

Departament de Didàctica de la Llengua i la Literatura,  
i de les Ciències Socials  
Universitat Autònoma de Barcelona

Tesi doctoral

**LA SEGMENTACIÓ DEL TEXT ESCRIT EN  
INFANTS DE PRIMER CURS DE PRIMÀRIA**  
Estudi evolutiu

Montserrat Bigas Salvador

Direcció: Dra. Anna Camps i Mundó

Bellaterra, 2006

***"Vine, làvia, vine"***

**Joana, 2 anys i dos mesos**

## AGRAÏMENTS

Per sort, aquesta primera pàgina és l'última que s'escriu, cosa que té l'avantatge d'oferir una certa perspectiva i permetre reflexionar sobre el procés que ara es tanca.

I per sort, també, aquest és l'espai on puc, en primera persona, agrair el suport, l'ajuda, la col·laboració, la crítica i els ànims d'amics i amigues, companys i companyes, família. Les seves aportacions, sense cap mena de dubte, han ajudat a fer arribar el carro a final de trajecte.

He trobat, entre els companys del Departament, l'ambient favorable i el recolzament necessari per a la realització de la tasca. La seva disposició a escoltar les meves disquisicions i fer-me suggeriments adequats ha estat una contribució valuosa.

Agraeixo, molt especialment, el guiatge d'Anna Camps, que ha estat crític i rigorós, però a la vegada amb idees suggerents i aportacions aclaridores. La seva capacitat de situar-se pel damunt dels problemes, petits i concrets, en què moltes vegades m'entrebancava ha estat una ajuda inestimable: els seus comentaris m'obrien perspectives més generals i àmplies que em permetien superar els esculls.

I com que som el que som i sabem el que sabem, he d'agrair especialment a l'Anna Cuxart, que m'hagi ajudat a interpretar amb una mica més de rigor del que jo sola hauria estat capaç, les dades numèriques al moment de l'anàlisi quantitativa.

A la Núria Vilà li deuré tota la vida el munt d'hores que ha dedicat a l'edició: marges, negretes, taules i un llarg etcètera de problemes que m'ha resolt amb dedicació benevolent.

Agraeixo el suport de l'Alfons Valenzuela que m'ha conduït, amb la seva solvència, paciència i discreció, pels camins intricats de la burocràcia administrativa.

No vull passar per alt el nom de les mestres que em van aportar, generosament, tots els textos que conformen el corpus analitzat: la Montserrat Fons, de l'Escola Estel de Barcelona, i la Montserrat Font, de l'Escola Bellaterra. Com passa amb freqüència, la nostra feina no seria possible ni tindria sentit sense la col·laboració desinteressada de les mestres que ens obren les portes de les seves aules.

A l'Eva i en Sasha, a en Martí i l'Elena i a en Joan pel seu afecte i suport.

A tots i a totes, moltes gràcies

INTRODUCCIÓ .....	3
I MARC TEÒRIC .....	13
1. EL CONCEPTE DE PARAULA.....	15
1.1. Les dificultats per a una definició.....	15
1.2. Les definicions proposades .....	20
1.3. La noció de paraula gràfica .....	28
2. LES UNITATS DE L'ORAL I LES DE L'ESCRIT .....	31
3. LA SEGMENTACIÓ DE LA PARLA I LA CONSTRUCCIÓ DEL CONCEPTE DE PARAULA .....	37
3.1. La segmentació de la parla i la percepció de les unitats.....	39
3.2. El coneixement de la paraula: naturalesa i evolució.....	44
3.3. La construcció del lexicó i l'accés al lexic.....	54
3.4. La representació de les unitats. El model de redescrípció representacional de Karmiloff-Smith de 1986(1994).....	57
4. ESTUDIS PREVIS .....	63
II. LA RECERCA: LA SEGMENTACIÓ DEL TEXT D'UN GRUP D'INFANTS DE PRIMER CURS DE PRIMÀRIA .....	71
5. OBJECTIU GENERAL I PREGUNTES DE RECERCA.....	73
6. DISSENY DE LA RECERCA.....	85
6.1. Establiment del corpus.....	88
6.1.1. Les escoles .....	88
6.1.2. Els subjectes i la mostra recollida .....	89
6.2. Tractament de les dades .....	92
6.2.1. Agrupacions dels textos.....	93
6.2.2. Tipus d'error i decisions durant el procés de categorització.....	93
6.2.3. Les unitats de les seqüències no segmentades (SHO).....	98
6.2.4. Agrupacions segons l'organització de les unitats.....	102
6.2.5. L'estructura sintàctica de les seqüències hiposegmentades .....	107
7. ANÀLISI I RESULTATS .....	114
7.1. Quins problemes de segmentació són més freqüents en els textos? .....	115
7.1.1. Anàlisi i resultats .....	115
7.2. Com progressa la llargada dels textos dels infants? .....	120
7.2.1. Evolució en la llargada dels textos dels infants .....	120
7.2.2. Correlació entre les llargades dels textos, per infants, al llarg de tot el període estudiat .....	127
7.2.3. S'intueix algun patró de comportament? .....	128
7.3. Com evoluciona la comissió d'errors? Disminueix el percentatge d'errors al llarg del curs? .....	129
7.3.1. Evolució individual. ....	129
7.3.2. Evolució dels errors en conjunt .....	132
7.4. Existeix una correlació entre la llargada dels textos i la comissió d'errors? .....	135

7.4.1. Anàlisi i resultats .....	135
7.5. La distribució dels errors en els textos llargs .....	137
7.5.1. Anàlisi i resultats .....	137
7.6. Anàlisi de la hipersegmentació .....	143
7.6.1. Anàlisi i resultats .....	145
7.7. Quina estructura tenen, les seqüències hiposegmentades? .....	150
7.7.1. Anàlisi i resultats .....	151
7.8. Es transgredeixen els límits sintagmàtics en les seqüències hiposegmentades? .....	155
7.8.1. Anàlisi i resultats .....	156
7.9. L'estructura sintàctica de les SHO. Anàlisi dels components. ....	157
7.9.1. Estructura [art+nom].....	159
7.9.2. Les estructures amb la preposició A i amb la preposició DE.....	169
7.9.3. Les estructures de nucli verbal.....	194
8. CONCLUSIONS .....	217
8.1. L'anàlisi de les dades.....	217
8.2. Les hipòtesis i els resultats.....	220
8.3. Conclusió general .....	228
9. BIBLIOGRAFIA .....	237
ANNEXOS .....	249
1. Mostra de textos originals.....	251
2. Relació completa dels textos transcrits. Corpus analitzat.....	265
3. Taules d'anàlisi quantitativa .....	285
4. Dades de les estructures analitzades .....	295
4.1. Les estructures amb article .....	295
4.2. Les estructures amb la preposició A .....	306
4.3. Les estructures amb la preposició DE.....	316

## INTRODUCCIÓ

El treball que presentem se centra en l'estudi d'un aspecte del sistema d'escriptura: la segmentació del text en paraules. Els dubtes i els errors relacionats amb la segmentació afecten una proporció notable de les primeres produccions escrites dels infants als inicis de l'escolarització obligatòria. Els primers textos que els infants escriuen són encara molt lluny d'ajustar-se a la convenció i, a més de mostrar les dificultats per a la gestió de l'espai i per a la correcta grafia de les lletres, presenten desajustaments ortogràfics de molts tipus: errors en la correspondència grafofònica, en la utilització dels accents i els signes de puntuació i, també, en l'establiment dels blancs convencionals que delimiten segments gràfics.

El sistema d'escriptura del català, com el de totes les llengües romàniques i de la majoria de les de l'àmbit cultural que ens és pròxim, és un sistema alfabètic. Contràriament a la creença més estesa, els sistemes alfabètics no són sistemes de representació que es basin només en els principis de representació fònica, bé sigui el fonema la unitat representada, bé sigui la síl.laba. Jaffré (1995) explica que tots els sistemes d'escriptura giren a l'entorn de dos principis: el principi fonogràfic i el semiogràfic, en la mesura que donen informacions de tipus fònic, però també de significat de les unitats lingüístiques. Les escriptures fonogràfiques alfabètiques es basen en la relació grafofònica, en què cada una de les unitats mínimes -les lletres- representen unitats lingüístiques. Però aquesta és una representació força complexa, en el sentit que les unitats gràfiques mínimes, i també els altres signes que formen part del sistema d'escriptura, a més d'ajustar-se al principi fonogràfic a què al.ludíem al principi, apel.len a altres nivells del sistema lingüístic: el sistema morfològic, el sintàctic o el lèxic que responen al principi semiogràfic en la mesura que les lletres o grups de lletres són portadores de significat.

En aquest sentit, els espais són una forma de notació que respon al principi semiogràfic de l'escriptura. La gestió dels espais entre les paraules per part dels aprenents -és a dir, la segmentació- dona informació dels criteris amb què es guien i permet saber si aquests

critèris responen a comportaments aleatoris o si indiquen la projecció d'algun coneixement gramatical. Els segments entre blancs es corresponen a unitats lingüístiques significatives, les paraules, que, malgrat la discussió sobre la seva validesa científica, són considerades els elements centrals d'una llengua.

Per aquest motiu la segmentació d'un text escrit es pot considerar una qüestió diferent a d'altres qüestions ortogràfiques, com ara l'ús de les majúscules, alguns signes de puntuació, algunes relacions grafofòniques arbitràries<sup>1</sup>, algunes lletres sense representació fònica o l'accentuació. En quin sentit es pot afirmar que són qüestions diferents? En el sentit que el sistema d'escriptura del català prioritza l'aïllament entre blancs d'aquestes unitats centrals de la llengua: les paraules, i es minimitza, en la segmentació, el grau d'arbitrarietat inherent als sistemes gràfics. Aquesta afirmació implica acceptar que la relació entre la paraula gràfica i la paraula del sistema és una relació motivada. La mateixa qüestió es planteja Reichler-Béguelin quan aborda el tema de l'estatus de l'escrit dins la ciència lingüística:

"En effet, du point de vue adopté à ce sujet [es refereix a l'estatus de la llengua escrita], dépend étroitement le crédit qu'il est possible d'accorder à la notion de "mot" graphique, la question étant posée de savoir si, oui ou non, il s'agit d'une notion arbitraire en langue" (1992:146).

La relativa rapidesa amb què els infants aprenen a segmentar encara reforça més la idea que existeix alguna cosa de "natural" en les unitats tal com estan delimitades per la normativa<sup>2</sup>.

Malgrat que l'ús del terme *paraula* com a terme científic és moltes vegades qüestionat per la Lingüística, se n'accepta l'existència d'una noció intuïtiva i, pel que fa a la llengua escrita, el pes de la tradició

---

<sup>1</sup> El concepte d'ortografia arbitrària i natural va ser proposat per A. Galí per diferenciar aquells aspectes ortogràfics relacionats "amb una norma fonètica derivada, com hem expressat, de les primeres lletres" dels altres "que són els que podem anomenar pròpiament ortogràfics, on predomina el convencionalisme" (1980[1926]:14).

<sup>2</sup> Galí (op.cit.) planteja la qüestió de la separació de mots assumint dos principis: és un coneixement que depèn de l'aprenentatge gramatical –entenen aquesta noció com una pràctica més que com un saber- i també de la intuïció de la unitat paraula (pp.24-27).

afavoreix i justifica l'ús de *paraula* per referir-se als segments gràfics entre blancs.

També, la majoria d'estudis consultats esmenten, d'una manera o d'una altra, el terme *paraula* per referir-se als elements que queden delimitats per blancs, tot associant *paraula* i *paraula gràfica*. Cárdenas y Báez estableixen clarament al capítol 3, la qüestió de la relació entre els segments i la paraula en el títol mateix: "*La segmentación gráfica del texto y la noción de palabra: un problema psicolingüístico*" (1999:25). Les autores segueixen el camí iniciat anteriorment per d'altres investigadors, com Ferreiro, coautora d'un dels estudis de referència sobre segmentació, explicat més endavant. Aquesta autora titula també el capítol 2: "*Los límites entre las palabras*" (Ferreiro et al. 1995:45).

Ens hem interessat per la segmentació com a objecte d'estudi perquè, a més del coneixement ortogràfic convencional, *la nostra suposició és que els segments erronis indiquen algun tipus de coneixement gramatical i que l'establiment de blancs no segueix un procés aleatori sinó que, justament, està guiat per aquest coneixement.*

Per tal d'obtenir resultats versemblants, semblava més adequat treballar amb un corpus sorgit dels textos realitzats lliurement pels subjectes que no pas amb textos que provinguessin de l'aplicació d'un protocol establert. Aquest tipus de corpus, difícil d'analitzar, val a dir, informa d'una manera clara de les decisions que els infants prenen amb relació a la segmentació, tant les decisions equivocades com les convencionals. Cada text és un testimoni real de l'estadi de coneixement del sistema gràfic, en aquest cas sobre l'aspecte concret de la paraula gràfica.

És probable que, fa unes dècades, no fos possible d'obtenir un corpus d'aquestes característiques. Les activitats d'escriptura estaven molt pautades, amb l'objectiu, justament d'evitar l'error. Actualment, però, en algunes escoles, s'ha produït un canvi en el plantejament de l'ensenyament de la llengua escrita; no solament en els aspectes metodològics, sinó també en la pròpia consideració del procés de llegir i escriure i en els objectius que l'escola s'ha de plantejar amb relació a l'alfabetització.



En aquests canvis hi ha tingut un pes fonamental -entre d'altres factors- la informació que els resultats de les recerques en les diferents disciplines relacionades amb l'aprenentatge inicial de la llengua escrita han aportat. Aquestes recerques han aportat dades suficients que han servit de suport teòric a molts dels canvis metodològics que s'estan produint a les aules. Les investigacions sobre llegir i escriure, en el camp de la psicologia i la psicolingüística o en d'altres, han posat de relleu que l'eix fonamental d'aquestes activitats és la recerca i creació de significat, i que la construcció d'aquest significat, tant en la recepció com en la producció d'un text escrit, no depèn tant d'habilitats relacionades amb el coneixement del codi com d'activitats cognitives que en desencadenen moltes d'altres, amb el propòsit de crear significat (Solé, 1987).

Les aportacions científiques sobre el procés d'aprenentatge de la lectura i l'escriptura ens han proporcionat una informació precisa sobre com els infants construeixen aquests sabers i com els desajustaments que es produeixen durant l'aprenentatge responen a les idees que ells tenen, en un moment determinat, sobre el que aprenen (Ferreiro i Teberosky, 1979; Uta Frith, 1989; Solé, 1987). S'accepta que una intervenció educativa ajustada a les possibilitats de l'infant justifica la possibilitat d'anticipar el procés d'alfabetització sense que aquesta anticipació suposi una forma de forçar aprenentatges.

Aquests canvis en el camp del coneixement científic han tingut repercussions en la tasca escolar i han portat a revisar tant els continguts com la metodologia a l'hora d'iniciar l'alfabetització a l'escola. L'adopció de maneres de fer relacionades amb la difusió del model constructivista s'ha traduït a l'escola en pràctiques centrades en una visió de l'escrit com un fet comunicatiu, funcional i significatiu més que no pas com una suma d'habilitats que parteixen del coneixement del codi i que estan desvinculades del procés cognitiu de comprensió i creació de significat. La concepció que l'ensenyament del sistema d'escriptura s'havia de produir per fases i procedint des de les unitats més petites -més senzilles, es deia-, *les lletres*, fins a les més complexes -més difícils, es pensava-, com *el text*, ja no és una veritat indiscutible. En la pràctica escolar, no era concebible ensenyar una unitat suposadament més difícil fins que s'hagués après i consolidat el

coneixement de les anteriors. L'aprenentatge era vist, en aquest cas, com un procés molt controlat, amb poc marge d'experimentació i d'error.

Actualment, les propostes d'escriptura -i també de lectura-, són molt més obertes, menys pautades i controlades: els infants proven d'escriure mots, frases i textos dels quals no han tingut una informació prèvia. Una proposta educativa d'aquest tipus prioritza la funció comunicativa de la llengua escrita per damunt de l'ortografia i cal·ligrafia correctes, aspectes que en un plantejament pedagògic diferent haurien estat els eixos centrals de la tasca. Una de les conseqüències de l'espontaneïtat en l'escriptura es pot apreciar en la producció de textos llargs però, també, evidentment, de textos amb ortografies força allunyades de la convenció.

En aquest marc metodològic -que no és un marc generalitzat-, els infants han desenvolupat una actitud davant de la producció d'un text desinhibida i desacomplexada en el sentit que, com que el que els mou no és la producció de paraules o de lletres, sinó la producció d'un text que té per a ells una finalitat comunicativa clara, la falta de competència en els diferents nivells gramaticals, com ara el coneixement suficient del sistema gràfic i morfològic, no són impediments per a la tasca, i els problemes que apareixen es resolen durant el procés de redacció.

Així doncs, no hi ha una presentació seqüenciada i jerarquitzada dels continguts en funció de la seva suposada dificultat, sinó que la tasca es presenta de forma holística, guiada per la finalitat comunicativa. Hi ha una manera de procedir menys controlada per part del mestre, la qual cosa dona com a resultat textos amb un desajustament ortogràfic potser més gran que les manifestacions escrites que es produïen amb una metodologia més guiada. Utilitzem, conscientment, l'expressió *manifestació escrita* i no producció o redacció de textos, per subratllar que, en el marc de la perspectiva metodològica que optava per la seqüenciació i graduació dels continguts, l'escriptura de textos arribava després que els infants haguessin practicat amb l'escriptura de paraules i frases, no de textos. La idea de text com a unitat de comunicació que pot ser produït des dels inicis del procés

d'alfabetització és recent, encara que en trobem els antecedents en les propostes de Freinet (Freinet, 1979).

No és propòsit d'aquestes ratlles valorar els diversos plantejaments metodològics d'ensenyament de la llengua escrita, ni tampoc afirmar que el coneixement ortogràfic és *millor* en una metodologia que assegura la pràctica de l'escriptura de paraules per sobre de la producció de textos. Potser, però, es pot afirmar que en un plantejament de l'ensenyament de l'escriptura més funcional i comunicatiu, els infants produeixen textos molt més llargs des de l'inici de l'alfabetització a l'escola, i que aquests textos mostren el coneixement ortogràfic en construcció. Podríem dir que aquest coneixement ortogràfic es fa visible sobre el paper i pot fer pensar, a mestres i educadors, que les metodologies menys controlades provoquen que els infants cometin més errors i que, en conseqüència, actualment, el nivell ortogràfic dels infants sigui *pitjor* que abans. Justament, la manifestació gràfica de les idees dels infants sobre el sistema d'escriptura, amb els errors que hi apareixen, és una informació necessària que permet explicar l'evolució del seu coneixement.

L'objectiu de l'estudi és analitzar un aspecte del coneixement ortogràfic: la competència a segmentar els textos d'acord amb la normativa, que mostren deu infants de 1r. curs de Primària de dues escoles, al llarg d'un període de sis mesos. Per guiar l'anàlisi i obtenir informació que permeti assolir aquest objectiu s'han formulat 9 preguntes que proporcionen les dades bàsiques. És un estudi evolutiu sobre l'adquisició d'aquesta capacitat, que parteix de l'acceptació de la paraula gràfica com a unitat d'anàlisi, tal com ha estat establerta en els estudis consultats. És un concepte relativament fàcil d'establir, com afirmen Tolchinsky i Cintas (2001:79).

És un estudi en la línia de les recerques de tipus natural i no pas un treball experimental com d'altres estudis sobre aquest tema que han estat consultats. El corpus està format per textos escrits en situació natural d'aula i no hi ha hagut restriccions per a la seva escriptura. La decisió de treballar amb els textos complets és interessant perquè

dóna la possibilitat d'analitzar els errors i posar-los en relació amb els no errors; és a dir, amb totes les ocurrències.

L'anàlisi acurada de les dades obtingudes ha de permetre conèixer els tipus d'errors comesos i observar l'evolució de la competència dels infants al llarg del període. Però, també, i aquest és un aspecte rellevant, ha de donar informació sobre els processos d'adquisició de de l'escriptura, en el cas dels infants a l'inici de l'escolarització. És per aquest motiu que ens ha semblat imprescindible no considerar només els errors, sinó el conjunt dels textos escrits.

...

L'estudi s'insereix en una de les línies de recerca de la unitat de Didàctica de la Llengua i la Literatura del Departament de Didàctica de la Llengua, la Literatura i les Ciències Socials, de la Universitat Autònoma de Barcelona. Es particularitza en el primer cicle de primària, per diversos motius: a) el fet que l'autora ha desenvolupat, principalment, la seva docència tant en la formació inicial com la formació permanent, en els temes relacionats amb els inicis de l'alfabetització, i b) més important encara, l'interès social que s'atorga als nivells d'alfabetització de la població; en aquest sentit, obtenir dades corresponents al procés inicial de l'alfabetització és considerat un objectiu d'estudi prioritari en les línies de treball del Departament de Didàctica al qual el present estudi està inscrit.

El treball s'organitza en dues parts: la primera, dedicada a l'exposició del marc teòric de referència; i la segona, a la presentació de la investigació feta: les dades, l'anàlisi i les conclusions que se'n desprenen.

I. El marc de referència està constituït per dos apartats clarament diferenciats:

En el primer, s'aborden les qüestions de tipus epistemològic relacionades amb el tema de la investigació:

- a) La pertinència del terme *paraula* per referir-se als segments gràfics delimitats per blancs. Sobre aquest punt ens hem basat en les aportacions de la lingüística i de la psicolingüística que es mouen entre dos punts de vista: un, més científic, que renuncia a l'ús del *paraula* com a terme científic de referència, i un altre, que defensa l'existència d'una noció intuïtivament acceptada entre els parlants, d'aquest terme.
- b) La relació entre les unitats de l'escrit amb les unitats de l'oral i, més específicament, si es pot parlar de paraules de la llengua oral, ateses les característiques del flux de la parla, o si aquestes són fruit de la influència de la cultura escrita.
- c) La segmentació de la cadena parlada, la descoberta i emmagatzement de les unitats i la seva disponibilitat per a les tasques de producció oral i escrita.
- d) La justificació de la tria del model de redescrípció representacional (RR), de Karmiloff-Smith (1994), com un model que permetria explicar per què les unitats paraula, emmagatzemades en el lexicó dels infants i disponibles per a la comprensió i producció de la parla, encara no ho estan per a les tasques d'escriptura als inicis de l'alfabetització.

En el segon apartat, es fa referència a treballs sobre el mateix tema que, publicats o no, han pogut ser consultats durant la realització de l'estudi.

- II. A la segona part, s'hi especifiquen els objectius, el disseny de la recerca, l'anàlisi i les conclusions.
  - a) S'ha formulat un objectiu general i uns objectius específics, en forma de preguntes, amb la finalitat que el conjunt de respostes obtingudes configuri un panorama sobre el coneixement de la segmentació en el grup estudiat.
  - b) El disseny de la recerca conté la descripció dels subjectes estudiats i del corpus amb què s'ha treballat: periodicitat,

consignes i condicions de recollida, consideracions metodològiques de la transcripció, categoritzacions dels errors i consideracions metodològiques en l'aplicació.

- c) L'anàlisi es presenta en apartats corresponents a cada objectiu específic. De cada aspecte, se n'ha fet una anàlisi quantitativa, que ha servit per donar informació de les freqüències d'errors o d'altres dades, i una altra de qualitativa en què es considera el conjunt de les ocurrències: tant les formes errònies com les convencionals. L'observació de cada aspecte i categoria analitzats amb relació a tots els casos ha facilitat un coneixement aprofundit del fenomen i trobar una via d'interpretació més enllà de la descripció i constatació dels fets.
- d) Les conclusions fan referència a tres àmbits diferents: en primer lloc, es fa una reflexió sobre l'interès i a la vegada la complexitat del treball amb corpus complets, reals, sense consigna prèvia. Es sintetitzen, breument, els resultats de cada apartat analitzat i es posen amb relació a les hipòtesis de partida. Finalment, en un tercer apartat, s'esbossa una interpretació dels fets, més enllà de les informacions que els resultats han aportat.



# I MARC TEÒRIC





## 1. EL CONCEPTE DE PARAULA

Malgrat el consens en la utilització del terme paraula ¿podem parlar de paraules per referir-nos a aquests segments gràfics? Els lingüistes no han arribat a establir una noció clara i inequívoca del concepte de *paraula*, tot i que sembla que aquest concepte sigui indiscutible i estigui a l'abast de tots els parlants. Voldríem exposar algunes de les dificultats més rellevants per arribar a una definició única i consensuada.

### 1.1. Les dificultats per a una definició

Una d'aquestes dificultats, com tot just acabem de dir, prové del fet que el terme *paraula* sigui usat en la parla quotidiana, ben al contrari del que passa amb d'altres termes gramaticals, com ara fonema, síl.laba o sintagma, que tenen una referència inequívoca. Gabriel Ferrater ho explica d'una forma molt poètica:

...podríem dir que un fantasma corre per la gramàtica: el fantasma del mot, i que tots els lingüistes coaligats no són prou per caçar-lo, és a dir, a definir-lo. Parlem constantment de mots, i entre tots fórem ben incapaços de dir què és un mot. (...) El mal és que el mot té un estatut especial: mentre que el lingüista pot pensar que tots els altres conceptes de la llista precedent [*l'autor es refereix a: fonemes, morfemes, mots, categories*] són propietat seva, i cadascú fa el que vol amb les seves coses, amb el mot no va així. Tota persona, fins i tot un analfabet total, parla de mots, pregunta i discuteix què volen dir, com s'han de pronunciar o com els pronuncia tal o tal poble. (Ferrater, 1981:47)

Aquesta és una de les raons per la qual molts lingüistes no accepten la paraula com una categoria lingüística que tingui valor científic<sup>3</sup>. De fet, que l'ús ordinari no es correspongui amb el terme científic no és una raresa: en d'altres camps de la ciència passen fenòmens semblants. Per a moltes persones les balenes o els calamars són peixos, simplement perquè viuen en un medi marí; o els ratpenats són aus, pel fet que volen; és a dir, en el procés de conceptualització, el parlant ha seleccionat una nota: el medi on viuen aquests animals, ometent altres

---

<sup>3</sup> Vegeu Blanche-Benveniste (1998:65).

característiques que són les que permeten decidir la classe a la qual pertanyen. El coneixement poc precís dels termes científics no impedeix a les persones d'usar-los en les situacions de comunicació. Així doncs, també en el camp de la Lingüística -segurament més que en d'altres camps- la utilització de termes metalingüístics, com *paraula* o *so*, en l'ús quotidià no implica que els usuaris comparteixin la mateixa referència que els lingüistes, sobretot quan des de la Lingüística no hi ha acord en la definició del terme.

No obstant el que acabem de dir, i encara que pugui semblar una paradoxa, entre els lingüistes hi ha un cert consens en l'acceptació de l'existència d'una noció intuïtiva en els parlants del concepte paraula. Es planteja un debat entre aquells qui creuen que aquesta no és una noció científica que pugui ser utilitzada per la Lingüística -concretament, els lingüistes estructuralistes- i aquells que defensen la preexistència d'una noció o idea de paraula -en general, els generativistes- molt abans de cap descripció lingüística i que forma part de la consciència col·lectiva de la humanitat. Segurament, no tots els lingüistes estarien d'acord amb aquesta afirmació. Goody<sup>4</sup>, per exemple, explica que els termes de paraula i de frase sembla ser que no existeixen al marge de les societats que tenen un sistema d'escriptura.

La idea que la paraula és una entitat psicològica real, al marge de si pot ser considerada una categoria científica o no, la trobem recollida tant per alguns lingüistes com per psicòlegs i psicolingüistes. Saussure diu:

Il faudrait chercher sur quoi se fonde la division en mots -car le mot, malgré la difficulté qu'on a à le définir, est une unité qui s'impose à l'esprit, quelque chose de central dans le mécanisme de la langue.  
(Saussure. 1967:154)

Sapir, després d'acceptar que la paraula és el primer element de la parla del qual podem dir que realment existeix (Sapir, 1985:29), creu que és necessari definir què és, realment, una paraula, però una de les definicions que proposa s'ajusta ben poc a les característiques d'una

---

<sup>4</sup> Goody, citat per Blanche-Benveniste (1998).

definició feta per un lingüista: "la paraula és la unitat existent de la parla viva, respon a la unitat de l'experiència realment percebuda, de la història, de l'art" (Sapir, 1985:34). És una definició filosòfica i antropològica, però ben poc lingüística.

També Michèle Fruyt insisteix en l'existència d'un ús intuïtiu del terme i en la dificultat de definir-lo, sense que aquest fet suposi un obstacle a l'ús.

Le mot, en premier lieu, a un caractère d'évidence, tels ces postulats qui fondent certaines sciences sans pouvoir être eux-mêmes fondés sur un raisonnement et faire l'objet d'une démonstration. Les opérations les plus simples en apparence sont souvent les plus simples à décrire d'un point de vue théorique: on peut faire de l'arithmétique sans avoir une idée exacte de la définition théorique d'une opération fondamentale comme l'addition. Or, avec le mot, nous touchons la nature du signe, l'ensemble de la sémiologie et sommes aux confins des rapports entre langue et pensée (1992: 113).

Les paraules de Lyons (1987) són prou eloqüents del caràcter intuïtiu del mot i mostren la paradoxa que representa usar un terme que apareix a bastament en el llenguatge quotidià i, a la vegada, en els llibres de gramàtica, en la llengua oral i en l'escrita i entre persones alfabetitzades i analfabetes:

A word is what you think is a word (Lyons, a Fruyt, 1992).

Un segon tipus de dificultat prové de les característiques estructurals de la pròpia llengua. En les llengües flexives, com el català, les unitats gràfiques són representacions complexes que inclouen unitats de categories diverses. Els segments entre blancs dels exemples següents són una mostra del que diem: *tub*, *prehistòria*, *mengem*, *pitroig*, *Aiguablava*. També les diferents formes d'una mateixa paraula creen confusió a l'hora de determinar si ens trobem davant d'una o més paraules. ¿Són dues paraules diferents *dormo* i *dormirem*? ¿o *la*, *l'* i *les*? ¿o *em m'* -*m* i *'m*? Segons el criteri que s'apliqui en l'anàlisi -semàntic, sintàctic, morfològic, fonètic o gràfic-, podem donar-hi respostes diferents. Segons el criteri semàntic o sintàctic, és acceptable de considerar que ens trobem davant d'una mateixa paraula; però segons un criteri fonètic, morfològic o gràfic, sembla clar que són paraules diferents. Bloomfield ho explica d'una forma molt enraonada:

En la nostra tradició escolar parlem de vegades de formes tals com *book*, *books* "llibre, llibres", o *do*, *does*, *did*, *done* "fer, fa, feia, fet, com 'formes diferents d'un mateix mot'. No cal dir que això és inexacte, des del moment que hi ha diferències de forma i de sentit entre els membres d'aquests conjunts: formes com aquestes que hem citat són formes lingüístiques diferents i, doncs, mots diferents (1977:197).

Una altra de les característiques de la pròpia llengua és la que es refereix al tipus de paraula. En català, com en d'altres llengües romàniques, les paraules poden ser classificades en dos grups segons el contingut semàntic i la funció sintàctica que realitzin en les frases. En el primer grup hi situem les *paraules lèxiques* o de contingut, que pertanyen a una classe oberta: noms, adjectius, verbs i adverbis; i en el segon grup, hi agrupem les *paraules funció*<sup>5</sup> o gramaticals, que pertanyen a una classe tancada: les preposicions, els pronoms, les conjuncions o els articles.

Així, doncs, els al·lomorfs, les paraules funció i les formes compostes<sup>6</sup> com *tauleta de nit o Sant Andreu de la Barca* són igualment una font de discrepàncies a l'hora de determinar si una unitat és o no una paraula. En les investigacions fetes en població adulta en què es demana a persones si una unitat lingüística és o no una paraula, el resultat mostra que els dubtes apareixen en la consideració de les paraules funció i en les expressions compostes. Fins i tot, si diferenciem el tipus de coneixement desplegat pels parlants: o bé un coneixement explícit, que pot ser verbalitzat, o bé, simplement un coneixement procedimental -que és el que es posa en marxa en la resolució de tasques de decisió- es pot observar que hi ha respostes diverses davant d'un mateix problema. Per exemple, tal com acabem d'explicar, en tasques de resolució en què es demana a persones de nivells de formació diversos que responguin si algun dels exemples citats unes ratlles més amunt són paraules, les respostes acostumen a

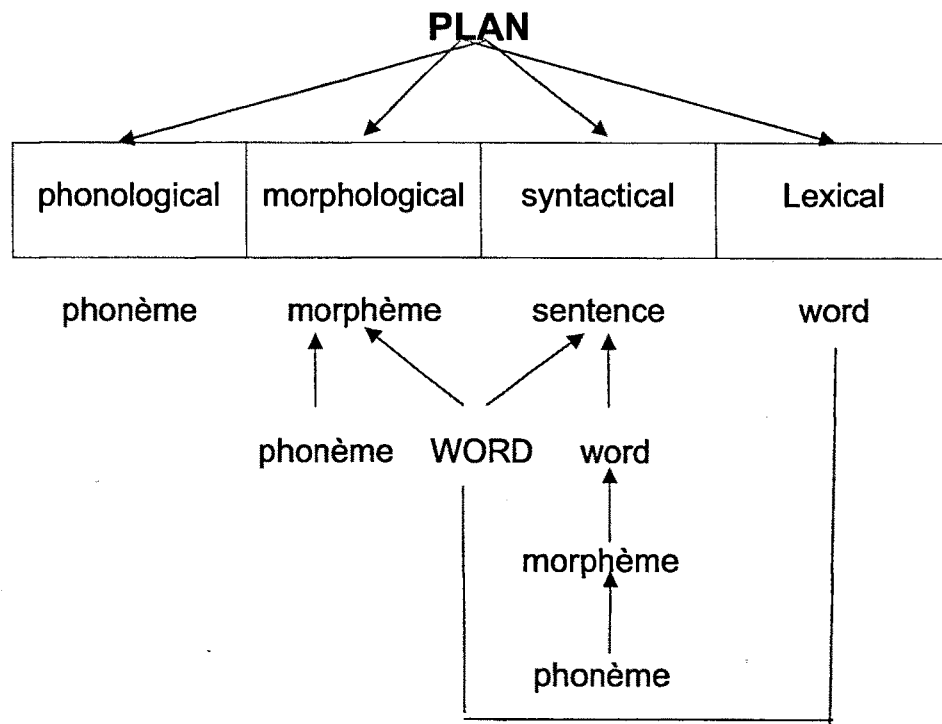
---

<sup>5</sup> Utilitzem els termes **paraules funció** o **gramatical**, en el sentit que els usa Halliday (1985): **grammatical items** o **function words**, en oposició a **content words** o **lexical items** (paraules o elements lèxics). Aquesta terminologia ha estat usada també per Tolchisky i Cintas (2001). Es poden trobar d'altres termes per als mateixos referents. GR Cardona (1991) cita, com a sinònims: paraules autosemàntiques/ sinsemàntiques; principals / accessòries; autònomes / partícules, etc.

<sup>6</sup> Per a una explicació detallada de la complexitat de la composició de les paraules, vegeu G. Ferrater (1981).

divergir. En canvi, si se'ls demana d'incloure una paraula en un enunciat, no es produeix cap cas de partició incorrecta de paraula<sup>7</sup>. Sembla clar, doncs, que l'ús habitual i corrent del terme i l'habilitat per manipular les unitats lingüístiques, no pressuposen una noció clarament establerta per part dels parlants, sense que aquesta indefinició representi un obstacle per al seu ús en tots els àmbits ni per a la seva manipulació.

La tercera de les dificultats fa referència a la situació de la paraula en el conjunt del sistema gramatical de la llengua i la relació amb els diferents subsistemes o nivells: amb el fonològic, el morfològic, el sintàctic i el lèxic. Jirí Krámský (1969), lingüista pertanyent al Cercle Lingüístic de Praga, és autor d'un dels estudis més interessants sobre la naturalesa de la paraula. Diu: "*...the word belongs both: to the morphological and to the syntactic and lexical plans*" (pp.16). Proposa el següent esquema per mostrar gràficament la posició de la paraula en el sistema lingüístic:



Krámský (1969:16)

<sup>7</sup> La indivisibilitat és un dels cinc principis que Crystal anomena per determinar si una unitat és o no paraula (1994:90).

Aquesta posició a cavall entre més d'un sistema és un punt de vista compartit, també, per d'altres lingüistes. Crystal afirma: "*las palabras se asientan de modo inseguro entre la morfología y la sintaxis*" (Crystal, 1994:90). De tota manera, introdueix un matis diferenciador segons la tipologia a la qual pertanyen les llengües: en les llengües aïllants les paraules són fàcilment identificables, en les llengües flexives la morfologia és un element distorsionador i en les llengües polisintètiques, la dificultat per a l'establiment del límit de la paraula és encara molt més difícil, atès que les estructures són molt complexes i poden equivaldre a oracions. Resulta curiós, i revelador de la complexitat del tema que tractem, que Crystal, per parlar de la dificultat d'establir el concepte de paraula en llengües aglutinants o polisintètiques, s'expressi en aquests termes: "*en las lenguas polisintéticas, las palabras son formas altamente complejas que equivalen a oraciones algunas veces*" (1994:90). I posa el següent exemple en què s'hi observa una certa contradicció:

Palyamunurrinkutjamunurtu (llengua australiana carduchara)

*Él/ella definitivamente no se volvió malo/a*

Si l'enunciat és com una oració, per què el continuem anomenant paraula? Sembla força evident la influència de la imatge gràfica en la consideració del terme. Realment és difícil en aquest exemple o en d'altres de llengües polisintètiques decidir què és una paraula.

## 1.2. Les definicions proposades

Malgrat les dificultats, que acabo d'exposar breument, per acordar una definició precisa del terme, molts lingüistes continuen acceptant el terme paraula com una unitat lingüística i en proposen definicions; d'altres, com els estructuralistes, no accepten una unitat tan polièdrica i amb tanta història com a unitat científica i proposen d'altres unitats d'anàlisi que tinguin un referent més clarament determinat.

Ens ha semblat útil en aquest apartat fer una presentació, necessàriament breu, de definicions o explicacions sobre el terme paraula. Com que les referències són moltes i aquest no és el tema

central de l'estudi, hem seleccionat només alguns exemples, basant-nos en els criteris següents: en primer lloc, l'interès de consultar alguns diccionaris de lingüística, atesa la seva funció de divulgació; en segon lloc, la importància de les aportacions de Sapir i Bloomfield, com a autors de dos estudis generals sobre el llenguatge, ja clàssics, al costat del ja citat de Saussure; i el de J. Krámsky, estructuralista representant del Cercle Lingüístic de Praga, que hi va dedicar un estudi complet. En tercer lloc, l'aportació fonamental d'una lexicòloga, en aquest cas, de Teresa Cabré, com a representant d'un camp, el de la lexicologia, que té com objecte d'estudi la paraula.

### *Diccionaris de lingüística*

Els diccionaris de lingüística estan citats per ordre cronològic i no se n'ha transcrit l'explicació completa, que en alguns casos és molt llarga, sinó que han estat seleccionats aquells fragments que semblaven més il·lustratius:

**Jean Dubois (1979)** fa una definició ajustada a la concepció tradicional, bastant àmplia: *"Elemento lingüístico significativo compuesto por uno o varios fonemas; esta secuencia es susceptible de una transcripción escrita (ideogramática, silábica o alfabética) comprendida entre dos espacios; conserva su forma, total o parcialmente, en sus empleos sintagmáticos; la palabra denota un objeto (sustantivo) una acción o un estado (verbo) o una cualidad ..."*.

Segons l'autor, la lingüística estructural evita d'usar aquest terme per la seva falta de rigor. Martinet -diu- proposa de substituir paraula per monema i sintagma. En l'escrit, accepta el criteri que utilitza l'estadística lèxica: els segments limitats per blancs.

Els generativistes, encara que integren en part l'anàlisi estructural en matèria de lèxic, estan més interessats en la interpretació semàntica dels enunciat i donen una importància molt relativa a les unitats lèxiques que apareixen en l'actuació.

**Werner Abraham (1981)**: selecciona les aportacions d'altres lingüistes. En cito dues, la de Funkkolleg (1971), que vol ser molt funcional: *"Es una unidad desplazable en la oración, aislable en una manifestación"*



*por medio de una pausa opcional*". És curiós que el significat no hi apareix per enlloc. Una observació: potser aquesta definició no seria vàlida per a les llengües aglutinants i les polisintètiques...

L'altra definició que cita és la ja coneguda, de Bloomfield: *"Una palabra es una forma mínima libre"*.

I finalment, cita la de Reichmann, de 1969: *"La palabra es la unidad mínima de la lengua, escrita junta, fija fonológica y morfológicamente, destacable prosódicamente, que realiza una determinada función sintáctica en cuanto que es aislable de otras unidades de este tipo"*. En aquesta proposta tampoc no es fa referència al significat i s'exclou categòricament la possibilitat de considerar paraula les formes compostes.

Georges Mounin (1982): Per a aquest lingüista, la paraula és la *"unidad significativa empírica de la gramática tradicional, que no corresponde al criterio de separabilidad funcional (...) ni a los criterios de delimitación entonativa.(...)*. De hecho, en la conciencia lingüística de los no analfabetos, la palabra está vinculada con su forma escrita. La palabra no es una realidad de lingüística general". És una definició molt poc precisa i extensa, sense voluntat de ser completa. En destaquen les dues últimes idees: la força de la llengua escrita en la creació del concepte i el poc valor que té com a unitat de referència científica.

Giorgio R. Cardona (1991) defineix paraula de la següent manera: *"Palabra: término del lenguaje corriente y de la gramática tradicional usado para indicar la unidad mínima significativa de la lengua"*; i a continuació cita Boas (1911), que fa una definició basada en el significat, com la de Cardona, però tenint en compte altres aspectes: *"un grupo fonético que por su forma, claridad de significado y autonomía fonética puede ser inmediatamente separado de la totalidad de la frase"*. Aquesta proposta té punts que són realment molt discutibles. En aquest cas sí que el criteri gràfic no hi apareix, però, en canvi, queda molt poc clar el que ell anomena autonomia fonètica d'un grup, sobretot per ser separat del conjunt de la frase.

David Crystal (1994:68): El lingüista utilitza el terme paraula com la categoria clau en tractar de la morfologia i la sintaxi. No n'esbossa cap definició, simplement en parla. Admet la dificultat de parlar de paraula en el cas de llengües que no tenen escriptura i també que els sistemes escripturals de les llengües no determinen, però ajuden a trobar les unitats paraula de la parla.

Enrique Alcaraz i M. Antonia Martínez (1997). Aquest és el més recent dels diccionaris consultats i, més que una definició fa una explicació extensa, en la qual, després d'acceptar la paraula com la unitat lèxica tradicional per excel·lència, fa un repàs a la consideració d'aquest terme en els diferents moments i corrents de la lingüística.

*Sapir, Bloomfield i Krámský.*

Sapir (1985), de qui ha estat citada en pàgines anteriors una definició molt àmplia i potser poc lingüística, dedica bona part del capítol: *Els elements de la parla* a la paraula. Es mou entre l'acceptació indiscutible de la seva existència: "*la paraula és el primer element de la parla que hem trobat del qual podem dir que realment existeix*" (p.28) i intents de definicions més precises: "*el millor que podem fer és dir que la paraula és un dels trossets més petits i totalment satisfactoris de "significat" aïllat en què es resol la frase*" (p.36). O aquesta altra: "*La paraula és simplement una forma, una entitat amb una estructura precisa que inclou tant o tan poc del material conceptual de tota la idea com ho permet l'esperit de la llengua [no queda clar el referent de llengua: és una de concreta?]*". Són definicions ben poc precises.

Admet que l'estructura de les llengües pot ser un factor que dificulti saber si un element determinat d'una llengua s'ha d'interpretar com a paraula independent o no, però "*aquests casos de transició no debiliten la validesa psicològica de la paraula*" (p.35). Explica una anècdota<sup>8</sup> pròpia que reforça la idea que la paraula és un constructe

---

<sup>8</sup> L'anècdota fa referència a un indi a qui va demanar que li dictés a poc a poc el que li deia, perquè tingués temps d'escriure-ho. En aquest dictat lent, va constatar que l'indi s'aturava respectant els límits entre paraules.

no dependent de la llengua escrita, contràriament al que sostenen altres lingüistes<sup>9</sup>, com ja hem explicat més amunt.

Per a Bloomfield (1977), "*el mot és una forma lliure (per oposició a travada) mínima de la llengua*" (p.197). En la vida ordinària, el mot és la unitat mínima de la parla. Els nostres diccionaris donen llistes dels mots d'una llengua; per a tots els fins que no són l'estudi lingüístic sistemàtic, no hi ha dubte que aquest procediment és més útil que no ho fóra una llista de morfemes. L'anàlisi de les formes lingüístiques en mots ens és familiar, perquè escrivint i imprimint tenim el costum de deixar espais entre els mots. Les persones que no han après de llegir i d'escriure troben algunes dificultats quan, casualment, se'ls demana que separin en mots. (...) El fet d'espaiar mots ha passat a formar part de la nostra tradició d'escriptura, de tota manera, tendeix a mostrar que als parlants no els és antinatural d'identificar el mot com una unitat de la parla; i de fet excepte en alguns casos dubtosos, la gent aprèn fàcilment a fer aquesta anàlisi".

Jiří Krámský (1969), citat unes ratlles més amunt, precisa que l'objectiu d'un estudi dedicat a la paraula deriva de l'interès per analitzar la frase, formada de morfemes, que són les unitats de significat. Però, quan s'analitzen els enunciats de la parla, el que s'hi troba -afirma- no són morfemes, sinó paraules, que són les unitats naturals i indiscutibles de la llengua. Per a l'autor, totes les llengües tenen paraules, perquè és un universal: "*Our definition of the word presupposes the existence of the word as a universal fact*" (p. 74). Afirmació no compartida per tothom, com ell mateix indica, citant un altre lingüista col·lega, estudiós de llengües americanes -Tadeusz Milewski- que sosté que en moltes d'aquestes llengües el que compta són morfemes, grups sintàctics i oracions, però no paraules.

Exposa també els criteris per definir què és una paraula, alguns dels quals coincideixen amb els que formula Reichler-Béguelin (1992). Es proposa, també, l'objectiu d'enunciar una definició prou àmplia

---

<sup>9</sup> Per exemple, Goody, citat per Blanche-Benveniste (1998). De fet, Goody no nega la intuïció de paraula en les societats àgrafes; el que diu és que només tenen el terme **paraula**, les societats que posseeixen un sistema d'escriptura.

perquè convingui a una majoria de llengües, independentment de la seva estructura. I la definició que elabora és la següent:

The word is the smallest independent unit of language referring to a certain extralinguistic reality or to a relation of such realities and characterized by certain formal features (acoustic, morphemic) either actually (as an independent component of the context) or potentially (as a unit of the lexical plan) (p. 67).

*Teresa Cabré*

Al llibre A l'entorn de la paraula (1994:33), l'autora atribueix la dificultat de l'estudi del lèxic a la complexitat i polivalència de les paraules. Fa una presentació àmplia del concepte perquè parteix d'una visió no simplement lingüística, sinó que destaca la paraula com una unitat de referència, entre la realitat i el concepte; també com una unitat de comunicació, que identifica culturalment i socialment els parlants. I, finalment, la considera una unitat del component lèxic de la gramàtica que constitueix una entrada de diccionari. És una unitat de forma, de funció i de significat, que està relacionada també amb els altres components de la gramàtica: el fonològic, el morfològic, el sintàctic i el semàntic. En aquest sentit, admet, com Krámský i Crystal, la pertinença de la paraula a més d'un subsistema o nivell gramatical.

Potser el fet que la paraula manifesti aquesta polivalència és el que explica la diversitat de criteris en què es basen els intents de definició. Sobre la qüestió dels criteris, Reichler-Béguelin (1992) proposa l'agrupació que sembla més clara i completa. Segons aquesta autora, les definicions responen a l'aplicació dels criteris següents: entonatiu, sintàctic, flexiu, semàntico-referencial i gràfic. És interessant la inclusió de l'últim criteri, el gràfic, que no apareix en d'altres classificacions anteriors, per exemple, la de Jirí Krámský, citat unes ratlles més amunt. Aquest fet posa de manifest el que molts lingüistes i estudiosos d'altres camps revelen: la influència de l'escriptura en el coneixement gramatical -en aquest cas, en l'establiment de la

categoria de paraula<sup>10</sup>. Teberosky recull en aquesta cita la idea que expressem:

(...) Però també podríem parlar d'efecte de *modelització* en el cas dels parlants que han estat alfabetitzats. Modelització de l'escriptura sobre alguns aspectes de segmentació de la cadena sonora i sobre la reflexió gramatical dels aprenents. Per als usuaris de la llengua escrita el coneixement de les unitats lingüístiques és, en bona part, determinada per la modelització rudimentària que en permet l'escriptura. (Teberosky, 1994:140)

També en l'exemple de Crystal citat i comentat unes ratlles més amunt, s'entreveu que és el criteri gràfic el que ha primat a l'hora d'usar el terme paraula, malgrat la complexitat del segment gràfic de la llengua carduchara pres com a exemple. Un altre exemple de la influència de l'escrit en la consideració de paraula ens el forneix Blanche-Benveniste (1998:78) quan reproduïx una cita de Cornillac: "*La palabra esquimal [com ja hem dit abans, llengua polisintètica] se pronuncia, en efecto, con un solo golpe de aire, cualquiera que sea su longitud*" i posa, com a exemple, un enunciat complex que equival a una oració.

Les paraules, com afirmen Tolchinsky i Cintas (2001), tenen en comú amb molts altres constructes que escapen a criteris de definició clars.

M. Fruyt fa una revisió i síntesi de les diferents propostes que els lingüistes estructuralistes fan per substituir el terme paraula<sup>11</sup>, i el resultat mostra que, tampoc des del camp científic hi ha una unanimitat a triar els elements que han de ser objecte d'estudi i a acordar uns mateixos termes per a designar-los. Així, Martinet proposa el terme *sintagma* per salvar el terme paraula: "on désigne sous ce terme tout groupe de plusieurs signes minima" (1966:51); i en aquest sentit *un énorme rocher* és un sintagma igual que *donn-er-ons*<sup>12</sup>. El

<sup>10</sup> D'altres autors que comparteixen aquesta idea: Catach, 1990; Fruyt, 1989; Reichler-Béguelin, 1990. Totes tres autores, citades a Teberosky (1994).

<sup>11</sup> Vegeu M. Fruyt (1992)

<sup>12</sup> "Se da el nombre de sintagma a toda combinación de monemas [en oposició a *monema autónomo*]. (...) Un sintagma autónomo es la combinación de dos o más monemas cuya función no depende de su posición en el enunciado (...). Un sintagma autónomo formado por monemas no separables es lo que se llama generalmente palabra (Martinet, 1974:143).

lingüista francès utilitza els termes de monema, amb els dos tipus de *monema lèxic* i *monema gramatical* o *gramema*, amb el significat que d'altres donen a *morfema lèxic* o *lexema* i *morfema gramatical* (Lyons, (1963:11-12, citat a Fruyt, 1992:115), que són els termes actualment més estesos. Pottier (1973:109-118) parla de *lexia*<sup>13</sup>, i en distingeix de tres menes: la *lexia simple* - com nit o lluna-, la *composta* -com extraordinari o pica-soques-, i la *complexa* -com tauleta de nit- i, en tots tres casos els dóna el valor d'unitat de designació; i el problema rau, aleshores, en les paraules gramaticals o funció, que no cabrien en la definició proposada. És a dir, només citant alguns autors, ja es fa evident les discrepàncies a l'hora de crear altres termes que substituïssin la noció vague, poc precisa, però a la vegada acceptada, de paraula.

Es pot sintetitzar el que acabem de dir en aquest apartat sobre el concepte de paraula i les dificultats per elaborar-ne una definició, en els següents punts:

- ▶ és un terme que pertany al llenguatge habitual, no específic, cosa que el fa poc útil i poc precís per donar-li un ús científic i utilitzar-la com a unitat lingüística d'anàlisi,
- ▶ és una unitat que depèn de l'estructura lingüística de les llengües, segons si son aïllants, flexives, aglutinants o polisintètiques,
- ▶ en el cas de les flexives, que es el grup al qual pertany el català, els problemes solen concentrar-se en les paraules funció i en les expressions compostes,
- ▶ és una unitat que està molt influïda per l'escriptura i per l'existència mateixa de la cultura escrita. Sembla, com alguns antropòlegs afirmen, que les cultures àgrafes no tenen aquest terme en el seu lèxic. Seria una creació artificial de les societats alfabetitzades.

---

<sup>13</sup> El terme *lexia* és usat, també, per Martinet (1968:288-289) per referir-se a la unitat lèxica, a la vegada que associa *lexema* a *monema lèxic*, terme que prefereix al de *semantema*.

La conclusió és la mateixa del punt de partida: no hi ha una definició teòrica clara de la noció de paraula i la majoria de lingüistes no l'accepten com una categoria gramatical que permeti estudiar la llengua. Malgrat tot, s'accepta aquesta unitat com una realitat psicològica, com una unitat natural de la parla i com una unitat cultural per als individus alfabetitzats de les societats que posseeixen un sistema d'escriptura. És més, encara que la noció de paraula sigui tan difícil d'establir i no tingui un referent clar i únic, es continua utilitzant com una categoria en l'ensenyament gramatical de les llengües, així com per parlar dels problemes de segmentació dels textos.

### 1.3. La noció de paraula gràfica

Com a referència dels segments gràfics delimitats per blancs, que són la base de l'estudi que presento, s'utilitza el terme *paraula gràfica*. Es podria optar per continuar utilitzant segment gràfic, que també apareix en la literatura sobre el tema, però ha semblat més adequat el terme finalment escollit, per la raó que és també el terme emprat en la majoria dels estudis sobre segmentació consultats (Reichler-Béguelin, 1992; Ferreiro et alii, 1996; Báez y Cárdenas, 1999; Tolchinsky y Cintas, 2001), encara que coexisteix amb el de paraula sense cap tipus de qualificació (Ferreiro y Teberosky, 1979; Ferreiro y Pontecorvo, 1993). El fet d'utilitzar aquest terme indica l'assumpció de determinats supòsits teòrics i a la vegada, també, alguns problemes conceptuals que es presenten a continuació i s'explicaran al llarg d'aquest marc teòric.

La paraula gràfica com a unitat de segmentació és la base dels estudis sobre la segmentació del text escrit i amb l'aplicació d'aquest criteri les unitats gràfiques són fàcils de delimitar. Malgrat aquesta facilitat, persisteixen algunes qüestions per resoldre. En primer lloc, perquè és possible que l'estructura lingüística de cada llengua influeixi en la determinació d'aquesta unitat lingüística i també en la seva presentació gràfica i, en aquests casos, seria abusiu i poc precís pronunciar-se sobre la noció de paraula i de paraula gràfica de forma general, ignorant que l'estructura de les llengües i el seu sistema de representació gràfica és divers i pot condicionar tant la construcció de

la noció de paraula com la seva manifestació escrita. Estem molt habituats a pensar en l'estructura de les llengües des de la perspectiva d'aquelles que ens són familiars, però hi ha diferències molt marcades entre l'estructura de les llengües aïllants com el vietnamita, sense morfologia; les flexives com el català i les altres llengües romàniques; les aglutinants com el turc o les polisintètiques com l'inuït. I també hi ha diferències segons el sistema d'escriptura escollit per cada llengua, de base alfabètica o logogràfica. Com s'entén, la paraula gràfica, en les llengües segons l'estructura gramatical i segons el sistema d'escriptura?

A banda de les qüestions que acabem de plantejar, derivades de la peculiar estructura de les llengües, l'aplicació del criteri gràfic també presenta algunes incongruències en un mateix tipus de llengua; per exemple, en llengües flexives romàniques molt pròximes trobem: *deprisa/ de presa; sobre todo/sobretot; a la playa/alla piagge/ a la platja*. I en una mateixa llengua, també: *antes de ayer, anteayer; Rhône-Alpes/ rhônalpin; comióselo/ se lo comió; se'l va menjar/ va menjar-se'l*.

Tot i l'existència de desajustaments, el criteri de paraula gràfica s'ha imposat a l'hora de determinar les paraules d'una llengua en tasques de comptatge o d'anàlisi de la segmentació. I no solament s'ha imposat en el tractament de les unitats de l'escrit, sinó que, atesa la influència de la llengua escrita en les societats alfabetitzades actuals, la paraula gràfica ha esdevingut sinònim de paraula i així, d'una forma intuïtiva i potser poc precisa, solem referir-nos a les qüestions de segmentació del text escrit com aquelles que involucren l'escriptura de paraules: els nens encara ajunten paraules o separen malament les paraules, etc. són expressions freqüents entre els mestres.

Hem començat presentant les dificultats teòriques de la utilització del terme paraula, perquè ens semblava un pas imprescindible en un estudi sobre la segmentació de l'escrit que utilitza com a terme de referència per denominar les unitats gràfiques entre blancs, justament el de paraula.



El següent objectiu serà el d'intentar esbrinar el grau d'arbitrarietat en l'establiment dels blancs per delimitar les unitats gràfiques. Són, les paraules gràfiques, segments establerts de forma arbitrària? Tenen alguna relació amb unitats de la parla?

## 2. LES UNITATS DE L'ORAL I LES DE L'ESCRIT

L'interès per la relació entre les unitats de l'oral i de l'escrit sorgeix -tal com hem plantejat a la introducció- de la necessitat de saber fins a quin punt les paraules gràfiques que estan clarament determinades a l'escrit tenen alguna relació amb unitats de la parla o, ben al contrari, són segments gràfics establerts arbitràriament. En el cas d'una resposta afirmativa sobre la motivació de les unitats gràfiques ens podríem plantejar si la competència adquirida en aprendre a parlar té alguna influència en l'aprenentatge de l'escriptura. De fet, que la relació no fos del tot arbitrària explicaria la rapidesa amb què els infants aprenen a posar blancs entre les paraules, és a dir, la rapidesa amb què aprenen un aspecte de l'ortografia amb relació amb la dificultat i lentitud amb què n'aprenen d'altres.

Si acceptem la no arbitrariedad de la paraula gràfica estem abocats a trobar la unitat amb què aquesta paraula gràfica es relaciona i és fàcilment pensable que la unitat representada en el text escrit correspongui a la paraula de la llengua parlada.

Existeixen paraules en la llengua parlada? És clar! dirà tothom, almenys tothom qui no té res a veure amb la lingüística, ni amb la gramàtica... Parlem amb paraules. I fins i tot ens diran que ens aturem a cada paraula, que paraula és allà on fem una pausa. Així s'afirma a la Gramàtica de Port-Royal, en una cita molt coneguda, recollida per Blanche-Benveniste: "*Se llama palabra aquello que se pronuncia separado y se escribe separado*" (Blanche-Benveniste, 1998:71).

Aquesta acceptació tan contundent de la paraula com una unitat clarament distingible en la llengua parlada topa amb la realitat quan se'n fa una anàlisi una mica més aprofundida. Una observació menys apassionada de la parla de les persones fa adonar de la inconsistència d'aquesta acceptació. És molt evident que, fins i tot sense tenir cap aparell ni instrument especial, només amb una observació acurada, qualsevol persona pot adonar-se que el flux de la parla és un continuum en el qual percebem les paraules perquè les hi volem trobar i ja les coneixem d'abans. Si, a més d'una simple observació

desapassionada, utilitzem una gravadora que ens permeti reproduir la parla pròpia, o millor encara, un utillatge més sofisticat, com ara un espectrògraf, l'evidència serà irrefutable. Fins i tot podrem comprovar que no solament les pauses que els parlants fan no coincideixen amb els límits de les paraules, sinó que ni els sons són el que semblen, tenyits i modificats pel contacte amb els sons veïns, en la cadena parlada.

La parla es produeix en un flux continu en el qual es fa difícil de trobar-hi unitats. Aquesta és una idea universalment compartida pels estudiosos de la llengua, d'àmbits diversos, i que apareix reiteradament en els documents consultats (Flores d'Arcais, 1992:125; B. Roberts, 1992:125; Blanche-Benveniste, 1998:67). La dificultat per distingir les unitats paraula -i les altres: fonemes, sintagmes, frases- ha fet en alguns moments afirmar, a alguns lingüistes, que, en realitat aquestes unitats eren fruit de la influència de la llengua escrita, tal com hem explicat a l'apartat anterior. Des d'aquest punt de vista, la paraula seria un "*simple artefacto debido a la escritura*", com diu Beck, citat per Blanche-Benveniste (1998:67). Aquest plantejament ens remet a un tema que en alguns moments de la història de la lingüística ha estat objecte d'atenció: el de l'estatus de la llengua oral i la relació d'ambues manifestacions, l'oral i l'escrita.

Els estudis sobre la relació entre la paraula de la parla i la de l'escrit constitueixen només una part d'aquest conjunt més ampli d'estudis, investigacions i debats sobre la relació entre oralitat i escriptura. Tot i que aquest és un camp ampli i que té un interès tangencial per a un treball de segmentació de l'escrit, és cert que les nombroses aportacions des de disciplines diverses han estat una font interessant d'informacions útils per a l'estudi realitzat.

Entre aquest conjunt d'estudis s'hi troben els treballs de base antropològica i etnològica, moltes vegades historicista, que indaguen en l'aparició i la funció de l'escrit en les diferents societats. Les aportacions de G.R. Cardona, J. Goody o D. Olson en són exemples. Temes com la relació entre cultura escrita i desenvolupament intel.lectual i tecnològic de les societats, relació entre cultura escrita i desenvolupament cognitiu han estat d'interès per a aquest camps

científics. La lingüística i la psicolingüística s'han ocupat d'altres aspectes, com el de la relació entre discurs oral i discurs escrit, els usos d'una i altra manifestació, el predomini de l'escrit o de l'oral en la descripció lingüística segons les èpoques i tendències, i les diferències en el procés d'aprenentatge.

L'interès per la relació entre les dues manifestacions del llenguatge és relativament recent. A principis del segle XX es produeix un gir important en els estudis de lingüística. Si el segle XIX s'havia caracteritzat per un interès envers la filiació de les llengües -de la mà de l'establiment de les colònies per part de moltes nacions europees-, els inicis del XX estableixen les bases de la moderna lingüística, que es desvincula dels aspectes històrics i dels estrictament normatius per interessar-se pel fet lingüístic en ell mateix, pel llenguatge. El coneixement d'altres realitats fora de l'òrbita europea, de pobles que parlen llengües desconegudes fins al moment, sense representació escrita, desvetlla l'interès de lingüistes i antropòlegs. La lingüística de la primera meitat del segle XX reivindica el valor de la llengua oral com la manifestació lingüística rellevant, la que ha de ser la base de l'estudi lingüístic, i considera l'escrit com un codi de segon ordre, secundari, subsidiari, que té com a finalitat la transcripció de la parla. És el moviment estructuralista qui situa la llengua oral com la manifestació primigènia del llenguatge i redueix, d'aquesta manera, l'escriptura a la funció de representació o, encara més, de simple traducció de l'oral. Saussure<sup>14</sup> manifesta:

"Langue et écriture sont deux systèmes de signes distincts; l'unique raison d'être du second est de représenter le premier; l'object linguistique n'est pas défini par la combinaison du mot écrit et du mot parlé; ce dernier constitue à lui seul cet objet. Mais le mot écrit se mêle si intimement au mot parlé dont il est l'image, qu'il fini par usurper le rôle principale" (1916:45).

---

<sup>14</sup> La utilització del terme *llengua* com a sinònim de llengua parlada es presta a una certa confusió amb el terme *llengua* com a sistema. Al llarg del tercer capítol, Saussure deixa ben clara la distinció entre llengua i parla. Però al capítol quart, d'on és la cita d'aquesta pàgina, utilitza moltes vegades el terme *llengua* com a sinònim de *parla*.

En un altre passatge del seu llibre, Saussure afirma que l'escriptura pot expressar idees directament, sense passar per l'oralització, tal com cita Tolchinky (1991:46). Aquesta interpretació hauria deixat la porta oberta a la posició desenvolupada, més endavant, per l'escola glossemàtica, que defensa la relació directa entre llengua (com a sistema) i llengua escrita.

Sapir es refereix a l'escriptura com el més important dels simbolismes visuals de la parla -entre els quals parla del llenguatge de sords i el del sistema Morse- i explica que la característica més significant és que:

"cada element (lletra o paraula escrita del sistema correspon a un element específic (so o grup de sons o paraula parlada) del sistema primari. El llenguatge escrit és, doncs, una equivalència simètrica (...) de la seva rèplica parlada. Les formes escrites són símbols secundaris dels parlats -símbols de símbols (...)" (1985:23).

Bloomfield afirma que "*l'escriptura no és el llenguatge, sinó una manera d'enregistrar el llenguatge*" (1978:39). En aquesta afirmació tan categòrica, sense els matisos i precisions de Saussure, no queda gens clar, amb la utilització de la paraula *llenguatge*, en la traducció catalana, si l'autor fa una distinció entre llengua i parla, en el sentit saussurià, o no.

En tots els casos, però, queda clara la idea de representació, de secundarietat de la llengua escrita respecte de la parlada<sup>15</sup>. Sembla, a més, que la comparació entre la llengua parlada i l'escrita es limita a la paraula i, per tant, es parla de la transcripció no tant de la parla (el discurs oral), com de la paraula i les unitats fòniques que la formen. És una perspectiva estrictament gramatical, entenent aquest terme en el sentit de gramàtica de les categories. Però no solament és una qüestió de perspectiva, sinó que, retornant al que dèiem en començar aquest paràgraf, la paraula de la parla a la qual es refereixen els lingüistes és més una il·lusió que una realitat. Potser hauríem de considerar que la unitat a la qual realment es fa referència és la paraula del sistema, atès que la paraula oral, entesa com una unitat aïllada, és infreqüent.

---

<sup>15</sup> Scholles & Willis (1995:285) registren aquest desinterès dels lingüistes de la primera part del segle XX, per l'escriptura.

Blanche-Benveniste<sup>16</sup> es proposa superar la dificultat i trobar una explicació que permeti combinar dos fets: l'existència de la paraula i el fet que aquesta unitat realment no es percep en la parla. La seva proposta<sup>17</sup> es basa en la distinció entre dos nivells de la llengua: el nivell de l'enunciat i el nivell del sistema. La paraula, en la parla, apareix com un "*fragmento de enunciado*" i aïllar-la demana "*ciertas operaciones de 'segmentación' bastante complejas*" (1998:67). La paraula del sistema de la llengua és com una unitat abstracta que representa totes les formes possibles en què es pot presentar segons els contextos. Aquesta distinció és aclaridora d'un dels aspectes que ens preocupava, però no d'un altre. Si bé ofereix una via de solució per al problema de la distinció de les paraules en la parla, no resol la qüestió de l'equivalència entre aquestes i les de l'escrit. És a dir, seria més fàcil pensar que les unitats de l'escrit es corresponen amb les unitats del sistema, però això no és el que diu Blanche-Benveniste. Les unitats del sistema, tal com l'autora les entén, són una espècies de *paraules abstractes o paraules potencials*:

Este nivel [es refereix al nivell del sistema de la llengua] correspondería a las "palabras separadas", tal como se las encuentra en las listas de repertorios, palabras fuera de uso, y potencialmente disponibles para diversos empleos. Son *palabras abstractas*, técnicamente llamadas *lemas*, que pueden subsumir en una sola forma diversas realizaciones morfológicas." (...) Esta palabra potencial representa todo su *paradigma* de formas " (Blanche-Benveniste, 1998:68).

Queda pendent, per tant, a quines unitats representen les paraules gràfiques. Ja sabem que no a les de la parla, però tampoc no és clar que sigui una representació de les del sistema en tots els casos.

Ens trobem en una espècie de paradoxa: en un sentit, les unitats de l'oral -entre les quals la paraula, objecte del nostre estudi- no es corresponen amb les unitats de l'escrit, en l'altre, no es pot negar que en el període de desenvolupament del llenguatge, els infants seleccionen les mateixes unitats que els adults, encara que no es

---

<sup>16</sup> És autora d'un treball, molt conegut: *La escritura del lenguaje dominguero*, a Ferreiro y Gómez Palacio (1982), en què aborda la relació entre el discurs oral i l'escrit.

<sup>17</sup> Publicada en el volum de Gedisa, va ser formulada en una conferència i recollida a: ESF, Network on Written Language and Literacy, Proceedings, Siena, 1992.

presentin de forma aïllada en l'*input* que reben. Quines són les unitats que els infants seleccionen en el corrent continu de la parla? Com s'ho fan els nens per trobar aquestes unitats?

En l'apartat següent abordarem la qüestió de la percepció de la paraula, fent referència a la percepció d'altres unitats del llenguatge. La paraula és una de les unitats privilegiades, tal com explicarem a continuació.

### 3. LA SEGMENTACIÓ DE LA PARLA I LA CONSTRUCCIÓ DEL CONCEPTE DE PARAULA

Els infants, en el desenvolupament del llenguatge i per a l'adquisició de la llengua o llengües de l'entorn, han de resoldre diversos problemes: en primer lloc, el de la segmentació del senyal acústic i la percepció de les unitats lingüístiques significatives; en segon lloc, han de poder analitzar el món en objectes i fets que puguin ser codificats lingüísticament i, finalment, han de fer correspondre les unitats lingüístiques percebudes amb aquests objectes seleccionats del món real. La descoberta de les unitats lingüístiques com a formes de representació de la realitat no substitueixen les primeres representacions que els infants ja han construït sobre els objectes, els fets i les relacions, a partir de l'experiència, sinó que es conserven totes dues (Clark, 2004).

Per a l'estudi que presentem, ens centrarem en el primer dels aspectes citats: la descoberta de les unitats lingüístiques de la parla, amb la finalitat de poder considerar amb posterioritat si es pot establir algun tipus de relació entre el procés que es produeix en la parla i el que es produeix en l'escrit. El coneixement de com els infants descobreixen aquestes unitats és un tema complex que ha estat abordat pels psicòlegs, els psicolingüistes i els especialistes en fonètica acústica, amb intensitat, des de la segona meitat del segle XX, principalment. En els dos últims decennis, en els camps més específics de la neurociència i de la ciència cognitiva, s'han realitzat nombrosos treballs que han contribuït a un coneixement més profund dels processos que permeten l'eclosió de la parla.

En la descoberta de les unitats hi intervenen dos factors fonamentals: d'una banda la percepció de l'entrada lingüística i la naturalesa mateixa d'aquesta entrada, i de l'altra la posada en funcionament del programa biològic de la parla que aquesta entrada provoca. El concepte de percepció no és, però, interpretat per tothom en el mateix sentit. En aquest treball, ens hem basat en la interpretació de Flores d'Arcais (1992), el qual explica amb deteniment les diferents accepcions del



terme percepció, depenent de les disciplines que se n'ocupen. Segons aquest autor, "*percibir el lenguaje significa llevar a cabo varias operaciones psicológicas, tales como aislar y segmentar palabras, frases y unidades más largas, además de atribuirles un significado*" (1992:123). La percepció del llenguatge és, vist així, el procés psicològic que permet segmentar, interpretar, donar una representació mental a una seqüència de la llengua parlada o a un fragment del text escrit.

A parer seu, la percepció del llenguatge es sinònim de reconeixement però també de comprensió i admet que és difícil establir els límits entre els dos conceptes: on acaba la percepció i comença la comprensió? es pregunta. Proposa, des de l'àmbit de la psicolingüística, la següent definició per al terme percepció.

"Consideramos el estudio de la percepción del lenguaje como aquella esfera de la psicolingüística que intenta explicar los procesos mediante los cuales segmentamos, interpretamos y damos una representación mental a una secuencia de la lengua hablada o a un fragmento de un texto escrito" (Flores d'Arcais, 1992:124)

Aquesta definició és interessant perquè presenta un punt de vista integrador entre els processos físics i els mentals; entre la segmentació i l'elaboració d'una representació mental de les unitats que s'emmagatzemen en memòria i que han d'estar disponibles en els processos de comprensió i de producció de la parla. La qüestió fonamental és veure quines són les unitats lingüístiques que són percebudes per l'infant, quins són els indicis que en guien la percepció, com s'emmagatzemen i com es recuperen per a les tasques de comprensió i producció.

Tanmateix, però, la percepció, encara que en la definició que acceptem tingui un sentit ampli que superi l'estrictament físic, no explica per ella sola com els infants s'apropien de la paraula. Encara que sabem que el desenvolupament del llenguatge comença des del moment del naixement, l'atenció dels psicòlegs i psicolingüistes s'adreça a les primeres manifestacions que poden ser reconegudes com a formes lingüístiques de la llengua de l'entorn. I aquestes primeres manifestacions, després de les etapes descrites del balboteig i

l'ecolàlia, consisteixen en produccions d'una única paraula; és el que coneixem com a període de l'holofrase. Sembla, per tant, que la paraula s'erigeix, d'entre totes les altres, en una unitat privilegiada a l'hora d'analitzar l'adquisició de la llengua. No ens estendrem en l'anàlisi de les produccions i les fases i característiques del desenvolupament perquè aquest no és un estudi sobre el desenvolupament de la parla. Però sí que es poden establir paral·lelismes o relacions entre els processos de la parla i els de l'escriptura; i és per això que hi fem referència.

L'interès en la percepció de la paraula, més que no pas cap altra de les unitats lingüístiques, deriva del tema del treball, però també d'altres factors; és cert que de totes les unitats: el fonema, la síl·laba, el sintagma o la frase, la paraula és la unitat privilegiada, tal com s'ha dit unes ratlles més amunt. La raó d'aquest privilegi és la importància que té en l'aprenentatge i en el coneixement d'una llengua. Diu Mehler (1992:189): "*Para aprender una lengua hay que reconocer las palabras*" i Flores d'Arcais, després d'analitzar les possibilitats de percepció de les altres unitats esmentades, diu: "*Las palabras son, a primera vista, los candidatos más probables para ser unidades estructurales y funcionales en la percepción*" (1992:127). Pel que fa a la importància del coneixement de les diferents unitats lingüístiques amb relació a la competència dels parlants, el coneixement de les paraules és considerant com un coneixement rellevant. E. Clark (1993:1), de forma contundent, afirma, just al començar el llibre que dedica a l'adquisició del lèxic: "*Words make a language*". Resumint, les paraules serien elements rellevants tant en la percepció, com en la fase d'aprenentatge o com elements de la competència lingüística d'un parlant adult, independentment que, durant l'aprenentatge de la llengua, hi hagi d'altres unitats que siguin també percebudes.

### 3.1. La segmentació de la parla i la percepció de les unitats

En el camp de les investigacions en ciència cognitiva, el desenvolupament de tècniques noves per a l'anàlisi del comportament i el desenvolupament infantil, ha propiciat tenir molta més informació sobre els processos d'adquisició del llenguatge des del naixement, molt

abans doncs que aquest desenvolupament es faci patent en la producció de paraules reconeixibles.

Així, existeix actualment un corpus d'estudis<sup>18</sup> important sobre les característiques del desenvolupament lingüístic. Estudis que demostren que els infants posseeixen, en néixer, capacitats acústiques desenvolupades i capacitats innates per reconèixer els sons propis de l'espècie, per a la distinció soroll / llenguatge, per al reconeixement de les veus conegudes, per a la distinció de dues llengües diferents presents en l'entorn ja des dels quatre dies de néixer<sup>19</sup>; els infants mostren preferència en la discriminació de la síl.laba sobre el fonema, com la unitat acústicament rellevant de la parla; aquesta discriminació seria una de les estratègies usades per trobar la paraula dins el corrent fònic.

Se sap que els bebès són sensibles, també, a l'entonació i prefereixen el llenguatge ben estructurat i coherent que el que no ho és; així, reaccionen de manera diferent segons si se'ls parla amb una entonació marcada i aguda -característica del llenguatge maternès (*motheress o baby talk*)- que si es fa amb una expressió molt monòtona; es neguitegen si, enmig d'una explicació s'introdueixen elements estranys en la parla, que distorsionen el significat i violenten l'estructura sintàctica.

Abans de saber parlar són sensibles als contorns entonatius que delimiten sintagmes i frases (Hirsh-Pasek i Golinkoff, 1996) i a les fronteres entre paraules (Gleitman i al. 1992) i als quatre mesos són capaços de distingir les fronteres oracionals de dues llengües diferents (Jusczyk, 1989, a Karmiloff-Smith, 1994:58).

---

<sup>18</sup> Incloem una relació d'aquests estudis, sense estendre'ns en l'explicació, atès que aquest és un treball basat en la segmentació de l'escrit i tota referència al desenvolupament oral té una importància molt complementària. Moltes de les referències han estat tretes de Mehler i Dupoux (1992) i de Karmiloff-Smith (1994); d'altres han estat consultades directament de les fonts, tal com constarà a les referències en cada cas.

<sup>19</sup> Laura Bosch & N. Sebastián-Gallés (1997), Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments, *Cognition*, 65(1).

Totes aquestes investigacions, que mostren les capacitats dels infants per a desenvolupar el llenguatge des del naixement, se situen en la perspectiva de considerar que naixem equipats biològicament per parlar i pressuposen un infant amb unes predisposicions lingüístiques innates i unes inclinacions atencionals també innatament especificades (Karmiloff-Smith, 1994). Proporcionen informacions molt precises sobre el procés de desenvolupament i sobre l'accés a les unitats lingüístiques, abans que es puguin produir formes lingüístiques reconeixibles pertanyents a la o les llengües de l'entorn.

L'altra qüestió que ens plantejàvem era saber en quins indicis es basen, els infants, per accedir a les unitats de la llengua. En aquest punt concret, de la bibliografia sobre el tema es desprèn que caldria tenir en compte la intervenció de factors diversos. Els resultats d'algunes investigacions apunten a l'atenció que els bebès fan a les característiques acústicoprosòdiques de les paraules i a l'entonació d'unitats superiors a la paraula. Segons Mehler, els infants utilitzen dades acústiques i es basen en propietats complexes i dinàmiques del senyal sonor, com l'entonació. Aquesta perspectiva explicaria que la síl.laba accentuada és la primera que es percep i la primera que es diu com a expressió de la paraula sencera<sup>20</sup> (Gleitman et al., 1992); i explicaria, també, que la detecció de la síl.laba tònica doni indicatiu de les paraules en el corrent fònic de la parla. Però no donaria indicació d'on acaba i comença la paraula. Gleitman considera que l'establiment dels límits és una tasca no gens fàcil i es pregunta com s'ho fan els nens per quedar-se amb la paraula i no amb mitges paraules. De fet, la diferent influència de la síl.laba accentuada sobre l'àtona en la percepció és un factor que té la seva importància en d'altres aspectes de l'aprenentatge lingüístic; per exemple, permet explicar el diferent tractament que reben les paraules de la classe tancada -la majoria àtones- en el desenvolupament de la parla, en comparació a les de la classe oberta; però també en l'aprenentatge de l'escrit, pel fet que són aquestes unitats les més susceptibles de no ser segmentades i d'aparèixer unides a la paraula de classe oberta. De la mateixa manera

---

<sup>20</sup> Els exemples citats a Gleitman et al. (1992): [raf] per [giraffe], [e-fant] per [elephant] són recollits, també per Teberosky (1994:137).

que, com mostren els estudis de Ferreiro i Teberosky, en el procés d'adquisició de l'escriptura, la síl.laba apareix com la unitat perceptible en la primera anàlisi que els aprenents fan i és representada la majoria de vegades pel nucli sil.làbic: la vocal, encara que no sempre.

L'accent, però, no ho explica tot pel que fa a l'habilitat que manifesten els nens per establir els límits entre unitats tal com ho fan els parlants adults. Hi ha estudis que mostren els errors que cometem els nens quan en una tasca de segmentació basada en picar per cada mot que senten (*word tapping task*), s'equivoquen davant de paraules compostes que tenen més d'un accent (Ehri, 1975, citat a Gombert, 1990:90). Segurament s'han de tenir en compte moltes activitats que aquests adults fan amb els infants, entre les quals podríem esmentar l'ús d'expressions en què, prosòdicament, l'adult destaca algunes paraules que són clau per a la comprensió del missatge. Les activitats d'etiquetatge aporten, també, informació de l'estructura i els límits de les paraules.

En síntesi, es pot dir que:

- ▶ la sensibilitat que mostren els infants, els primers sis mesos de vida, per a la distinció fonemàtica i per a la detecció de la síl.laba tònica en paraules que segueixen el patró de la llengua a la qual estan exposats són habilitats que li permeten descobrir la unitat paraula en el corrent continu de la parla,
- ▶ la sensibilitat envers la prosòdia corresponent als límits sintagmàtics, els permet descobrir segments de la parla superiors a la paraula, com ara els sintagmes, el nucli dels quals és una o més paraules de la classe oberta, acompanyades d'altres elements de la classe tancada, àtones i monosil.lábiques la majoria, que actuen com a elements dependents. D'aquesta manera els infants es familiaritzen amb l'organització sintàctica del llenguatge, que no és lineal sinó clarament jerarquitzada,
- ▶ de totes les unitats lingüístiques envers les quals el bebè és sensible, només són les paraules les que apareixen en les primeres manifestacions lingüístiques que fa el nen a partir de

l'any. Totes les capacitats que ha manifestat, i que es posen de relleu en les investigacions fetes, han servit perquè pogués discriminar, reconèixer i aïllar la paraula. És per això que dèiem, al començament d'aquest apartat, que la paraula és la unitat lingüística privilegiada.

Hem revisat de forma sumària com els infants desenvolupen el llenguatge i es fan amb la llengua de l'entorn i hem destacat sobretot la importància de la paraula com la unitat troncal del sistema lingüístic. El coneixement que sobre la paraula tenen, en aquest moment del desenvolupament, anirà evolucionant i tornant-se més abstracte a mesura que s'avanci en el coneixement de la llengua com un sistema de signes.

Explicarem, en el següent apartat, la naturalesa i l'evolució d'aquest coneixement, a la llum dels resultats de les moltes investigacions fetes. Generalment, en la literatura sobre aquest tema, les investigacions sobre el concepte de paraula es basen en la interacció verbal entre l'experimentador i els subjectes. Per mitjà del diàleg, de preguntes o de resolució de tasques, es busca saber el grau de coneixement que l'infant construeix sobre el concepte de paraula. En canvi, tot i que es reconeix i s'esmenta sovint la influència del procés d'alfabetització en la construcció d'aquest coneixement -sobretot pel que fa a la lectura-<sup>21</sup>, els estudis sobre la segmentació de l'escrit no són analitzats, generalment, des d'aquesta perspectiva. Els errors en la segmentació de l'escrit i la progressió cap a l'escriptura convencional es posen poques vegades en relació amb el coneixement de la paraula i no s'analitzen com una mostra de la construcció d'aquest coneixement, sinó des del punt de vista de l'adquisició de la competència gramatical i específicament la competència ortogràfica. Els estudis sobre la segmentació de l'escrit pertanyen a l'àmbit de la psicogènesi de l'escriptura més que no pas al dels estudis sobre la construcció del concepte de paraula o als de l'emergència de la capacitat de reflexió metalingüística.

---

<sup>21</sup> Els treballs de Downing i els de Hirsh-Pasek citats a Gombert, (1990:106); Roberts, 1992:125,135); Bowey & Tunmer, 1984, en són alguns dels exemples.

### 3.2. El coneixement de la paraula: naturalesa i evolució

Són nombroses les investigacions que indaguen la mena de coneixement sobre la paraula que els infants han construït i com aquest va evolucionant. Des de les observacions que indiquen -com acabem d'explicar- la sensibilitat precoç dels infants per distingir alguns aspectes lingüístics fins a la possibilitat de fer una definició de paraula, amb el que això implica de coneixement metalingüístic, hi ha un recorregut molt llarg durant el qual es produeixen canvis interns i la influència decisiva de factors externs. Els canvis van en el sentit de la presa progressiva de consciència del sistema lingüístic. Es passa d'un estadi exclusivament procedimental, que permet l'ús de les paraules, a un altre en què la llengua esdevé un objecte opac<sup>22</sup> i pot ser observada i analitzada. Kolinsky (1986), d'acord amb Piaget, creu que la presa de consciència no és un fenomen que apareix de cop sinó que és fruit d'un procés de conceptualització, procés que implica l'existència de graus diferents de consciència. La consciència metalingüística comporta reflexió conscient, capacitat d'analitzar el llenguatge i control sobre aspectes diversos del llenguatge.

Presentarem a continuació alguns dels estudis sobre la construcció del concepte de paraula, el gruix més important dels quals va ser realitzat sobretot a les dècades dels seixanta als vuitanta. Ens hem basat en dos treballs de síntesi: el de Bowey & Tunmer, de 1984, i el de Gombert, de 1990, i en una investigació sobre el mateix tema, dirigida per Karmiloff-Smith<sup>23</sup>, que fa una referència molt extensa dels treballs anteriors, i justifica una metodologia i uns resultats una mica diferents dels que la majoria d'autors han trobat. En el primer cas, tots dos autors havien realitzat estudis previs (1981, 1983 i 1984) i recullen les informacions aportades per investigacions anteriors. En el segon cas, la referència al desenvolupament de la consciència de paraula forma part

---

<sup>22</sup> La idea d'opacitat, la trobem a Cazden (1976): "Cazden define la metalingüística como 'la capacidad de opacar las formas y prestarles atención en y por sí mismas'" (Olson, 1991:305). També Tolchinsky recull la mateixa idea (Tolchinsky, 1995:43).

<sup>23</sup> Karmiloff-Smith, K, Julia Grant, Kerry Sims, Marie-Claude Jones, Pat Cuckle. Estudi publicat a *Cognition*, 1996, però realitzat el 1991. Citat per la mateixa autora a "Beyond the modularity" (1992) [Trad. cast. "Más allá de la modularidad" (Alianza, 1994)].

d'un estudi sobre el desenvolupament metalingüístic, en general. Hem considerat, també, les aportacions de Roberts, que explicita clarament la influència de l'escrit en la construcció de la consciència de paraula com a unitat lingüística. L'autora dóna referències de treballs que van en la mateixa línia, i diferencia clarament entre el que és la influència de l'escrit en l'establiment dels límits de la paraula i el que és una tasca cognitiva més profunda com és elaborar el concepte metalingüístic de paraula (Roberts, 1992:127).

Els estudis realitzats i ressenyats, tot i perseguir objectius semblants, presenten diferències importants pel que fa a les tasques concretes que es demanen als infants i a la manera de plantejar-les. En el seu treball, Bowey i Tunmer distingeixen tres components imprescindibles per desenvolupar completament la consciència de paraula:

- a) consciència de paraula com una unitat del llenguatge,
- b) consciència de paraula com una etiqueta fonològica arbitrària,
- c) comprensió del terme metalingüístic de paraula.

Defensen la idea que a) i c) poden manifestar-se de forma relativament independent i que, a l'interior de cada un d'aquests components, es poden donar diferents graus de consciència. Segons aquest plantejament, pot haver-hi infants que resolguin tasques de segmentar la parla en paraules sense que hagin desenvolupat el concepte de paraula. I es pot donar el cas de nens que responguin a preguntes sobre el concepte de paraula sense que això impliqui que comparteixin la mateixa noció que els adults<sup>24</sup>. A parer seu, alguns investigadors no han tingut en compte la distinció clara entre a) i c), fet que ha conduït a confusions entre la paraula com a unitat lingüística i el concepte metalingüístic de paraula i, en conseqüència, ha afectat els resultats de les investigacions. El que sí que afirmen és que el descobriment de la unitat paraula precedeix òbviament la possibilitat de desenvolupar el concepte de paraula.

---

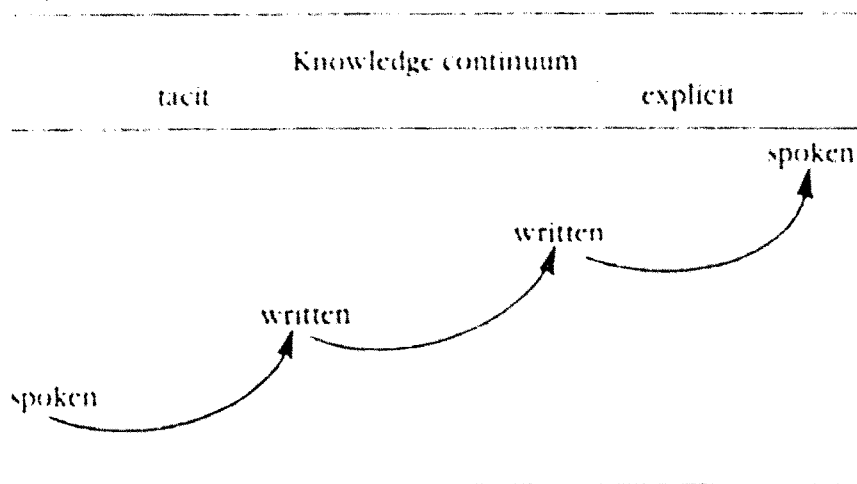
<sup>24</sup> Així ho indiquen els exemples de Bohn, amb un infant de 1:9 anys (1914) i el d'Slobin, amb la seva filla Heida, de 2:10 anys (1978), citats a Bowey & Tunmer (1984:75) i a Gombert (1990:105).



Gombert proposa una classificació diferent, però fàcilment assimilable a la que proposen Bowey & Tunmer, basada en la distinció entre els conceptes d'habilitat metalèxica i habilitat metasemàntica. El primer cas correspondria, aproximadament, a l'apartat 'consciència de la paraula com a unitat del llenguatge', de Bowey, però hi inclouria el tema del control deliberat d'accés al lèxic. El segon cas, l'habilitat metasemàntica, correspondria, també aproximadament, a la 'comprensió del terme metalingüístic paraula' i inclouria el tema de la diferenciació entre paraula i referent, amb referències als estudis de Piaget (1929).

Roberts presenta un esquema d'adquisició del concepte de paraula interessant perquè integra la parla, des del moment del desenvolupament, i l'escriptura en un continuum que va des del coneixement tàcit a l'explícit:

Figure 1 Development of concept of word



Roberts. B (1992:177)

Presentem, a continuació, alguns dels estudis sobre el concepte de paraula, la majoria ressenyats pels autors citats, amb alguns comentaris i reflexions.

1. Hi ha un grup d'estudis sobre les activitats que mostren les habilitats dels infants per al joc lingüístic, les rimes. Roberts cita el treball de Clark, de l'any 1978, en què es posa de manifest aquesta habilitat (Roberts, 1992:125). Bowey i Tunmer (1984:79) citen els treballs de Bohn (1914), Snyder (1914) i Weir (1962) d'observació de jocs en què els infants substitueixen una paraula per una altra en la cadena parlada. Les manipulacions, alteracions i canvis que poden fer els subjectes amb les paraules mostren un coneixement intuïtiu d'aquestes unitats i indiquen poc o molt un cert nivell de consciència de la paraula com a unitat de la parla; correspondria al concepte de coneixement tàcit<sup>25</sup>, de Roberts (B. Roberts, 1992: 125). Aquests tipus d'activitats lingüístiques no responen a cap consigna, sinó que són manifestacions espontànies dels infants. La introducció d'una consigna exigeix un nivell de reflexió i un esforç cognitiu superiors.

En aquest grup s'hi podrien incloure, també, els exemples dels nens que viuen en contextos o estan en situacions de contacte de llengües i mostren un cert nivell de reconeixement de les paraules que pertanyen a les dues llengües del seu entorn. Aquests infants es fan conscients precoçment de l'arbitrarietat del signe, tal com afirma Bialystok (1986 :17), reprenent el que ja havia afirmat Vigotsky (Vigotsky, 1962).

2. Un grup força nombrós d'investigacions són les que apelen a la segmentació de la parla i al reconeixement de la paraula en tasques diverses: de picar per cada paraula sentida (*word tapping task*) (Kingston, 1972; Holden & Mc Ginitie, 1972; Ehri, 1975; Evans, 1979; Tunmer & Bowey, 1981); de comptatge de paraules (Karpova, 1966,

---

<sup>25</sup> Som conscients de les dificultats en la utilització del terme consciència. En la literatura anglesa existeixen dos termes: *awareness* i *consciousness* que entren en concurrència i en dificulten un ús precís en la traducció. En la literatura francesa, almenys en els textos d'E. Gombert que ens han servit de referència, el mot *awareness* ha estat traduït per *maîtrise*, que en català correspondria millor a 'habilitat'. Com a mostra, dos exemples: A la p.14, Gombert diu: "(...) ce type d'activité demande une maîtrise ("awareness") métalinguistique définie comme (...)" (Gombert, 1990:14). I, en fer referència als components que defineixen la consciència completa de mot, tal com els presenten Bowey & Tunmer (1984), tradueix per *maîtrise* allò que a l'original és *awareness*.

citada a Bowey i Tunmer, 1984 i a Gombert, p. 91), d'enumeració de paraules (Bialystok, 1986); de decisió sobre l'"últim bocí sentit" (Fox & Routh, 1975).

Malgrat que els estudis d'aquest grup demanen del subjecte una activitat de reflexió semblant, s'hi observen diferències per la manera que es formula la demanda i per algunes altres qüestions de tipus metodològic. Per exemple, en alguns casos, l'investigador proporciona exemples als infants de com realitzar la tasca (Holden & Mc. Ginitie, 1972; Fox & Routh, 1975; Tunmer & Bowey, 1981), o els entrena prèviament, mentre que en d'altres casos, no. En la majoria de casos, l'investigador utilitza el terme paraula per fer la demanda, qüestió que dificulta l'obtenció d'un resultat fiable, perquè pressuposa el coneixement del terme<sup>26</sup>. El baix nombre de resultats positius en aquests tipus de treballs podrien venir donats més per la incomprensió de la consigna que per la incapacitat real de segmentar en unitats lèxiques.

És per salvar aquest obstacle, que Fox & Routh, tal com ja hem explicat, substitueixen paraula per "l'últim bocí que has sentit"; però, en aquest cas concret, la decisió de superar la dificultat metodològica amb la supressió del terme paraula, no dona uns resultats gaire creïbles perquè, si la resposta dels subjectes és una unitat superior a la paraula, els investigadors els demanen de continuar fins que diuen un terme equivalent a la paraula. Aquest procediment emmascara els resultats. En l'estudi de Bialystok ja citat, es proposen als infants quatre tasques per resoldre, una de les quals consisteix a demanar-los que trobin les paraules d'una frase. Aquestes frases poden ser dites de forma gramatical o amb les paraules desordenades. En els resultats sorprèn que, sobretot als nens més grans (7 anys), els és més fàcil detectar les unitats en les frases desordenades que en les que no ho són. Una explicació possible, proposada per Gombert (p. 92) seria que, en les sèries

---

<sup>26</sup> La utilització del terme paraula en la majoria d'investigacions és criticat també per Gombert: "(...) pour le fait que les consignes utilisent généralement le terme métalinguistique 'mot' dont la compréhension est soupçonnée être difficile et très dépendante de l'instruction formelle" (1990:105).

desordenades, les paraules funció o gramaticals, en aparèixer desvinculades de les paraules plenes amb les quals formen un sintagma, emergeixen clarament i són més fàcilment perceptibles; és més, probablement, el fet que la seqüència no tingui cap significat permet al subjecte focalitzar l'atenció en les unitats preses aïlladament.

3. Un altre tipus d'investigacions tenen com a objectiu indagar sobre l'evolució del concepte de paraula i sobre el propi grau de consciència de paraula. Generalment, els investigadors treballen amb nens d'edats diferents, la qual cosa permet analitzar l'evolució del concepte. Karpova, en l'estudi ja citat, demanava a nens russos entre 3 i 7 anys quantes paraules hi havia en una frase que els deia (activitat semblant a les del grup anterior) i, a continuació, els preguntava quina era la primera, la segona, la tercera, etc. Els resultats mostraven que els nens més petits resolien amb poc èxit la tasca i que els grans, més experts, mostraven més competència però també algunes confusions entre les paraules funció i les plenes. Downing proposava tasques de decisió entre paraula i elements de categoria diversa: síl.labes, frases, sorolls, etc. (Downing, 1970, 1971), a nens de 5 anys. Va obtenir molt pocs resultats positius; molts pocs nens podien distingir la unitat paraula. Aquests treballs varen ser criticats des d'un punt de vista metodològic per la dificultat que comportaven, atès que la tria entre elements de classes molt diferents -síl.labes, sorolls, pseudoparaules, paraules, etc.- dificultava encara més una tasca de per si difícil. En tot cas, però, els resultats mostraven -com de fet mostren els altres treballs- que a 5 anys no s'ha elaborat encara un concepte de paraula semblant al dels adults.

Citem especialment, per la seva amplitud, els treballs de Papandropoulou & Sinclair (1974) i Berthoud-Papandropoulou (1978) sobre l'evolució del concepte paraula. En tots dos casos, les autores justifiquen els treballs fent notar que s'han fet molts estudis per explicar el desenvolupament del llenguatge però consideren que n'hi ha molt pocs que investiguin com evoluciona la capacitat dels infants per reflexionar sobre el llenguatge més enllà de l'ús que en

fan per a la comunicació. La seva atenció se centra en l'evolució de les idees dels infants sobre el concepte paraula, sobre la comprensió de l'arbitrarietat de la relació entre les paraules i els referents, i sobre el reconeixement de les paraules, entre d'altres aspectes estudiats. Tant la investigació de 1974 com la del 1978 són longitudinals i abracen edats compreses entre els 4 anys i els 10 (1974) i els 4 i els 12 anys (1978).

Aquestes són algunes de les conclusions :

- ▶ Els més petits confonen realitat amb llenguatge. Un exemple freqüentment citat: La investigadora pregunta: "sis nens juguen" quantes paraules hi ha?: Sis. Per què? Jo, el meu germà, la Maria, etc. (i compten les persones amb els dits)
- ▶ A 5 i 6 anys, els infants no diferencien entre les paraules i els objectes o les accions que aquestes representen ("una paraula és una cosa que existeix") o les paraules comparteixen característiques dels objectes ( "la maduixa és una paraula perquè creix al jardí")
- ▶ Alguns, de 6 anys, relacionen paraula amb parlar o dir: "una paraula és quan parlem, quan diem alguna cosa".
- ▶ Cap a 6 anys, llarg i curt està relacionat amb enunciat llarg/ curt: se'ls demana digues una paraula llarga i diuen: "Entro a casa per treure'm les sabates"
- ▶ Cap als 6-7 anys, cada vegada més les paraules són vistes com etiquetes, separades de la realitat que representen. Però han de tenir una quantitat mínima de lletres (això és el que diuen quan se'ls demana si l'article és o no una paraula). Això fa que les paraules funcionals no siguin considerades paraules; en canvi, noms curts, sí, perquè existeixen.
- ▶ Als 7-8 anys, comencen a usar termes gramaticals per referir-se a les paraules: "paraules són adjectius, noms i verbs"

- Només cap als 11 anys, els infants són capaços d'esbossar una definició lingüística del terme paraula i mostrar el nivell més alt de coneixement metalingüístic.

Hem explicat amb més deteniment algunes de les conclusions a què arriba Berthould-Papandropoulou, -que, de fet, són molt semblants a les d'altres investigacions-, perquè els dos estudis esmentats han tingut una repercussió considerable per la seva amplitud en les edats i en els aspectes investigats.

Tanmateix, però, algunes de les seves conclusions són revisades per Karmiloff-Smith en treball més recent (1996, op.cit.) en què la investigadora i els seus col.laboradors varen aplicar un procediment d'experimentació diferent.

Els autors d'aquest estudi assumeixen en la seva formulació teòrica que les conclusions sobre el desenvolupament del concepte de paraula depenen, en part, del significat que s'atribueixi al terme conèixer. No hi ha cap dubte -diu- que cap als 3 anys, quan els infants ja tenen una paraula força desenvolupada, han de tenir internament representats els límits de les paraules, malgrat les dificultats de percebre les unitats en la cadena parlada. Aquesta representació ja és una forma de coneixement.

El procediment que segueixen és en part en directe (*on-line*) i en part en diferit (*off-line*)<sup>27</sup>, i consisteix a interrompre l'explicació d'una història en algun punt que coincideixi amb una paraula de la classe tancada o oberta i demanar als infants -de 4 i 5 anys- que repeteixin l'"última paraula" que s'ha dit [Les cometes són nostres i volen subratllar el fet que les investigadores no han obviat el terme paraula]. Aquest procediment seria semblant al que Fox & Routh (op.cit.) van utilitzar, amb la diferència que els investigadors, en la consigna donada als infants, substituïen el terme paraula per "l'últim bocí que has sentit".

---

<sup>27</sup> Per a una explicació de les tasques cognitives involucrades en els dos tipus de tasques, vegeu Karmiloff-Smith (1994:76;1996:200).

A diferència dels estudis citats anteriorment, els seus resultats mostren que els infants no tenen dificultats per trobar la paraula en l'explicació d'una història -i no en una tasca experimental en què la sintaxi i el significat en són absents- i que no mostren massa més dificultats per trobar les paraules gramaticals que les lèxiques. Les competències mostrades pels subjectes de 4-5 anys que participaven en l'estudi corresponen a les de nens més grans en d'altres estudis. Les investigadores atribueixen els resultats al procediment utilitzat, tant per la manera de formular la pregunta com pel tipus de text triat i el context en què es realitza la prova. Es podrien aventurar, igualment, altres hipòtesis relacionades amb la familiarització amb tasques d'alfabetització.

En síntesi, els resultats dels estudis ressenyats, malgrat algunes diferències referides a les edats en què les capacitats es manifesten, permeten aquestes afirmacions:

*Pel que fa a l'ús inconscient del llenguatge*

Els infants, des de ben petits, ofereixen exemples de manipulació del llenguatge; poden distingir dues llengües i, fins i tot, usar el terme paraula, sense que aquests fets indiquin que han desenvolupat un concepte de paraula; ni tan sols que siguin capaços de reconèixer aquesta unitat o segmentar la parla en paraules.

*Pel que fa al reconeixement de les unitats o a la segmentació de la parla en tasques de demanda explícita*

Els infants més petits, entre 3 i 4 anys, mostren dificultats per a la segmentació de la cadena parlada. Cap als 5-6, les tasques de segmentació presenten més resultats amb èxit; i es constata que els elements que primer aïllen són els que corresponen a les paraules plenes, pel fet -segons Van Kleeck- que són semànticament autònomes i, pensem nosaltres, perquè tenen generalment més d'una sil.laba i amb una sil.laba tònica com a mínim. De fet hi ha autors (Ehri, 1975) que sostenen que l'accent és l'element que permet en molts casos la discriminació de les unitats en les tasques de picar paraules (*word tapping task*). L'exemple de Bialystok, citat una ratlles més amunt, confirma també la dificultat d'aïllar les paraules funció en una frase

ordenada, pel fet que aquestes paraules queden subsumides a la unitat significativa superior: el sintagma.

*Pel que fa a l'elaboració del concepte de paraula*

Abans de començar l'alfabetització, el concepte de paraula no està encara ben establert. Les investigacions basades en tasques de decisió (Karpova, 1966; Downing, 1970, 1971 i 1973) deixen bastant clar que els nens de preescolar tenen un coneixement baix del concepte de paraula. I si la pregunta que se'ls adreça és, encara, més directa, com: "Què et sembla que és una paraula?" (Papandropoulou & Sinclair, 1974, i Berthoud-Papandropoulou, 1978), els resultats indiquen la no diferenciació entre la paraula i la realitat o la seva resposta es basa en algunes característiques de la realitat representada per la paraula, tal com els estudis de Piaget i de Vigotsky ja havien mostrat.

Segons Berthoud-Papandropoulou, fins als 6-7 anys, no es comença a tenir consciència que la paraula és una etiqueta, un signe arbitrari. En aquesta fase, els infants admeten que una seqüència de lletres és una paraula si té una quantitat mínima de lletres. Això fa que les paraules funcionals no siguin considerades paraula; en canvi, noms curts, sí, perquè existeixen<sup>28</sup>. És força clar que el procés d'alfabetització contribueix a establir la unitat paraula, perquè, encara que les demandes es facin sobre la llengua parlada, els infants hi apliquen el coneixement que estan construint sobre les lletres. Aquestes dades es veuen confirmades per Ferreiro (Ferreiro & Teberosky, 1979).

A partir de 7-8 anys, els infants usen termes gramaticals per referir-se a les paraules: "*paraules són adjectius, noms i verbs*", fet que mostra la influència de les tasques d'alfabetització. Aprendre a llegir i a escriure són factors decisius en la possibilitat d'aïllar les paraules i establir-ne els límits, sobretot per l'*input* gràfic que els nens i les nenes reben durant les activitats d'alfabetització, però no solament per aquest *input*, sinó per les informacions, comentaris i demandes que la mestra fa sobre la llengua escrita durant la realització d'aquestes

---

<sup>28</sup> Els exemples de Berthoud-Papandropoulou són contundents: "Un mot, c'est des lettres"; "Il faut assez de lettres pour faire un mot, pas trop, ni trop peu" (1978:59).



tasques. Aquestes tasques conformen un marc general de desenvolupament de les habilitats metalingüístiques, les quals són vistes no com un resultat discontinu o aleatori, sinó com habilitats emergents reflex d'un procés gradual de desenvolupament no solament lingüístic sinó també cognitiu (Bialystok, 1986:14; Bigas i al. 2001).

Després d'haver abordat el tema de la percepció i detecció de la paraula en el flux de la parla i de la construcció del coneixement sobre aquesta unitat, ens interessem per la manera que els infants emmagatzemen la informació d'aquestes unitats, que hauran d'estar disponibles per a la comprensió oral i escrita, i per a la producció de la parla i de l'escriptura.

### **3.3. La construcció del lèxic i l'accés al lèxic**

Aprendre a parlar implica aprendre les paraules d'una llengua, les quals permeten referir-se a la realitat i crear representacions dels objectes, els fets i les relacions. Les paraules són les unitats lingüístiques que primer apareixen en el llenguatge i per tal que puguin ser reutilitzades bé sigui per a la comprensió, bé per a la producció, oral o escrita, han de poder ser emmagatzemades d'alguna forma. El conjunt de les paraules conegudes per un subjecte en cada moment configura el seu lèxic. El lèxic és central tant en el coneixement d'una llengua com en la seva adquisició (Clark, 1993:1).

A fi que el lèxic sigui accessible als parlants, aquests han de ser capaços d'identificar les paraules per l'oïda o per la vista i recuperar-les amb la forma adequada al discurs del moment.

Quina forma tenen les unitats en el lèxic? La forma de les paraules de l'enunciat? La de les unitats del sistema? Contenen la informació semàntica i fonològica associada? És a dir, com està organitzat, el lèxic?

Clark explica que cada entrada lèxica del nostre lèxic mental ha d'incloure les següents informacions:

- ▶ semàntica

- ▶ sintàctica
- ▶ morfològica
- ▶ fonològica

Les dues primeres constitueixen el *lema (lemma)*, i les dues últimes, la *forma (form)* (Levelt, 1989, a Clark, 1993:3).

Per exemple, la paraula *futbolista* inclou:

- ▶ Informació del significat de 'futbol',
- ▶ Informació del significat morfològic: '-ista' relacionada amb el significat del sufix, que és la d'agent/ofici,
- ▶ Informació de la categoria sintàctica: nom, comptable, pot fer de subjecte o d'objecte directe, etc.,
- ▶ Informació fonològica, lligat a l'especificació dels fonemes, de les síl.labes i de l'accent,
- ▶ Informació de l'estatus de la paraula, en relació als registres: és una paraula vulgar o no, voc. específic, dialectal, adaptació d'una altra llengua, etc.

Els nens han d'aprendre la forma de les paraules tal com les usen els parlants i el significat que tenen. Per a cada paraula els nens han d'emmagatzemar tot tipus d'informació en memòria: els significats, la categoria de la paraula (nom, adjectiu...) i el funcionament sintàctic, l'estructura interna i com es pronuncia. Hi ha d'altres autors que, com Clark, defensen també que el lèxic està organitzat de forma multidimensional i que l'accés a la informació lèxica es pot produir per vies diferents (Seguí, 1992:71).

Seguí sosté que allò que és essencial dels nostres coneixements sobre el lèxic és molt poc accessible a la consciència -"à l'inspection consciente", com diu (1992:71). Malgrat aquesta dificultat d'accés conscient, el gruix d'aquests coneixements abstractes ens permet reconèixer si una seqüència de lletres és una paraula possible de la

nostra llengua, o com s'ha de pronunciar, o escriure, una paraula que sentim o veiem per primera vegada.

Aquesta informació és elaborada i reelaborada pels infants a partir de l'*input* que reben en la interacció amb els altres parlants, queda representada en la memòria i és la que permet el reconeixement de les unitats en ocasions posteriors. Aquest tipus de representacions són les que Clark anomena *C-representations*, és dir representacions a partir de la comprensió. Quan tornen a sentir la mateixa paraula, no solament la reconeixen com una paraula que ja han sentit abans, sinó que accedeixen a la informació (fonètica, de significat, etc.) que han emmagatzemat. Els infants es guien per uns principis a l'hora de seleccionar la informació rellevant de les unitats que segmenten, clarament explicats per Clark<sup>29</sup>, sobre els quals no ens estendrem.

A més de les representacions que els infants es fan de les unitats relatives a la comprensió, hi ha també les relatives a la producció; el que Clark anomena *P-representations*, que proporcionen al parlant informació sobre com es produeix una paraula: els òrgans que hi intervenen, l'ordre dels sons, els accents, els sons que combinen, etc. Només després que els infants hagin emmagatzemat representacions d'unitats, podran iniciar-ne la producció oral. Amb relació a la llengua escrita, es podria suposar que els infants, quan comencen el procés d'alfabetització, creen també representacions relacionades amb el sistema gràfic.

La llengua escrita també implica dos processos: un de recepció, la lectura, (a partir d'un estímul sensorial) i un altre de producció, l'escriptura, (a partir d'una activitat cognitiva) de les unitats. Mentre que les investigacions d'accés al lèxic són molt nombroses pel que fa a la lectura, el cas de l'escriptura ha estat molt poc estudiat. En el cas de la lectura, les explicacions sobre l'accés al lèxic es divideixen segons la prioritat que es dona a la mediació fonològica o a la percepció/ reconeixement global d'unitats. És un tema que cau fora de l'àmbit d'aquest estudi, però que ha tingut transcendència a l'hora de

---

<sup>29</sup> Aquests principis són: contrast i convencionalitat; transparència, simplicitat i productivitat (Clark, 1993: 84, 109, 126).

defensar un tipus de metodologia o un altre d'ensenyament de la lectura.

En aquest punt, el problema plantejat al començament retorna: si els infants han après a parlar, procés que ha implicat haver segmentat el corrent de la parla, aïllat les paraules i fer-se'n representacions per a la comprensió i la producció, per què en les primeres manifestacions escrites mostren un cert desconeixement dels límits entre les unitats? De quina manera aquestes representacions de les paraules que permeten comprendre i produir la parla estan disponibles per a la producció de l'escrit? Com es fan accessibles a l'infant al moment d'escriure un text?

### 3.4. La representació de les unitats. El model de redescrípció representacional de Karmiloff-Smith de 1986(1994)

Una hipòtesi ben plausible és que hi ha una distància, podríem dir cognitiva, entre la naturalesa del coneixement que els nens tenen de les unitats paraula per poder parlar i el que necessiten per poder escriure. Entre un moment i l'altre del desenvolupament lingüístic, entre una competència i una altra, es produeixen canvis el coneixement dels quals és el que fa possible trobar una explicació dels fets. El model de redescrípció representacional (RR), formulat per Karmiloff-Smith el 1986<sup>30</sup> era el que semblava oferir una informació més detallada del procés que condueix els infants des de les primeres representacions fins a un coneixement verbalitzable, manifestació de la capacitat de reflexionar sobre el llenguatge. A la vegada era un model general d'interpretació de la transformació del coneixement, no solament del coneixement lingüístic. Gombert, al seu extens i profund estudi sobre el desenvolupament metalingüístic, explica que hi havia pocs models que intentessin fer una teoria de tot el procés i que superessin les dicotomies implícit/explicit, procedimental/declaratiu, inconscient/conscient. Karmiloff-Smith, diu l'autor, s'atreveix a

---

<sup>30</sup> El model que es descriu és el que va formular l'autora el 1986 i publicat a *Cognition*, 23, 95-147. Per a aquest estudi ha estat consultat el model tal com apareix a la publicació del 1992, *Beyond the modularity*, Massachussets, MIT Press, traduït per Alianza Editorial, el 1994, tal com consta a la bibliografia.

proposar un model que pugui donar compte de la complexitat del procés (Gombert, 1990:240).

Segons Karmiloff-Smith, les representacions de les paraules que construeixen els nens quan aprenen a parlar els serveixen per això: per parlar; són una mostra de destresa lingüística, però no ho són de reflexió lingüística ni metalingüística. Tal com diu l'autora: "*son información en la mente, pero no para la mente*" (p. 71), és a dir, no estan disponibles per a ser usades com a objecte de reflexió. La destresa procedimental o conductual no seria suficient per abordar tasques que impliquen un cert nivell de reflexió i de conceptualització; tasques com ara, segurament, escriure. El model RR intenta explicar de quina manera les representacions inicials que l'infant ha construït es tornen més flexibles i, per tant, més accessibles a la consciència; i aquesta explicació, l'autora la basa en un principi que entén que és específic de la ment humana: la capacitat d'explotar internament el coneixement que s'ha adquirit o que es té de forma innata. Entès així, les primeres representacions que l'infant ha construït estarien en un nivell implícit (NI) i, al llarg de successives redescrípcions, es convertirien en coneixement explícit i conscient.

Karmiloff sosté que, una vegada els nens, al voltant dels tres anys, són capaços de parlar utilitzant tot tipus de paraules tant de classe oberta com tancada<sup>31</sup>, en contextos nous, és que n'han d'haver construït una representació, que estaria, segons l'autora, en el format de nivell I (o coneixement implícit). El coneixement de les paraules que els infants tindrien en aquest nivell seria de tipus procedimental o conductual i no els permetria la realització de tasques que impliquessin accés conscient a les unitats. Per això, Karmiloff anomena aquest nivell corresponent a la fase 1, nivell implícit. L'infant se centraria en dades externes, en informacions que provenen de l'entorn, i els diferents coneixements implícits, és a dir, les representacions de les unitats lingüístiques que aniria construint, s'emmagatzemarien en unitats independents, sense relació entre elles. Dit d'una altra manera, aquestes unitats no constituïrien, encara, un sistema. La producció

---

<sup>31</sup> O lexical / grammatical items; o content words / functional words.

d'una forma lingüística particular dependria d'estímuls externs. En aquesta fase o fase 1 no s'ha produït encara cap redescrípció.

La fase 2 es caracteritza per ser una fase guiada internament, segons l'autora, en què l'infant no estaria tant atent a les dades externes ni a la correcció dels adults com a redescríure representacions anteriors, cosa que explicaria retrocessos aparents en el coneixement i aparició d'errors que en moments anteriors no havien aparegut, com també de regularitzacions. Segons Gombert, "*le moteur du développement serait la recherche d'un control sur l'organisation interne des connaissances implicites accumulées lors de la phase 1*" (1990:239). Aquesta fase té dos nivells, tots dos de coneixement explícit: E1 i E2, que es diferencien pel grau d'explicitud i d'accés a la consciència. Les representacions del nivell E1 són el resultat de la redescrípció de les representacions codificades procedimentalment en el nivell I. Són més flexibles, poden ser usades per a d'altres finalitats que requereixen coneixement explícit. De fet, però, és en el nivell E2, que les representacions es fan accessibles a la consciència.

La fase 3, representada pel nivell E3, es caracteritza perquè les representacions, a més de ser accessibles a la consciència, són verbalitzables. Constitueixen el que solem anomenar el coneixement metalingüístic del subjecte<sup>32</sup>. Durant aquesta fase, els lligams establerts durant la fase anterior serien reconsiderats en funció dels estímuls externs. Karmiloff-Smith no precisa gaire l'abast del significat d'"estímuls externs". Amb relació als experiments que ella mateixa va dirigir<sup>33</sup>, tant en tasques en directe [*on-line*] com en diferit [*off-line*], per determinar el nivell de representació de les paraules, en nens des de 3 fins als 7 anys, diu:

"Y en tercer lugar, algo debe ocurrir de nuevo internamente después de los cinco o seis años para explicar por qué, hacia esa edad, los niños pueden entregarse a la construcción de teorías acerca de qué son *las palabras*, teorías que resultan más fácilmente accesibles a la conciencia y cuyo conocimiento es utilizado en tareas en diferido. Para ello,

---

<sup>32</sup> L'autora diu explícitament que no fa distinció entre el nivell E2 i E3, perquè no hi ha investigacions centrades en el nivell E2. (p.43).

<sup>33</sup> Karmiloff-Smith, Grant, Jones i Cuckle, 1996, (op.cit.)

sostengo, es necesaria una nueva redescrpción en formato E2/3.”  
(Karmiloff-Smith, 1994:78).

Observem la inconcreció de l'expressió "*algo debe ocurrir...*". Sembla que l'autora se centri més en els aspectes interns, relacionats amb el desenvolupament de la cognició, i no en els factors que afavoreixen el pas d'un nivell i d'una fase a l'altre. En aquest aspecte, Gombert és molt més explícit, com veurem a continuació.

Gombert es mostra força d'acord amb el model proposat per Karmiloff-Smith, sobretot pel seu gran valor explicatiu i, com s'ha dit més amunt, per l'esforç de bastir una teoria general. Malgrat aquesta acceptació, destaca dos aspectes que considera dos punt febles en la seva teoria, un dels quals és, justament, la influència dels factors externs. Diu Gombert:

"Si nous sommes convaincu de la très grande valeur explicative de ce modèle -il constituera l'armature de notre explication du développement métalinguistique- nous n'en sommes pas moins réservé sur deux points : d'une part, son caractère récursif ; de l'autre, *la stricte alternance qu'il postule entre la prise en compte de facteurs exogènes et endogènes* " (Gombert, 1990 :240). [la cursiva és nostra]

L'autor defensa que el pas de la fase 1 a la 2, que -com hem dit més amunt- Karmiloff interpreta com una fase de reorganització interna de les representacions provocada per factors endògens, està sotmès a la influència, també, de factors externs. Des del seu punt de vista, la complexitat creixent dels enunciats, necessària per les exigències de comunicació que els infants experimenten, desencadenaria la redescrpción de les representacions inicials. En aquest sentit, els errors que als inicis de la fase 2 els infants mostren apareixen perquè aquests han de resoldre problemes de comunicació cada vegada més complexos.

La influència de factors externs no es faria sentir només, també segons l'autor, en el pas de la fase 1 a la 2, sinó que tindria un paper decisiu en el desenvolupament del funcionament metalingüístic. Malgrat que ni el model de Karmiloff-Smith, ni tampoc el que proposa Gombert, és un model d'estadis, aquest autor creu que es poden donar algunes referències de l'edat en què els canvis es produeixen i, en aquest

sentit, apunta que els indicis de desenvolupament lingüístic els podem trobar a l'entorn dels 6-7 anys, vinculats a l'aprenentatge de la lectura i l'escriptura.

D'altres investigadors valoren les aportacions del model cognitiu de Karmiloff-Smith. Kleek (1994), per exemple, en una exposició dels diferents models interpretatius del desenvolupament metalingüístic, fa una explicació molt acurada del model de Karmiloff i en valora la flexibilitat amb relació al model d'estadis proposat per Piaget. Una mostra d'aquesta flexibilitat és que el model permet entendre que hi ha nivells de representació diferents per a les unitats lingüístiques i que aquests nivells no són homogenis per a tots els infants. Podria haver-hi, per exemple, unitats del sistema fonològic representades a un nivell i d'altres, a un altre nivell, en funció de factors externs relacionats amb l'experiència de l'aprenent. Aquest model tindria en compte tant els aspectes evolutius endògens - els que correspondrien a la fase 2, en els nivells E1 i E2, descrita per Karmiloff-, com els aspectes que dependrien de la influència de l'entorn -tant el nivell I, com l'E3- amb la qual cosa deixaria obert el camí cap a la integració d'un model clarament cognitiu amb altres models relacionats amb el constructivisme social que posen l'accent en la influència dels factors socials en el desenvolupament lingüístic (Kleek, 1994:64).

**En síntesi**, es podria dir que l'interès del model de Karmiloff per a l'estudi sobre la segmentació de l'escrit es trobaria en el fet que permet

- ▶ explicar que l'accés al coneixement lingüístic és gradual i asimètric, i que les unitats són redescrites i *rerepresentades* d'una manera més flexible en un procés que va des de l'explicitació (nivell E1) fins a la consciència i a la manifestació d'aquesta consciència per mitjà de la verbalització (nivells E2/E3),
- ▶ suposar que els canvis que es produeixen són fruit d'activitats lingüístiques que tenen a veure amb el procés d'alfabetització. Encara que l'autora no precisa l'abast i la influència dels factors externs que afavoreixen el canvi, la seva proposta



permet fer la hipòtesi que l'aprenentatge de la lectura i l'escriptura serien activitats desencadenants del coneixement metalingüístic,

- ▶ considerar que la representació de les paraules que els infants han construït en el desenvolupament de la parla és de naturalesa diferent que la que necessiten per escriure i llegir, i que les unitats que ja estan disponibles per a la parla no ho estan per a l'escriptura. Aquest és un punt crucial sobre el qual retornarem.

## 4. ESTUDIS PREVIS

Tal com hem explicat en els apartats anteriors, l'interès envers els estudis sobre la capacitat d'escriure, a diferència dels relacionats amb el desenvolupament de la parla, és relativament recent. Es podria afirmar segurament que la investigació sobre la psicogènesi de l'escriptura, realitzada per E. Ferreiro i publicada l'any 1979 (Ferreiro & Teberosky, op.cit.) marca l'inici d'un seguit de treballs orientats a investigar el desenvolupament de la competència escrita. En dir competència escrita, som conscients que estem utilitzant un terme excessivament ampli que inclou aspectes tan diversos com el sistema escriptural o la producció de textos, però en realitat tant els uns com els altres han començat a ser objecte d'interès i d'estudi, de forma preferent, en les últimes dècades.

El cas de la segmentació se situa en l'àmbit de l'aprenentatge del sistema d'escriptura, considerant que aquest, fins i tot en els sistemes alfabètics com és el cas de la llengua dels textos analitzats, mostra una dimensió grafofònica, però també una altra d'ideogràfica, la qual recull la representació gràfica dels aspectes prosòdics, suprasegmentals, de disposició gràfica segons els diferents tipus de text i dels límits de les paraules gràfiques (Jaffré, 1996) .

Recórrer a d'altres estudis fets i publicats amb anterioritat sobre el tema de la segmentació, per veure els objectius plantejats i els aspectes analitzats, és necessari perquè, malgrat les diferències en alguns aspectes, els resultats obtinguts en aquestes investigacions poden tenir concomitàncies amb els de la nostra recerca. En aquest sentit, presentem alguns dels estudis consultats, dels quals explicarem aquells aspectes que puguin tenir relació amb els que han estat estudiats en el present treball.

Encara que totes aquestes investigacions tenen com a objectiu la competència en la segmentació del text escrit, difereixen en alguns aspectes: en els subjectes, en el tipus d'estudi, en les consignes donades i el context de realització, en els aspectes analitzats, etc.

Seguint l'ordre cronològic, els estudis consultats són els de: Glenda Bissex (1980); Rosa Clemente (1984); Martin Nystrand (1986); Arturo Gomes de Moraes (1995); E. Ferreiro i al. (1996); Mónica Báez i Viviana Cárdenas (1999) i Concha Cintas (2000). Els treballs no són del tot comparables, a més dels factors anomenats unes ratlles més amunt, perquè són de naturalesa diversa: dos són tesis doctorals (Gomes de Moraes i Cintas), consultades de forma completa; un altre és un article la informació del qual correspon a una part d'una tesi sobre un tema més ampli (Clemente), mentre que els estudis de Ferreiro i al. i de Báez & Cárdenas no tenen el format de tesi i han estat consultats en la publicació corresponent. Finalment, el treball de Bissex correspon al tipus d'estudi de cas. Aquestes diferències fan que les informacions aportades per cada document siguin, també, de naturalesa diversa. En el cas de les tesis, òbviament, la informació sobre els objectius, les hipòtesis i la metodologia presenta un grau de detall que ha facilitat l'accés a un coneixement més complet i precís. En els altres documents, sobretot en el cas de Clemente, no s'ha tingut accés a la informació completa. Malgrat tot, en allò que ens interessava, les aportacions ens han estat útils.

#### *Pel que fa al tipus d'estudi i a la població analitzada*

La majoria d'investigacions són de tipus evolutiu, en el sentit que analitzen els resultats de subjectes de diferents edats, encara que no es faci un seguiment del mateix subjecte sinó que se seleccionin i s'analitzin simultàniament infants d'edats diferents, que estan en cursos diferents també i, fins i tot, de països i llengües diferents. L'únic estudi evolutiu de tipus longitudinal és el que va fer Glenda Bissex, amb el seguiment de Paul, el seu fill, des de l'edat de 5 anys fins als 11. El material recollit per Bissex va servir per a un estudi posterior conduït per M. Nystrand, el 1986. Rosa Clemente va estudiar la mateixa franja d'edat; Ferreiro i al. es van centrar en infants de 2n. i 3r. de primària; Gomes de Moraes, va ampliar la mostra a 4rt. de primària; Báez i Cárdenas, treballen amb nens des de 1r. fins a 7è i Cintas, amb nens i nenes des de parvulari 5 anys fins a tercer de primària: 8 anys, aproximadament.

L'edat és un factor que permet classificar els infants, però és obvi que no determina una competència homogènia en el coneixement lingüístic: hi ha molts factors en joc en els aprenentatges i els nivells assolits pels infants. En alguns dels estudis, es diu de forma explícita que s'han tingut en consideració, a l'hora d'agrupar els subjectes, els factors socials que podien influir en els resultats (Ferreiro i al.), la procedència rural o urbana (Báez & Cárdenas), o el tipus d'escola -privada/pública- a la qual assistien els nens i nenes relacionada amb la classe social: mitja o treballadora (Gomes de Morais). També és cert que les experiències familiars, encara que difícils de categoritzar, tenen repercussions en els aprenentatges dels infants i podrien matisar aquestes variables, però aquesta variable no ha estat controlada en cap dels casos esmentats.

#### *Pel que fa a la situació i a les tasques plantejades*

Amb l'excepció de l'estudi de Bissex i de Ferreiro i al., en la majoria de casos s'ha creat una situació experimental en què s'ha proposat als infants la realització d'uns textos de característiques concretes, prèviament planificades. Les anomenem situacions experimentals perquè, al marge que haguessin estat conduïdes en les aules escolars, amb la mestra o el mestre de l'aula, responien a demandes: dictats, rescriptura de textos, entrevistes clíniques, en què el material proposat havia estat dissenyat pensant en els objectius de la recerca.

Clemente dissenya dues proves: en l'una, els nois i noies han d'escriure una petita narració molt breu amb unes consignes també molt concretes i restrictives. En l'altra, la investigadora els demana de retallar en fragments, el "més petits que puguin", amb l'objectiu d'observar quins criteris guien el retallat.

En l'estudi de Gomes de Morais, les tasques dissenyades eren dues: un dictat i una tasca de transgressió. El dictat havia estat *preparat* tenint en compte aspectes de normativa (basats en la diferent casuística de les relacions grafofòniques) i aspectes gramaticals (si les paraules formaven part del lèxic bàsic del portuguès). Per acabar, hi havia una entrevista clínica, realitzada uns dies després de la prova de transgressió, en un context fora de l'aula i individualment, amb

l'objectiu d'observar, a través de les respostes dels infants, el nivell d'explicitació verbal de les qüestions ortogràfiques analitzades.

En la investigació de Cintas, són quatre, les tasques proposades: escriptura de frases de dues paraules, dictades, i de frases de quatre paraules; rescriptura d'un conte conegut i dictat prèviament unes quantes vegades i, finalment, una tasca de recompte de paraules a l'oral i a l'escrit. Cada una de les propostes havia estat prèviament pensada amb la finalitat que hi apareguessin elements de característiques gramaticals diverses, en posició diversa i en contextos sintàctics diferents, que poguessin donar resposta als objectius perseguits.

El treball coordinat per Ferreiro i al. (1996), en canvi, fa un plantejament que podríem anomenar *natural* o no experimental, per la raó que la tasca proposada als nois i noies que participen en la investigació és l'escriptura d'un conte conegut per tots, en una situació d'aula habitual, amb les úniques restriccions següents: que hi apareguin els tres personatges principals: la Caputxeta, l'àvia i el llop; i que no hi apareguin personatges d'altres històries. La justificació del tipus d'investigació està extensament explicada per dues de les investigadores: Ferreiro i Pontecorvo (p:22). En aquest aspecte concret, l'estudi que presentem fa un plantejament que s'hi aproximaria.

#### *Pel que fa als objectius perseguits*

Les diferències entre els estudis mencionats són notables. En algun cas, com en la tesi de Gomes de Morais, la segmentació és un aspecte del sistema d'escriptura analitzat al costat d'altres, però no és -ni els altres tampoc- l'objectiu final de l'estudi, sinó que aquest té a veure amb les representacions dels infants i les redescrípcions d'aquestes representacions sobre alguns aspectes ortogràfics, tenint en compte algunes variables com l'edat i la procedència. Malgrat tot, però, els resultats de l'anàlisi pel que fa a l'apartat de la segmentació poden ser comparats amb els de la nostra recerca, i és per aquest motiu que hi fem referència.

El cas de l'estudi de Clemente és semblant al de Gomes de Morais, amb la salvetat que l'apartat referit a l'anàlisi específica de l'escrit només és un fragment, presentat de forma aïllada i publicada en forma d'article, cosa que fa que les informacions no siguin massa completes ni precises. De la lectura de l'article es desprèn que els objectius eren indagar en la comprensió metalingüística de les unitats del llenguatge escrit i en la influència del procés d'alfabetització en el desenvolupament de la capacitat de reflexió metalingüística.

Ferreiro i al. analitzen la segmentació, agrupant els textos segons si són convencionals, si presenten només EHO, EHR o tots dos tipus d'errors. En fan una anàlisi quantitativa, que completen amb una altra més qualitativa en què aborden qüestions referides a determinants, articles, preposicions, pronoms. Les hiposegmentacions amb la preposició A en les seqüències *a la* i *ala* són objecte d'una atenció preferent. En la interpretació s'ha donat més importància al criteri de la atonicitat i el nombre de lletres que no pas a d'altres criteris com el sintàctic i el semàntic. Aquest és un punt de vista diferent del que s'ha considerat en el treball que presentem.

La tesi de Cintas proposa quatre objectius sobre la segmentació, succintament exposats a continuació (2000:50): a) l'evolució del desenvolupament gràfic i la influència que en aquest desenvolupament pugui tenir el procés d'alfabetització; b) explorar la influència de la categoria gramatical en la segmentació, analitzant, sobretot, la categoria de les paraules: lèxiques o gramaticals; c) analitzar la influència que el tipus de text i la llargada dels textos poden tenir en la segmentació i d) explorar si la relació entre la identificació de les paraules a l'oral és posterior o no a la identificació en l'escrit. Per a cada un dels objectius té en compte algunes altres variables, a més de l'edat.

Ens hem entretingut a enunciar tots els objectius, perquè alguns dels aspectes analitzats ho han estat també en l'estudi que presentem, malgrat que en el nostre cas els objectius plantejats siguin diferents.

*Pel que fa als resultats*

Amb relació al tipus d'error, es pot veure que hi ha en tots els casos força coincidències: es constata un nombre més gran d'errors d'hiposegmentació que d'hipersegmentació, però cal introduir algunes matisacions, relacionades amb els resultats en els errors d'hipersegmentació de l'estudi de Báez & Cárdenas i en el de Gomes de Morais; en el primer cas, les investigadores constaten que els nens que han estat agrupats en el nivell II (NII), el qual representa un grau de competència més alt, cometem més errors d'hipersegmentació en el dictat que no pas d'hiposegmentació. En el mateix sentit, Gomes de Morais troba que els nens i nenes que tenen un nivell de competència més alt, mostren més errors d'hipersegmentació que no pas d'hiposegmentació. La interpretació que l'autor mateix proposa està basada en la idea que la major competència lingüística dels nois i noies els fa prendre les decisions de segmentació basant-se en els aspectes morfològics i, d'aquesta manera, es produeixen casos d'hipersementació en què els elements aïllats corresponen moltes vegades a morfemes del portugués (p.130-131).

Els resultats en l'estudi de Ferreiro i al. van en el mateix sentit: en l'anàlisi quantitativa, s'observa que en totes les llengües menys en el cas de l'italià els textos amb ortografia no convencional són més nombrosos que els convencionals; i també, s'hi veu un domini de la hiposegmentació sobre la hipersegmentació, al marge de les llengües i dels *scripts* utilitzats en els textos (p. 58). De l'anàlisi qualitativa, en destacarem només alguns resultat: per exemple, les investigadores troben que els elements proclítics apareixen més freqüentment hiposegmentats sobretot amb la paraula que els segueix; que els elements homòfons: *la* (article) i *la* (pronome), la hiposegmentació és més freqüent en el cas de l'article; que la preposició *a* també s'hiposegmenta en les construccions [*a + infinitiu*] o [*a + la*] o [*a + donde*].

De l'extens treball de Cintas, en seleccionarem també només alguns aspectes, que són aquells que poden tenir relació amb els que nosaltres hem analitzat. Pel que fa a la quantitat d'errors de cada tipus, coincidint també amb els resultats obtinguts en els treballs previs,

l'autora troba un nombre superior d'errors d'hiposegmentació en totes les tasques realitzades. S'observa la influència de la categoria gramatical en el fenomen de la hiposegmentació, en totes les tasques menys en la de segmentació de grups de dues paraules en què aquesta influència no es nota. Aquesta influència es manifesta en un nombre més alt d'hiposegmentacions si les frases comencen amb un element gramatical en lloc d'un element lèxic. Una altra dada molt interessant és la que deriva de l'última tasca demanada: la del recompte de paraules. Les conclusions a què arriba Cintas amb els resultats obtinguts és que *"para poder realizar la tarea metalingüística de contar palabras de la cadena oral, niños y niñas tienen que saber escribir convencionalmente"* (2000:236). Finalment, destaca la influència de l'escolarització en l'adquisició de la competència a segmentar convencionalment el text.

Tal com dèiem al començament, aquesta revisió ni pretén servir de base comparativa per al nostre estudi. Les diferències en la majoria d'aspectes tractats en fan difícil la comparació, però malgrat tot alguns dels resultats poden servir-nos de referència en el moment de les conclusions i discussió finals.





## **II. LA RECERCA: LA SEGMENTACIÓ DEL TEXT D'UN GRUP D'INFANTS DE PRIMER CURS DE PRIMÀRIA**



## 5. OBJECTIU GENERAL I PREGUNTES DE RECERCA

Tal com ha estat explicat a la Introducció, els infants que aprenen de llegir i d'escriure segons plantejaments metodològics relacionats amb el model constructivista aborden la tasca d'escriure guiats per l'objectiu comunicatiu de produir un text significatiu. L'abordatge d'una tasca complexa que mobilitza coneixements de naturalesa diversa, escollint la producció d'un missatge com a fil conductor, té com a conseqüència la producció d'escrits amb errors ortogràfics de diversa mena, els quals mostren moments diferents en la construcció del coneixement sobre el sistema d'escriptura.

La segmentació, tema escollit com a objecte d'estudi, és un dels aspectes rellevants en les primeres produccions dels infants perquè la majoria presenta errors en l'establiment dels blancs convencionals. Estudiar els errors que es produeixen en el procés d'aprendre a segmentar convencionalment té un interès que va més enllà de la simple consideració ortogràfica que podrien tenir altres qüestions com l'escriptura de les vocals i consonants, o de l'accentuació. Podríem considerar que aquest interès deriva de la pròpia naturalesa dels segments delimitats pels blancs convencionals, que solem anomenar *paraules*. La paraula, malgrat la controvèrsia que suscita entre els lingüistes, tal com ha estat explicat a la primera part, és considerada una de les unitats bàsiques de la percepció de la parla, al costat del fonemes, les síl.labes o d'altres segments superiors, els quals han estat proposats com a unitats estructurals i funcionals de la percepció (Flores d'Arcais, 1992:127).

Per assolir la competència en el que anomenem ortografia convencional d'un text -al marge, doncs, de consideracions relatives a la seva producció-, els infants que s'inicien en l'escriptura han d'elaborar coneixement relatiu a tres àmbits diversos:

a) El primer àmbit és el grafofònic i abraça, en primer lloc, les qüestions referides al coneixement del valor sonor de les lletres i en segon lloc, el coneixement de les convencions que regeixen per a una llengua en qüestió i que fan, per exemple, que calgui saber una norma

per seleccionar correctament la *g* o la *j* per representar un mateix fonema. En aquest cas, estem parlant de la selecció de la lletra adequada, que suposa la capacitat d'analitzar el llenguatge oral en les unitats mínimes no significatives<sup>34</sup>. Ens movem, en aquest cas, en l'àmbit de l'escriptura dels elements que componen una paraula.

b) Un altre àmbit és el dels signes de puntuació, que ofereix dos camps una mica diferenciats: per una banda, els signes que representen els elements prosòdics del discurs, és a dir, els signes d'admiració i interrogació, principalment, i, per l'altra, els signes que, si bé en alguns casos tendeixen a ser relacionats amb la prosòdia, tenen la funció d'organitzar sintàcticament el discurs escrit.

c) El tercer àmbit, el que hem escollit com a objecte de l'estudi, és el que recull les qüestions ortogràfiques centrades en la capacitat d'establir blancs per delimitar les unitats gràfiques tal com la norma ortogràfica de cada llengua té establert. Els àmbits b) i c) estan relacionats amb la dimensió ideogràfica del sistema d'escriptura.

Els problemes de segmentació que mostren els textos dels infants que inicien l'alfabetització poques vegades superen el cicle inicial de primària. De forma força sorprenent i una mica diferent del que sol passar amb altres qüestions ortogràfiques, la segmentació convencional s'assoleix molt aviat. Aquesta rapidesa ha estat, en part, el motor que ens ha empès a centrar l'objectiu del present estudi en la segmentació, no solament per observar com es produeix el fenomen en la població estudiada, sinó per esbossar-ne alguna interpretació.

**L'objectiu general del present estudi és analitzar la segmentació dels textos, escrits en situació natural d'aula, d'un grup d'infants de 1r. curs de Primària, amb un doble objectiu: a) conèixer els errors i la seva evolució al llarg d'un període de sis mesos i b) de manera molt especial, detectar els factors diversos que incideixen en les solucions gràfiques adoptades.**

---

<sup>34</sup> El coneixement de la correspondència grafofònica i la possibilitat d'analitzar el llenguatge oral són coneixements indispensables –però no suficients– per afirmar que una persona està alfabetitzada (Solé, 1987)

Per a aquest objectiu, hem formulat les preguntes següents, orientades a obtenir informació del fenomen de la segmentació tal com es manifesta en els textos i a observar els aspectes evolutius, tant dels errors com de la llargada.

1. *Quins problemes de segmentació són més freqüents en els textos?*

- els errors d'hiposegmentació (EHO)

- els errors d'hipersegmentació (EHR)

2. *Com evoluciona la llargada dels textos?*

3. *Com evoluciona la competència en la segmentació al llarg d'un període de sis mesos?*

4. *Es pot trobar alguna correlació entre la llargada dels textos i el nombre d'errors?*

5. *Com es distribueixen els errors en el text escrit? Se'n produeix una concentració més gran a la segona part a causa de l'efecte cansament?*

6. *Es pot observar algun patró o tendència en els casos d'hipersegmentació (HR)?*

7. *A quin tipus d'estructura pertanyen, les seqüències hiposegmentades?*

8. *Es transgredeixen els límits sintagmàtics en les seqüències hiposegmentades?.*

9. *Anàlisi de l'estructura de les seqüències hiposegmentades.*

En les apreciacions ràpides dels textos dels infants, sembla que el tipus d'error que cometem més sovint és per la falta de segmentació; és a dir errors d'hiposegmentació (EHO). Aquesta impressió inicial és una dada corroborada, també, en l'estudi de referència ja citat (Ferreiro i al.,

1996) i en d'altres (Gomes de Morais, 1995; Báez & Cárdenas, 1999; Cintas, 2000).

Una de les raons que poden explicar aquest fet és de tipus cognitiu: escriure per separat cada una de les unitats que anomenem paraula - amb totes les reserves que la lingüística imposa amb relació al terme paraula-, exigeix una capacitat de reflexió sobre la llengua que els infants tot just estan construint<sup>35</sup>. D'una banda, podem afirmar que, des dels 3 anys aproximadament, els infants segmenten amb competència la cadena parlada (Teberosky, 1994:137) i tracten les unitats de forma aïllada, tant pel que fa a paraules de base lèxica (LX) com a paraules gramaticals (PG). Fins i tot coneixen el valor de l'article determinat, indeterminat o de l'absència d'article. És a dir, podríem dir que mostren un cert coneixement gramatical, encara que sigui un coneixement implícit no accessible a la consciència, el qual han assolit per un procés inconscient durant el desenvolupament del llenguatge. De l'altra banda, però, les primeres produccions escrites mostren errors en la segmentació, talment com si el coneixement implícit de les unitats, que els ha permès de parlar, no fos suficient o no fos directament accessible en el moment d'escriure. L'aprenentatge de l'escrit demanarà la reflexió sobre el llenguatge i el tractament conscient de les unitats.

Hem de suposar que, perquè les unitats siguin tractades aïlladament a l'escrit, caldrà que l'infant en tingui una certa representació, encara que sigui intuïtiva. I els coneixements, implícits, a què els infants recorren per segmentar l'escrit, són, segurament, diversos: el coneixement d'unitats -inferiors o superiors a la paraula-, la distinció grafofònica, els coneixements prosòdics, els coneixements sintàctics, etc. Les estratègies que cada infant utilitza poden veure's condicionades per la prioritització d'alguns d'aquests coneixements.

Les activitats d'alfabetització desencadenen i afavoreixen, probablement, el coneixement de les unitats, a través d'un procés en què es combinen diversos procediments: si bé per una banda la producció gràfica ajuda a l'establiment i fixació d'unes unitats, les

---

<sup>35</sup> Vegeu apartat 3 del Marc teòric.

*paraules gràfiques*, segons la convenció de cada llengua, per l'altra, es potencia l'anàlisi del flux de la parla per poder parlar de les lletres i les correspondències sonores. Les investigacions sobre la percepció de les paraules, o sobre el concepte de paraula, realitzades amb infants que no han iniciat el període d'alfabetització, mostren que les partícules gramaticals no són percebudes com unitats aïllades o no són considerades paraules (Papandropoulou & Sinclair, 1974; Berthould-Papandropoulou, 1978; Ferreiro & Teberosky, 1979)<sup>36</sup>. Tot això ens permet suposar que el fenomen de la hiposegmentació (HO), que entenem com una qüestió de no segmentació d'unitats, més que no pas d'annexió, és, probablement, més freqüent que el de la hipersegmentació (HR), que s'explicarà, segurament, per causes diferents.

Així doncs, en la primera pregunta partim de la hipòtesi que les seqüències hiposegmentades (SHO) són superiors en nombre a les hipersegmentades (SHR). En un altre moment ens demanarem per la naturalesa de la seqüències.

Quan haurem determinat la quantitat d'EHO i d'EHR al llarg del període estudiat, i la seva evolució, analitzats cas per cas i en grup, tractarem de trobar alguna explicació dels errors: es poden interpretar com la manifestació d'alguna teoria de l'infant? O són errors sense cap lògica i, per tant, no s'hi ha de buscar cap justificació?

Per respondre a la pregunta que ens formulem, partim de la base que els errors mostren l'estadi de construcció de coneixement del sistema d'escriptura, però en la mesura que aquest sistema es pot relacionar amb el nivell de representació del funcionament de la llengua que els aprenents han elaborat fins aquell moment. Els estudis sobre processos d'aprenentatge i, concretament, sobre l'aprenentatge de l'escrit, ens han conduït a considerar els errors com una manifestació no solament d'allò que els infants no saben, sinó de l'estadi en què es troba la construcció dels seus coneixements sobre algun aspecte del saber. Els

---

<sup>36</sup> Hem de considerar, igualment, els resultats d'una altra recerca (Karmiloff-Smith, 1986) que, tal com ha estat indicat, no mostren diferències entre la percepció de les partícules gramaticals i les paraules plenes.



errors són les solucions que, en un moment determinat, adopten davant dels problemes que se'ls plantegen.

En el cas de l'aprenentatge de la segmentació convencional, la hipòtesi que sostenim és que els errors comesos no són aleatoris sinó que són el reflex de les teories dels infants sobre el funcionament del sistema d'escriptura. Quan diem *no són aleatoris* ens estem referint a errors que no són casuals, sense lògica, sinó errors que afecten un determinat tipus d'unitat, que es manifesten en un tipus d'estructura i que es presenten amb una certa regularitat. Es podria dir que tendeixen a ajustar-se a un patró de funcionament.

De fet, en el treball ja citat de Ferreiro i col.laboradores (1996), les investigadores també tenen una hipòtesi de partida, encara que no hagi estat formulada com a tal. Opten per partir directament de l'estudi de les partícules que més freqüentment apareixen en les seqüències no segmentades. En primer lloc trien les d'una lletra: A, Y; en segon lloc, les de dues lletres: DE, EL, LA. La tria d'aquestes partícules, tal com recull la cita següent, no està feta amb un criteri morfològic o sintàctic, sinó estrictament gràfic.

"...podemos anticipar que la mayor parte de las hiposegmentaciones (uniones indebidas entre palabras, según la norma ortográfica vigente) deben aparecer cuando se deberían aislar secuencias *de pocas letras (dos o tres)* y, *sobre todo, cuando debería quedar una sola letra funcionando como palabra*, ya que sabemos que los niños hacen una cuidadosa (y muy sabia) distinción entre partes-letras, que pueden ser nombrables pero no son interpretables, y las totalidades interpretables, compuestas por varios elementos diferentes ordenados (generalmente tres como mínimo) (Ver Ferreiro, 1988)". (Ferreiro i al. 1996:62) [la cursiva és de l'autora].

La justificació d'aquesta opció l'hem de trobar en les aportacions de Ferreiro (1982) i Ferreiro i Teberosky (1979) als estudis sobre la psicogènesi de l'escriptura, tot i que aquests, a diferència de l'estudi de referència, estaven fets sobre població que tot just encetava el període d'escolarització.<sup>37</sup>

---

<sup>37</sup> Ferreiro & Teberosky (1979: 42)

Aquesta autora explica que els infants exigeixen un mínim de grafies per sota del qual una escriptura no és *llegible*, però ella mateixa situa aquesta exigència a l'etapa anterior a l'escriptura sil·làbica. En aquest sentit, doncs, així com és pensable que aquest criteri guïi les primeres produccions dels infants, en el cas d'infants que han arribat a una etapa alfabètica és més difícil de sostenir que aquesta sigui l'única explicació possible per explicar la hiposegmentació de moltes partícules gramaticals, i, probablement, s'haurà de pensar en la intervenció de més d'un factor: fònic, lèxic i sintàctic.

Així, si bé en alguns casos es parla de "*la preposició A*", o de "*la conjunció Y*", no hi ha, almenys en els documents consultats, la manifestació que cap altre criteri -ni morfològic ni semàntic- hagi estat tingut en consideració. Només una breu referència que obre el camí cap a un altre punt de vista:

"En conclusión, el tipo de letra que la norma ortográfica aísla parece jugar un rol decisivo en la frecuencia de hiposegmentación que es dable esperar. Está claro que hay, además, una diferencia en la categoría sintáctica imposible de controlar en este caso." (Ferreiro, 1996:63)

Aquesta breu referència a la sintaxi prové d'alguns dels resultats obtinguts en la no segmentació de la preposició A i de la conjunció Y en l'estudi citat, els qual fan pensar a Ferreiro que hi ha d'haver alguna altra raó a més de la gràfica que ha estat esmentada.

Al marge d'aquesta petita consideració, no s'hi esmenta la possible influència que el coneixement gramatical que els infants han construït de manera implícita durant el desenvolupament del llenguatge pugui tenir en l'aprenentatge del sistema d'escriptura. És a dir, tal com ja s'ha explicat, la qüestió de la segmentació és tractada des d'un punt de vista exclusivament gràfic i, així doncs, es parla sempre de lletres i poques vegades de preposicions o pronoms o articles. I s'analitza, només, la unitat de la seqüència que, suposadament, és -o són- la que no ha estat segregada del conjunt. Amb aquesta perspectiva es perd la possibilitat de projectar una mirada sobre el conjunt no segmentat per saber si aquesta seqüència té algunes característiques que expliquen la no segmentació.

La proposta de l'estudi és analitzar les *paraules gràfiques* en tant que formen part del conjunt més ampli on estan inserides, en lloc de considerar-les aïlladament al marge d'aquest context. Molt probablement, les decisions que impulsen els infants a situar els blancs no estan guiades només pel nombre de lletres, en tant que *lletres*, sinó en tant que *paraules*, o, millor dit, en tant que segments que poden no correspondre's amb les paraules gràfiques convencionals, però que tenen algunes característiques: en l'eix sintagmàtic es presenten estretament relacionades amb les altres paraules amb què formen una seqüència funcional i significativa.

És des d'aquesta perspectiva doncs, que en aquest estudi es defensa que l'objecte d'anàlisi l'han de constituir tant les unitats enteses com a elements gramaticals -no solament en les seves característiques gràfiques- com el conjunt de la seqüència de la qual formen part.

Quins arguments sostenen aquesta hipòtesi?

En primer lloc, una observació ràpida dels textos analitzats indica que hi ha dades que reforcen aquesta suposició. Una d'aquestes dades, per exemple, és que els elements d'una lletra o de dues, quan no apareixen de forma aïllada, formen gairebé sempre un agregat amb la paraula que precedeixen -en el cas de les preposicions, articles, auxiliars verbals i pronoms proclítics i conjuncions- o amb la paraula a la qual segueixen -en el cas dels pronoms enclítics. És a dir, sembla que l'establiment de blancs respecta unitats superiors a la paraula no de forma aleatòria; unitats que podrien tenir relació amb el sintagma o amb agrupacions que tenen coherència sintàctica i semàntica. Per posar un exemple, sembla que mai no apareixen seqüències del tipus: [LEX+ prep] o [INF+article]: *\*vaig anara Cadaqués* o *\*vaig menjaruna coca*.

De fet, doncs, encara que d'una manera intuïtiva, sembla que no hi ha problemes d'hiposegmentació entre sintagmes<sup>38</sup>, sinó que quasi

---

<sup>38</sup> El terme sintagma ha estat utilitzat amb sentits diversos en la lingüística i en els manuals de gramàtica (vegeu nota 12). En aquest estudi, utilitzarem la definició del Diccionari de la Llengua Catalana (1995): "*Conjunt d'unitats lingüístiques que mantenen una relació de combinació dins la cadena parlada, a diversos nivells*

sempre els errors apareixen entre els elements que constitueixen un sintagma. És a dir, a primer cop d'ull sembla que en les SHO no hi apareixen unitats que pertanyin a dos sintagmes diferents, com si no es transgredissin els límits del sintagma, unitat estructural i funcional, superior a la paraula. Aquesta suposició estaria relacionada amb les tesis d'alguns psicolingüistes que defensen que el sintagma és una de les unitats de percepció en la segmentació de l'oral, al costat de la síl·laba i la paraula (G. B. Flores d'Arcais, 1988), (Hirsh-Pasek et al. 1987, a Karmiloff-Smith, 1994:59, a Mehler, 1992:186).

Finalment, la idea apuntada per Ferreiro sobre una possible influència de la categoria sintàctica dels elements no segregats ens estimula a seguir el fil de la seva suposició. De fet, pel que fa al llenguatge oral, en el període de desenvolupament, els infants mostren una preferència cap a les manifestacions lingüístiques que s'ajusten al patró lingüístic de la llengua de l'entorn i, pel que fa a la sintaxi, envers els enunciats que respecten l'estructura sintàctica de la llengua (Mehler, 1992).

Pel que fa als EHR, i seguint la hipòtesi que la majoria d'errors que els infants cometem no són casuals, es pot suposar que la segmentació no convencional realitzada reflecteix alguna idea sobre les unitats lingüístiques.

Seguint el fil d'aquestes argumentacions, en aquest estudi, l'anàlisi s'adreçarà al conjunt de la seqüència, entenent que l'error -o errors- no s'ha comès sobre una unitat que ha estat transcrita gràficament de forma errònia, sinó sobre un conjunt que té un significat, una estructura i una funció determinats, i que les SHO i les SHR responen a l'aplicació, segurament intuïtiva, d'algun tipus de coneixement gramatical i estan en relació amb estructures lingüístiques superiors a la paraula, que podrien ser els sintagmes o, simplement, agrupacions - grups o seqüències- amb coherència sintàctica i semàntica.

---

*d'anàlisi*". Hem aplicat el terme a aquelles seqüències o estructures o conjunts que tenen un valor semàntic i sintàctic, el nucli de les quals és un nom o un verb i només un. La resta d'elements o són relacionals o en són dependents. En aquesta definició no es considera la relació sintàctica amb els altres sintagmes que formen l'oració i no s'associa sintagma a constituent immediat de l'oració.

Per respondre a aquesta pregunta serà necessari comprovar si es confirma la intuïció que les seqüències no segmentades respecten els límits entre els sintagmes. Una resposta afirmativa indicaria un cert coneixement gramatical implícit, d'acord amb les teories de Karmiloff-Smith (Karmiloff-Smith, 1994).

Una anàlisi més aprofundida d'aquestes seqüències proporcionarà informació de quines són les estructures on majoritàriament es concentren els errors i quines són les paraules gramaticals involucrades en les seqüències errònies.

Si, d'acord amb aquesta idea, sostenim que hi ha d'haver factors de tipus semàntic, morfològic i sintàctic, caldrà triar una unitat que doni informació de la naturalesa morfològica o semàntica dels elements i que permeti, a la vegada, analitzar l'organització interna i les relacions entre els components de la seqüència no segmentada. Per tant, doncs, les unitats d'anàlisi que s'utilitzaran hauran de ser forçosament diferents de les que es van triar en l'estudi ja citat, perquè puguin donar compte dels coneixements que han guiat la solució gràfica dels infants.

Pel que fa als aspectes evolutius s'observa, encara que de forma aproximada, que la llargada dels textos augmenta de forma notable, però cal saber-ne l'increment de forma exacta, tant en xifres absolutes com en percentatges.

L'evolució en la capacitat d'escriure textos més llargs és una dada interessant en ella mateixa, perquè informa dels progressos dels infants, però també perquè serà una dada que es posarà en relació amb l'evolució dels errors. És a dir, ¿com es presenta en cada infant l'evolució de les dues mesures: la de la llargada i la dels errors? Hi ha algun tipus de relació?

La hipòtesi de partida és que, a mesura que avança el curs i els infants augmenten la seva competència en l'escrit, han de ser més capaços de segmentar convencionalment, tractant les unitats aïlladament i que, per tant, tant els EHO com els d'EHR han de disminuir percentualment, encara que no en xifres absolutes. S'analitzaran dos aspectes de la qüestió: l'evolució en cada subjecte i la comparació

entre els subjectes, per mirar si el progrés es produeix de forma semblant en tots els casos, amb relació al punt de partida.

Malgrat tot, una altra hipòtesi és possible: atès que escriure textos més llargs i més complexos requereix un esforç cognitiu més alt, adreçat a resoldre qüestions de gestió del text, podria ser que l'atenció disminuís i el percentatge d'errors es mantingués o, fins i tot, augmentés.

Els textos del tercer període<sup>39</sup> són clarament més llargs que els del començament. La llargada implica, també, complexitat en la gestió del text. No és el mateix escriure un text d'una frase, com molts dels primers exemples, que fer-ne un altre que quadruplici el nombre de paraules i de frases. No es tracta només d'una qüestió de nombre de paraules, sinó de complexitat, sobretot pel que fa a l'organització dels fets que s'expliquen i a la pròpia estructura sintàctica, que introdueix la utilització de més complementació i de subordinació. La complexitat sintàctica porta aparellada la utilització de paraules gramaticals - preposicions i conjuncions- que són més difícils de percebre i d'aïllar i poden, per tant, ser causa de més errors.

Així mateix, també podria ser que, durant la realització d'un text llarg, es pogués apreciar una diferència en la distribució dels errors al llarg del text perquè aquests es concentrassin en la segona part del text. Aquest fet indicaria que es pot parlar d'un factor cansament, fruit de l'esforç cognitiu exigit per l'escriptura d'un text llarg i complex i que tindria repercussions en la comissió d'errors.

Segurament, doncs, és interessant observar l'evolució en el nombre d'errors, així com la distribució, perquè el resultat pot confirmar la hipòtesi de la disminució dels errors o bé conduir-nos cap a una altra interpretació.

En resum, les hipòtesis de partida, que s'haurien de poder confirmar o rebutjar al llarg de l'estudi, són les següents.

---

<sup>39</sup> Tal com s'explicarà al punt 6. Disseny de la recerca, el total dels sis mesos estudiats ha estat dividit en tres períodes de dos mesos cada un, que s'indiquen de la següent forma: 1P, 2P, 3P.

*Pel que fa als tipus, estructures i freqüències d'errors:*

- ▶ El nombre d'errors d'hiposegmentació és més elevat que els d'hipersegmentació.
- ▶ La seqüència hiposegmentada no és aleatòria, sinó que els elements aglutinats pertanyen a un segment, superior a la paraula, que té una unitat semàntica i sintàctica i que anomenem sintagma.
- ▶ Les seqüències no segmentades més freqüents són probablement les que responen al model [PG+PG] o [LEX+PG], per damunt de l'estructura [LEX+LEX]. Dit d'una altra manera, les PG són les unitats més difícils d'aïllar i són les que apareixen més sovint en les seqüències hiposegmentades.
- ▶ En els casos d'hipersegmentació, els fragments erròniament aïllats corresponen, molt probablement, a partícules gramaticals o a lexemes existents, la qual cosa indicaria que la separació no és completament aleatòria, sinó que respon a la interpretació que el nen o nena fan de la gramàtica de la seva llengua en un moment determinat.
- ▶ És probable que en els textos més llargs, s'observi una certa concentració d'errors en la segona part del text, fruit de la complexitat en la gestió de textos més llargs, que exigeix un esforç cognitiu més alt.

*Pel que fa als aspectes evolutius*

- ▶ Segurament es produeix una disminució del nombre d'errors de segmentació al llarg del període analitzat, malgrat que augmenti la llargada dels textos.

## 6. DISSENY DE LA RECERCA

Les recerques sobre temes com el que presentem poden ser abordades des de perspectives diverses, en funció de la disciplina en què s'emmarca la recerca i dels objectius perseguits. En el nostre cas, el tema de la segmentació és, com la majoria de temes, polièdric en el sentit que pot ser objecte d'estudi de disciplines diferents -la lingüística, la psicolingüística, la història, l'antropologia, etc.- que, a la vegada, es plantegen objectius diferents. L'estudi que presentem s'inscriu en l'àmbit de la recerca en didàctica de la llengua, si bé presenta alguns matisos que caldrà explicar.

En primer lloc, la recerca en aquesta àrea parteix de l'assumpció que no té sentit considerar els fenòmens relatius a l'ensenyament i l'aprenentatge de la llengua si no és en el si de la interacció de tots els elements que hi estan implicats. Avançar en el coneixement de la incidència que tenen aquests diferents elements en l'aprenentatge escolar de la llengua es dibuixa com un dels objectius prioritaris per a l'àrea d'aquesta didàctica específica (Camps, 1998). I per avançar en aquest coneixement cal partir d'un replantejament de la relació que s'estableix entre la teoria i la pràctica. Segons Milian i Camps (1990: 24), *"el camp d'observació i d'acció de la didàctica (...) és la realitat dels contextos educatius. (...) L'interès principal és comprendre el funcionament dels aprenentatges (...)"*. En conseqüència, aquestes dues autores assenyalen l'objectiu de la didàctica com *"l'afany de descobrir i comprendre, per actuar-hi, la relació entre els factors d'ordre divers que hi incideixen"*.

Aquest plantejament entén que la realitat socioeducativa és complexa i que les investigacions que s'hi realitzen han d'assumir aquesta complexitat, així com *"la imprescindible participació de la subjectivitat en el procés d'explicació-intervenció en la realitat"* (Noguerol, 1998:63). És un model que respon al que se sol qualificar com a recerca natural pel fet que les dades s'obtenen en una situació habitual d'aula, amb la mestra de l'aula, i realitzant tasques que no representen cap canvi en relació al que els nois i noies estan



acostumats a fer. En el nostre cas, a més, no hi ha hagut la presència de cap observador extern i ha estat la mateixa mestra qui ha conduït la tasca i ha recollit les dades. Si ens atenem als quatre tipus de recerca en didàctica que assenyala Jaffré<sup>40</sup>, el present estudi s'ajusta a les característiques de les recerques descriptives, però amb una finalitat interpretativa:

"[recherches]qui ont pour but d'observer les donnés de la réalité éducative (pratiques, faits, comportements) afin de les décrire. On a dans ce cas affaire à une recherche descriptive dont les résultats peuvent également apporter des explications" (Jaffré, 1992:155).

Subratllem la segona part de la cita, en la mesura que el resultat de l'anàlisi permetrà trobar explicacions del fenomen de la segmentació.

La majoria de recerques consultades, excepció feta de la realitzada per Ferreiro, Pontecorvo, Ribeiro Moreira i García Hidalgo (Ferreiro i al., 1996), responen a un altre tipus de model, que s'acosta més al de la recerca experimental, en el sentit que s'han dissenyat tasques específiques per als objectius plantejats, partint d'unes hipòtesis que caldrà verificar experimentalment. En aquest tipus de recerca, pot ser que les tasques es realitzin a fora l'aula, com ara les entrevistes individuals, o a l'aula, amb la pròpia mestra, però amb un disseny pautat. Algunes de les seves característiques és que hi ha un control més estricte de les diferents variables que poden influir en els resultats; variables referides a la població, com el nivell social i cultural dels nens i nenes o el seu rendiment escolar; o les que es refereixen a la pròpia tasca, com ara el tipus de text, les paraules o les estructures sintàctiques, etc.

Les dificultats i els desavantatges que pot comportar la recerca de tipus natural es concentren principalment en la diversitat de les dades obtingudes, cosa que n'exigeix un tractament diferent que tingui en compte la complexitat de la situació. Es corre el perill de perdre's en un munt de dades, fruit d'una descripció minuciosa i aprofundida, sense que aquestes condueixin a l'establiment de tendències o a patrons

---

<sup>40</sup> Els altres tres tipus són, segons Jaffré: la recerca teòrica, la recerca experimental i la recerca-acció.

explicatius. Com que no es parteix d'una tasca dissenyada prèviament per a l'objectiu de la recerca ni hi ha un control de les variables tan estrictes com en el model anterior, això explicaria la diversitat de les dades. Però, malgrat aquestes dificultats, la recerca de tipus natural i basada en dades observacionals ha de permetre accedir a una visió més general del conjunt de la tasca i emmarcar aquestes dades en una situació real d'ensenyament i aprenentatge, amb una activitat i amb uns agents concrets. I no solament això, sinó que cal arribar a l'establiment d'algun model o patró interpretatiu que sigui útil per a d'altres casos similars.

En aquest sentit, els resultats no són vàlids només per a una situació concreta sinó que en la mesura que aquesta situació té molts punts de coincidència amb moltes d'altres (per nivell escolar, tipus d'escola, metodologies, inici d'alfabetització, etc.) les dades obtingudes mostren una certa homogeneïtat i els resultats es poden fer extensibles a d'altres situacions similars.

Després d'haver afirmat que l'estudi que presentem és una recerca en didàctica i que s'ajusta a un model de recerca natural, hem dit que calia fer algunes precisions. En el nostre cas, no s'han tingut en compte tots els elements que configuren una situació real d'ensenyament i aprenentatge, perquè, si bé, la situació en la qual s'ha produït l'objecte d'anàlisi -els textos realitzats pels infants-, és l'habitual de l'aula, no tots els elements han estat analitzats: ni el procés de producció, ni les intervencions de la mestra durant l'execució de la tasca, ni els intercanvis entre iguals. L'atenció s'ha posat en l'anàlisi del producte acabat per determinar la competència dels infants en el cas de la segmentació i l'evolució d'aquesta competència.

Una recerca conduïda en aquestes condicions hauria de poder donar compte de com els infants construeixen el seu saber en una situació educativa concreta i donar als mestres informació d'aquest procés. El nostre interès tant per a la formació inicial com per a la continuada és poder millorar les situacions educatives.

## 6.1. Establiment del corpus

En aquest apartat donem algunes informacions sobre els centres amb els quals hem treballat i els infants escollits, justificarem els criteris de la tria i les característiques d'alguns aspectes que poden ser rellevants, com ara la metodologia d'aprenentatge de la llengua escrita.

### 6.1.1. Les escoles

Els nens i nenes estudiats pertanyen a dos centres públics: l'Escola A, pertany al municipi de Cerdanyola del Vallès, encara que està situat als afores, i l'Escola B, està ubicada en ple barri del Guinardó de Barcelona.

L'Escola A acull preferentment nois i noies que es desplacen a aquesta escola, mentre que l'Escola B atén la població de l'entorn. En tots dos casos, la majoria famílies són de classe mitjana, algunes tenen com a llengua familiar el català i d'altres el castellà, i amb un percentatge molt baix de població immigrada.

En tots dos centres, la llengua vehicular és el català, en la qual es realitzen els aprenentatges i s'inicia el procés d'alfabetització.

El medi cultural en què creixen els infants és un factor que pot influir en els resultats que s'obtenen i, per tant, es buscava que els centres no fossin marcadament diferents entre ells ni representessin grups situats en els extrems del ventall social. Les dues escoles eren centres públics amb una població no gaire diferent des d'un punt de vista econòmic, social i cultural; les famílies pertanyien a un estàndard que podríem qualificar de mitjà: ni eren classe baixa o deprimida, ni classe alta o d'elit. Ens vam basar en el tipus d'escola -pública- i en la ubicació, perquè pensàvem que aquests dos aspectes podrien garantir un tipus de població semblant. Aquest criteri pot semblar excessivament intuïtiu i ser, per tant, discutible, perquè no està basat en cap classificació rigorosa de "classe social", però en el nostre cas resultava suficient. De la mateixa manera, els subjectes triats responien a un nivell mitjà de coneixements i de desenvolupament

segons el criteri de la mestra, sense que hi hagués hagut cap mena de prova que permetés corroborar-ho d'una manera objectiva <sup>41</sup>.

Pel que fa als plantejaments metodològics per a l'aprenentatge de la lectura i l'escriptura, es constaten algunes diferències entre els dos centres al moment de la recollida de dades. Mentre que l'Escola A, seguia una metodologia molt més centrada en el coneixement de la correspondència grafofònica, més pautada i amb una introducció més jerarquitzada dels continguts, amb més control sobre la producció, l'Escola B havia introduït alguns canvis metodològics. La tradició de l'escola era, en línies generals, la mateixa que la de l'Escola A, però l'aula dels infants de l'estudi estava conduïda per una mestra que aquell any va introduir una manera de fer més global, no tant pautada, amb un estímul cap a una producció més lliure. En l'anàlisi es veurà si aquestes lleus diferències metodològiques es tradueixen en diferències en els resultats. En tot cas, però, durant el primer curs es produeix la reflexió explícita sobre el sistema d'escriptura, tant pel que fa a les relacions grafofòniques, com als aspectes ideogràfics del sistema: signe de puntuació, disposició en el paper segons el tipus de text, etc.

### **6.1.2. Els subjectes i la mostra recollida**

Han estat estudiats els textos de dos grups de 5 nens cada un, al llarg d'un curs, des d'octubre fins a març<sup>42</sup>. La selecció dels infants va ser feta per les mestres i les úniques demandes van ser que hi participessin nens i nenes i que no hi hagués cap infant que destaqués sobre els altres en la competència en llengua escrita ni en cap altre aspecte. No es va fer cap prova objectiva, sinó que el criteri de la mestra va ser el que va guiar la tria. L'objectiu de treballar amb nens que

---

<sup>41</sup> En molts dels estudis consultats, els infants estaven classificats segons el nivell escolar, i segons dos nivells socials i culturals: classe baixa i classe mitjana. Els criteris per classificar la població entre classe baixa i mitjana es basaven en el tipus d'escola (pública/privada) o en la ubicació del centre i la professió dels pares (Ferreiro, 1996; Gomes de Morais, 1995). En alguns casos, s'ha tingut en compte la competència dels subjectes i aquests han estat separats en grups diferents (Clemente, 1984; Báez & Cárdenas, 1999; Cintas, 2000).

<sup>42</sup> Els textos estudiats formaven part d'una recerca anterior sobre el coneixement ortogràfic: *Evolució de les estratègies ortogràfiques dels nens de primer de Primària*. EMER 1995-8 (CIRIT).UAB.

representessin un nivell mig era per poder suposar que el coneixement que mostraven podia ser més representatiu del conjunt de la població de primer, que no pas si la mostra hagués considerat nivells de coneixements molt dispars. En la mesura que representaven un nivell mitjà, els resultats obtinguts podien ser més fàcilment extrapolables.

Els nens escrivien, aproximadament, quatre textos per mes. La consigna de producció era escriure un conte o un altre tipus de text sobre qualsevol tema que els interessés: un fet que hagués passat o una pel·lícula que haguessin vist, etc. No hi havia cap indicació ni de temps ni de llargada, cosa que s'observa clarament perquè les diferències en aquest ítem són notables. La producció va ser individual, sense intervenció de la mestra ni dels companys.

Al moment de començar l'estudi, al mes d'octubre de primer curs, era la primera vegada que els infants escrivien un text d'una certa llargada per explicar algun fet, però no la primera que es trobaven confrontats a la tasca d'escriure. Això vol dir que podem considerar que tenien uns coneixements sobre l'activitat d'escriure, així com sobre les habilitats que hi estan implicades: tenien nocions del sistema d'escriptura i del que és un text. Els textos que havien escrit eren, quasi exclusivament, el nom propi a les produccions pròpies, l'escriptura de la data, algun rètol; però aquest curs era la primera vegada que se'ls demanava d'escriure un text de més complexitat i llargada, i no preparat prèviament. I, òbviament, sabien com es fa per escriure, com s'agafa el llapis, la direcció de l'escriptura, etc. Per tant, encara que era per a ells la primera vegada que escrivien un text, ja tenien molts coneixements que els permetien abordar la tasca sense haver de dedicar l'atenció a aspectes motrius i de traç ni d'organització de l'escrit sobre el paper.

Alguns dels seus textos mostraven encara omissions i substitucions de lletres, per la raó que encara estaven en fase de consolidació del sistema alfabètic. Aquest fet no es tenia en compte si l'objectiu de l'anàlisi quedava clar, però va obligar a fer suposicions i interpretacions. Com a exemple:

*quema gade* [que m'agrada] (Joel, 19)

En aquest cas, hi ha l'omissió de la lletra R.

la casa qua ansem *qum em qum prat* és Mas gran (Ferran, 3)

En aquest segon exemple, s'aprecia el dubte en l'escriptura que porta a en Ferran a repetir un segment similar: qua ens hem comprat; que hem comprat. Només s'han fet aquelles interpretacions que era possible fer si el text i el fragment ho permetien. En els altres casos, s'ha eliminat el fragment.

Aquest aspecte no apareix en algunes de les investigacions consultades, ja que en la majoria de casos la població està ja en el segon curs de primària i el sistema alfabètic està assolit. Només en el cas de Clemente (1984), Báez & Cárdenas (1999) i Cintas (2000), s'analitzen les produccions d'infants més petits. Concretament, en aquest últim cas, es diu explícitament que els subjectes més petits estan encara en una fase de transició cap al domini de l'escriptura alfabètica.

La majoria de textos han estat escrits en paper blanc i alguns, molt pocs, corresponents a l'Escola A, ho han estat en pauta Montessori. La lletra utilitzada pels infants és la manuscrita lligada -a diferència d'alguns casos en les investigacions consultades- fet que ha facilitat la tasca interpretativa, ja que no hi ha hagut massa dificultat per decidir si la separació entre les unitats corresponia a una simple separació de lletres o marcava la separació conscient entre unitats lingüístiques superiors<sup>43</sup>. En alguns casos, pocs, l'infant deixava molt d'espai entre les unitats, sense que aparentment, ni pel sentit, ni per la gramàtica, semblés que hi faltava cap paraula. Aquest espai ha estat recollit en forma de punts suspensius i aquests casos no han estat comptabilitzats. S'ha interpretat com una conseqüència del procés de domini gràfic de l'espai del full. Com a exemple:

"ivexfeat.....una canumes era an Joan" (Joel, 7)

*i [en] vaig fer una [de foto]que només era en Joan*

---

<sup>43</sup> En la majoria d'estudis consultats hi ha mostres dels dos tipus de lletra; o bé per la tradició escolar, o bé perquè l'ús d'una lletra o d'una altra varia de curs en curs. Evidentment, les mostres d'infants a partir de segon curs estan escrites ja en lletra manuscrita lligada.

En la recollida van ser desestimats 6 textos en total, perquè no havien estat acabats; els mesos de gener i de desembre no són complets a causa de les vacances de Nadal i també en falta algun per motiu de malaltia. Tot plegat va fer que no s'hagués obtingut el nombre teòric de textos previst (24, quatre per mes) i que la xifra per mes variés bastant o molt entre els infants, tant en el recompte final com en els recomptes parcials de cada període; per exemple, de l'Escola B, la Clara té 8 textos al tercer període, mentre que en Ferran només en té 5 i la Júlia, 10. Totes aquestes circumstàncies van condicionar les agrupacions de textos que es van realitzar, tal com s'explica més endavant. En total s'ha treballat amb 200 textos. Del conjunt dels textos finalment seleccionats, han estat desestimades 11 paraules -o segments buits que podien correspondre a una o més paraules. A l'apartat 6.2.2., explicarem les decisions preses al moment de categoritzar sobre aquelles paraules o seqüències poc intel·ligibles.

També hi ha hagut alguns problemes en l'establiment de les dates en què havien estat realitzats els textos: en alguns casos no hi havia data i aquesta s'havia d'establir amb relació a l'anterior i a la posterior, o segons el tema: per exemple, quan tornaven de vacances escrivien textos que parlaven dels Regals de Reis; això permetia situar-los a les primeres dates de gener. En d'altres casos, no coincidien les dates dels nens i les nenes de la mateixa escola; per exemple: la Clara té tres textos de desembre, amb data 7, 8 i 9, mentre que l'Emma, de la mateixa escola, en té dos, amb dates 6 i 7. ¿És un error en l'escriptura de la data? ¿o és que els infants tenien llibertat per escriure els textos en algun moment al llarg del dia? La veritat és que no teníem aquesta informació, però ens va semblar que aquesta manca no tindria una repercussió important en l'estudi.

## 6.2. Tractament de les dades

Aquest apartat informa de les decisions preses en relació a les agrupacions dels textos, tenint en compte com es va produir la recollida dels textos; els tipus d'errors considerats, els criteris per establir-los i les decisions preses en el moment de categoritzar; la classificació de les seqüències no segmentades (SHO) segons els

formants (LX o PG) i, finalment, l'organització sintàctica per veure quines són les partícules involucrades en la seqüència no segmentada.

### **6.2.1. Agrupacions dels textos**

Per les causes explicades al punt 6.1.2., que van fer que la consigna inicial d'escriure quatre textos per mes no fos sempre respectada i, per tant, no pogués comptar amb un nombre de textos idèntic per cada mes i nen, es va decidir d'agrupar els textos per tractar les dades en tres blocs només, corresponents a tres períodes, en lloc d'analitzar-les text per text. Entre la possibilitat de dividir, aleatòriament, el conjunt de textos de cada nen per tres, o bé constituir tres blocs que corresponguessin a agrupacions de dos mesos cada una, es va optar per aquesta segona. L'objectiu era que, d'aquesta manera, els resultats fossin més fàcilment comparables. Per tant, els sis mesos varen ser repartits en tres períodes de dos mesos cada un: el primer (1P) correspon als mesos d'octubre i novembre; el segon (2P), als de desembre i gener i el tercer (3P) a febrer i març. Com que cada agrupació bimestral pot tenir un nombre diferent de textos, s'ha treballat amb la mitjana de paraules i d'errors de cada període. D'aquesta forma, els resultats obtinguts donen igualment la visió evolutiva que es buscava sense haver de treballar amb xifres molt baixes. Hem de pensar que, sobretot als primers textos, el nombre de paraules és molt baix: per sota de 10 paraules (Clara, text 1: 6 paraules; Emma, text 2: 4 paraules).

### **6.2.2. Tipus d'error i decisions durant el procés de categorització**

#### **6.2.2.1. Tipus d'error**

S'ha considerat error aquella escriptura que no respecta la segmentació gràfica normativa del català. El nombre d'errors, expressat en percentatges, sorgeix de la diferència entre el nombre teòric de paraules i el nombre real en cada text. El nombre teòric de paraules és el que correspon a una segmentació convencional -unitats entre blancs o relacionades amb un apòstrof o un guionet-, mentre que el nombre real correspon a les seqüències limitades per blancs tal com apareixen al text. El percentatge d'errors és, per tant, el quocient entre el nombre d'errors i el nombre teòric de paraules. Més endavant,



s'expliquen amb més detall les decisions que s'han pres per tal de poder reflectir el nombre d'errors total de cada text amb el nombre d'errors de les dues categories amb què els hem classificat.

Per a l'anàlisi dels errors s'han establert dues grans categories d'errors: els errors d'*hiposegmentació* (EHO) i els d'*hipersegmentació* (EHR).

**Hiposegmentació:** s'han considerat en aquesta categoria aquells errors que consisteixen a no separar amb blancs -o guions o apòstrofs- dues unitats lingüístiques, tal com la normativa preveu. L'apòstrof i el guió són dos signes de puntuació l'ús dels quals segons la normativa de la llengua catalana requereix una explicació.

la disfressa malabran

puereta als reis

(La disfressa me la van portar els reis)

**Hipersegmentació:** aquesta categoria recull els exemples en què una unitat lingüística apareix dividida en més blancs -o apòstrofs o guions- del que és normatiu.

Ferran

Un Senyor  
Ba la ma a  
a la muntanya

---

(Un senyor va anar a la muntanya)

Per a l'anàlisi de la segmentació cal considerar la influència de l'apòstrof i el guió, dos signes que, en català, tenen incidència en les qüestions de segmentació, perquè podrien ser causa d'error més que no pas les fomes plenes de l'article:

### L'apòstrof

S'usa en català i en d'altres llengües romàniques -francès, italià, portuguès- per marcar a l'escrit l'elisió a l'oral d'alguna vocal, en situacions en què n'hi ha dues en contacte, entre una paraula gramatical i una paraula lèxica, o dues de gramaticals. Des del punt de vista de la segmentació, l'apòstrof només representa una manifestació gràfica diferent de l'escriptura d'un blanc. En català, l'apòstrof s'utilitza entre l'article i el nom, entre els pronoms febles i el verb, tant en posició proclítica com enclítica i entre la preposició DE i el nom o verb que introdueix<sup>44</sup>.

L'apòstrof, però, no representa totes les diferents parles d'una llengua, sinó que ha acabat essent, com d'altres grafies o signes, una convenció. No és gens estrany de sentir combinacions d'article i nom en què l'article és pronunciat en la seva forma plena i no elidida.

L'apòstrof s'usa tant en les paraules masculines com en la majoria de les femenines, en singular, -excepció feta dels casos de paraules que comencen amb *i (hi)* o *u (hu)* àtones. Les modificacions de les dues formes, però, no són les mateixes: per a l'article femení, l'apostrofació implica elidir la vocal que està en contacte amb la vocal del nom que precedeix. En el cas de l'article masculí, en canvi, es produeix l'elisió d'una vocal, la *e*, que no està directament en contacte amb la vocal del nom que precedeix. Si consideréssim que el factor pronunciació és el determinant a l'hora d'escriure, semblaria, a priori, que s'haurien de produir més casos d'hiposegmentació en les formes femenines que no pas en el de masculines.

No afecta totes les paraules gramaticals ni, aquelles que afecta, ho són en totes les situacions. La norma està clarament determinada i la seva aplicació és estable des de la promulgació de les Normes ortogràfiques, l'any 1913.

S'han considerat cinc possibilitats d'escriptures no convencionals en l'apostrofació:

---

<sup>44</sup> Per un coneixement de la normativa de l'apòstrof i del guió, consulteu Joan Solà (2002), *Gramàtica del català contemporani*, Barcelona, Empúries.

1. Supressió de l'apòstrof i al mateix temps de la vocal elidida, la qual cosa és mostra d'hiposegmentació: *lalbum* (Clara, 17).
2. Supressió de l'apòstrof i annexió de l'article a la primera vocal de la paraula següent: *la quip*.
3. Supressió de l'apòstrof, amb manteniment de blanc: *l os* (Clara, 17).
4. Utilització de les formes plenes, sense apòstrof: *el os* (Clara, 17)
5. Utilització indeguda de l'apòstrof: *de'els anjels* (Clàudia, 12).

Per a l'estudi de la hiposegmentació només interessen els casos del grup 1). En el cas de la hipersegmentació, interessen els exemples del grup 2), tal com s'explica a Decisions sobre el procés de categorització.

### *El guió*

És utilitzat entre pronoms i el verb, en posició enclítica, i entre alguns prefixos i el nom o en formacions complexes. La normativa sobre l'ús del guió, sobretot en aquest últim cas, ha sofert canvis al llarg dels temps, cosa que dóna una idea del grau d'arbitrarietat a què està subjecte. De fet, els textos dels adults alfabetitzats mostren moltes fluctuacions i un cert desconeixement de la normativa. En els textos analitzats, la presència de guions en els casos de prefixos o paraules compostes és molt baixa.

En el cas del guió hem considerat com a hiposegmentació la supressió del guió en els casos del pronom enclític: *dientli* (Emma, 17).

#### *6.2.2.2. Decisions durant el procés de categorització*

Durant el procés de categorització es van haver de prendre un seguit de decisions d'acord amb la casuística amb què ens trobàvem. Són aquestes:

- L'atribució a totes dues categories dels casos com *as tapagan* (Jaume, 2), *escum barti an*, (Joel, 21), per tal de poder fer un

recompte ajustat, ja que en aquests casos, l'escriptura presentava totes dues categories d'errors<sup>45</sup>.

- ▶ La comptabilització de tots els errors comesos sobre una unitat lingüística en el cas que se n'hi hagi comès més d'un; per ex. *Alan trada* (Jaume, 13) mostra dos errors d'hiposegmentació i un d'hipersegmentació.
- ▶ Les construccions amb apòstrof han estat considerades com a seqüències de dos elements i l'apòstrof, com un blanc.
- ▶ Els casos en què l'escriptura de les paraules era poc intel·ligible s'han resolt de dues maneres: les que no podien ser interpretades de cap manera han estat eliminades del recompte. Aquelles que podien ser interpretades han estat: a) comptades com a paraules, i b) categoritzades si s'hi podia interpretar un error. Exemple:

"Abans *a...i* vag na dascursio i .....al sac al tera i ..... " (Emma, 5)

En aquest text, la seqüència *a...i* ha estat comptabilitzada i categoritzada, perquè queda clar que es tracta de la paraula *ahir*; en canvi, els segments marcats amb punts suspensius, eren intel·ligibles i no han estat ni tan sols comptats.

El nombre de segments desestimats, en el conjunt de tots els textos, és d'11. En conjunt són un nombre irrellevant.

- ▶ Alguns casos han estat categoritzats de manera que pot semblar que el criteri aplicat sigui contradictori; per exemple: *la Eva* no ha estat considerat error i els casos de no contracció, com ara *de els* (Clàudia, 12), sí. Al primer cas, segons la normativa, hi correspondria un apòstrof, és a dir, un espai en blanc i, per tant, el recompte final no varia. En el segon cas, en canvi, varia el nombre d'unitats respecte de la solució normativa.

---

<sup>45</sup> Aquest criteri ha estat utilitzat en d'altres estudis, com a Gomes de Morais (1995).

Si bé el nombre d'errors sorgeix de la diferència entre el nombre de paraules teòriques i el nombre de paraules reals en cada text, pot semblar en alguns casos que s'hagi produït un error en el càlcul. Per exemple,

QUIQUE	Nombre teòric paraules	Nombre real paraules	Errors HIPO	Error HIPER
5 novembre	9	9	1	1

Si el nombre de paraules teòric és idèntic al nombre de paraules real pot semblar que no hauria d'haver-hi cap xifra a la casella dels errors. Però no sempre és així perquè, en alguns casos, el nombre d'EHO i el d'EHR és el mateix i, per tant, el resultat s'equilibra. Així doncs, el fet que el nombre teòric i el nombre real coincideixi no és indicador d'error zero sinó que cal comptabilitzar l'error produït i incloure'l a la casella corresponent.

### 6.2.3. Les unitats de les seqüències no segmentades (SHO)

En analitzar la segmentació i definir els fenòmens de la HO i la HR, assumíem, implícitament, que aquests fenòmens es produeixen entre paraules, tant si és per segmentació insuficient, com per una segmentació excessiva<sup>46</sup>. A banda de conèixer la quantitat i els tipus d'errors que es produeixen, un altre dels objectius era conèixer la naturalesa de la seqüència afectada pels errors i per saber-ho calia trobar un criteri de classificar les paraules per poder analitzar les diferents formes de composició i de relació entre elles en una seqüència no segmentada; i aquesta decisió ha representat una certa dificultat. Com que la hipòtesi és que en les SHO hi intervenen factors de tipus semàntic, sintàctic, fonètic i prosòdic, el criteri de classificació de les paraules que es triava havia de donar informació pertinent sobre aquests factors.

---

<sup>46</sup> Utilitzarem el terme de paraula -o paraula gràfica- per referir-nos a les unitats que convencionalment es presenten entre blancs, al marge, doncs, de les objeccions que han estat explicades a l'apartat 1 del Marc teòric.

Era evident, doncs, que el criteri no podia ser el del nombre de lletres, semblantment a com han enfocat la qüestió les investigadores d'un dels estudis consultats (Ferreiro i al., 1996), perquè, tal com ja s'ha explicat, en el seu cas tenia coherència perquè es vinculava a les seves línies d'investigació sobre el desenvolupament de l'escriptura, però en el present cas, no aportava la informació necessària per a les preguntes que ens fèiem i el plantejament del qual partíem. L'altra opció hauria estat utilitzar les categories gramaticals: nom, verb, etc. Però aleshores els inconvenients eren que apareixien massa categories.

Finalment, l'opció triada va ser classificar les unitats en dos grans grups, segons la seva càrrega semàntica i la seva funció en l'estructura del missatge. Les unitats amb una càrrega semàntica més forta i amb funció de nucli, com són els noms, adjectius, verbs: *paraules de base lèxica* (LX); i les unitats amb una càrrega semàntica més feble i amb una funció no de nucli, com els articles, determinants, clítics, preposicions i conjuncions: *paraules gramaticals* (PG). Les primeres pertanyen a classes obertes en el sentit que formen una classe teòricament il·limitada i que admet noves formacions. A la bibliografia consultada, es troben de vegades referenciades amb el nom de "paraules contingut" *content words* o *lexical items*, (Tolchinsky i Cintas, 2001). Les segones pertanyen a classes tancades, cosa que significa que és una llista limitada d'unitats. A l'estudi que acabem de citar, apareixen com a *grammatical words* o *functional items* i fan referència tant als morfemes lliures com als travats<sup>47</sup>; en el present treball, el terme se circumscriu exclusivament als morfemes lliures, de les categories que acabem de citar i tant si són monosíl·labes (àtones o tòniques) com bisíl·labes.

Som conscients que aquesta opció no deixa de ser limitadora, com sempre que es vol reduir la diversitat a unes poques categories. No tots els noms i els verbs que pertanyen a la classe oberta de paraules i que

---

<sup>47</sup> En paraules de les autores: "The first (content words) denoting 'material content' includes nouns and verbs, and usually adjectives; the second, denoting a 'relation' includes free morphemes such as conjunctions, prepositions and bound morphemes such as affixes marking categories of number, case, tense, and so forth (Sapir, 1958)" (Tolchinsky i Cintas, 2001:85).  
Vegeu, també, la nota 5.

poden actuar com a nucli de sintagma o d'agrupacions superiors a la paraula tenen més càrrega semàntica que alguns determinants que, pel fet de no ser nucli i de pertànyer a una classe tancada de paraules queden classificats al grup de PG. Per exemple, els determinants *tot*, *tota*, *tots*, *totes* o *molt*, *molta*, *molts*, *moltes*, per citar només alguns casos, podria considerar-se que tenen una aportació semàntica tan important com el ver *fer*, classificat com a LX. En el moment de l'aplicació del criteri triat, s'han pres algunes decisions que s'expliquen més endavant, en parlar del tipus d'agrupacions segons les unitats.

Malgrat les precisions que acabem de fer, el criteri és útil perquè:

- ▶ indica la funció que aquestes paraules poden tenir en el conjunt de la SHO: de nucli, o de partícula relacional o marca de dependència. La combinació i organització d'aquestes unitats caracteritzen sintàcticament la seqüència no segmentada i, d'aquesta manera, es pot saber si aquestes seqüències responen a algun criteri gramatical o, al contrari, no s'hi pot trobar cap tipus de regularitat que ens ho faci pensar. A la vegada, triant aquest criteri, es té en compte un altre aspecte que pot ser rellevant: el del nombre de lletres; les PG solen ser paraules mono o bisíl·labes, la qual cosa coincideix amb l'anàlisi feta per Ferreiro, en el sentit que les partícules que apareixien no segmentades en els textos del seu estudi, tenien poques lletres i corresponien a PG (articles, conjuncions, clítics, etc) (Ferreiro, 1996:62).
- ▶ aquestes partícules tenen, a més, poca càrrega fònica -la majoria són àtones-; característica molt important en el cas de la llengua oral<sup>48</sup>. Sabem, pels estudis fets sobre el

---

<sup>48</sup> La relació que es pugui establir entre paraules gramaticals i material innacentuat depèn de les llengües. Tal com ens indiquen els treballs de Slobin (1982), citat per Gleitman (1992), hi ha llengües com ara el turc en què els elements flexius són unitats accentuades, que no deformen les paraules del seu entorn i no són clítics; en aquest cas, pertanyen a una classe oberta. En el cas del català i en d'altres llengües romàniques que tenen una estructura diferent del turc, una part dels elements flexius com moltes paraules gramaticals -menys els pronoms personals forts, els adverbis i els determinants- que es presenten de forma aïllada, són formes innacentuades i es classifiquen en el grup de paraules tancades.

desenvolupament del llenguatge, que en les tasques d'identificació d'unitats a l'oral, o en les primeres produccions lingüístiques, els infants tarden més temps a reconèixer el material no accentuat que el que ho és. Entre les explicacions de les causes, n'hi ha que es basen en criteris fonètics i prosòdics i d'altres, en criteris sintacticosemàntics. Com diu L. Gleitman: "*Resumiendo, los datos interlingüísticos apoyan la idea de que la sílaba acentuada es perceptivamente muy relevante*" (Gleitman et al. a Newmeyer, 1992:182)

Ens podríem demanar si hi ha alguna relació entre aquesta dada i una altra que ens indica que les paraules gramaticals triguen més temps a ser percebudes i a ser utilitzades pels infants en comparació amb les paraules de base lèxica. Brown i Bellugi (a L. Gleitman, 1992:184), mostren que la parla dels infants sona "telegràfica" perquè ometen les paraules funcionals. Karmiloff-Smith afegeix que tant en tasques en directe (*on-line*) com en diferit (*off-line*), els infants no identifiquen les paraules gramaticals com a tals:

" Hay numerosos estudios que han mostrado cómo hasta aproximadamente los seis años de edad (y, en el caso de algunas tareas, incluso hasta más tarde) los niños no saben explícitamente que tanto las palabras de clase abierta («niño», «silla», «silencio», «correr», «pensar») como las de clase cerrada («el», «un», «a», «en», «cuando», «de») son *palabras*". (1994:75).

Els resultats d'altres investigacions s'expressen en el mateix sentit (Van Kleeck, 1984; Friederici, 1983, citats a Gombert, 1990), (Papandropoulou i Sinclair, 1974) entre d'altres. Llegim:

"También hay coincidencia en los datos en lo que concierne a la mayor facilidad para distinguir los sustantivos, verbos y adjetivos y una gran dificultad para aceptar que artículos, conjunciones, preposiciones y otros elementos de enlace sean palabras" (Ferreiro, 1996:52)

Aquesta dificultat es pot explicar tant per la poca càrrega semàntica d'aquestes unitats com per la seva atonicitat, que les fa menys perceptibles i, en definitiva, més difícils de reconèixer; o per ambdós factors.



El que acabem d'explicar amb relació al criteri per triar les unitats és aplicable al cas de les SHO, però pot ser que no ho sigui per analitzar els errors d'HR, en què la segmentació pot estar guiada per factors diferents dels que han guiat la HO. Si en aquest últim cas, la hipòtesi està fonamentada en el coneixement intuïtiu<sup>49</sup> que l'infant té del sintagma, és evident que en el cas de la segmentació excessiva, aquest argument no és vàlid. El sintagma representa una de les unitats de percepció, al costat de la paraula o de la síl.laba i, en aquest cas, l'escriptura normativa de les unitats i la no segmentació queden justificades per la funcionalitat de la seqüència. En canvi, és difícil saber el factor -o factors, segurament,- que explica la solució gràfica adoptada per l'infant en el cas de la HR. Per tant, creiem que l'explicació s'ha de buscar en l'anàlisi dels fragments segregats, per veure si s'hi pot suposar algun criteri gramatical o no<sup>50</sup>. Com que, si la nostra hipòtesi es confirma, el nombre d'errors serà notablement baix, la tasca d'analitzar cas per cas per mirar de trobar alguna explicació a les unitats hipersegmentades no serà difícil.

#### 6.2.4. Agrupacions segons l'organització de les unitats

De la combinació d'aquestes dues classes d'unitats -tal com han estat descrites a l'apartat 2.2.3.- en les seqüències no segmentades, en van sortir els següents grups:

##### 6.2.4.1. Subcategorització dels errors d'hiposegmentació

###### Grup 1

1a: LX+LX+(nLX) +(PG) (*scriptio continua*). Inclou les estructures agregades que contenen dues o més paraules lexemàtiques (LX), amb o sense paraula/es gramaticals (PG): *juganamlasmevasamigas* (Paula, 1) *felligas* (Rubèn, 18.), *amvacauraunaden* (Quique, 1), *fagratis* (Júlia, 10); *cullibulets*;

<sup>49</sup> Hirsh-Pasek i Golinkoff (1996), ja citats.

<sup>50</sup> En els resultats d'alguns dels estudis citats, s'apunta com a hipòtesi que, en els casos d'HR, els fragments segregats serien fragments interpretables com a paraules independents (G. de Morais, (1995); Clemente, 1984).

. estic jugant amb les meves amigues

(Estic jugant amb les meves amigues)

Grup 2

LX+PG+(PG). Inclou les estructures d'una única paraula lexemàtica (LX) i paraula/es gramaticals: *manpurtat* (Júlia, 12) *cabinges* (Quique, 8); *sapagan* (Clara, 17), *dor* (Jaume, 13), *vacust* (va acostar) ( ):

bosc i se li vacust serp.

(...bosc i se li va acostar [una] serp)

Grup 3

PG+PG +(PG). Engloba les annexions formades només per paraules gramaticals: *asba* (Paula, 16); *lem* (Quique, 15), *sa* (s'ha acabat), (Clàudia, 15), *ques* (cun vartian), (Joel, 20), *su pasan* (s'ho passen) (Clara, ):

avui fa sol les lletres surten a jugar i s'ho passen molt bé jugant.

(Avui fa sol les lletres surten a jugar i s'ho passen molt bé jugant)

L'element en joc que condiona la classificació és justament el que hem anomenat *paraula de base lèxica* (LX) o *paraula plena*, a l'entorn de la qual, en una estructura sintagmàtica, s'organitzen els altres elements.

L'únic element que justifica la separació de les estructures en G2 i en G3 és la presència o no de la paraula de base lèxica (LX) en la seqüència no segmentada. Es pot interpretar que en una seqüència com *manpurtat* l'aprenent no ha reconegut cap dels tres formants: el pronom, l'auxiliar i el verb, mentre que en *sa acabat* el subjecte ha aïllat -perquè l'ha reconegut- el formant *acabat* i no els altres dos. Aquest segon cas indicaria una competència més gran, pel fet que sembla raonable de pensar que els elements més difícils de percebre i, per tant, els últims problemes que es resolen són els que afecten a l'escriptura de les paraules gramaticals, justament per la seva atonicitat i baixa densitat semàntica.

Seguint el fil d'aquest raonament, considerarem que els exemples de segmentació no convencional podrien ser interpretats de manera diferents: així, les mostres d'*scriptio continua* indicarien el nivell més baix de competència en la capacitat de segmentar, pel fet que el subjecte encara no aïlla dues paraules lexemàtiques; el segon nivell estaria representat pels exemples del segon grup, en què s'associen paraules gramaticals a una paraula lexemàtica i el tercer nivell, correspondria als casos en què els elements no segmentats pertanyen a la categoria de paraula gramatical.

#### 6.2.4.2. Decisions durant l'agrupació de les dades

Com ha estat explicat al punt anterior, en l'aplicació d'aquests criteris ha calgut prendre algunes decisions sobre qüestions que podrien haver canviat la interpretació dels resultats.

- En relació als verbs que poden actuar o com a auxiliar o com a paraula lèxica, hem considerat PG les formes dels verbs *haver*, *anar* i *estar*, quan actuaven com auxiliars i LX en els altres casos. Així doncs, a *sanva* (a París) (Clàudia, 17. Escoleta), el verb *anar* no és un auxiliar i, per tant, aquesta forma ha estat categoritzat en el G2. En canvi, a *casva* (ufaga) (Jaume, 12. Escoleta) el verb *anar* és clarament la forma auxiliar del perfet perifràstic i en aquest cas, la forma ha estat categoritzada en el G3

- ▶ Alguns conjunts, malgrat la seva composició, ens ha semblat que, a més de classificar-los en els grups corresponents atenent a l'estructura morfològica, calia analitzar-los com una unitat lèxica: *alescondite*, (Clàudia, 12. Escoleta), *parapjempla*, (ídem).
- ▶ Els adverbis han estat considerats LX. La relació dels que apareixen als textos és: *aquí, allà, dins, avui, ahir, sempre, mai, després, abans, ara, també*.
- ▶ Les locucions preposicionals, com *a dalt de, a punt de, a sota de i al cap de* han estat considerades PG, independentment de la proximitat amb algunes paraules que pertanyen a la categoria dels adverbis, com en el cas de *a dins de* o *dins*.
- ▶ Els determinants han estat classificats a la categoria de PG.
- ▶ L'estructura *hi ha / hi havia* ha estat inclosa en el grup 2: PG+LX

#### 6.2.4.3. Subcategorització dels errors d'hipersegmentació

L'anàlisi dels EHR seguirà una altra pauta. Segons els resultats obtinguts en els estudis de referència, aquest tipus d'error és poc freqüent. Sembla doncs, que establir un criteri de classificació semblant al que hem plantejat per als EHO no caldria i que seria millor fer una anàlisi cas per cas, amb posterioritat, intentant aleshores de trobar els possibles criteris que hagin seguit els infants. Per exemple, si els fragments hipersegmentats corresponen a alguna paraula real de la llengua.

Malgrat tot, per una qüestió de facilitació metodològica, hem establert, *a priori*, les següents agrupacions:

##### Grup 1

Paraules que presenten un error d'hipersegmentació simplement pel fet que aquesta segmentació contravé la normativa. És el cas de les formes *per que* conjunció causal (Clara, 4)), les no contraccions: *a el*

cap de setmana (Rubèn,11), per al (el), en lloc de pel (Clàudia, 5), i algun cas de paraula composta: *superr garres* (Rubèn,19 ):

*ba dir per que*

(va dir perquè)

Grup 2

Separació errònia de la paraula: *a van dunada*, ( Júlia, 8), *e ni mal* (Joel, 21), *has tava* (Ferran, ):

*i has tava molt trist*

(i estava molt trist)

Grup 3

Conjunt de dues o més unitats que presenta problemes associats d'hiposegmentació i d'hipersegmentació: *la limentu* (Jaume, 9); *la basadari* (Emma, 16). Aquest grup de casos han estat comptabilitzats i analitzats per duplicat i categoritzats en els dos grans grups. En la presentació s'ha optat per col.locar-los a l'apartat d'hipersegmentació perquè, en general, representa un problema de trencament del lexema per desconeixement de la forma.

*i havia la basadari*

(hi havia l'abecedari)

El G2 i el G3 han estat separats només perquè presenten o no dos tipus d'error, però quan s'analitzaran des d'un punt de vista qualitatiu, és evident que seran tractats com un sol grup. És a dir, el cas de *la Gusti* pot ser explicat també com el d'*a van dunada*. En tots dos casos, les partícules inicials incorrectament segregades corresponen a PG: l'article determinat femení, la preposició *a* i l'auxiliar *van*. En tots dos casos s'ha fet prevaldre la imatge gràfica i el valor lèxic d'aquestes unitats, per damunt del valor gramatical.

El G1, en canvi, recull els casos en què s'ha transgredit la convencionalitat establerta, però no un criteri gramatical.

#### **6.2.5. L'estructura sintàctica de les seqüències hiposegmentades**

Conèixer l'estructura sintàctica de les SHO i la naturalesa de les unitats que les integren és una informació necessària per veure si els errors d'hiposegmentació són aleatoris o s'ajusten a un cert patró. El coneixement de les SHO donarà informació de si es produeixen més errors en unes estructures que en d'altres i de quines són les partícules que hi apareixen més freqüentment.

Dos han estat els aspectes analitzats: per una banda, es vol saber si, tal com les observacions prèvies indicaven, no es produeixen transgressions dels límits dels sintagmes; per l'altra, es persegueix saber el tipus de partícula gramatical involucrada més freqüentment en les seqüències hiposegmentades.

En el cas de la transgressió de les seqüències, es parteix de la hipòtesi que els infants tenen un cert coneixement intuïtiu de l'estructura i l'organització de la frase i que seria aquest coneixement el que faria que es cometessin pocs errors d'hiposegmentació entre elements de sintagmes diferents. Els sintagmes s'entenen, en aquest cas, com macroorganitzacions a l'entorn d'un element nominal o verbal, en general<sup>51</sup>.

En el cas de l'anàlisi de la composició de la seqüència hiposegmentada, també es parteix de la hipòtesi que els errors no són del tot aleatoris,

---

<sup>51</sup> Vegeu nota 38.

sinó que són una manifestació de la construcció del coneixement gramatical, i es considera a més que les decisions gràfiques que els infants prenen són fruit de multiplicitat de factors.

Trobar una manera d'agrupar les dades per obtenir una explicació significativa era difícil. Si s'enfocava l'anàlisi de les diferents estructures no segmentades tenint en compte la categoria gramatical, amb l'intent de veure si s'hi podia observar alguna regularitat, la variació era tan alta que no permetia trobar-hi cap estructura predominant. Posarem com exemple les estructures que apareixen en l'anàlisi dels textos d'un infant:

Clàudia

Nº text	GRUP 1	GRUP 2	Estructures	GRUP 3	Estructures
1				<i>Comun purcet</i>	Adv+Det
5		Agafa al seumenja iavia	Det +Nom Pf+Verb		
8		tucarla	Inf+Pf		
10		parapjempla	Prep+Art+Nom		
11				<i>Sa acabat</i>	Pf+Aux
12		Parapjempla Alescondite Alaterra Amart Iaspai	Prep+Nom Prep+Art+Nom Prep+Art+Nom Prep+Nom Art+Nom	Atot Iaspai Atots	Prep+Det. Art+Nom Prep+Det.
13				<i>Ma caigut Atots als nens</i>	Pf+Aux Prep+Det.
14		marivava	Pf+verb	<i>Asota de la tela Lem fet</i>	Pf+Aux
15				<i>Sa (acabat)</i>	Pf+Aux
16				<i>Sa (acavat)</i>	Pf+Aux
17		Dirli Sanva (a Paris)	Inf+Pf Pf+Pf+Verb		
TOTAL		12		10	

És a dir, del Grup 2, sobre 12 casos de no segmentació hi ha 6 estructures diferents. I només en tres casos es repeteix la mateixa estructura.

En el cas del Grup 3, sobre 11 casos de no segmentació hi ha 4 estructures diferents.

Aquest exemple és una mostra de la varietat d'estructures possibles, cosa que feia difícil de trobar una forma d'agrupar les dades; una anàlisi de totes les estructures seria una anàlisi excessivament fragmentada que acabaria sense donar una informació significativa.

Per evitar el problema de la fragmentació excessiva i considerant que l'interès se centrava en l'anàlisi de les partícules en el conjunt d'una seqüència sintàctico-semàntica, es va optar per seleccionar només algunes de les partícules que prèviament es consideraven susceptibles de ser més freqüentment hiposegmentades. Els criteris de selecció es van basar en la atonicitat, en el nombre de síl·labes i la densitat semàntica de les partícules. Les partícules finalment seleccionades han estat: totes les formes d'article, les preposicions A i DE, en totes les combinacions possibles, els pronoms febles, i els auxiliars *haver* i les formes del verb *anar* corresponents al perfet perifràstic, que, encara que no siguin àtons, són de baixa densitat semàntica. Les seqüències hiposegmentades (SHO) analitzades han estat: a) les que estaven organitzades a l'entorn d'un nom i b) les que ho estaven a l'entorn del verb en forma personal.

El criteri per seleccionar les estructures ha estat, principalment, el de quantitat. Són les que apareixen més freqüentment als textos. En el cas de [art+nom] és força evident perquè és l'estructura que, a més d'altres funcions, exerceix la de subjecte. En el cas de les estructures [prep+nom], les més freqüents en els textos analitzats són les introduïdes per les preposicions A i DE. Les raons les hauríem de trobar, segurament, en el tipus de text que els infants han produït: relats, que exigien marques de situació en el temps i en l'espai. Per exemple: *ai vag na al·soo* (Emma, 4) o: *Els fantasmas nomes surten a la nit* (Clara, 8). Encara que la temporalitat i la situació són les expressions més freqüents, hi ha exemples d'altres funcions sintàctiques: *Un monstra que anpaitave a tota la jent* (Clàudia, 15) *Ban juga a veibol* (Rubèn, 21) i d'altres.



Del cas de la preposició DE, n'hi ha també moltes mostres, que serveixen per indicar la pertinença: *La jen del popla* (Jaume, 12), l'atribució: *barret de pape* (Clàudia, 13), algun exemple de recció verbal: *Em vaig disfressar de pastanaga* (Clàudia, 14) i formant part d'una locució prepositiva: *I acap de 3 dias* (Jaume, 17).

En el cas de les estructures de nucli verbal, només s'ha seleccionat la combinació dels tres components: pronom(s) feble(s), auxiliar i verb, atès que l'objectiu d'estudi s'ha focalitzat en el comportament d'aquestes tres peces.

D'aquesta manera, l'estudi de les partícules es realitzarà conjugant dos criteris: un, el criteri de la pròpia naturalesa de la partícula i dos, l'organització sintàctica i semàntica de la seqüència en què la partícula està integrada. Aquest plantejament, que considerem coherent en relació als objectius perseguits, ha comportat una certa complexitat metodològica, tal com anirem explicant.

Finalment, les partícules i les estructures analitzades han estat les que es mostren a la següent taula d'exemple:

#### *Partícules i estructures analitzades (\*)*

	Nucli nominal	Nucli verbal	Altres possibilitats
Article	[art+nom]		
Preposició A	[A+nom [A+art.+nom]	[A+inf]	locucions preposicionals
Preposició DE	[DE+nom] [DE+art+nom]	[DE+infinitiu]	locucions preposicionals
Pronoms febles		[PF+aux+verb] [PF+aux] [PF+verb] [PF+PF]	
Auxiliars		[Aux+verb]	

(\*) Totes les formes de l'article, de la preposició DE i dels pronoms febles; l'auxiliar haver i les formes del perfet perifràstic; les contraccions.

L'aplicació de les estructures triades no ha estat exempta de dificultats, tal com dèiem. El fet de combinar l'estudi de les partícules

d'una banda i de l'altra, l'estudi de les estructures de les qual les partícules triades formen part ha provocat que en alguns casos es produïssin encavalcaments i que algunes partícules fossin analitzades en més d'una estructura. L'estructura [art+nom], per exemple, pot aparèixer sola, sense estar marcada per cap altra unitat o partícula, o bé formant part d'una altra estructura com [prep+art+nom].

En el primer cas, els únics problemes se circumscriuen a les diferents formes dels articles segons el gènere, el nombre i la lletra de la vocal inicial del nom que precedeixen.

Però, si el grup [art+nom] va precedit d'una marca de dependència, com les preposicions *a* o *de*, i forma part d'una estructura més àmplia, l'anàlisi és més complexa.

- a) *abui fan titllas i as titula La Caputxeta Varmella* (Clara, 3)
- b) *a la nit durmim per que tanim son* (Clara, 4)
- c) *atrapades a dins dalacoba* (Rubèn, 19)
- d) *van surti dala coba* (Rubèn, 19)

L'article *la*, a l'exemple b), pot ser analitzat en l'apartat [art+nom] o en el de [prep+art+nom]. La nostra opció ha estat considerar-lo en totes dues estructures perquè ha permès tenir informació del comportament de les partícules en els diferents contextos. Els casos en què, segons la normativa, la preposició i l'article formen una única unitat: *al*, *als* i *de*, *dels*, només han estat considerats dintre l'estructura [prep+art+nom], mentre que els casos en què l'article es presenta segregat de la preposició: *a la*, *a l'*, *a les*; *de la* i *de l'*, *de les* han estat considerats en totes dues estructures, tenint en compte el comportament de l'article i l'estructura de la qual forma part.

La perspectiva adoptada ha fet que en alguns casos, les mostres hagin estat comptabilitzades en dos apartats. Per aquest motiu no ha estat possible conèixer el percentatge d'errors d'hiposegmentació d'una estructura amb relació al conjunt d'errors d'hiposegmentació i només s'ha comptat el percentatge d'errors amb relació al nombre d'ocurrències en una mateixa estructura.

Per dir-ho utilitzant dos exemples: el percentatge d'errors d'hiposegmentació en l'estructura [art+nom] només podia ser comparat amb el nombre d'ocurrències d'aquesta estructura però no podia comparar-se amb el percentatge d'errors de l'estructura [prep+nom+art] ni tampoc amb el conjunt de tots els errors d'hiposegmentació calculats, tal com consten a la taula d'exemple.

En el cas de l'estructura de nucli verbal, ha passat un fet similar: el nombre tan elevat de combinacions possibles entre els tres tipus d'unitats que configuren l'estructura ha fet difícil establir paral·lelismes entre les seqüències hiposegmentades i les ben segmentades. Així, una SHO com *mal bindran a beura* (Paula, 7), era classificada en el grup [PF+PF], però *se la menjaba* (Paula, 12), podia estar classificada com a estructura correcta en dos grups: [PF+PF] i [PF+PF+verb].

Vista la dificultat de poder establir una relació entre les SHO i les ben segmentades, s'ha pres la decisió d'analitzar els següents aspectes:

- a) una anàlisi quantitativa de les SHO, en conjunt, i les ben segmentades. Aquestes dades han permès establir un percentatge d'errors comparable amb les altres estructures analitzades.
- b) una comptabilització, en xifres absolutes, de les SHO per cada un dels grups establerts. La informació que aquests valors proporciona és molt relativa, atès que el nombre de seqüències que apareixen és diferent per cada subjecte i no se'n pot establir cap proporció, així com tampoc, pel que acabem d'explicar, ens permet establir proporció entre les SHO de cada grup i les ben segmentades.
- c) una observació aprofundida de les SHO, focalitzant l'atenció en:
  - i) el comportament dels pronoms febles, fent atenció especialment en la incidència en l'apostrofació -formes elidides i reduïdes- en la hiposegmentació; ii) l'escriptura de formes pronominals i d'auxiliars verbals en una sola lletra aïllada; el comportament dels auxiliars *haver* i les formes *anar* del perfet perifràstic; i, finalment, les formes *hi ha* i *hi havia*.

L'exigència de precisió ha comportat, com es pot veure, una certa complexitat de l'anàlisi, justificada per la necessitat d'interpretar més afinadament el fenomen de la hiposegmentació.

Voldríem precisar, però, que, en totes les estructures analitzades, les dades quantitatives són el marc per a una anàlisi en profunditat, qualitativa, que facilitarà una observació i interpretació acurada de casos particulars. Un estudi amb textos reals és una font d'informació extraordinària, malgrat que a la vegada la mateixa riquesa de les dades en faci més difícil la sistematització.

De fet, un dels objectius d'un estudi d'aquestes característiques és trobar regularitats, explicacions, enmig d'un munt -aparentment caòtic- de dades. A base d'observar els textos amb minuciositat, s'acaba descobrint que les decisions gràfiques dels infants no són fruit de l'atzar sinó que estan guiades per coneixements de diversa mena sobre l'estructura de la seva llengua i sobre un sistema d'escriptura del qual, tot just, enceten l'aprenentatge.