicos en el uso de una segunda lengua", *Revista de Filología* 22, 25-36.
- CASANOVAS CATALÀ, Montse. 2005. *Español y catalán en contacto. La expresión deíctica en el castellano hablado en Lleida*. Aachen: Shaker.

- CASANOVAS CATALÀ, Montse. 2008. "Patrones léxicos en el español de los catalanohablantes: aproximación cualitativa" en SINNER, Carsten y Andreas WESCH (eds.). 2008. *El castellano en tierras de habla catalana*. Frankfurt am Main/Madrid: Iberoamericana/Vervuert, 181-198.
- CASANOVAS, Montserrat e Imma CREUS. 1999. "Apunts sobre el lleidatà: fonètica i morfosintaxi", *Zeitschrift für Katalanistik. Revista d'Estudis Catalans* 12, 83-108.
- CEDERGREN, Henrietta J. 1983. "Sociolingüística" en LÓPEZ MORALES, Humberto (ed.). *Introducción a la lingüística actual*. Madrid: Playor, 147-165.
- CERDÀ, Ramón. 1984. "Comentarios en torno a la influencia léxica del castellano sobre el catalán actual", *Beiträge zur Romanischen Philologie* 23, 2, 275-281.
- CHOMSKY, Noam. 1957. *Syntactic structures*. The Hague: Mouton.
- CLYNE, Michael G. 1967. *Transference and Triggering*. The Hague: Nihoff.
- COLOMINA, Jordi. 1999. *Dialectologia catalana. Introducció i guia bibliogràfica*. Alacant: Universitat d'Alacant.
- COLÓN DOMÉNECH, Germán. 1967. "Elementos constitutivos del español: catalanismos" en ALVAR, Manuel *et al.* (eds.). *Enciclopedia Lingüística Hispánica*. Madrid: CSIC, vol. II, 193-238.
- COLL, Pep. 1991. *El parlar del Pallars*. Barcelona: Empúries.
- COSERIU, Eugenio. 1978. *Sincronía, diacronía e historia*. Madrid: Gredos.
- CREA = Real Academia Española. Banco de datos (CREA). *Corpus de referencia del español actual*. <http://corpus.rae.es/creanet.html>
- CRUZ ALONSO, Raquel. *El léxico disponible de Castilla y León*. Salamanca: Universidad de Salamanca. Tesis doctoral inédita.
- DCVB = ALCOVER, Antoni M^a, Francesc de BORJA MOLL. 2002. *Diccionari català-valencià-balear*. Barcelona/Palma de Mallorca: Institut d'Estudis Catalans. <http://dcvb.iecat.net>
- DEA= SECO, Manuel, Olimpia ANDRÉS y Gabido RAMOS. 1999. *Diccionario del español actual*. Madrid: Aguilar.

- DIEC = INSTITUT D'ESTUDIS CATALANS. 2007. *Diccionari de la llengua catalana*. Barcelona: Edicions 62. <http://dlc.iec.cat>
- DIMITRIJEVIC, Naum R. 1969. *Lexical availability: A New Aspect of the Lexical Availability of Secondary School Children*. Heidelberg: Julius Gross Verlag.
- DIMITRIJEVIC, Naum R. 1981. "A comparative study of the lexical availability of monolingual and bilingual schoolchildren", *Studia Anglica Posnaniensa* 13, 109-130.
- DLL= MASSANA, Josep M^a. 2004. *Diccionari de lleidatanismes*. Lleida: Pagès editors.
- DRAE= REAL ACADEMIA ESPAÑOLA. 2001. *Diccionario de la Lengua Española*. Madrid: Espasa-Calpe, 22^a ed. <http://lema.rae.es/drae>
- DUE = MOLINER, María. 2007. *Diccionario de uso del español*. Madrid: Gredos.
- ECHEVERRÍA, Max S. *et al.* 1987. "Disponibilidad léxica en educación Media", *Revista de Lingüística Teórica y Aplicada* 25, 55-115.
- ECHEVERRÍA, Max S. 1991. "Crecimiento de la disponibilidad léxica en estudiantes chilenos de nivel básico y medio" en LÓPEZ MORALES, Humberto (ed.). *La enseñanza del español como lengua materna*. Río Piedras: Universidad de Puerto Rico, 61-78.
- ECHEVERRÍA, Max *et al.* 2008. "DispoGrafo: una nueva herramienta computacional para el análisis de las relaciones semánticas en el léxico disponible", *Revista de Lingüística teórica y Aplicada* 46, I, 81-91.
- ENGUITA UTRILLA, José María. 2008. "Norma culta y norma rural en el léxico disponible de los jóvenes aragoneses" en ARNAL, M^a Luisa (ed.). *Estudios sobre disponibilidad léxica en los jóvenes aragoneses*. Zaragoza: Institución "Fernando el Católico", 75-101.
- ESCORIZA MORERA, Luis. 2003. "Lenguas en contacto y disponibilidad léxica: la situación lingüística e intercultural de Ceuta y Gibraltar", *Linred: Revista electrónica de lingüística*, 1.
- ESCORIZA MORERA, Luis. 2007. "Disponibilidad léxica y multilingüismo: el contacto entre inglés y español en Gibraltar" en CANO LÓPEZ, P. (coord.). *Actas del VI Congreso de Lingüística General*. Madrid: Arco Libros, vol. II, tomo 2, 2477-2484.

- ESCORIZA MORERA, Luis. 2009. "Consideraciones sobre el estudio de la variación léxica" en VEYRAT RIGAT, M. y E. SERRA ALEGRA (eds.). *La lingüística como reto epistemológico y como acción social. Estudios dedicados al profesor Ángel López García con ocasión de su sexagésimo aniversario*. Madrid: Arco/Libros, vol. II, 785-793
- ESCORIZA MORERA, Luis. 2012. "La variación de expresión en el plano léxico. Dificultades y perspectivas", *Lingüística* 28, 247-273.
- ESPAÑOL GIRALT, M. Teresa. 1996. "Algunas insuficiencias de la clasificación de la interferencia lingüística" en BRIZ GÓMEZ, Antonio *et al.* (eds.). *Pragmática y gramática del español hablado. Actas del II Simposio sobre análisis del discurso oral (Valencia, 14-22 de noviembre de 1995)*. Valencia/Zaragoza: Universitat de València/ Pórtico, 299-303.
- ETXEBARRIA AROSTEGUI, Maitena 1996a. "Variación e interferencia en una comunidad bilingüe: vasco y español en contacto" en ARJONA IGLESIAS, Marina *et al.* (eds.). *Actas del X Congreso Internacional de ALFAL*. México: UNAM, 908-912.
- ETXEBARRIA AROSTEGUI, Maitena 1996b. "Disponibilidad léxica en escolares del País Vasco: variación sociolingüística y modelos de enseñanza bilingüe", *REL* 26, 2, 301- 325.
- ETXEBARRIA AROSTEGUI, Maitena 1997. "Disponibilidad léxica y caracterización sociolingüística: variación y diversidad" en MOLINA REDONDO, J.A. y J. DE DIOS LUQUE DURAN, *Estudios de lingüística general (II)*. Granada: Método Ediciones, 65-83.
- ETXEBARRIA AROSTEGUI, Maitena. 1999. "Disponibilidad léxica y enseñanza de la lengua materna en el área metropolitana del Gran Bilbao" en SAMPER PADILLA, José Antonio *et al.* (eds.). *Actas del XI Congreso Internacional de ALFAL*. Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria y Librería Nogal, vol II, 1479-1494.
- ETXEBARRIA AROSTEGUI, Maitena. 2002. *La diversidad de lenguas en España*. Madrid: Espasa Calpe.
- FABÀ, Albert. 2008. "Les llengües a Catalunya" en BOIX, Emili (ed.). *Els futurs del català*. Barcelona: Edicions Universitat de Barcelona, 27-49.

- FALL, K. 1976. *L'univers Culturel de l'enfant wolophone de Dakar, d'après la disponibilité des concepts dans la langue maternelle et la langue d'enseignement (cas du wolof et du français)*. Université de Laval. Thèse de maîtrise.
- FASOLD, Ralph. 1996. *La sociolingüística de la sociedad. Introducción a la sociolingüística*. Madrid: Visor Libros.
- FERNÁNDEZ JUNCAL, Carmen. 2008. *El léxico disponible de Burgos*. Burgos: Fundación Instituto Castellano y Leonés de la Lengua.
- FERNÁNDEZ JUNCAL, Carmen. 2013. *El léxico disponible de Cantabria. Estudio sociolingüístico*. Santander-Salamanca: Ediciones de la Universidad de Salamanca y Editorial de la Universidad de Cantabria.
- FERNÁNDEZ MERINO, Pablo Vicente. 2011. "Presencia del léxico disponible de inmigrantes en glosarios específicos de vocabulario", *Cuadernos Comillas* 2, 1-18.
- FERNÁNDEZ MERINO, Pablo Vicente. 2013. *Disponibilidad léxica de inmigrantes en Castilla y León*. Universidad de Valladolid. Tesis doctoral inédita.
- FERNÁNDEZ RODRÍGUEZ, Mauro. 1997. "Los orígenes de la sociolingüística" en CASAS GÓMEZ, Miguel y Jacinto ESPINOSA GARCÍA (coords.). *II Jornadas de Lingüística*. Cádiz: Universidad de Cádiz, 105-132.
- FERNÁNDEZ SMITH, G., A. M^a RICO MARTÍN, M^a J. MOLINA GARCÍA y M. A. JIMÉNEZ JIMÉNEZ. 2006. "Proyecto sobre la disponibilidad léxica en alumnos preuniversitarios de Melilla", en BLAS ARROYO *et al.* (eds.) *Discurso y sociedad. Contribuciones al estudio de la lengua en su contexto social*. Castellón de la Plana: Universitat Jaume I, 767-777.
- FERNÁNDEZ SMITH, G., A. M^a RICO MARTÍN, M^a J. MOLINA GARCÍA, M^a A. JIMÉNEZ JIMÉNEZ. 2008. *El léxico disponible de Melilla: estudio sociolingüístico y repertorios léxicos*. Madrid: Arco/Libros.
- FERNÁNDEZ SMITH, Gerard, Marta SÁNCHEZ-SAUS LASERNA y Luis ESCORIZA MORERA. 2012. "Studies on Lexical Availability: The Current Situation and Some Future Prospects" en EIZAGA REBOLLAR, Bárbara (ed.). *Studies in Linguistics and Cognition*, Bern: Peter Lang, 35-56.

- FISHMAN, Joshua A. 1979. *Sociología del lenguaje*. Madrid: Cátedra.
- GALÁN, Josep y Héctor MORET. 1995. *Estudi descriptiu de la llengua de Fraga*. Fraga: Institut d'Estudis del Baix Cinca.
- GALINDO SOLÉ, Mireia. 2006. *Les llengües a l'hora del pati. Usos lingüístics en les converses dels infants de primària a Catalunya*. Barcelona: Universitat de Barcelona. Tesis doctoral.
- GALLISON, Robert. 1979. *Lexicologie et enseignement des langues (essais méthodologiques)*. France: Hachette.
- GALLOSO CAMACHO, M^a Victoria. 1997. "Tratamiento y perspectivas del "léxico disponible": el caso de la provincia de Zamora", *Revista de estudios de adquisición de lengua española (REALE)*8, 23-33.
- GALLOSO CAMACHO, M^a Victoria. 1998. "El léxico disponible en el nivel preuniversitario (provincia de Zamora)", *Interlingüística* 7, 75-80.
- GALLOSO CAMACHO, M^a Victoria. 1999a. "Hacia una definición de léxico disponible: posibles aplicaciones" en FERNÁNDEZ, J. *et al.* (eds.). *Actas III Congreso de Lingüística General. Lingüística para el Siglo XXI*. Salamanca: Ediciones de la Universidad de Salamanca, 711-717.
- GALLOSO CAMACHO, M^a Victoria. 1999b. "La heterogeneidad en los trabajos de Sociolingüística: el caso de la disponibilidad léxica", *Interlingüística* 10, 139-143.
- GALLOSO CAMACHO, M^a Victoria. 2002. "¿Para qué sirven los estudios de léxico disponible?", en DOLORES MUÑOZ, M^a T. *et al.* (eds.). *IV Congreso de Lingüística General*. Cádiz: Universidad de Cádiz y Universidad de Alcalá, vol. III, 1143-1151.
- GALLOSO CAMACHO, M^a Victoria 2003a. *El léxico de los estudiantes preuniversitarios en el distrito universitario de Salamanca (Ávila, Salamanca y Zamora)*. Burgos: Fundación Instituto Castellano y Leonés de la Lengua.
- GALLOSO CAMACHO, M^a Victoria. 2003b. "Lematización en masculino y femenino de los nombres de profesiones y oficios", *Documentos de Español Actual* 5, 33-46.

- GARCÍA DOMÍNGUEZ, María Jesús, Vicente MARRERO PULIDO, José Alberto PÉREZ MARTÍN y Gracia PIÑERO PIÑERO. “Estudio de la disponibilidad léxica en Gran Canaria. La variante geográfica y el tipo de educación”, *Revista de estudios de adquisición de lengua española (REALE)* 2, 65-72.
- GARCÍA GONDAR, Francisco. 2011. “La creatividad léxica a través de recursos morfológicos en el léxico disponible del español de Galicia” en LÓPEZ MEIRAMA, Belén (ed.). *Estudios sobre disponibilidad léxica en el español de Galicia*. Santiago de Compostela: Universidade de Santiago de Compostela, 217-266.
- GARCÍA MARCOS, Francisco y María Victoria MATEO. 1997. “Resultados de las encuestas sobre disponibilidad léxica realizadas en Almería”, *Revista de estudios de adquisición de lengua española (REALE)* 7, 57-68.
- GARCÍA MARCOS, Francisco Joaquín. 1993. *Nociones de sociolingüística*. Barcelona: Octaedro.
- GARCÍA MARCOS, Francisco Joaquín. 1999. *Fundamentos críticos de sociolingüística*. Almería: Universidad de Almería.
- GARCÍA MEGIA, Antonio. 2003. *La disponibilidad léxica en la ciudad de Almería entre los nueve y los doce años*. Almería: Universidad de Almería. Tesis doctoral inédita.
- GDLC = ENCICLOPÈDIA CATALANA. 1998. Gran diccionari de la llengua catalana. Barcelona: Enciclopèdia catalana. <<http://www.diccionari.cat>>
- GIMENO MENÉNDEZ, Francisco y M^a Victoria GIMENO MENÉNDEZ. 2003. *El desplazamiento lingüístico del español por el inglés*. Madrid: Cátedra.
- GIMENO MENÉNDEZ, Francisco y Esther MARTÍNEZ OLMOS. 2008a. “La transición graduada de la frontera catalano-castellana meridional” en CASANOVA, Emili (ed.). *Els altres parlars valencians. I Jornada de parlars valencians de base castellano-aragonesa*. Paiporta (Valencia): Denes editorial, 269-287.
- GIMENO MENÉNDEZ, Francisco y Esther MARTÍNEZ OLMOS. 2008b. “Estándar y comunidad de habla en los estudios de disponibilidad léxica” en ÁLVAREZ,

- Antonio (ed.). 2008. *Lengua Viva: estudios ofrecidos a César Hernández Alonso*. Valladolid: Universidad de Valladolid, 373-387.
- GÓMEZ CAPUZ, Juan. 1998. *El préstamo lingüístico. Conceptos, problemas y métodos*. Cuadernos de Filología. Valencia: Universitat de València.
- GÓMEZ CAPUZ, Juan. 2004. *Préstamos del español: lengua y sociedad*. Madrid: Arco Libros.
- GÓMEZ DEVIS, Begoña 1995. *Estudi descriptiu sobre disponibilitat lèxica del valencià en la ciutat de València i àrea metropolitana*. Pla Eva. Trabajo inédito.
- GÓMEZ DEVÍS, Begoña. 2005. “Aprovechamiento del léxico disponible”, *Lenguaje y Textos* 23, 69-74.
- GÓMEZ DEVÍS, Begoña. 2007. “Una cuestión metodológica en el Test de Disponibilidad léxica”, *Lenguaje y Textos* 25, 267-727.
- GÓMEZ DEVÍS, Begoña. 2010. “Aportes de la disponibilidad léxica a la investigación sobre fenómenos de contacto lingüístico” en GARCÍA, Manuel *et al.* (eds.). *Interacció comunicativa i ensenyament de llengües*. Universitat de València, 233-240.
- GÓMEZ DEVÍS, Begoña. 2013. “Fenómenos frecuentes del contacto lingüístico en el castellano-catalán (variedad valenciano)” en GÓMEZ MOLINA, José Ramón (coord.). *El español de Valencia. Estudio sociolingüístico*. Peter Lang, 347-370.
- GÓMEZ DEVÍS, Begoña y Francesc LLOPIS RODRIGO. 2006. “La disponibilitat lèxica a València: criteris d’edició” en BLAS ARROYO, José Luis *et al.* (ed.). *Discurso y sociedad. Contribuciones al estudio de la lengua en su contexto social*. Castellón de la Plana: Universitat Jaume I, 779-790.
- GÓMEZ DEVÍS, Begoña y Francesc LLOPIS RODRIGO. 2010a. “El léxico disponible de Valencia: convergencia conceptual”, *Revista de Lingüística y Lenguas aplicadas* 5, 53-76.
- GÓMEZ DEVÍS, Begoña y Francesc LLOPIS RODRIGO. 2010b. “El léxico disponible del catalán de Valencia. Hacia la necesaria planificación lingüística de una lengua minoritaria”, *Romanitas, lenguas y literaturas romances* 4, 2, 1-28.

- GÓMEZ DEVÍS, Begoña y Francesc LLOPIS RODRIGO. 2011. “Com es pot millorar la competència léxica en valencià: una proposta d’aula per a secundària”, *Lenguaje y Textos* 33, 83-91.
- GÓMEZ DEVÍS, Begoña y Francesc LLOPIS RODRIGO. 2012. “Enseñar vocabulario a lo largo de la Educación Obligatoria” en NAVARRO PABLO, Macarena (ed.). *Selección, desarrollo y evaluación de competencias en Didáctica de la lengua y la literatura*. Sevilla: Universidad de Sevilla. Sociedad Española de Didáctica de la Lengua y la Literatura, 410-423.
- GÓMEZ MOLINA, José Ramón. 1986. *Estudio sociolingüístico de la comunidad de habla de Sagunto (Valencia)*. Valencia: IAM Investigación
- GÓMEZ MOLINA, José Ramón. 1999-2000. “Transferencia y cambio de código en una comunidad bilingüe: Área metropolitana de Valencia (I y II)”, *Contextos* XVII-XVIII/33-36, 309-360.
- GÓMEZ MOLINA, José Ramón. 2000. “Consecuencias del contacto de lenguas” en BRIZ, Antonio. *¿Cómo se comenta un texto coloquial?* Barcelona: Ariel, 287-313.
- GÓMEZ MOLINA, José Ramón. 2002a. “Norma i ús en el valencià central: nivell lèxic”, *Treballs de sociolingüística catalana* 16, 169-194.
- GÓMEZ MOLINA, José Ramón. 2002b. “Préstamos léxicos del castellano en el valenciano central: inserción, evaluación y actualización” en ECHENIQUE, M. T. y J. SÁNCHEZ (eds.). *Actas del V Congreso Internacional de Historia de la lengua española*. Madrid: Gredos, 1739-1764.
- GÓMEZ MOLINA, José Ramón. 2006. “La variable ‘sexo’ en los estudios de disponibilidad léxica”, *Revista de Lingüística Teórica y Aplicada* 44/1, 47-65.
- GÓMEZ MOLINA, José Ramón. 2009. “Una aplicación psicolingüística de la disponibilidad léxica: la categoría nocional ‘animales’” en VEYRAT RIGAT, M. y E. SERRA ALEGRA (eds.). *La lingüística como reto epistemológico y como acción social. Estudios dedicados al profesor Ángel López García con ocasión de su sexagésimo aniversario*. Madrid: Arco/Libros, 1047-1060.

- GÓMEZ MOLINA, José Ramón y M^a Begoña GÓMEZ DEVÍS. 2004. *La disponibilidad léxica de los estudiantes preuniversitarios valencianos*. Valencia: Universitat de València.
- GONZÁLEZ MARTÍNEZ, Adolfo 1999a. “Andalucismos en el léxico disponible de Cádiz”, *Tavira* 17, 181-193.
- GONZÁLEZ MARTÍNEZ, Adolfo y Pepa ORELLANA RAMÍREZ. 1999b. “Análisis del comportamiento de la variable sexo en el léxico disponible”, *Revista de estudios de adquisición de lengua española (REALE)* 11, 65-73.
- GONZÁLEZ MARTÍNEZ, Adolfo y Pepa ORELLANA RAMÍREZ. 2000. “Cotejo de los léxicos disponibles de Cádiz y Zamora”, *DEA (Documentos de Español Actual)* 2, 139-160.
- GONZÁLEZ MARTÍNEZ, Adolfo y Pepa ORELLANA RAMÍREZ. 2000-2001. “Compatibilidad y discrepancia entre los léxicos disponibles de Puerto Rico y Cádiz”, *Pragmalingüística* 8-9, 179-196.
- GONZÁLEZ MARTÍNEZ, Adolfo 2002. *La disponibilidad léxica de los alumnos preuniversitarios de la provincia de Cádiz*. Cádiz: Universidad de Cádiz.
- GONZÁLEZ MARTÍNEZ, Adolfo y ORELLANA RAMÍREZ, Pepa. 2006. “Anglicismos en el léxico disponible de la provincia de Cádiz”, *Boletín de Lingüística* 18, 3-21.
- GOUGENHEIM, Georges, René MICHÉA, Paul RIVENC y Aurélien SAUVEGOT. 1964. *L'élaboration du français fondamental (I degré). Étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. París: Didier.
- GROSJEAN, François. 1982. *Life with two languages*. Cambridge: Harvard University Press.
- GRUPO INTERALIA. 1998. “Lenguas pensinsulares en contacto: castellano y catalán en la producción de estudiantes universitarios bilingües”. Ponencia presentada en la *7th University of New Mexico Conference on Ibero-American Culture and Society: 'Spanish and Portuguese in Contact with Other Languages' and 16th Conference on Spanish in the United States*. Mexico, ms. Inédito.
- GUMPERZ, John J. 1982. *Discourse strategies*. Cambridge: Cambridge University Press.

- HAUGEN, Einar. 1969. *The Norwegian Language in America: A study in bilingual behaviour*. Bloomington-London: Indiana University Press [1ª ed. 1953].
- HERNÁNDEZ CABRERA, Clara Eugenia y José Antonio SAMPER PADILLA. 2003. “Los dialectalismos en el léxico disponible de Gran Canaria. Análisis de un centro de interés” en MORENO FERNÁNDEZ, F. *et al.* (coords.). *Lengua, variación y contexto. Estudios dedicados a Humberto López Morales*. Madrid: Arco Libros, vol. I, 339-394.
- HERNÁNDEZ GARCÍA, Carmen. 1996. “Algunas propuestas didácticas para trabajar la interferencia catalán-español en el ámbito universitario (a partir del análisis de errores)” en CANTERO, Francisco José *et al.* (eds.). *Didáctica de la lengua y la literatura para una sociedad plurilingüe del Siglo XXI. Actas del IV Congreso Internacional de la Sociedad Española de Didáctica de la Lengua y la Literatura; 27-29 de noviembre de 1996*. Barcelona: Universitat de Barcelona, 633-639.
- HERNÁNDEZ GARCÍA, Carmen. 1998a. *Algunas cuestiones más sobre el contacto de lenguas: Estudio de la interferencia lingüística del catalán en el español de Cataluña*. Barcelona: Universitat de Barcelona. Tesis doctoral inédita.
- HERNÁNDEZ GARCÍA, Carmen. 1998b. “Una propuesta de clasificación de la interferencia lingüística a partir de dos lenguas en contacto: el catalán y el español”, *Hesperia* 1, 61-79.
- HERNÁNDEZ MUÑOZ, Natividad. 2004. *El léxico disponible de los estudiantes conquenses*. Salamanca: Ediciones Universidad de Salamanca.
- HERNÁNDEZ MUÑOZ, Natividad. 2006. *Hacia una teoría integrada de la disponibilidad léxica: el léxico disponible de los estudiantes castellanomanchegos*. Salamanca: ediciones de la Universidad de Salamanca.
- HERNÁNDEZ MUÑOZ, Natividad y Julio BORREGO NIETO. 2004. “Cuestiones metodológicas sobre los estudios de disponibilidad léxica” en VILLAYANDRE LLAMAZARES, Milka (ed.). *Actas del V Congreso de Lingüística General*. Madrid: Arco Libros, 1519-1527.
- HERNÁNDEZ MUÑOZ, Natividad, Cristina IZURA y Carmela TOMÉ. 2014. “Cognitive Factors of Lexical Availability in a Second Language” en JIMÉNEZ CATALÁN, Rosa María (ed.). *Lexical Availability in English and Spanish as a*

- Second Language*. Educational Linguistics, 17. Springer Science+Business Media Dordrecht.
- HOFFMAN, Charlotte. 1991. *An introduction to bilingualism*. New York: Longman.
- HUDSON, Richard. 1981. *La sociolingüística*. Barcelona: Anagrama.
- HUMBLEY, John. 1974. “Vers une typologie de l’emprunt linguistique”, *Cahiers de lexicologie* 25, 46-70.
- INSTITUT D’ESTADÍSTICA DE CATALUNYA. 2004. *Estadística d’Usos Lingüístics a Catalunya 2003*. Barcelona: Generalitat de Catalunya. <http://idescat.cat>
- INSTITUT D’ESTADÍSTICA DE CATALUNYA. 2009. *Enquesta d’usos lingüístics de la població 2008*. Barcelona: Generalitat de Catalunya. <http://idescat.cat>
- INSTITUT DE SOCIOLINGÜÍSTICA CATALANA. s.d. *El català a Catalunya (Espanya)*, disponible en: <http://www.uoc.es/euromosaic/web/document/catala/ca/i1.html>
- ILLAMOLA, Cristina. 2008. “La expresión del futuro en el español de Cataluña. Un posible caso de confluencia lingüística” en BLAS ARROYO, José Luis *et al.* (eds.). *Discurso y sociedad II. Nuevas contribuciones al estudio de la lengua en contexto social*. Castellón de la Plana: Universitat Jaume I, 703-715.
- JIMÉNEZ BERRIO, Felipe. 2013. *Léxico disponible de inmigrantes escolares no hispanohablantes*. Pamplona: Servicio de Publicaciones de la Universidad de Navarra.
- JULIÀ MUNÉ, Joan. *et al.* 2004. *El català nordoccidental: descripcions i orientacions ortoèpiques*. Lleida: Pagès.
- JUSTO HERNÁNDEZ, Hortensia. 1986. *Disponibilidad léxica en colores*. México: Universidad Nacional Autónoma de México. Memoria de licenciatura.
- KÄUPER, Anja y Carlos J. GUERRERO. 2008. “El habla de los graueros: peculiaridades del castellano de los jóvenes del distrito portuario de Castellón” en SINNER, Carsten y Andreas WESCH (ed.). *El castellano en tierras de habla catalana*. Frankfurt am Main/Madrid: Iberoamericana/Vervuert, 133-154.

- KRUSKAL, William y Allen WALLIS. 1952. "Use of Ranks in One-Criterion Variance Analysis", *Journal of the American Statistical Association* 47 (260), 583-621.
- LABOV, William. 1983. *Modelos sociolingüísticos*. Madrid: Cátedra.
- LABOV, William. 1996. *Principios del cambio lingüístico*. Madrid: Gredos.
- LAGÜENS GRACIA, Vicente. 2008. "La variable sexo en el léxico disponible de los jóvenes aragoneses" en ARNAL, María Luisa (ed.). *Estudios sobre disponibilidad léxica en los jóvenes aragoneses*. Zaragoza: Institución "Fernando el Católico", 103-162.
- LANG, Mervyn F. 1992. *Formación de palabras en español: morfología derivativa creativa en el léxico moderno*. Madrid: Cátedra.
- LLOPIS RODRIGO, Francesc. 2005. "Transferències lèxiques en una comunitat de llengües en contacte", *Lenguaje y Textos* 23, 59-68.
- LLOPIS RODRIGO, Francesc. 2008. "Disponibilitat lèxica del valencià: classificació dels fenòmens de contacte lingüístic" en BLAS ARROYO, José Luis *et al.* (eds.). *Discurso y sociedad II. Nuevas contribuciones al estudio de la lengua en contexto social*. Castellón de la Plana: Universitat Jaume I, 457-469.
- LLOPIS RODRIGO, Francesc. 2009. *El lèxic del valencià central: un estudi de disponibilitat léxica*. Valencia: Servei de Publicacions de la Universitat de València.
- LLOPIS RODRIGO, Francesc. 2010. "Convergència semántica en comunitats de parla bilingües" en GARCÍA, Manuel *et al.* (eds.). *Interacció comunicativa i ensenyament de llengües*. Valencia: Universitat de València, 257-264.
- LLOPIS RODRIGO, Francesc. 2011. "El lèxic valencià en els estudiants", *Aula de Lletres Valencianes. Revista Valenciana de Filologia* 1, 113-135
- LLOPIS RODRIGO, Francesc y Begoña GÓMEZ DEVÍS. 2010. *El lèxic disponible de València*. Valencia: Denes.
- LLOPIS RODRIGO, Francesc y Joaquim LÓPEZ RÍO. 2013. "Transferències lèxiques en el estudiants valencians", en CASANOVA, Emili y Cesáreo CALVO (eds.). *Actas del XXVI Congreso internacional de Lingüística y Filología Románicas (Valencia 6-11 de septiembre de 2010)*. De Gruyter, vol. 8, 4493-5004.

- LÓPEZ CHÁVEZ, Juan y C. STRASSBURGUER. 1991. "Un modelo más para el cálculo de disponibilidad léxica individual", en LÓPEZ MORALES, Humberto (ed.). *La enseñanza del español como lengua materna*. Río Piedras: Universidad de Puerto Rico.
- LÓPEZ CHÁVEZ, Juan. 1992. "Alcances panhispánicos del léxico disponible", *Lingüística* 4, 26-124.
- LÓPEZ CHÁVEZ, Juan. 1994. "Comportamiento sintáctico de algunos verbos ordenados según su grado de disponibilidad léxica", *Revista de estudios de adquisición de lengua española (REALE)* 1, 67-84.
- LÓPEZ DEL CASTILLO, Lluís. 1975. "Les interferències català-castellà" en *Bilingüismo y educación en Cataluña. Ponencias y comunicaciones presentadas al primer Seminario sobre Educación Bilingüe en Cataluña organizado por el Instituto de Ciencias de la Educación de la Universidad de Barcelona*. Barcelona: Teide, 85-91.
- LÓPEZ MEIRAMA, Belén. 2008. *Léxico disponible en el español de Galicia*. Santiago de Compostela: Universidade de Santiago de Compostela.
- LÓPEZ MEIRAMA, Belén (ed.). 2011. *Estudios sobre disponibilidad léxica en el español de Galicia*. Santiago de Compostela: Universidade de Santiago de Compostela.
- LÓPEZ MEIRAMA, Belén. 2011. "Léxico rural y léxico urbano" en LÓPEZ MEIRAMA (ed.). *Estudios sobre disponibilidad léxica en el español de Galicia*. Santiago de Compostela: Universidade de Santiago de Compostela, 103-156.
- LÓPEZ MORALES, Humberto 1978. "Frecuencia léxica, disponibilidad y programación curricular", *Boletín de la Academia Puertorriqueña de la Lengua Española* 6/1, 73-86.
- LÓPEZ MORALES, Humberto 1979. "Disponibilidad léxica y estratificación socioeconómica" en *Dialectología y Sociolingüística. Temas puertorriqueños*. Madrid: Hispanova de Ediciones, 173-181.
- LÓPEZ MORALES, Humberto. 1983. "Lingüística Estadística" en *Introducción a la lingüística actual*. Madrid: Playor, 209-225.

- LÓPEZ MORALES, Humberto. 1986. *La enseñanza de la lengua materna. Lingüística para maestros de español*. Madrid: Playor.
- LÓPEZ MORALES, Humberto. 1989. *Sociolingüística*, Madrid: Gredos.
- LÓPEZ MORALES, Humberto. 1994. *Métodos de investigación lingüística*. Salamanca: Colegio de España.
- LÓPEZ MORALES, Humberto. 1995-1996. “Los estudios de disponibilidad léxica: pasado y presente”, *Boletín de Filología de la Universidad de Chile XXXV*, 245-259.
- LÓPEZ MORALES, Humberto. 1999a. *Léxico disponible de Puerto Rico*. Madrid: Arco Libros.
- LÓPEZ MORALES, Humberto. 1999b. “Anglicismos en el léxico disponible de Puerto Rico” en ORTIZ LÓPEZ, Luis A. (ed.). *El caribe hispánico: perspectivas lingüísticas actuales. Homenaje a Manuel Álvarez Nazario*. Vervuert Iberoamericana, 147-170.
- LÓPEZ MORALES, Humberto. 2008. “Un alto en el camino: el léxico disponible de Aragón” en ARNAL, M^a Luisa (ed.). 2008. *Estudios sobre disponibilidad léxica en los jóvenes aragoneses*. Zaragoza: Institución “Fernando el Católico”, 7-15.
- LÓPEZ MORALES, Humberto y Francisco GARCÍA MARCOS. 1995. “Disponibilidad léxica en Andalucía. Proyecto de investigación”, *Revista de estudios de adquisición de lengua española (REALE)*3, 65-76.
- LOZANO DOMINGO, Irene. 1995. *Lenguaje femenino, lenguaje masculino. ¿Condiciona nuestro sexo la manera de hablar?* Madrid: Minerva ediciones.
- LÜDI, George. 1987. “Les marques transcodiques: regards nouveaux sur le bilinguisme” en *Devenir bilingue-parler bilingue. Actes du 2ème colloque sur le bilinguisme, 20-22 septembre 1984*. Université de Neuchâtel. Tübingen: Niemeyer, 1-19.
- MACKEY William F. 1971a. *Le vocabulaire disponible du Français*, vol I y II, París-Bruxelles-Montréal: Didier.
- MACKEY, William F. 1971b. *Le sondage dans les enquêtes de disponibilité*. Québec: CIRB, Université de Laval.
- MACKEY, William. 1976. *Bilinguisme et contact des langues*. Paris: Klincksieck.

- MANGADO CRUZ, Marta y María ARETA LARA. 2008. "Procesamiento informático de datos para la elaboración de diccionarios de disponibilidad léxica" en OLZA, I. et al. (eds.). 2008. *Actas del XXXVII Simposio Internacional de la Sociedad Española de Lingüística*. Pamplona: Universidad de Navarra, 479-493.
- MANJÓN-CABEZA, Antonio. 1999. "Datos léxicos y reflexión docente", *Docencia e Investigación* 24, 93-128.
- MARTÍNEZ OLMOS, Esther. 2008. *Disponibilidad léxica en las comunidades de habla alicantinas*. Alicante: Universidad de Alicante. Tesis doctoral.
- MARSÀ, Francisco. 1986. "Sobre concurrencia lingüística en Cataluña", en GARCÍA DE LA CONCHA, Víctor et al. (eds.). *El castellano actual en las comunidades bilingües de España*. Salamanca: Junta de Castilla y León, Consejería de Educación y Cultura, 93-104.
- MASSANA, Josep M^a. 2004. *Diccionari de lleidatanismes*. Lleida: Pagès editors.
- MASIP, Robert, Ferran MONTARDIT y David PRENAFETA. 2010. *Lo nou diccionari català-lleidatà*. Lleida: Alfazeta.
- MATEO, María Victoria. 1996. "Disponibilidad léxica en Andalucía", *Interlingüística* 5, 103-107.
- MATEO, María Victoria. 1998. *Disponibilidad léxica en el COU almeriense. Estudio de estratificación social*. Almería: Universidad de Almería.
- MAYORAL, Dolors. 1994. *Lenguaje y estatus sociocultural: un estudio empírico del lenguaje infantil en los barrios de Lleida*. Lleida: Universitat de Lleida. Tesis doctoral.
- MENA OSORIO, Mónica. 1986. *Disponibilidad léxica infantil en tres niveles de enseñanza básica*. Chile: Universidad de Concepción. Tesis de maestría.
- MESA CANALES, Rosa M. 1989. *Disponibilidad léxica en preescolares*. México: Universidad Nacional Autónoma de México. Memoria de licenciatura inédita.
- MILROY, Lesley. 1987. *Language and Social Networks*. Oxford: Blackwell.
- MIR, Pere. 1995. *Recull estadístic de la ciutat de Lleida*. Lleida: La Paeria-Ajuntament de Lleida.

- MOLL, Francesc de Borja. 1961. "El castellano en Mallorca" en *Homenaje ofrecido a Dámaso Alonso por sus amigos y discípulos con ocasión de su 60.º aniversario*. Madrid: Gredos, vol. 2, 469-474.
- MOLLÀ, Toni i Carles PALANCA. 1987. *Curs de sociolingüística*. Alzira: Bromera.
- MORENO FERNÁNDEZ, Francisco. 1990. *Metodología sociolingüística*. Madrid: Gredos.
- MORENO FERNÁNDEZ, Francisco, José Enrique MORENO FERNÁNDEZ y Antonio J. GARCÍA DE LAS HERAS. 1995. "Cálculo de disponibilidad léxica. El programa Lexidisp", *Lingüística* 7, 243-249.
- MORENO FERNÁNDEZ, Francisco. 1998. *Principios de sociolingüística y sociología del lenguaje*. Barcelona: Ariel.
- MORENO FERNÁNDEZ, Francisco. 2007. "Anglicismos en el léxico disponible de los adolescentes hispanos de Chicago" en POTOWSKI, K. y R. CAMERON (eds.), *Spanish in contact: Policy, Social and Linguistic Inquires*. Chicago: University of Illinois, 41-58.
- MOYER, Melissa G. 1991. "La parla dels immigrants andalusos al barri de Sant Andreu", *Treballs de Sociolingüística* 9, 83-104.
- MURILLO ROJAS, Marielos. 1993. "Disponibilidad léxica en los preescolares: estudio de cinco campos semánticos", *Káñina* 17/2, 117-127.
- MURILLO ROJAS, Marielos 1994. "Comidas y bebidas: un estudio de disponibilidad léxica en preescolares", *Káñina* 18/2, 117-133.
- NAVARRO, Pere. 2005. *Aproximació geolingüística als parlars del Matarranya*. Calaceit: Associació cultural del Matarranya.
- NDLLC = MASIP, Robert, Ferran MONTARDIT y David PRENAFETA. 2010. *Lo nou diccionari català-lleidatà*. Lleida: Alfazeta.
- NINYOLES, Rafael. 1969. *Conflicte lingüístic valencià: substitució lingüística i ideologies diglòssiques*. València: Elisseu Climent.
- NJOCK, P. E. 1978. *L'univers familial de l'enfant africain*, Québec: CIRB.

- O'HUALLACHAIN, C.L.1966. *Buntus Gaeilge*. Dublin: Oifig Dhiolta Foilseachan Rialtais (Government Publications Sales Office), An Stuaire, Ard-Oifig an Phoist (The Arcade, General Post-Office), Vaile Atha Cliath.
- ORTEGA OJEDA, Gonzalo. 1996. "Implicaciones dialectológicas y didácticas del léxico disponible" en SEGOVIANO, C. (ed.). *La enseñanza del léxico español como lengua extranjera*. Madrid- Frankfurt am Main: Iberoamericana- Vervuert, 22-34.
- PAREDES, Florentino. 1999. "La ortografía en las encuestas de disponibilidad léxica", *Revista de estudios de adquisición de lengua española (REALE)* 11, 75-97.
- PAREDES, Florentino. 2001a. "Disponibilidad léxica de alumnos de enseñanza secundaria de Alcalá y su comarca: resultados generales" en CRUZ, Isabel *et al.* (eds.). *La Lingüística Aplicada a finales del siglo XX. Ensayos y propuestas. Actas del XVII Congreso de AESLA*. Universidad de Alcalá, vol. 2, 721-728.
- PAREDES, Florentino. 2001b. "Disponibilidad de los extranjerismos en estudiantes de educación secundaria" en MARTÍN ZORRAQUINO, M^a A. y C. DÍEZ PELEGRÍN (eds.). *¿Qué español enseñar? Norma y variación lingüísticas en la enseñanza del español a extranjeros. Actas del XI Congreso Internacional de ASELE (Zaragoza, 13-16 de septiembre 2001)*. Zaragoza: ASELE Universidad de Zaragoza/Instituto Fernando el Católico / Gobierno de Aragón, 567-576.
- PAREDES, Florentino. 2012. "Desarrollos teóricos y metodológicos recientes de los estudios de disponibilidad léxica", *Revista Nebrija de Lingüística Aplicada* 11 (6), 78-100.
- PASTOR MILLÁN, M^a Ángeles y Francisco José SÁNCHEZ GARCÍA. 2008. *El léxico disponible de Granada y su provincia*. Granada: Universidad de Granada.
- PAYRATÓ, Lluís. 1984. "Barbarismes, manlleus i interferències. Sobre la terminologia dels contactes interlingüístics", *Els Marges* 32, 45-58.
- PAYRATÓ, Ll. 1985. *La interferencia lingüística. Comentaris i exemples català-castellà*. Barcelona: Curial Edicions catalanes-Publicacions de l'Abadia de Montserrat.
- PAYRATÓ, Lluís. 1996. *El català col·loquial. Aspectes de l'ús corrent de la llengua catalana*. València: Universitat de València.

- PÉREZ EDO, M^a Ángeles. 2007. “La disponibilidad léxica en la provincia de Gerona y en las Islas Baleares”, comunicación presentada en el *XXXVIII Simposio Internacional de la Sociedad Española de Lingüística*, Pamplona, 17-20 de diciembre de 2007.
- PFEFFER, J.A. 1964. *Grundeutsch Basic (spoken) German Word list Grundstufe*. New Jersey: Prentice Hall Inc. Englewood Cliffs.
- POMARES, Joaquim. 1997. *Diccionari del català popular i d'argot*. Barcelona: edicions 62.
- POPLACK, S. 1980. “Sometimes I’ll start a sentence in Spanish y termino en español”: toward a typology of code-switching”, *Linguistics* 18, 581-618.
- POPLACK, S. 1981. “Bilingualism and the vernacular” en HARTFORD, B. y A. VALDMAN (ed.). *Issues in international bilingual education. The Role of the Vernacular*. New York and London: Plenum Press, 1-23.
- POPLACK, S. 1987. “Contrasting patterns of code-switching in two communities” en WANDE, E. et al. (eds.). *Aspects of Multilingualism*. Uppsala: Borgströms, 51-77.
- POPLACK, S. 1993. “Variation theory and language contact” en PRESTON, D. (ed.). *American dialect research: An anthology celebrating the 100th anniversary of American Dialect Society*. Amsterdam: John Benjamins, 251-286.
- PRADO ARAGONÉS, Josefina y M^a Victoria GALLOSO CAMACHO. 2005. *Léxico Disponible de Huelva.- Nivel preuniversitario*. Huelva: Servicio de Publicaciones la Universidad de Huelva.
- PRADO ARAGONÉS, Josefina y M^a Victoria GALLOSO CAMACHO. 2006a. *Diccionario del léxico disponible de alumnos de 6º curso de Educación Primaria de Huelva*. Huelva: Universidad de Huelva.
- PRADO ARAGONÉS, Josefina, M^a Victoria GALLOSO CAMACHO. 2006b. *Diccionario del léxico disponible de alumnos de 2º curso de Bachillerato de Huelva*. Huelva: Universidad de Huelva.
- PRADO ARAGONÉS, Josefina y M^a Victoria GALLOSO CAMACHO. 2008. “La variable sexo en el léxico disponible de alumnos de Primaria y Secundaria de Huelva” en BLAS ARROYO José Luis et al. (eds.). 2008. *Discurso y sociedad*

- II. *Nuevas contribuciones al estudio de la lengua en contexto social*. Castellón de la Plana: Universitat Jaume I, 583-595.
- PRADO ARAGONÉS, Josefina, M^a Victoria GALLOSO CAMACHO y Manuel CÉLIO CONCEIÇÃO. 2009. *La disponibilidad léxica en situación de contacto de lenguas en las zonas limítrofes de Andalucía y Extremadura (España) y Algarve y Alentejo (Portugal)*. Huelva: Universidad de Huelva.
- PRAT SABATER, Marta. 2003. *Préstamos del catalán en el léxico español*. Barcelona: Universitat Autònoma de Barcelona. Tesis doctoral.
- PRAT SABATER, Marta. 2005. “La influència del català sobre el lèxic castellà: visió diacrònica”, *Llengua & Literatura* 16, 363–387.
- PUEYO, Miquel. 1991. *Llengües en contacte a la comunitat lingüística catalana*. Valencia: Universitat de València.
- PUJOLAR, Joan, Isaac GONZÁLEZ, Anna FONT y Roger MARTÍNEZ. 2010. *Llengua i joves. Usos i percepcions lingüístics de la joventut catalana*. Barcelona: Generalitat de Catalunya, Departament d'Acció Social i Ciutadania.
- QUEROL, Ernest (coord.). 2007. *Llengua i societat als territoris de parla catalana a l'inici del segle XXI*. Barcelona: Departament de Política Lingüística de la Generalitat de Catalunya.
- RAMONEDA CARRERA, Rosa (coord.). 2007. *Aïna: vocabulari temàtic del dialecte pallarès*. Lleida: Grup d'Estudis de Llengua i Literatura de Ponent i del Pirineu.
- RODRÍGUEZ BOU, Ismael. 1952. *Recuento de vocabulario español*. Río Piedras: Universidad de Puerto Rico.
- ROMAINE, Suzanne. 1995. *Bilingualism*. New York: Oxford.
- ROMAINE, Suzanne. 1996. *El lenguaje en la sociedad*. Barcelona: Ariel.
- ROMÁN, Belén. 1985. *Disponibilidad léxica en escolares de Dorado*. Río Piedras: Universidad de Puerto Rico. Memoria de licenciatura inédita.
- ROMERA, Magdalena. 2003. “La variedad del castellano actual en Baleares”, *Moenia* 9, 359-381.
- ROMERO RUBILAR, Carlos. 2000. “Variabilidad y prototipicidad en vocabulario disponible”, *RLA. Revista de Lingüística Teórica y Aplicada*, vol. 38, 139-148.

- ROTAETXE, K. 1990. *Sociolingüística*. Madrid: Síntesis.
- RUIZ BASTO, Araceli. 1987. *Disponibilidad léxica de los alumnos de primer ingreso en el Colegio de Ciencias y Humanidades Plantel Naucalpán*. México: UNAM. Tesis Profesional inédita.
- SAMPER, Lluís, Ramon MORELL, Fidel MOLINA, Jordi GARRETA y Dolors MAYORAL. 2000. *Continuïtats i canvi en l'estructura social a Lleida (1960-1990)*. Lleida: Pagès Editors.
- SAMPER HERNÁNDEZ, Marta 2001. "Dificultades de los estudios de disponibilidad léxica en ELE: los criterios de edición de los materiales" en BARTOL, José Antonio *et al.* (eds.). *Nuevas aportaciones al estudio de la lengua española. Investigaciones filológicas*. Salamanca: Luso-Española Ediciones, 277-285.
- SAMPER HERNÁNDEZ, Marta. 2001-2002. "La variable "lengua materna" en la disponibilidad léxica de estudiantes de ELE", *Documentos de Español Actual* 3-4, 147-172.
- SAMPER HERNÁNDEZ, M. 2002. *Disponibilidad léxica en alumnos de español como lengua extranjera*. Málaga: ASELE.
- SAMPER HERNÁNDEZ, M. 2008. "Datos comparativos entre dos léxicos disponibles: Aragón y Gran Canaria" en ARNAL, M^a Luisa (ed.). *Estudios sobre disponibilidad léxica en los jóvenes aragoneses*. Zaragoza: Institución "Fernando el Católico", 245-282.
- SAMPER HERNÁNDEZ, Marta. 2009. *Evolución de la disponibilidad léxica en estudiantes grancanarios de enseñanza primaria y secundaria*. Las Palmas de Gran Canaria: Cabildo de Gran Canaria.
- SAMPER HERNÁNDEZ, Marta y Rosa María JIMÉNEZ CATALÁN. 2014. "Researching Lexical Availability in L2: Some Methodological Issues" en JIMÉNEZ CATALÁN, Rosa María (ed.). *Lexical Availability in English and Spanish as a Second Language*. Educational Linguistics, 17. Springer Science+Business Media Dordrecht.
- SAMPER PADILLA, José Antonio. 1990. *Estudio sociolingüístico del español de Las Palmas de Gran Canaria*. Las Palmas: La Caja de Canarias.

- SAMPER PADILLA, José Antonio. 1998. "Criterios de edición del léxico disponible", *Lingüística* 10, 311-333.
- SAMPER PADILLA, José Antonio. 1999. "Léxico disponible y variación dialectal: datos de Puerto Rico y Gran Canaria" en MORALES, A. *et al.* (eds.). *Estudios de lingüística hispánica (Homenaje a María Vaquero)*. San Juan de Puerto Rico: Universidad, 550-573.
- SAMPER PADILLA, José Antonio. 2006. "Disponibilidad léxica y sociolingüística" en BLAS ARROYO, José Luis *et al.* (eds.). 2006. *Discurso y sociedad. Contribuciones al estudio de la lengua en su contexto social*. Castellón de la Plana: Universitat Jaume I, 99-120.
- SAMPER PADILLA, José Antonio y Clara Eugenia HERNÁNDEZ CABRERA. 1997 "Estudio de la disponibilidad léxica en Gran Canaria. Datos iniciales y variación sociolingüística" en DORTA, J. y M.V. ALMEIDA. *Contribuciones al estudio de la lingüística hispánica (Homenaje a Ramón Trujillo Carreño)*. La Laguna: Universidad de la Laguna, vol II, 229-239.
- SAMPER PADILLA, José Antonio, Juan José BELLÓN FERNÁNDEZ y Marta SAMPER HERNÁNDEZ. 2003. "El proyecto de estudio de la disponibilidad léxica en español" en ÁVILA, R. *et al.* *Pautas y pistas en el análisis del léxico hispano(americano)*. Madrid-Frankfurt: Iberoamericana-Verbuert, 27-139.
- SAMPER PADILLA, José Antonio y Marta SAMPER HERNÁNDEZ. 2006. "Aportaciones recientes de los estudios de disponibilidad léxica", *Lynx. Panorámica de Estudios Lingüísticos* 5, 5-95.
- SANCHO SÁNCHEZ, Miriam. 2006. "Disponibilidad léxica de los hispanos en EEUU", *Cuadernos Cervantes de la Lengua Española* 63, 22-31.
- SÁNCHEZ-SAUS LASERNA, Marta. 2011. *Bases semánticas para el estudio de los centros de interés del léxico disponible. Disponibilidad léxica de informantes extranjeros en las universidades andaluzas*. Universidad de Cádiz. Tesis doctoral inédita.
- SANKOFF, David (ed.). 1978. *Linguistic Variation. Models and methods*. Nueva York: Academic Press.

- SANTIAGO GUERVÓS, Javier de. 2008. *El léxico disponible de Segovia*. Burgos: Instituto Castellano y Leonés de la lengua.
- SANZ ALCALÀ, Cristina. 1991. "Diglòssia a Catalunya?: Teoria i realitat als anys seixanta i actualment", *Sintagma* 3, 49-61.
- SARALEGUI, Carmen y Cristina TABERNERO. 2008. "Aportación al proyecto panhispánico de léxico disponible: Navarra" en OLZA, I. et al. (ed.). *Actas del XXXVII Simposio Internacional de la Sociedad Española de Lingüística*. Pamplona: Universidad de Navarra, 745-761.
- SAUSSURE, Ferdinand de. 1945. *Curso de lingüística general*. Buenos Aires: Losada.
- SEIB, Jörg. 2001. *La variedad bilingüe del español hablado en Cataluña caracterizada por interferencias y convergencias con el catalán*. Mannheim: Universität Mannheim. Tesina inédita.
- SERRANO VÁZQUEZ, María del Carmen. 1996. "Interferencias léxicas y semánticas en una situación de contacto entre dos lenguas, catalán y castellano", *Diálogos Hispánicos* 18, 375-394.
- SERRANO ZAPATA, Maribel. 2003a. "Disponibilidad léxica en la provincia de Lleida: datos iniciales", *Interlingüística* 14, 929-937.
- SERRANO ZAPATA, Maribel. 2003b. *Análisis sociolingüístico del léxico castellano disponible en la ciudad de Lleida*. Lleida: Universidad de Lleida. Proyecto de investigación inédito.
- SERRANO ZAPATA, Maribel. 2004. "Aspectos sociolingüísticos del léxico disponible castellano de los preuniversitarios leridanos", *Pragmalingüística* 12, 147-165.
- SERRANO ZAPATA, Maribel. 2006. "Consecuencias del contacto de lenguas en Lérida: interferencias detectadas en las encuestas de disponibilidad léxica", en BLAS ARROYO, José Luis et al. (eds.). 2006. *Discurso y sociedad. Contribuciones al estudio de la lengua en su contexto social*. Castellón de la Plana: Universitat Jaume I, 811-829.
- SILVA-CORVALÁN, C. 1989. *Sociolingüística: Teoría y Análisis*. Madrid: Alhambra.
- SILVA-CORVALÁN, Carmen. 1995. "The study of Language Contact: An Overview of the Issues" en SILVA-CORVALÁN, Carmen (ed.). *Spanish in Four Continents:*

- Studies in Language Contact and Bilingualism*. Washington D.C.: Georgetown University Press.
- SILVA-CORVALÁN, C. 2001. *Sociolingüística y Pragmática del español*. Washington, D.C.: Georgetown University Press.
- SIGUÁN, M. 1999. *Conocimiento y uso de las lenguas. Investigación sobre el conocimiento y uso de las lenguas cooficiales en las comunidades autónomas bilingües*. Madrid: Centro de Investigaciones Sociológicas.
- SIGUÁN, M. y W. MACKEY. 1986. *Educación y bilingüismo*. Madrid: Santillana/Unesco.
- SINNER, Carsten. 2001. *Corpus oral de profesionales de la lengua castellana en Barcelona*.
- SINNER, Carsten. 2004. *El castellano de Cataluña. Estudio empírico de aspectos léxicos, morfosintácticos y metalingüísticos*. Tübingen: Niemeyer.
- SINNER, Carsten y Andreas WESCH. 2008. “El castellano en las tierras de lengua catalana: estado de la cuestión” en SINNER, Carsten y Andreas WESCH (eds.). *El castellano en tierras de habla catalana*. Frankfurt am Main/Madrid: Iberoamericana/Vervuert, 11-57.
- SISTAC, Ramon. 1993. *El Ribargorçà a l’Alta Llitera: els parlars de la vall de la Sosa de Peralta*. Barcelona: Institut d’Estudis Catalans.
- SISTAC, Ramon. 1998. *El català d’Àneu: reflexions a l’entorn dels dialectes contemporanis*. Barcelona: Consell Cultural de les Valls d’Àneu.
- SOLÀ, Joan. 1980. “Tractats de catalanismes” en ALARCOS LLORACH, Emilio *et al.* (eds.). *Miscel·lània Aramon i Serra: estudis de llengua i literatura catalanes oferts a Ramon Aramon i Serra en el seu setantè aniversari*. Barcelona: Curial, vol. 2, 559–582.
- SOLÉ CAMARDONS, Jordi. 2001. *El poliedre sociolingüístic*. València: Contextos 3/4.
- SUÏLS, Jordi. 2004. “L’occità, llengua de la Vall d’Aran” en PAYRATÓ, Lluís y Francesc Xavier VILA (dirs.). *Les llengües a Catalunya*. Cicle Joan Coromines III. Sabadell: Fundació Caixa de Sabadell, 53-66.
- SZIGETVÁRI, Mónika. 1994. *Catalanismos en el español actual (Katalán elemek a mai spanyol nyelvben)*. Budapest: Eötvös Loránd Tudományegyetem.

- SZIGETVÁRI, Mónika. 2002. *Algunos aspectos de la influencia del catalán en el español de Barcelona*. Budapest: Eötvös Loránd Tudományegyetem. Tesis doctoral inédita.
- TABERNERO, Cristina. 2008a. “El léxico disponible como fuente de aproximación al estudio de regionalismos” en OLZA, I. *et al.* (ed.). *Actas del XXXVII Simposio Internacional de la Sociedad Española de Lingüística*. Pamplona: Universidad de Navarra, 811-824.
- TABERNERO, Cristina. 2008b. “Disponibilidad léxica y contacto de lenguas”, *Oihenart: cuadernos de lengua y literatura* 23, 545-565.
- TERMCAT = <http://termcat.cat>
- TERRÁDEZ GURREA, Marcial. 1996. *La disponibilidad léxica del español hablado en Valencia*. Valencia: Universitat de València. Tesis de licenciatura inédita.
- TORRES GONZÁLEZ, Antonia Nelis y Ana Belén GARCÍA FALERO. 2009. “Planificación y primeros resultados de la disponibilidad léxica en Tenerife” en CORBELLA, Dolores y Josefa DORTA. *La investigación dialectológica en la actualidad*. Las Palmas de Gran Canaria: Agencia Canaria de Investigación, Innovación y Sociedad de la información, 75-101.
- TRIGO IBÁÑEZ, Ester. 2007. *El léxico disponible de la provincia de Sevilla: variación versus déficit*. Universidad de Cádiz. Tesis doctoral inédita.
- TURELL, M.T. 1995. *La sociolingüística de la variació*. Barcelona: PPU.
- TURULL, Albert. 1990. “Notes sobre el lleidatà”, *Ilerda Humanitats XLVIII*, Lleida: Institut d’Estudis Ilerdencs, 201-206.
- TRUDGILL, Peter. 1974. *Sociolinguistics: an introduction to language and society*, London: Penguin Books.
- VALENCIA, Alba y Max S. ECHEVERRÍA. 1999a. “El factor geográfico en la disponibilidad léxica chilena” en SAMPER PADILLA, José Antonio *et al.* (eds.). *Actas del XI Congreso Internacional de ALFAL*. Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria y Librería Nogal, tomo II, 1705-1711.

- VALENCIA, Alba y Max S. ECHEVERRÍA. 1999b. *Disponibilidad léxica en estudiantes chilenos*. Santiago de Chile: Universidad de Chile y Universidad de Concepción.
- VANN, Robert E. 2008. “Una derivación transcódica en el español de los *Països catalans* y el papel del conocimiento lingüístico en su génesis” en SINNER, C. y A. WESCH (eds.). *El castellano en tierras de habla catalana*. Frankfurt am Main/Madrid: Iberoamericana/Vervuert, 269-286.
- VANN, Robert E. 2009. *Materials for the sociolinguistic description and corpus-based study of Spanish in Barcelona*. New York: Edwin Mellen Press.
- VENY, Joan. 1984. *Els parlars catalans*. Palma de Mallorca: editorial Moll.
- VILA, Pep. 1990. *Bocavulvari eròtic de la llengua catalana*. Barcelona: edicions de la Magrana.
- VILA i MORENO, Francesc Xavier. 1996. *When classes are over. Language choice and Language contact in bilingual education in Catalonia*. Proefschrift aangeboden tot het behalen van de graad van Doctor in de Taal-en Letterkunde. Faculteit der Letteren en Wijsbegeerte, Vrije Universiteit Brussel. Tesis doctoral inédita.
- VILA i MORENO, Francesc Xavier. 2000. “Les llengües a la Vall d’Aran” en FARRÀS *et al.* *El coneixement del català 1996*. Barcelona: Generalitat de Catalunya.
- VILA i MORENO, Francesc Xavier. 2004. “El català i el castellà a començament del mil·leni a Catalunya: condicionants i tendències” en PAYRATÓ, Lluís y Francesc Xavier VILA (dirs.). *Les llengües a Catalunya*. Cicle Joan Coromines III. Sabadell: Fundació Caixa de Sabadell, 29-51.
- VILA PUJOL, M. Rosa. 1996. “Consideraciones acerca de la interferencia del catalán en el español de Barcelona” en BRIZ GÓMEZ, Antonio *et al.* (eds.). *Pragmática y gramática del español hablado. Actas del II Simposio sobre análisis del discurso oral (Valencia, 14-22 de noviembre de 1995)*. Valencia / Zaragoza: Universitat de València, Departamento de Filología Española / Pórtico, 269-282.
- VILA PUJOL, M. Rosa. 2001. *Corpus del español conversacional de Barcelona y su área metropolitana*. Barcelona: Edicions Universitat de Barcelona.
- VILA PUJOL, Rosa. 2008. “El español en Cataluña” en MOYA CORRAL, Juan Antonio y Marcin SOSINKI (eds.). *El español en los territorios bilingües. Actas*

- de las XIII Jornadas sobre la lengua española y su enseñanza*. Granada: Universidad de Granada, 77-104.
- VILAGRASA, Joan. 2003. *Història de Lleida. Final del segle XX*. Lleida: Pagès editors. Vol. 9.
- VINYOLES i VIDAL, Joan J. 1989. *Diccionari eròtic i sexual*. Barcelona: edicions 62.
- WEINREICH, U. 1996. *Llengües en contacte. Alzira: editorial Bormera*. [1ªedición, 1953, *Languages in contact: Findings and Problems*, La Haya, Mouton.]
- WESCH, Andreas. 1997. “El castellano hablado de Barcelona y el influjo del catalán. Esbozo de un programa de investigación”, *Verba* 24, 287-312.
- WESCH, Andreas. 2002. “La investigación sobre variedades del español hablado en contacto con el catalán (particularmente en Cataluña y Baleares): estado de la cuestión y perspectivas para el futuro” en ECHENIQUE, María Teresa *et al.* (eds.). *Actas del V Congreso Internacional de Historia de la Lengua Española*. Madrid: Gredos, 1857-1872.
- WESCH, A. 2008. “Mujeres de hacer faenas que limpian racholas: sobre algunos catalanismos léxicos” en SINNER, Carsten y Andreas WESCH (eds.). *El castellano en tierras de habla catalana*. Frankfurt am Main/Madrid: Iberoamericana/Vervuert, 57-64.
- WOOLARD, Kathryn A. 1992. *Identitat i contacte de llengües a Barcelona*. Barcelona: La Magrana.