

Lidia Cámara de la Fuente

Comunicación técnica multilingüe:  
gestión del conocimiento y de recursos lingüísticos  
para proyectos de traducción

Tesis Doctoral (Diciembre de 2005)

Dirigida por: Dra. M. Teresa Cabré Castellví

Programa de doctorado: Lingüística Aplicada (lèxic i discurs), 1998-2000

Institut Universitari de Lingüística Aplicada

Universitat Pompeu Fabra

INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA

UNIVERSITAT POMPEU FABRA

Barcelona 2006

Sèrie Tesis, 18

Responsable de la edició: Lidia Cámara de la Fuente

Direcció de las Publicaciones del IULA: Mercè Lorente Casafont

Coordinació de las Publicaciones del IULA: Lluís Codina, Gemma Martínez

Edició en CD-ROM: diciembre de 2006

© la autora

© Institut Universitari de Lingüística Aplicada

Pl. de la Mercè, 10-12

08002 Barcelona

<http://www.iula.upf.edu>

Diseño de la carátula: Cass

Soporte informático: Jesús Carrasco

Impresión: Institut Universitari de Lingüística Aplicada

Depósito legal: B-48.208-2006

ISBN 10: 84-89782-18-0

ISBN 13: 978-84-89782-18-1

## **PRESENTACIÓN**

La *Sèrie Tesis* de las publicaciones del Institut Universitari de Lingüística Aplicada (IULA) de la Universitat Pompeu Fabra es una iniciativa que pretende difundir las tesis doctorales defendidas en el marco del programa de doctorado en Lingüística Aplicada llevadas a cabo desde el año 1996 en el IULA. Esta nueva colección acoge también las tesis defendidas desde 1994 (año de creación del IULA) por profesores y profesoras que son miembros del Instituto, aunque las hayan leído en otros centros, si el objeto de estudio de la tesis está relacionado con los temas propios de nuestra investigación: lingüística de corpus, lingüística computacional, ingeniería lingüística, terminología, morfología, neología, lexicografía, análisis del discurso general y especializado, variación lingüística o representación del conocimiento.

Todas las tesis de esta serie se publican en la lengua original en la que fueron defendidas por sus autores; y los textos han sido revisados y editados por ellos mismos. Las tareas de edición final y de coordinación de cada volumen han sido realizadas directamente por miembros del Institut Universitari de Lingüística Aplicada.

Esperamos que esta iniciativa, que deseamos que tenga continuidad, llegue a los investigadores, profesores y estudiantes interesados en la lingüística aplicada y contribuya a difundir en nuestra comunidad temáticas aún poco consolidadas.



## **Perfil de la autora**

Directora del departamento de formación e investigación de STAR Servicios Lingüísticos. Licenciada en Filología Alemana e Hispánica por la Ruhr-Universität Bochum (Alemania), Máster en tecnologías aplicadas al entorno de Internet por el Institut Catalá de Tecnología de Barcelona. Título DEA sobre herramientas para la gestión del conocimiento por la Universitat Pompeu Fabra. Experiencia profesional como docente y como investigadora en el área de la tradumática, terminótica, lexicografía, terminología y gestión del conocimiento y de la información. Participación activa en numerosas conferencias, congresos y publicación de artículos relacionados con las áreas de especialización indicadas.

A Marlén, por ella todo

A Michael, con él más

## 0. Prefacio

Si esta tesis se compara con la tesis paralela *que no se escribe* y que alude a lo vivido durante todos los años de trabajo que acompaña una investigación y que no se desprende de la tesis escrita, es, sin lugar a dudas, corta, muy corta.

Mi tesis personal paralela está minada de muchas alegrías y descubrimientos pero también de silepsis, anacolutos, pleonasmos, cacofonías... que me llevaron casi hasta el abandono del trabajo de investigación. Éste que empieza con una mezcla de azar, curiosidad y ambición llega a convertirse en un reto personal.

Entender que el camino para llegar a algunos resultados es la meta y no los resultados en sí no se entiende a lo Machado, sino que se torna en un problema existencial. Y ese problema existencial sólo se sobrelleva y reconduce con afectos y amores de aquellas personas que con su apoyo te sostienen y animan de una forma absolutamente incondicional, seguros de que

vas a lograr lo que te has propuesto. Gracias a Michael, a mis padres y a mis hermanas.

Pero más allá del apoyo emocional, sin la experiencia empresarial, la formación académica y la experiencia didáctica recogida no me habría sido posible llevar a cabo este proyecto. La sinergia desprendida en las interacciones de las tareas que paralelamente he ido desempeñando, ha generado una serie de conocimientos que han culminado en el diseño conceptual y técnico de un sistema que además de cumplir los objetivos propuestos, es una muestra de la metabolización interdisciplinar y de mi perspectiva panorámica.

Esta simbiosis intelectual se apoya también en personas, en muchas personas. Por ello, quedo especialmente agradecida a los profesionales que han hecho posible la realización de este trabajo, sin cuya participación no hubiese alcanzado los objetivos fijados desde un principio. En este sentido agradezco especialmente a Teresa Cabré, la directora de esta tesis, por aceptar dirigir este trabajo y a Michael Scholand, director de STAR SL, por ofrecer aval logístico y financiero para la implementación del sistema diseñado, ambos con su criterio académico y empresarial respectivamente participaron desde los prolegómenos en el proyecto hasta la finalización del mismo; a Grzegorz Chrupala y Pamela Núñez quienes lograron transformar en *scripts* todas mis propuestas; a los gestores de proyectos de STAR SL que evaluaron los primeros desarrollos y con los que sin duda no habríamos podido mejorar el sistema implementado, a Araceli Alonso quien pacientemente corrigió todo lo escrito y a Gabriel Quiroz quien se sigue ocupando de administrar el sistema implementado.



# ÍNDICE DE CONTENIDOS

<b>0. PREFACIO</b> .....	<b>1</b>
<b>1. INTRODUCCIÓN</b> .....	<b>13</b>
1.1. Problema.....	17
1.2. Objetivos .....	20
1.3. Objeto de análisis.....	21
1.4. Supuestos de partida.....	24
1.5. Delimitaciones .....	25
1.6. Marco teórico.....	26
1.7. Metodología.....	29
1.8. Organización del estudio .....	31
<b>2. LA DOCUMENTACIÓN TÉCNICA MULTILINGÜE: RECURSOS LINGÜÍSTICOS Y TÉCNICOS</b> .....	<b>33</b>
2.1. A propósito del término “documentación” .....	34
2.2. La documentación técnica.....	37
2.2.1. Ingenios: artefactos y sistemas.....	38
2.2.2. Conocimiento técnico representado.....	40
2.2.3. Recursos para representar conocimiento técnico.....	43
2.2.3.1. Recursos lingüísticos: texto y estructuras conceptuales .....	48
2.2.4. Técnicas y tecnología de desarrollo de contenido técnico (multilingüe) .....	52
2.2.5. Los tipos / módulos textuales de la documentación técnica .....	61
2.2.5.1. Documentación .....	63
2.2.5.2. Recursos de software.....	66
2.3. Gestión y traducción de la documentación técnica multilingüe .....	71
2.3.1. La documentación técnica en su ámbito global .....	72
2.3.2. Ciclo de producción de la documentación técnica (multilingüe).....	76
2.3.3. Fase de traducción .....	82

2.3.3.1. Planificación del proyecto.....	86
2.3.3.2. Programación de producción .....	88
2.3.3.3. Envío de los datos al traductor.....	88
2.3.3.3. Traducción .....	90
2.3.3.4. Recepción del proyecto traducido → Control de calidad.....	91
2.3.3.5. Cierre del encargo.....	92
2.3.3.6. Postmortem.....	92
2.4. Gestión y tipos de recursos lingüísticos multilingües .....	93
2.4.1. Las memorias de traducción.....	97
2.4.2. Las bases terminológicas: diccionarios específicos y especializados .....	100
2.5. Formatos de codificación e intercambio .....	110
2.5.1. Unicode.....	112
2.5.2. XML .....	113
2.5.3. TMX .....	115
2.5.4. Martif.....	117
2.5.5. TBX.....	119
2.5.6. OLIF.....	119
2.5.7. XLT .....	119
2.6. Herramientas de gestión de recursos lingüísticos para la documentación técnica multilingüe.....	121
2.6.1. Sistemas integrados de traducción asistida.....	123
2.6.1.1. Gestores terminológicos.....	142
2.6.1.1.1. Extractores terminológicos.....	146
2.6.2. Herramientas no integradas en los sistema de traducción asistida ..	149
2.6.2.1. Gestores de tesauros multilingües .....	150
2.6.2.2. Gestores documentales → Gestores de contenido .....	155
2.6.2.3. Gestores de proyectos de traducción.....	157
2.6.3. Tendencias para la gestión integral del flujo total de la traducción...	159
<b>3. AUDITORÍA DE LOS RECURSOS LINGÜÍSTICOS MULTILINGÜES EN UNA EMPRESA DE TRADUCCIÓN: ESTUDIO DE CASO.....</b>	<b>163</b>

3.1. Estudio de caso: STAR Servicios Lingüísticos .....	164
3.2. Auditoría de la información como método de análisis .....	169
3.2.1. Fases del método auditor .....	171
3.2.1.1. Implicación y decisión por parte de la dirección de la entidad en implantar una política de la información .....	172
3.2.1.2. Identificación de una necesidad y evaluación de la viabilidad del proyecto .....	173
3.2.1.3. Planificación de la auditoría de la información .....	175
3.2.1.4. Análisis e interpretación de los datos para modelar un catálogo de propuestas técnicas y conceptuales .....	177
3.2.1.5. Implementación de las propuestas técnicas y conceptuales .....	178
3.2.1.6. Seguimiento y mantenimiento .....	179
3.3. Auditoría de los recursos lingüísticos multilingües: almacenamiento y explotación .....	179
3.3.1. Implementación y aplicación de una política de la información .....	180
3.3.2. Planificación del proyecto auditor .....	181
3.3.3. Análisis e interpretación de los datos.....	182
3.3.3.1. Análisis de los recursos lingüísticos almacenados.....	183
3.3.3.1.1. El corpus paralelo de STAR SL .....	185
3.3.3.1.2. Gestión del corpus paralelo con <i>Transit</i> .....	187
3.3.3.2. Análisis de los flujos de los recursos lingüísticos (RRL).....	191
3.3.3.2.1. Flujo de almacenamiento y acceso físico a los documentos	192
3.3.3.2.2. Gestión de las memorias de traducción.....	196
3.3.3.2.3. Gestión de los diccionarios (recursos terminológicos).....	203
3.3.3.3. Necesidades de los usuarios .....	210
3.3.3.4. Protocolo de cuellos de botella, incidencias o problemas .....	211
3.3.3.4.1. Problemas de almacenamiento, categorización y diversidad .....	212
3.3.3.4.2. Problemas de acceso a los recursos lingüísticos almacenados .....	213
3.3.3.4.3. Problemas de gestión de los recursos lingüísticos por falta de protocolos .....	217

3.3.3.4.4. Problemas de explotación de los recursos lingüísticos existentes .....	219
3.4. Diseño e implementación de las propuestas metodológicas, conceptuales y tecnológicas .....	220
3.4.1. Establecimiento de protocolos .....	222
3.4.1.1. Nomenclatura de los proyectos de traducción .....	222
3.4.1.2. Nomenclatura de los diccionarios.....	227
3.4.1.2.1. Clasificación de diccionarios por estado.....	229
3.4.1.3. Estructura de las subcarpetas de proyectos .....	232
3.4.2. Modelado para la gestión de los diccionarios .....	234
3.4.2.1. Dotación de nuevos puntos de acceso.....	235
3.4.2.2. Lenguaje de indización.....	237
3.4.2.3. Diseño del gestor de diccionarios .....	242
3.4.3. Modelado para la gestión de las memorias de traducción .....	246
3.4.3.1. Dotación de nuevos puntos de acceso.....	246
3.4.3.2. Diseño del gestor de proyectos.....	247
3.4.3.3. Diseño del gestor de corpus paralelos .....	249

#### **4. DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA INTEGRADO DE GESTIÓN DE CONOCIMIENTO Y RECURSOS LINGÜÍSTICOS:**

<b>COGNIMÁTICA.....</b>	<b>255</b>
4.1. Fase de diseño: especificaciones técnicas y funcionales.....	257
4.1.1. Diseño e implementación de la arquitectura de la información.....	258
4.1.2. Bases técnicas y funcionales.....	263
4.1.2.1. Estructura modular .....	263
4.1.2.2. Técnicas documentales: recuperación y extracción.....	268
4.1.2.3. Tecnología web.....	271
4.1.2.4. Ecología de la información y diseño centrado en el usuario .....	275
4.2. Fase de desarrollo de la versión 1.0. ....	277
4.2.1. El gestor documental de diccionarios .....	277
4.2.2. El gestor de proyectos .....	286
4.2.3. El gestor de corpus paralelos .....	292

4.3. Fase de pruebas .....	296
4.3.1. Selección de betatesters.....	297
4.3.2. Objetivos de los betatesters.....	298
4.4. Fase de rediseño y desarrollo de la versión 1.1.....	304
4.4.1. Fomento de la gestión del conocimiento.....	305
4.4.1.1. Acceso a la información: documentación en línea .....	306
4.4.1.2. Fomento de la comunicación: herramientas colaborativas.....	308
4.4.2. Herramientas de control.....	311
4.5. Seguimiento y mantenimiento de los recursos lingüísticos .....	314
4.5.1. Perfiles y competencias de los usuarios .....	316
4.5.2. Unidades de información: los recursos lingüísticos y su información asociada .....	320
4.5.2.1. Perspectivas de mantenimiento y seguimiento a medio y largo plazo .....	322
4.5.3. Seguimiento de COGNlímica.....	324
4.5.3.1. Perspectivas técnicas de afinamiento a corto plazo.....	324
4.5.3.2. Perspectivas técnicas de afinamiento a medio plazo: explotación de archivos TMX .....	326
<b>5. CONCLUSIONES Y REFLEXIONES FINALES.....</b>	<b>331</b>
<b>6. BIBLIOGRAFÍA .....</b>	<b>339</b>
6.1. Herramientas utilizadas para el diseño e implementación de COGNlímica .....	363

## ÍNDICE DE FIGURAS

Fig. 1: Objeto de análisis .....	22
Fig. 2: Interfaz de búsqueda de Eurodicautom.....	40
Fig. 3: Representación esquemática y discursiva del conocimiento.....	41
Fig. 4: Dificultad de transferencia de los niveles de conocimiento .....	42
Fig. 5: Gráfico de un microscopio.....	45
Fig. 6: Gráfico de la carrocería de un automóvil.....	46
Fig. 7: Muestra de un término con representación visual .....	47
Fig. 8: Formas textuales para la representación del conocimiento técnico .....	50
Fig. 9: Identificación para la catalogación de documentos .....	51
Fig. 10: Ontología Mikrokosmos.....	51
Fig. 11: Esquema .....	52
Fig. 12: Interfaz gráfica de cajero automático.....	54
Fig. 13: Menú contextual .....	67
Fig. 14: Cuadro de diálogo con advertencia.....	68
Fig. 15: Ventana de un asistente de instalación .....	68
Fig. 16: Tipología de los módulos textuales posibles en la documentación de un producto .....	70
Fig. 17: Creación de contenido técnico .....	78
Fig. 18: Flujo de gestión de los proyectos de traducción.....	85
Fig. 19: Ejemplo parcial de hoja de encargo para el traductor .....	89
Fig. 20: Lista de valores con información gramatical.....	104
Fig. 21: Códigos ISO para la representación de idiomas .....	107
Fig. 22: Tipos de información que componen el documento .....	114
Fig. 23: Ejemplo de codificación de memorias en TMX.....	116
Fig. 24: Ejemplo de codificación de terminología en Martif .....	118
Fig. 25: Concordancias resultantes de <i>Altar</i> en un corpus paralelo .....	130
Fig. 26: Barra de estado integrada en el editor específica para la traducción	133
Fig. 27: Ahorro potencial de costes con herramientas de traducción asistida	134
Fig. 28: Formatos de conversión de STAR Transit y Trados.....	137
Fig. 29: Información general de un concepto .....	143
Fig. 30: Información específica de un término.....	143

Fig. 31: Extracción y creación de terminología asistida.....	148
Fig. 32: Extracto de captura del tesoro SPINES.....	153
Fig. 33: Captura del tesoro utilizado por Google Scholar .....	154
Fig. 34: Componentes de LTC Organizer.....	158
Fig. 35: Integración de la tecnología en flujos de trabajo de la traducción técnica .....	160
Fig. 36: Aumento progresivo del volumen de proyectos gestionados (2002- 2004).....	166
Fig. 37: Evolución de las memorias de traducción almacenadas.....	183
Fig. 38: Evolución de registros de los diccionarios almacenados.....	184
Fig. 39: Porcentaje del nº total de recursos según combinaciones lingüísticas .....	184
Fig. 40: Interfaz del entorno de trabajo de Transit.....	188
Fig. 41: Interfaz de la red asociativa de Transit con información de ruta definida por defecto .....	188
Fig. 42: Archivo *.prj abierto con el <i>Bloc de notas</i> .....	189
Fig. 43: Acceso a los recursos lingüísticos almacenados a través de los asistentes de <i>Transit</i> .....	193
Fig. 44: Organización de las memorias de traducción por clientes y por nº de proyecto .....	194
Fig. 45: Clasificación de subcarpetas de proyectos .....	195
Fig. 46: Fases del flujo de gestión de las memorias de traducción .....	196
Fig. 47: Recepción de proyectos de traducción en el entorno de <i>Transit</i> .....	198
Fig. 48: Flujo de producción de las traducciones con <i>Transit</i> .....	201
Fig. 49: Fases del flujo de gestión de los diccionarios .....	203
Fig. 50: Almacenamiento de diccionarios en carpetas de clientes .....	204
Fig. 51: Almacenamiento de los diccionarios por proyectos o productos en bases de datos de clientes.....	205
Fig. 52: Resultado de las capturas de concordancias mediante <i>Search &amp;</i> <i>Replace</i> .....	208
Fig. 53: Sincronización de las ventanas en <i>Transit</i> .....	209
Fig. 54: Tabla de diccionarios de una base de datos .....	216
Fig. 55: Ejemplos de diccionarios contenidos en una base de datos .....	216
Fig. 56: Información asociada a los recursos lingüísticos .....	221

Fig. 57: Sintaxis de proyectos especiales (1) .....	223
Fig. 58: Tabla de categorías de proyectos .....	224
Fig. 59: Sintaxis de proyectos especiales (2) .....	224
Fig. 60: Sintaxis obligatoria para proyectos de traducción .....	224
Fig. 61: Sintaxis para proyectos de traducción (1) .....	225
Fig. 62: Ejemplos de nomenclatura para proyectos de traducción (1).....	225
Fig. 63: Sintaxis para proyectos de traducción (2) .....	225
Fig. 64: Ejemplos de códigos identificativos de los módulos textuales.....	226
Fig. 65: Sintaxis para subproyectos de traducción (1).....	227
Fig. 66: Sintaxis para subproyectos de traducción (2).....	227
Fig. 67: Sintaxis para proyectos de traducción (3) .....	227
Fig. 68: Ejemplos de nomenclatura para diccionarios .....	228
Fig. 69: Sintaxis para diccionarios FRL.....	230
Fig. 70: Estado de diccionarios: siglas y ponderación.....	232
Fig. 71: Estructura de directorios estándar para un proyecto .....	233
Fig. 72: Leyenda de estructura de directorios .....	234
Fig. 73: Tesauro DOCUTES.....	235
Fig. 74: Tesauro <i>Spines</i> de acceso público.....	241
Fig. 75: Flujo propuesto para un sistema de gestión de diccionarios .....	244
Fig. 76: Sección de la estructura del encabezado de un registro terminográfico de <i>TermStar</i> .....	245
Fig. 77: Sistema integrado de gestión de conocimiento y recursos lingüísticos .....	252
Fig. 78: Flujo propuesto para un sistema de gestión de proyectos y de explotación de corpus paralelos.....	253
Fig. 79: Ecosistema informacional de COGNImática .....	260
Fig. 80: Arquitectura de la información principal.....	262
Fig. 81: Funciones de los módulos de COGNImática.....	268
Fig. 82: Técnicas documentales de COGNImática.....	270
Fig. 83: Tecnologías aplicadas en COGNImática.....	274
Fig. 84: Interfaz de búsqueda del gestor documental de diccionarios (1 <sup>a</sup> propuesta) .....	278
Fig. 85: Descripción de tipos de diccionarios .....	279
Fig. 86: Tipos de búsqueda en la interfaz de consulta .....	279



Fig. 87: Captura de diccionarios tras <i>resina</i> .....	280
Fig. 88: Menú desplegable de formatos de los diccionarios .....	280
Fig. 89: Captura de diccionarios tras <i>impresora</i> .....	281
Fig. 90: Ficha de un diccionario.....	282
Fig. 91: Sección (1) de la interfaz de alimentación del gestor documental de diccionarios .....	283
Fig. 92: Acceso restringido al administrador de COGNImática .....	284
Fig. 93: Sección (2) de la interfaz de alimentación del gestor documental de diccionarios .....	285
Fig. 94: Resultado parcial de la captura de los proyectos relacionados con el cliente de búsqueda (AGFA).....	286
Fig. 95: Resultado parcial de la captura de proyectos relacionados con el diccionario de búsqueda (Microsoft_2000).....	287
Fig. 96: Sección (lenguas) de la interfaz de búsqueda del catálogo de proyectos .....	288
Fig. 97: Interfaz de búsqueda del gestor de proyectos.....	289
Fig. 98: Interfaz de selección del gestor de proyectos .....	290
Fig. 99: Ventana emergente con información de recursos lingüísticos asociadas a un proyecto .....	291
Fig. 100: Interfaz de selección de las propuestas capturadas.....	292
Fig. 101: Muestra obtenida mediante la búsqueda por concordancias en un subcorpus paralelo concreto .....	295
Fig. 102: Fases de registro en el gestor documental de COGNImática .....	300
Fig. 103: Criterios de almacenamiento de los diccionarios en línea.....	307
Fig. 104: Hipervínculos de acceso a los contenidos de procedimiento .....	307
Fig. 105: Procedimientos de catalogación en línea .....	308
Fig. 106: Acceso a procedimientos del gestor documental .....	309
Fig. 107: Catálogo de errores y soluciones de COGNImática.....	310
Fig. 108: Informe de diccionarios no registrados.....	312
Fig. 109: Informe de diccionarios sin <i>Spines</i> .....	313
Fig. 110: Seguimiento y mantenimiento de una política de gestión de los RRLL .....	315
Fig. 111: Futuras tecnologías aplicadas en COGNImática.....	325
Fig. 112: Función de exportación a archivos TMX en <i>Transit</i> .....	327



## 1. Introducción

Durante las últimas dos décadas los países industrializados han empezado a tomar conciencia de que no se pueden desechar de forma masiva los productos manufacturados. Se han optimizado costes en la producción de diferentes productos elaborados reciclando, por ejemplo, vidrio, papel o metal. A su vez, se han ido creando tecnologías paralelas para lograr óptimos resultados en el proceso de reciclado o reutilización.

El sector industrial que produce documentación técnica multilingüe, es decir, la industria GILT cuyo acrónimo extendido se refiere a *globalización, internacionalización, localización y traducción*<sup>1</sup> también es un sector que vive su expansión aprovechando, entre otras cosas, el principio de reutilización de

---

<sup>1</sup> Tema que abordaremos en el apartado 2.3.1. La documentación técnica en su ámbito global

los productos que gestiona, es decir, reciclando contenidos multilingües ya elaborados.

La industria GILT también ha evolucionado considerablemente y, como en otros sectores, se han mejorado las herramientas para la gestión de contenidos multilingües. Nos referimos a las herramientas de traducción asistida (por ordenador) que incorporan en su diseño el concepto de *reutilización*, en una doble dimensión: por una parte, reciclan los almacenes de recursos lingüísticos existentes y, por otra, generan nuevos recursos susceptibles a ser reciclados y aplicados en otros entornos de trabajo, por ejemplo, en la creación de corpus paralelos de referencia.

Esta doble idea de reutilización es una tendencia que ha marcado también la investigación en el ámbito de la lingüística computacional y el procesamiento del lenguaje natural desde finales de los ochenta, como así muestran algunos de los proyectos que vienen siendo financiados por la Unión Europea. Algunas de las primeras propuestas sufragadas por la todavía Comunidad Económica Europea, como la de Aquilex ET-7 (CEC) (Moreno, 2000: 2.2.2) y, más recientemente, varios de los proyectos e-Content aprobados dentro del VI Programa Marco de la Unión Europea comparten la estrategia de sistematización del etiquetado de los recursos para lograr su posterior reutilización. Concretamente, los proyectos IMDI, OLAC o INTERA tratan de establecer estructuras de metadatos para la organización de los recursos lingüísticos. En el caso de IMDI (*Isle Metadata Initiative*) se elaboran codificaciones aplicables a los recursos lingüísticos combinados con otros recursos auditivos y visuales, conocidos como recursos multimedia o multimodales, para recrear un universo virtual mediante recursos accesibles a través de Internet<sup>1</sup>. Dentro del marco del proyecto OLAC (*Open Language Archives Community*) se está confeccionando una gran biblioteca virtual de recursos lingüísticos intentando llegar a un consenso en los procedimientos

---

<sup>1</sup> Para más información véase <http://www.mpi.nl/IMDI/>

de almacenamiento de recursos lingüísticos digitalizados para acceder mejor a ellos<sup>1</sup>. Asimismo, el proyecto INTERA (*Integrated European Language Data Repository Area*) se propone crear un espacio virtual de recursos lingüísticos europeos y producir nuevos recursos lingüísticos multilingües<sup>2</sup>.

La estrategia principal y genérica de las herramientas de traducción asistida por ordenador estriba en el acceso a las memorias de traducción, para su reutilización. Por una parte, se reutilizan los elementos traducidos y revisados por un humano y, por otra, se recuperan los elementos traducidos con un índice de coincidencia variable mediante la técnica de lógica difusa.

Asimismo, estos sistemas cuentan, por regla general, con un gestor de terminología que permite gestionar otro tipo de recursos lingüísticos, los diccionarios. Por este motivo, estas herramientas pueden considerarse como gestores de recursos lingüísticos que pueden ser reutilizados para diferentes proyectos o para otras herramientas diferentes a los de su entorno de creación.

Estas herramientas son sistemas de información que dan soporte al almacenamiento, navegación, extracción y creación de recursos lingüísticos del tipo memorias de traducción y diccionarios terminológicos.

El auge experimentado por los sistemas de traducción asistida por ordenador está vinculado al hecho de que *los planteamientos más ambiciosos de la traducción automática han pecado de considerable ingenuidad, que explican la desilusión de los logros alcanzados* (Abaitua, 1995: 17). De la insatisfacción de esos logros, por la escasa respuesta de implantación obtenida en el ámbito profesional (Rinsche, 1997: 132-140), se desarrollan herramientas de software que aspiran a la automatización de las tareas mecánicas, reduciendo las pretensiones de automatización de las tareas de alto nivel de implicación cognitiva.

---

<sup>1</sup> Para más información véase <http://www.language-archives.org/>

<sup>2</sup> Para más información véase <http://www.mpi.nl/INTERA/>

Los límites reales que existen para la automatización de los procesos cognitivos implicados en el trabajo meramente traductológico se deben suplir, siempre que se pueda, con la automatización del resto de las tareas implicadas en el flujo de trabajo de la traducción.

El presente trabajo propone una solución para la automatización de una de esas tareas.

Esta tesis doctoral se encuadra en los siguientes campos de conocimiento:

- ❑ la gestión de la información
- ❑ la gestión de conocimiento
- ❑ la lingüística de corpus
- ❑ la recuperación de información
- ❑ la terminología y
- ❑ la traducción

Los términos adicionales propuestos para la identificación del contenido y que proponemos como puntos de acceso para una posible futura indización en una base de datos son los que siguen a continuación:

- ❑ anotación semántica
- ❑ corpus paralelo
- ❑ localización
- ❑ recursos lingüísticos
- ❑ representación del conocimiento
- ❑ sistema de gestión de información
- ❑ terminótica
- ❑ traducción automática
- ❑ traducción asistida

- transferencia de conocimiento

## **1.1. Problema**

En el entorno de una entidad proveedora de servicios relacionados con la documentación técnica multilingüe se ha extendido la aplicación de los sistemas de traducción asistida que, si bien ofrecen grandes ventajas que optimizan el flujo de trabajo de la traducción, también pueden generar problemas que impiden la reutilización eficiente de los recursos lingüísticos.

La selección de los recursos lingüísticos pertinentes para un proyecto de traducción, esto es, diccionarios de terminología y memorias de traducción, es una de las tareas implicadas en el flujo de gestión de cualquier proyecto de traducción.

A medida que aumentan los recursos almacenados se hace más difícil su gestión, porque el acceso a los recursos lingüísticos ya elaborados puede producir silencio y ruido repercutiendo en su posterior reciclado. Por tanto, de la dificultad de acceso se deriva una infrautilización de los recursos lingüísticos almacenados.

Así, la tarea de selección de recursos lingüísticos, lejos de ser un asunto baladí, se ha convertido en un gran problema a merced del incremento exponencial del volumen almacenado de información digitalizada que experimenta cualquier empresa de traducción y de servicios afines.

Para una entidad dedicada a la producción de documentación técnica multilingüe —traducción, localización, gestión de terminología, redacción técnica, maquetación y publicación—, la disponibilidad y accesibilidad a los recursos lingüísticos almacenados constituye un capital tan relevante, como el *know-how* y los recursos humanos de la entidad poseedora.

Los problemas de acceso a la información almacenada es una cuestión generalizada. El interés constante de diferentes colectivos en desarrollar

instrumentos que nos faciliten el acceso a la información es reflejo de la universalidad del problema que planteamos desde la óptica concreta de una entidad.

El proceso de acceso a la información empieza con las necesidades del usuario. El usuario quiere acceder a una determinada información, y esto es porque carece de la información que va a buscar o desea contrastarla o confirmarla.

Se trata de una necesidad que requiere de un estímulo para obtener respuesta. Es un estado cognitivo que nosotros llamamos *estado de inquietud cognitiva* y que se ha denominado también *Anomalous State of Knowledge* (Belkin *et al.*, 1982: 61)<sup>1</sup>.

Tras la ratificación de la existencia de una necesidad de información, se debe describir e identificar; es decir, se debe formular la ecuación de búsqueda para que, mediante esa formalización, se pueda acceder a la información deseada. El último paso del procedimiento está relacionado con los criterios de adecuación de lo extraído con respecto a lo buscado. De la forma de resolver la búsqueda depende el índice de satisfacción que logre el usuario.

La insatisfacción de los usuarios producida por las respuestas de los sistemas de información está mayoritariamente vinculada a la organización previa para el almacenamiento e indización de los archivos que contienen la información. Éste es uno de las principales complicaciones del ámbito de trabajo de la recuperación de información y, asimismo, es una de las dificultades específicas vinculadas a nuestro principal problema de

---

<sup>1</sup> Este término no nos parece del todo acertado, ya que el epíteto *anómalo* indica algo extraño, raro, especial y, a juzgar por la experiencia, lo verdaderamente anómalo para el procesamiento mental es no tener ninguna necesidad de extraer, contrastar o confirmar información. Puede ser que la motivación de los *padres del término* lo acuñaron a la fuerza para que su acrónimo ASK cumpliera sin restricciones las estrategias nemotécnicas y de difusión publicitaria a las que nos tienen acostumbrados también los científicos que investigan en Estados Unidos.



infrautilización de recursos almacenados en el entorno de trabajo de nuestro caso de estudio.

De esta manera, los problemas de acceso a los recursos lingüísticos y, por tanto, de la infrautilización de los mismos, presumiblemente, están vinculados a la falta de planificación y organización para el almacenamiento de esos recursos. Por ello, se deben diseñar técnicas y procedimientos de almacenamiento que permitan la óptima gestión de esos recursos lingüísticos.

Del problema general planteado se derivan otros específicos como analizaremos en el capítulo 3. Auditoría de los recursos lingüísticos multilingües en una empresa de traducción: estudio de caso.

Así pues, la razón de esta investigación nace de la necesidad de optimizar los flujos de trabajo implicados en la gestión de los proyectos de traducción y localización. La necesidad de optimizarlos mediante la automatización es apremiante debido al precio que está dispuesto a pagar el cliente por las traducciones. Desde una perspectiva empresarial, las decisiones sobre las inversiones que se deben realizar deben tomarse en función de los subsiguientes aspectos:

- la rentabilidad económica
- la garantía de calidad del producto final

En ese sentido, en el entorno de una empresa la inversión en un proyecto de investigación y desarrollo no sólo debe aspirar a ser realista en cuanto a la viabilidad del binomio modelo teórico-modelo implementado, sino, además, debe aspirar a implementarse e incorporarse en el flujo de trabajo para así lograr cuanto antes el máximo retorno de inversión (ROI)<sup>1</sup>.

---

<sup>1</sup> Tema que abordaremos en el apartado 3.2.1.2. Identificación de una necesidad y evaluación de viabilidad del proyecto.

## 1.2. Objetivos

El objetivo general de nuestra investigación es diseñar e implementar un sistema para optimizar la gestión —almacenamiento, recuperación, selección, explotación— de los recursos lingüísticos idóneos para los proyectos de traducción y localización en una empresa de traducción. Este objetivo principal está vinculado a dar respuesta a los problemas de gestión que surgen como consecuencia del aumento creciente del volumen de recursos lingüísticos almacenados necesarios y generados en el entorno de entidades que ofrecen servicios de documentación técnica multilingüe. En este sentido nos proponemos:

En primer lugar, desarrollar un procedimiento —propuesta metodológica— de almacenamiento de los recursos lingüísticos que contribuya a mejorar y automatizar su recuperación y selección. Este propósito nos obligará a realizar un análisis de los recursos lingüísticos generados en la empresa, es decir, debemos llevar a cabo una *auditoría interna de la información* — término conocido también en la jerga empresarial como *auditoría de gestión* o *auditoría operativa* (Soy, 2003: 21)— para identificar:

- ❑ qué tipo de recursos se almacenan, se reciclan y consultan;
- ❑ de qué manera se llevan a cabo las tareas de gestión de los recursos lingüísticos (escenarios de los flujos de trabajo);
- ❑ de qué manera se organizan a nivel individual y corporativo las tareas de gestión;
- ❑ qué herramientas nos asisten en el proceso del almacenamiento y selección de los recursos lingüísticos;
- ❑ cuáles son los límites y posibilidades de afinamiento de las tareas de gestión;
- ❑ qué necesidades tienen los usuarios.

Tras el análisis, procederemos a proponer y mejorar:

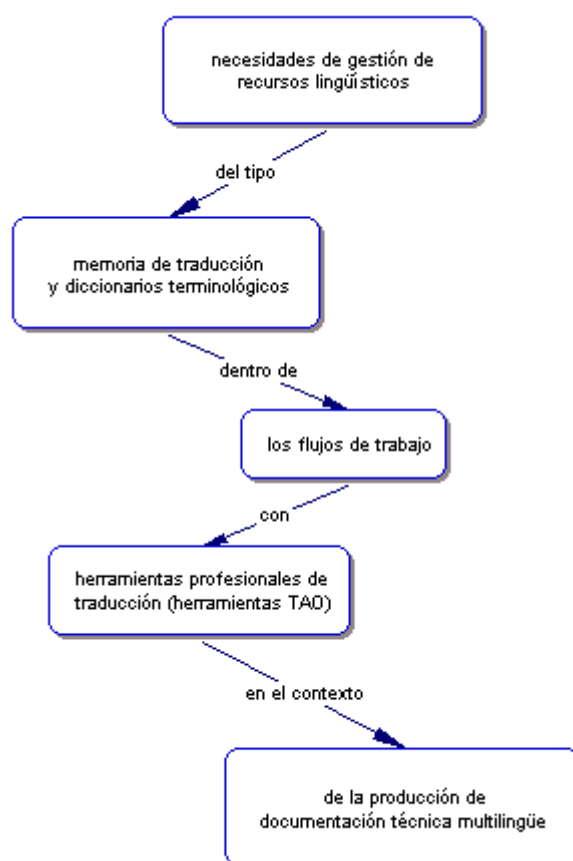
- los criterios para almacenar las memorias de traducción y otros recursos lingüísticos, como los terminológicos;
- el sistema conceptual para la organización sistemática de puntos de acceso de tipo semántico a los recursos lingüísticos almacenados;
- un protocolo de almacenamiento y de acceso para diferentes tipos de usuarios.

En segundo lugar, y como consecuencia de los objetivos enunciados, nos proponemos diseñar e implementar un sistema de gestión de recursos lingüísticos para la gestión de proyectos de traducción. Este sistema debe basarse en las necesidades de información de los usuarios en el entorno de su explotación para automatizar la gestión de los recursos lingüísticos reciclables y dar soporte a la gestión del conocimiento asociado a los recursos gestionados.

### ***1.3. Objeto de análisis***

El objeto de análisis de nuestro trabajo se centra en las necesidades específicas de gestión de los recursos lingüísticos dentro de los flujos de trabajo implicados en la producción de documentación técnica multilingüe mediante herramientas profesionales de traducción.

Estamos, pues, ante un objeto de análisis dinámico cuya dimensión se comprende en su contexto, como se muestra en el siguiente diagrama:



**Fig. 1: Objeto de análisis**

Así pues, los *recursos lingüísticos* son sólo una parte de nuestro objeto de análisis. En este sentido, los recursos lingüísticos que tratamos aquí son aquéllos que representan y transfieren conocimiento técnico, como veremos ampliamente en el apartado 2.2.5. Tipos de recursos lingüísticos para la documentación técnica multilingüe. Desde el punto de vista del sujeto que valora la información, los recursos lingüísticos del tipo memorias de traducción y de diccionarios terminológicos son los recursos que más beneficio ofrecen a los integrantes implicados en un proyecto de traducción. Ambos tipos de recursos se caracterizan por la utilización consecuyente de términos. Por tanto, los términos son la unidad mínima de representación de conocimiento de los recursos lingüísticos tratados.

Los términos, como unidades de representación de conocimiento de los recursos lingüísticos que nos ocupan, son unidades de naturaleza lingüística que al mismo tiempo representan conocimiento de un determinado dominio. Los términos son consensualizaciones para denominar conceptos. Y éstas componen, por una parte, la matriz cognitiva o sistema conceptual sobre la que se construye la parte textual de un documento técnico —siendo, en consecuencia, imprescindibles para la transferencia de conocimiento—. Asimismo, los términos son las unidades de registro que componen los diccionarios terminológicos y cualquier sistema conceptual (tesauros, ontologías, glosarios, etc.).

Sin perder de vista nuestra formación substancialmente en lingüística aplicada con especial atención en el ámbito de la tradumática y la terminótica, pretendemos tener en cuenta el estudio de los recursos lingüísticos también desde el ángulo económico-empresarial dentro del contexto de producción de documentación técnica. Esta perspectiva confiere a los recursos lingüísticos el valor de *activo*, por ser parte del conjunto de bienes de una empresa que pueden ser valorados en términos económicos y cuya protección y optimización de gestión repercute en los beneficios empresariales. Este concepto de activo específicamente relacionado con los recursos lingüísticos en el ámbito anglosajón se ha acuñado con el término *linguistic asset* como se muestra en los siguiente ejemplos:

MultiTrans™ Platform for Enterprise Linguistic Asset Management<sup>1</sup>.

Lionbridge delivers localisation services — translation, cultural adaptation, linguistic asset management and Web/desktop publishing<sup>2</sup>.

Transit XV, STAR offers your enterprise and translators the world's leading technology solution package for efficient linguistic asset management<sup>1</sup>.

---

<sup>1</sup> Para más información véase [http://www.multicorpora.ca/PRCSWeb\\_e.html](http://www.multicorpora.ca/PRCSWeb_e.html)

<sup>2</sup> Para más información véase <http://www.lionbridge.com/globalization/capabilities/content-localization.liox?intLangID=17>

Estos *activos lingüísticos* centran nuestra atención, que estudiamos y definimos como elementos integrantes de un flujo de conocimiento abocado a la gestión de proyectos de traducción y localización. De la relevancia que tienen estos recursos para optimizar el trabajo y de las implicaciones negativas que afectan al proceso de elaboración cuando se infrutilizan, nace la necesidad de que se gestionen y protejan en beneficio de la empresa. Esta consideración económica otorgada a los recursos lingüísticos supone un cambio indiscutible en la percepción no sólo de esos recursos, sino de los autores de esos recursos, esto es, los redactores o los traductores.

#### **1.4. Supuestos de partida**

Los supuestos, que consideramos como punto de partida que darán respuesta al problema expuesto, son los siguientes:

- si los puntos de acceso a la información digitalizada corresponden a la forma de organizar la información del usuario, mayor satisfacción obtendremos con el resultado de las búsquedas;
- si identificamos las memorias de traducción y los diccionarios con etiquetas semánticas —términos de identificación—, favorecemos las condiciones para mejorar su explotación;
- si construimos un sistema conceptual, es decir, una estructura armonizada de etiquetas semánticas, mejoraremos la indización y la recuperación de las memorias de traducción y los diccionarios;
- si empleamos este sistema conceptual modelado, facilitaremos la automatización de selección de los recursos lingüísticos en el flujo de trabajo previsto;

---

<sup>1</sup> Para más información véase <http://www.star-solutions.net/html/eng/produkte/summary.html>

- si aplicamos este sistema conceptual en la gestión de los recursos lingüísticos, fomentaremos la adquisición, contextualización y normalización de flujos de conocimiento relativos a los procesos;
- si utilizamos el sistema conceptual de forma consensuada, cimentaremos la creación de una base de conocimiento de estructura terminológica para gestionar mejor una parte fundamental del conocimiento corporativo de la entidad, como las bases de conocimiento traductológicas;
- si gestionamos mejor el conocimiento corporativo de la entidad, los recursos lingüísticos en cuanto a activos lingüísticos, adquirirán mayor valor como base de conocimiento para traductores, redactores técnicos y, en general, para todos los agentes implicados en la producción de documentación técnica multilingüe;
- si con el diseño propuesto para la organización de los recursos lingüísticos obtenemos buenos resultados, podremos desarrollar pautas genéricas (procedimientos comunes) de clasificación para las empresas de este sector.

### **1.5. Delimitaciones**

Aun cuando la selección de un caso concreto de empresa que se dedica a la oferta de servicios lingüísticos delimita automáticamente nuestro estudio de campo, necesitamos circunscribir aún más nuestro área de acción. El presente trabajo se concentra, pues, en el estudio de los problemas creados en proyectos de traducción y localización donde intervienen herramientas de traducción asistida, usuarias y generadoras de recursos lingüísticos.

La especialidad de nuestro caso de estudio, en cuanto a los contenidos que gestiona, se centra en la gestión de documentación técnica multilingüe, entorno consumidor y productor de recursos lingüísticos específicos. Por ello, nos centraremos en los recursos lingüísticos que primordialmente se

gestionan en el ámbito de la documentación técnica multilingüe: memorias de traducción y bases terminológicas.

Los límites de nuestra investigación están vinculados a las propias propuestas de solución. Esto es, pretendemos dar soluciones a los problemas de infrautilización de los recursos lingüísticos, diseñando una arquitectura para la gestión de la información que pueda servir de base de conocimiento corporativo. Y esta aspiración no está exenta de limitaciones.

El conocimiento, para poder gestionarlo, debe estar representado. El conocimiento es el resultado de un razonamiento que depende siempre de múltiples variables directamente vinculadas al agente y a su contexto. Por eso, intentar crear un conocimiento corporativo tiene dos limitaciones importantes que enunciaremos aquí:

- ❑ su representación formalizada para su posterior procesamiento y,
- ❑ la exigencia de idealizar el diseño de la información en beneficio de una comunidad y no tan solo de un individuo.

## **1.6. Marco teórico**

Nuestro trabajo se encuadra dentro del marco de la ingeniería lingüística, línea de investigación que enmarcamos en la lingüística aplicada. En este sentido, y teniendo en cuenta el campo de aplicación propuesto, se deben tener en cuenta los siguientes enfoques:

- ❑ la lingüística
- ❑ la terminología
- ❑ la computación

A partir de nuestro objeto de análisis, los enfoques previstos y los propósitos marcados para cumplir nuestros objetivos, fundamentamos nuestro trabajo en un marco teórico interdisciplinar. Estas disciplinas se recogen



mayoritariamente en la descripción que define la denominada lingüística documental:

La lingüística documental, si bien parte del cruce de términos que forman su denominación, es una disciplina impregnada de otros campos científicos, como son básicamente la Lógica, la Estadística y la Informática y, más concretamente, la Lexicología, la Archivística, la Biblioteconomía y la Telemática (García Gutiérrez, 1984: 137).

En esta lista de contribuciones disciplinarias asociadas a la lingüística documental, cabe añadir y destacar la influencia en este trabajo de la terminología y de la lingüística de corpus.

Concretamente, la dependencia de la terminología es crucial. Y esto es así porque el papel de la terminología se ve enfatizado en cuanto a suministradora de términos de los que se componen los productos de la lingüística documental, como los lenguajes documentales o cualquier otro sistema conceptual abocado a la indización o recuperación de la información.

La *Teoría comunicativa de la terminología* (Cabré, 1999: 69-92) nos ofrece el marco teórico terminológico que influye en la forma de entender los recursos lingüísticos tratados en cuanto a su función de representar y transferir conocimiento. En los postulados de esta teoría interdisciplinar, modular e integradora se reafirma la diversidad de representación del conocimiento y la necesidad de una variación denominativa y conceptual al servicio de la comunicación especializada. Esta teoría confiere a los términos —las unidades de su objeto de estudio— una dimensión lingüística, comunicativa y cognitiva. Estas unidades de representación terminológica son, por tanto, unidades lingüísticas de conocimiento especializado representadas mediante el lenguaje y unidades de comunicación cuya interacción es indisociable.

La influencia de la lingüística de corpus en este trabajo está vinculada al marco teórico lingüístico al que nos adherimos. En este sentido nuestra visión del lenguaje se describe en la *Teoría contextual del significado* (Firth,

citado en Pérez, 2002: 2.11) y se basa en la función social del lenguaje y la relevancia del contexto debido a que el significado de las palabras no puede separarse de su uso. La evidencia de que el contexto puede afectar a nuestra capacidad de razonar y de recuperar información ha sido aceptada e incluso considerada concluyente para entender la cognición humana (Westbury y Wilensky, 1998: 46). Así, el significado de las palabras reside en su uso, por ello el encuadre de su contexto es básico para la recogida sistemática de datos lingüísticos a partir de *almacenes de palabras* conocidos bajo el término *corpora textuales*, objeto de estudio de la disciplina de la lingüística de corpus.

Esta disciplina es, por una parte, deudora de la superación de una etapa estructuralista donde las piezas lingüísticas se observan como unidades simples relacionadas por un lenguaje sintáctico subyacente; y, por otra, de la evolución de dispositivos de tecnología hardware. La constante y creciente capacidad de almacenamiento de los soportes de memoria y el aumento progresivo de la velocidad de transferencia de bits de los microprocesadores repercuten en el tratamiento de volúmenes formidables de almacenes de palabras.

El marco computacional seleccionado se basa en el modelo teórico de las bases de datos relacionales con razonamiento basado en propiedades booleanas y en propiedades basadas en la probabilidad desarrollado por Codd (1970: 377-387). El planteamiento teórico de este matemático propone el reemplazo del modelo jerárquico de estructuración de la información por tablas que contienen columnas y que son susceptibles de cruzarse. Esta forma de estructurar la información ha marcado trascendentalmente la evolución de las aplicaciones informáticas desde los 70, ya que simplifica las especificaciones de las consultas para el acceso a la información permitiendo al usuario interactuar con el ordenador de una forma más sencilla.

Este marco teórico tiene su realización formal mediante la implementación de bases de datos relacionales orientadas a objetos a las que se accede y alimenta mediante interfaces persona-ordenador.

## **1.7. Metodología**

El método previsto es el *estudio de caso*, forma profundizada y exhaustiva de adentrarnos a nuestro objeto de análisis complejo que, como hemos visto en el apartado 1.3. Objeto de análisis, se centra en las necesidades específicas de gestión de los recursos lingüísticos dentro de los flujos de trabajo implicados en la producción de documentación técnica multilingüe.

Las necesidades específicas se han detectado en el entorno específico de una empresa proveedora de servicios lingüísticos, concretamente de STAR Servicios Lingüísticos SL —a partir de ahora STAR SL—, caso que tomamos como referencia.

Mediante el análisis de este caso pretendemos obtener un conocimiento amplio y detallado de la problemática que surge como consecuencia de la gestión de los recursos lingüísticos en sus diferentes flujos de trabajo, a fin de obtener datos concretos para tener una perspectiva y un punto de partida que pueda convertirse en genérica, a pesar de las limitaciones de una generalización a todo el universo de los hallazgos obtenidos.

Por tanto, partimos de la idea de que este método nos puede ayudar a:

- la formulación de supuestos de partida,
- el reconocimiento de las principales variables involucradas en un problema y,
- la formulación de propuestas que nos ofrezcan una mejora en los resultados.

El caso de STAR SL puede considerarse típico del sector industrial al que pertenece. Consideramos a STAR SL como un caso característico y representativo del sector porque:

- es una empresa proveedora de servicios lingüísticos específicos de documentación técnica multilingüe;
- aplica consecuentemente herramientas profesionales de traducción y localización asistida en su entorno de producción.

De esta manera, basándonos en este caso concreto y sus agentes implicados, podremos:

- identificar problemas adicionales, asociados a los ya apuntados en este mismo capítulo y,
- generar una metodología de procedimientos de acuerdo a situaciones *in vivo*, más coherentes —en cuanto a contextuales y en cuanto a su orientación a los usuarios— con las exigencias de los entornos de trabajo.

Desde el punto de vista del método para análisis de la información, la metodología de trabajo definida por Soy (2000) en su monografía acerca de la auditoría de la información nos sirve de referente metodológico para el análisis de nuestros datos. Soy reelabora y complementa las guías y manuales de referencia de los autores de mayor influencia en la auditoría de la información con perspectiva empresarial. A saber, *InfoMap: A Complete Guide To Discovering Corporate Information Resources* (Burk y Horton, 1988), *Practical Information Policies* (Orna, 1999) o *The information Audit: A Practical Guide* (Henczel, 2000). Todas estas propuestas enumeran los componentes, las fases y los métodos y ventajas en la aplicación de una auditoría de la información.

Estas técnicas obligan a preparar la información de tal forma que sea posible gestionar el conocimiento traducido en experiencia y *know-how* generado por

las personas que integran una entidad. Previsiblemente, mediante la aplicación de esta metodología de auditoría de la información se derivarán soluciones para la gestión de forma automatizada de los recursos lingüísticos y su información asociada.

### **1.8. Organización del estudio**

Tras el presente capítulo dedicado a describir el problema en su contexto general, los objetivos marcados, el objeto de análisis, los presupuestos de partida y los fundamentos del marco teórico y la metodología, este trabajo de investigación se estructura de la siguiente forma:

En el capítulo 2 se establecen las coordenadas básicas del entorno de trabajo en el que se contextualiza nuestra investigación. La diversidad de los entornos de traducción en cuanto al tipo de proyectos que pueden llegarse a gestionar nos obliga a dedicar un capítulo a los aspectos asociados a la gestión de documentación técnica multilingüe, dedicando especial atención a:

- el tipo de recursos lingüísticos gestionados y explotados;
- las herramientas de gestión de recursos lingüísticos implicadas en el proceso de producción.

En el capítulo 3 aplicamos el método auditor de análisis de la información para poder detectar de forma sistemática cuáles son los flujos de almacenamiento, creación, retención, transferencia y (re)utilización de los recursos lingüísticos, así como las dificultades y cuellos de botella detectados en su gestión. En el capítulo 3 se esbozan propuestas metodológicas y tecnológicas abocadas a la resolución de los problemas de gestión planteados.

Como resultado del capítulo anterior, en el capítulo 4 se implementa el diseño del sistema propuesto: un sistema de gestión de conocimiento y de recursos lingüísticos para entornos de traducción técnica que hemos

denominado COGNImática. En este capítulo, también se especifican los fundamentos técnicos y funcionales del sistema, con especial atención a las interfaces que establecen la comunicación entre el usuario y el ordenador. Por último, también se detallan las propuestas de seguimiento y mantenimiento de COGNImática para mejorar su rendimiento.

En el capítulo 5 se exponen las últimas reflexiones finales a modo de conclusión y una relación de las referencias bibliográficas —capítulo 6— cierra el presente proyecto de tesis.

## **2. La documentación técnica multilingüe: recursos lingüísticos y técnicos**

La documentación es un término que representa diferentes conceptos. En el ámbito de la *documentación técnica* adopta un significado concreto que se amplía en función de su carácter multilingüe. La trascendencia de la documentación técnica para establecer la comunicación y para gestionar el conocimiento es una consecuencia directa de la tecnificación y de la globalización.

El sector industrial que asume la producción de documentación técnica multilingüe sigue creciendo paralelamente con:

- ❑ la sistematización de los recursos de representación utilizados,
- ❑ el desarrollo de procedimientos y técnicas de gestión,

- la evolución cada vez más sofisticada de la tecnología que necesita ser descrita,
- la evolución de la tecnología que se utiliza para gestionar los contenidos multilingües.

Para entender mejor el flujo de trabajo de la traducción y la gestión de los recursos lingüísticos implicados, haremos hincapié en las fases de creación de la documentación técnica monolingüe que, como también abordamos, comparte algunas de las fases del entorno multilingüe.

Igualmente, la planificación, la codificación o el tratamiento en general que se hace de los recursos de representación lingüísticos utilizados tendrá implicaciones en su gestión multilingüe. A la par, las herramientas implicadas en la gestión de los contenidos multilingües repercuten en el flujo de trabajo.

### **2.1. A propósito del término “documentación”**

Antes de centrarnos en la producción de *documentación técnica multilingüe*, que, como ya hemos visto en el apartado 1.3. Objeto de análisis, es el contexto de nuestro objeto de análisis, cabe mencionar los diferentes conceptos que representa el término *documentación*.

Este término identifica un campo del saber que se encuadra dentro de las ciencias aglutinadoras de la información y biblioteconomía. La variedad denominativa y conceptual relacionada con el término *documentación* entendida como ciencia o disciplina muestra, por una parte, los diferentes enfoques y, por otra, la evolución de este campo del saber (véase López, 1996: 63-93).

Por tanto, estamos ante un concepto transformado fruto de su evolución, en especial, desde las primeras décadas de la segunda mitad del S. XX como consecuencia del desarrollo exponencial del conocimiento científico-técnico y, por tanto, de documentos e información relativa; y, desde la mitad de los



80 y 90 hasta nuestros días, derivado del instrumento más revolucionario de todos los tiempos, Internet, en cuanto a herramienta de difusión y transmisión de información (Delgado, 2002: 27-31). Actualmente los documentos, contenedores de información, se publican, transmiten, recuperan y consumen de forma inmediata y los ciclos de vida de estos documentos se acortan cada vez más. Por eso, la ciencia de la documentación se identifica no en pocas ocasiones con las ciencias de la información. Y esto es porque engloban un conjunto de disciplinas que estudian la teoría de la información, su estructura, sus propiedades y su gestión, además de desarrollar técnicas para su ejecución, con un objetivo concreto, tal y como se deriva de la siguiente definición:

La documentació és una disciplina científica que s'encarrega de la gestió d'ingents quantitats d'informació amb l'objectiu de difondre-la al nombre més elevat possible de persones (Recorder, 2003).

Esta definición que identifica la documentación con *gestión de información* es superada por la propuesta de Codina (1997: 23), que apuntamos a continuación:

La documentación es una disciplina científica que tiene como objeto de estudio la gestión eficiente del conocimiento social, a fin de ponerlo a disposición de un colectivo de usuarios o de la humanidad en su conjunto, con el objetivo de permitir y facilitar el proceso de obtención de nuevos conocimientos.

Esta última definición está más orientada a la función de la información que al producto contenedor de la información, esto es, el documento. Esta definición cuenta con un enfoque más amplio, donde se asume la función de gestionar conocimiento en todas sus facetas: adquisición, generación, representación y transferencia, tal y como hemos descrito en un estudio anterior (Cámara, 2004).

Consideramos el término *documentación* según la siguiente definición:

Proceso de creación, adquisición, ordenación, almacenamiento, recuperación, análisis y difusión de documentos o de información registrada en documentos (Tesauro Spines, 1984).

El tesauro multilingüe Spines, —acrónimo para *Science and Technology Policy Information Exchange System*—, de la UNESCO, de donde hemos extraído la definición, identifica el término en cuestión como una técnica para la gestión. En esa misma definición se lee la aclaración: *no confundir con "documentos" ni con "información"*. Esta nota muestra el uso adicional que se ha dado y sigue dando al término *documentación*.

Así, desde este punto de vista, la *documentación* que acompaña a un evento, procedimiento, gestión, producto, etc. son los *documentos* relacionados con lo que describe. Si, además, se trata de documentación específica de un producto, proceso o proyecto, en cuya concepción, representación o utilización está implicada la técnica, entonces se utiliza el término *documentación técnica*. En el contexto de producción de *documentación técnica* el término *documentación* identifica asimismo una tipología concreta de documentos. En este sentido, la *documentación* de un producto engloba los documentos relativos a la forma de explotación de un producto como, por ejemplo, instrucciones de instalación, manual de usuario, manual de reparación, guía rápida. Es decir, en general, se trata del material indispensable que precisa el usuario para poder comenzar a trabajar con el sistema o artefacto. Estos documentos se pueden encontrar impresos o en formato digital y, en conjunto, se agrupan bajo un solo término, *documentación*, para diferenciarlo de otros componentes textuales incluidos en el sistema como es el caso de los recursos textuales interactivos del software. Éstos requieren de un tratamiento textual y técnico determinado. Algunos de estos elementos considerados como parte del software que integra un programa son las opciones del menú, los cuadros de diálogo, los botones de las funciones, entre otros, tal y como exponemos más extensamente en el apartado 2.3.5. Los tipos / módulos textuales de la documentación técnica.

## **2.2. La documentación técnica**

Ya hemos visto que el uso extendido de *documentación* como término que identifica el conjunto de documentos o información relativa a algo, es aplicable también al término *documentación técnica*. Estos documentos, aun cuando son diferentes entre sí, tienen características comunes, debido a que:

- ❑ describen ingenios, esto es, artefactos, programas informáticos y ocio electrónico,
- ❑ representan conocimiento técnico,
- ❑ utilizan recursos de representación específicos,
- ❑ tienden a utilizar técnicas específicas de desarrollo de contenidos,
- ❑ tienden a utilizar tipos textuales concretos,
- ❑ cumplen la función de transferir conocimiento.

Muchos de los modernos aparatos y programas que utilizamos diariamente serían de nula utilidad sin explicaciones e instrucciones de cómo manejarlos; por ese motivo es absolutamente necesario crear un puente de comunicación entre el fabricante y el usuario. De la calidad que ofrezca la documentación técnica dependerá la eficiencia con la que se utilice la técnica en el ámbito doméstico y profesional. Es más, la relación entre los productos y el soporte documental técnico específico con el que se dota a cada producto puede ser determinante en la elección de una marca, además de jugar un papel importante en la imagen del producto. De ahí que hayan nacido nuevos perfiles de trabajo como el del redactor técnico, ilustrador técnico y el del traductor técnico. Concretamente las competencias y destrezas del traductor técnico se perfilan cada vez más diferenciadas con respecto al tradicional traductor científico-técnico (Jacob, 2000: 71).

### 2.2.1. Ingenios: artefactos y sistemas

En general la industria de cualquier sector asocia normalmente el concepto de *documentación técnica* con los manuales de usuario impresos o digitales que acompañan un aparato. Este tipo de documentos especifican y describen las funciones de estos adelantos técnicos que pueden ser:

- de uso común: ordenadores, coches, guitarras eléctricas, lavadoras, móviles, calculadoras, cepillos de dientes eléctrico, etc.
- de uso especializado: grúas, colectores solares, microscopios, antenas profesionales, cosechadoras, molinos eólicos, escáneres de uso médico, fresadoras, etc.

Todos estos artefactos precisan de documentación técnica para poder explotar todas las funcionalidades pensadas por sus desarrolladores.

Asimismo, también los sistemas informáticos, aunque no sean artefactos, tal y como se describe la categoría anterior que engloba otra clase, también requieren de especificaciones y descripciones técnicas y funcionales que ayuden al usuario a entender su funcionamiento.

Así, los programas de software, entendidos como el conjunto de instrucciones detalladas controlan la operación de un sistema computacional con funciones específicas y, en este sentido, requieren de documentación técnica tanto el software del sistema o sistema operativo, aglutinador de los programas que gestionan los recursos del ordenador —unidad central de proceso, dispositivos de comunicaciones y dispositivos periféricos— y que controlan los dispositivos de hardware, como también el software de aplicaciones.

Todas estas soluciones informáticas son también desarrollos del ingenio, vinculados a la industria de la tecnología de la información y de las telecomunicaciones.

Estos sistemas y aplicaciones, como los artefactos, pueden dividirse en dos categorías:

- de uso común: los programas de procesamiento de textos Word, hojas de cálculo como Excel, navegadores como el de Internet Explorer o Mozilla, los programas antivirus como Norton o Panda;
- de uso especializado: los programas de traducción asistida, los de gestión de finanzas, los de simulaciones para hacer prácticas aeronáuticas virtuales, entre muchos otros.

Paralela e íntimamente relacionado con la industria de la información y de las comunicaciones se desarrolla un sector, el del ocio electrónico. Este sector tiene una función lúdica o de entretenimiento explícito y, en algunos casos, tiene una orientación educativa, a la vez que entretiene. En el ámbito anglosajón esta doble función se representa con el término *edutainment*.

El sector del ocio electrónico desarrolla videojuegos, juegos de ordenador, juegos de Internet, software de entretenimiento interactivo y éstos requieren también de documentación técnica para que los usuarios sepan cómo explotar las funcionalidades. Asimismo, los programas de tipo lúdico, aun cuando tengan la intención de entretener, ofrecen también una gran diversidad de niveles de especialidad, siendo posible también su clasificación en más o menos especializados (Scholand, 2002).

Las pasarelas de información con las interfaces de usuario en línea, ya bien sean de navegación o de búsqueda, constituyen en sí una clase de desarrollo que también requiere de documentación técnica, aun cuando en algunos casos la interfaz de búsqueda sea tan intuitiva que haga innecesarias las instrucciones de navegación, como es el caso que mostramos a continuación:

**Fig. 2: Interfaz de búsqueda de Eurodicautom**

Sintetizando, los ingenios que precisan de documentación técnica los hemos clasificado en los siguientes tipos:

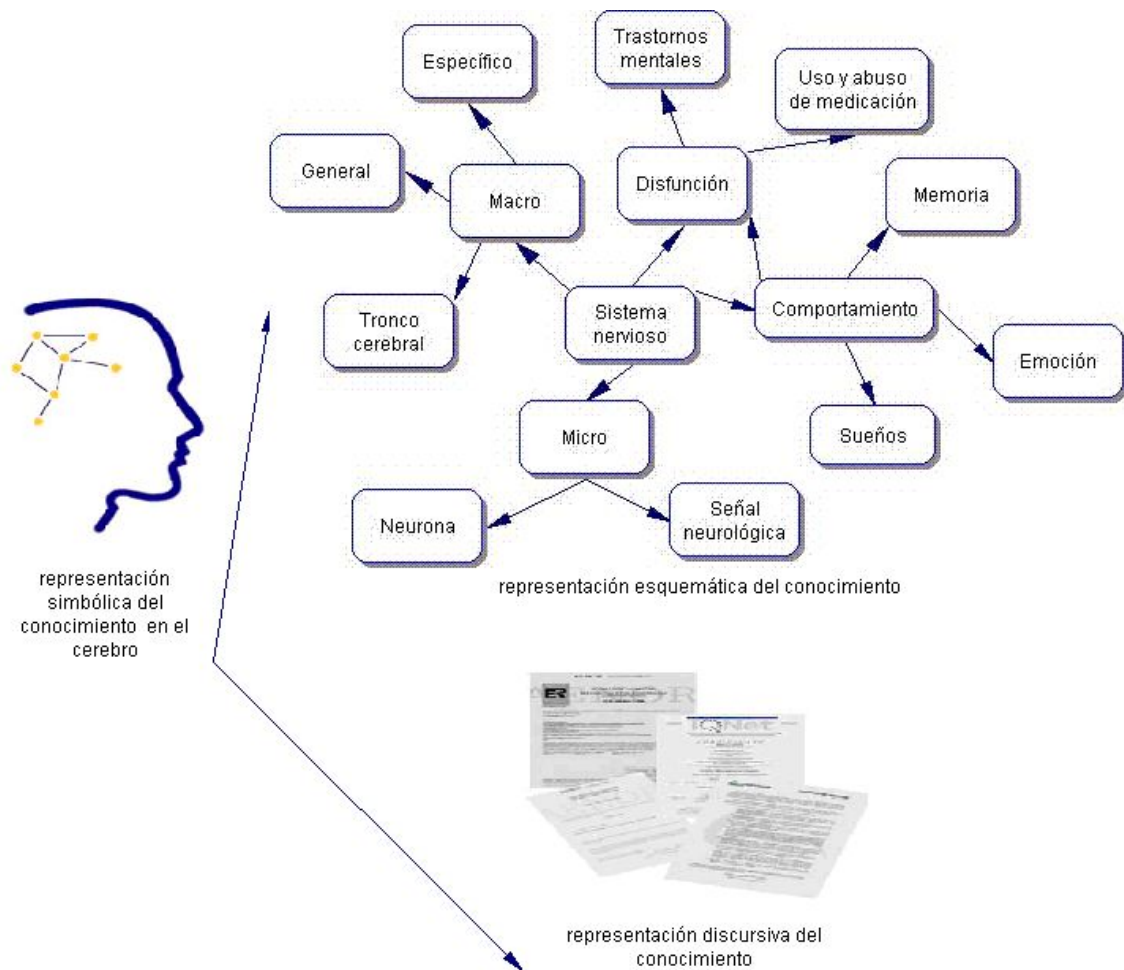
- ❑ artefactos
- ❑ programas informáticos
- ❑ programas de ocio o entretenimiento
- ❑ pasarelas de información en línea e interfaces

No obstante, esta clasificación que apuntamos no excluye la combinación de diversos ingenios en uno solo.

### **2.2.2. Conocimiento técnico representado**

Cualquiera que sea la función de la documentación técnica de un producto, ésta es, ante todo, conocimiento representado. El conocimiento representado o documentado se conoce como conocimiento explícito y se utiliza como medio para transmitir los procedimientos que se han de seguir para resolver de forma óptima cualquier función, problema o proceso de un ingenio. Este tipo de conocimiento explícito está orientado conscientemente a la resolución de problemas y se halla siempre representado, lo que supone poder acceder a él conscientemente. Todo el conocimiento representado es conocimiento

explícito y se genera a partir de conocimiento tácito o de conocimiento previo explícito. Esta representación del conocimiento es la interfaz entre el cerebro y el mundo exterior y puede representarse en forma esquemática o discursiva.



**Fig. 3: Representación esquemática y discursiva del conocimiento**

El conocimiento es un conjunto de conceptualizaciones que explican la naturaleza de nuestro mundo real y abstracto, pero el conocimiento en sí, no existe en la naturaleza (Galinsky, 2001). El conocimiento representa el contenido de nuestros razonamientos. Asimismo, la relación entre la información y el conocimiento es muy estrecha por dos razones:

- el conocimiento es un producto generado a partir de la metabolización de la información (Cornellá, 2000c) y,

- la información es conocimiento explícito.

El conocimiento técnico representado es un producto tangible, a diferencia del conocimiento que se crea, se infiere, se interpreta, se deduce, se desprende, se dilucida, y que engloba el conocimiento tácito. El conocimiento tácito es el saber que tiene un individuo o entidad compuesta de individuos. Este conocimiento responde a preguntas sobre cómo se hacen las cosas, cómo se resuelven problemas, cuándo resolverlos y de dónde se obtienen los recursos para poder solventarlos. Está orientado a la experiencia y tiene un motor consciente y otro inconsciente: por eso, es difícil hacerlo explícito en su totalidad. Este saber engloba todas las habilidades que tienen las personas y se transfiere de manera informal. En este sentido podemos afirmar que el contenido es una representación formal o explícita del conocimiento.

Según Reber (1995:15), el término *conocimiento tácito* aparece por primera vez en dos obras de Polanyi (1958/1974). Sin embargo, el concepto en cuanto a proceso intelectual (cognitivo) y emocional (de conducta) con dos niveles de percepción, el consciente y el inconsciente, es una idea formulada por Lashley (1956, 1-18).

	COGNITIVO	EMOCIONAL
CONSCIENTE	Fácil	Difícil
INCONSCIENTE	Difícil	Muy difícil

**Fig. 4: Dificultad de transferencia de los niveles de conocimiento**

Como ya apuntábamos en un trabajo anterior (Cámara, 2004), el conocimiento representado es más fácil de gestionar que el no representado. El conocimiento que no es representado y, por tanto, no es explícito, no se puede medir debido a su intangibilidad. La representación del saber puede someterse a un modelado relativo, es decir, se puede partir de términos,



estructuras conceptuales comunes o procedimientos que ayuden a codificar el conocimiento para desarrollar y gestionar de forma más eficiente el conocimiento individual y colectivo. Así, para optimizar la gestión del conocimiento explícito o también gestión de contenidos, se deben armonizar sus formas de representación, como veremos más adelante.

En este sentido, recordemos que el arte es también otra de las manifestaciones para la representación de conocimiento. Sin embargo, la transferencia de sus contenidos es mucho más difícil de consensuar. La razón de esta dificultad estriba en el papel que desempeña la percepción, es decir, el conocimiento sensorial o emocional, con respecto al peso que adquiere el conocimiento cognitivo o intelectual. Esto supone que los mecanismos argumentativos para defender el significado que contiene la obra representada, de alguna manera, estén sesgados igualmente por la percepción del agente que interpreta la pieza representada.

La poesía, la pintura, la escultura y, en general, todas las modalidades que se engloban en el concepto de arte son representaciones de conocimiento que se alimentan, a su vez, del conocimiento generado en todas las disciplinas científicas y que también se pueden servir de los desarrollos tecnológicos.

### **2.2.3. Recursos para representar conocimiento técnico**

La materia prima utilizada para la representación de contenido técnico puede ser de diferentes tipos: recursos lingüísticos y recursos no lingüísticos.

Los recursos lingüísticos se definen como:

Conjuntos de datos en lengua oral y escrita, así como de descripciones lingüísticas legibles por el ordenador, utilizadas para crear, mejorar o evaluar lenguaje natural y sistemas o algoritmos lingüísticos, o como materia prima para la localización de software y para la industria proveedora de servicios lingüísticos, para el estudio de la lengua, para la publicación electrónica, para

la ejecución de transacciones internacionales, para uso de especialistas y usuarios finales (ELRA, 2001-2003).

Si el soporte es el papel, los recursos utilizados para representar el conocimiento técnico pueden ser lingüísticos y visuales (ilustraciones, imágenes, gráficos). Si el soporte utilizado es el electrónico, además de la sofisticación de la representación de los recursos lingüísticos estructurados en forma de hipertexto, podemos encontrar recursos visuales en movimiento, en tres dimensiones con incorporaciones adicionales de efectos especiales tanto visuales como auditivas.

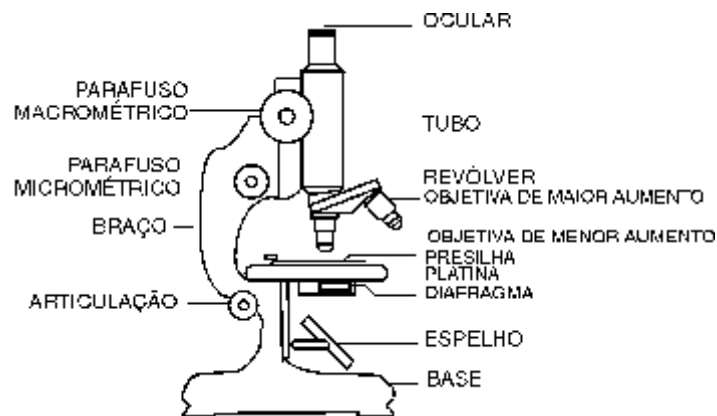
Para alcanzar la función atribuida al texto técnico se combinan con frecuencia los recursos lingüísticos con los visuales, incorporando recursos auditivos como señales auditivas para la identificación de acciones concretas en un programa informático, como por ejemplo, el sonido de una melodía identificadora de acciones como en el sistema operativo de Windows, donde se activa un determinado sonido para identificar que se inicia el PC o que el sistema genera un error.

La documentación técnica se compone principalmente de recursos lingüísticos que conforman textos escritos cuyos conceptos especializados se representan mediante términos. Los contenidos técnicos además se hallan representados y complementados con recursos multimedia, especialmente de tipo ilustrativo, como soporte explicativo e identificativo de las unidades de conocimiento de naturaleza lingüística, es decir, los términos.

Los recursos ilustrativos o visuales son fotos, figuras, dibujos, esquemas que aparecen en un trabajo, capítulo o artículo para explicarlos o para embellecerlos (Martínez de Sousa, 1999: 205). Los recursos ilustrativos utilizados en la documentación técnica van más allá del concepto que engloba la definición de Martínez. Sin olvidarnos de la función también estética que debe atribuirse a una buena maquetación o diseño gráfico, los recursos visuales ilustrativos tienen como objetivo optimizar la transferencia de

contenidos técnicos. En este sentido, las ilustraciones al igual que los términos son unidades de representación de conocimiento que contribuyen a la transferencia de conocimiento. Estas formas no son las únicas de representación de conocimiento, como aborda Monterde (2004a) en su modelo de representación del conocimiento donde incluye como formas no lingüísticas de representación los recursos acústicos, olfativos, táctiles y gustativos, junto a los visuales ya mencionados.

Dentro del ámbito de la documentación técnica multilingüe, concretamente en el proceso de gestión de proyectos de traducción o localización, son las formas lingüísticas y las no lingüísticas de tipo visual las que más relevancia tienen. Y esto es porque, mediante la ayuda de dibujos técnicos, gráficos, esquemas, mapas mentales, fotos, capturas de pantalla, iconos y símbolos se representan los objetos lo más exactamente posible, de forma parcial o total, como se ilustra en el siguiente ejemplo.



**Fig. 5: Gráfico de un microscopio**

En general, estos elementos de imagen sirven de apoyo para documentar las explicaciones discursivas de tipo lingüístico.

Las representaciones gráficas de algunos sectores industriales se guían por normas fijas y preestablecidas para poder describir de forma exacta y clara, dimensiones, formas, características y la construcción o funcionamiento de lo que se quiere reproducir. Por ejemplo, los dibujos técnicos emplean signos gráficos regidos por normas internacionales que los hacen más inteligibles.

Con el desarrollo industrial y los avances tecnológicos, el dibujo técnico aparece como un recurso ilustrativo casi indisociable de los recursos lingüísticos que contribuye a la transferencia de los contenidos técnicos. Así, por ejemplo, en documentación relacionada con ingeniería civil es común que se dote de planos de planta, fachadas, secciones, perspectivas, fundaciones, columnas, detalles y otros. Y en documentación relacionada con la ingeniería mecánica, los dibujos representan piezas o partes de máquinas, maquinarias, vehículos y máquinas industriales:

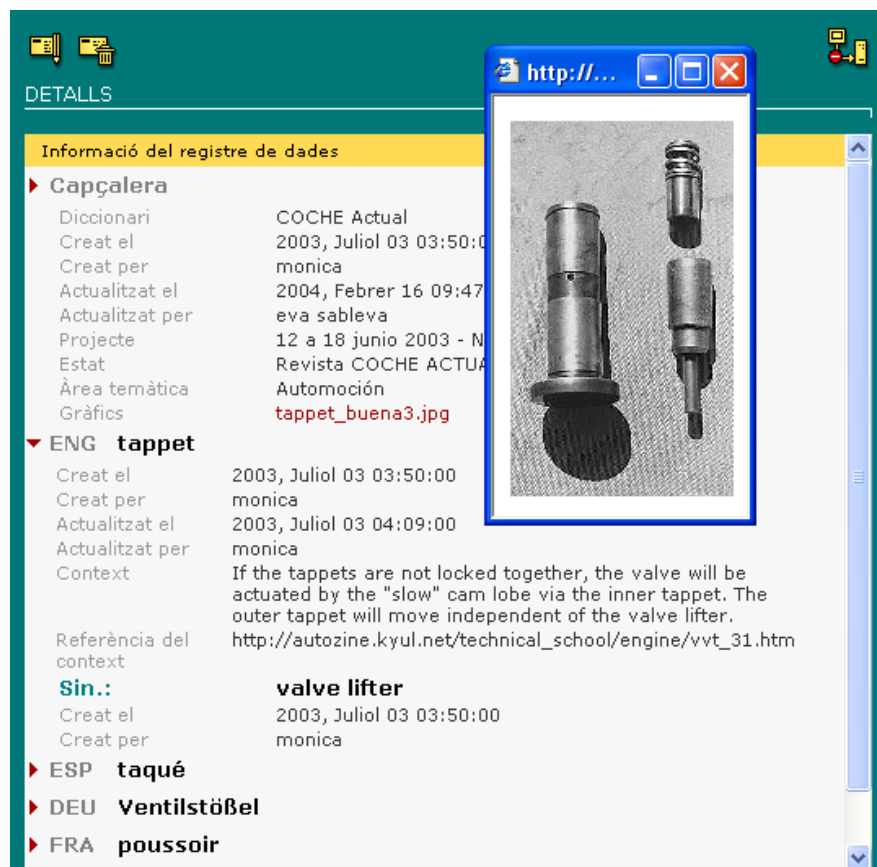


**Fig. 6: Gráfico de la carrocería de un automóvil**

Sin embargo y, a pesar de la relevancia que estos recursos visuales tienen para la representación del conocimiento técnicos con fines comunicativos, desde el punto de vista de la terminología este tema no ha sido tratado con la sistematicidad y exhaustividad con la que se debería (Galinsky y Budin, 1998: 16; Ghenghea, 1998: 783; Monterde, 2002b: 260).

En las recomendaciones para el diseño y configuración de bancos de datos terminológicos de la Gesellschaft für Terminologie und Wissenstransfer de Alemania (1996), se zanja el tema de los recursos visuales recomendando una sola categoría de datos, *símbolos*, que agrupan combinaciones de letras y números, y/o pictogramas, sin entrar en detalles o especificaciones que atiendan a los diferentes tipos.

Algunos de los programas informáticos de asistencia a la gestión terminológica, los llamados gestores terminológicos, ya incluyen la categoría *gráficos* —el nombre de este campo puede cambiarse manteniendo la función de acceso al archivo gráfico que documenta el concepto— donde es posible introducir un archivo de imagen identificativo de un término, como se muestra en la captura parcial de la información de un registro gestionado mediante la herramienta de gestión remota de terminología, WebTerm de la empresa STAR.



**Fig. 7: Muestra de un término con representación visual**

Estas herramientas profesionales concebidas para su empleo en el entorno de la traducción profesional se desarrollan en función de las necesidades de los usuarios. Éstos precisan con frecuencia poder visualizar de forma concreta lo expuesto de forma simbólica (estructuras lingüísticas).

A la par, otras iniciativas muestran la relevancia de la imagen para documentar el término, como la del banco de imágenes publicado por el Ministerio de Educación y Cultura (2004). Éste contiene un catálogo de fotografías e ilustraciones relacionadas con las áreas y materias de las distintas etapas y niveles de formación sobre ámbitos técnicos específicos como: electrónica, herramientas, etc., y que, aunque esté dirigido a aquellos que se están formando en este ámbito, puede servir de fuente documental para los mediadores del conocimiento como los traductores, aunque por ahora solo esté en castellano.

Concluyendo, la materia prima que da forma a la representación del conocimiento técnico es de tipo lingüístico, colmado de términos inmersos en un discurso similar al lenguaje natural. Asimismo, la evolución científica y tecnológica requiere de la representatividad que ofrecen otros recursos no lingüísticos, como las ilustraciones cualquiera que sea la estructura que adopte. La tendencia del uso de diferentes recursos de representación hace necesario establecer pautas sistemáticas y de mayor integración en el diseño del conocimiento representado de forma lingüística y no lingüística.

### ***2.2.3.1. Recursos lingüísticos: texto y estructuras conceptuales***

Como hemos visto, los recursos no lingüísticos son también transportadores de conocimiento y contribuyen junto a los recursos lingüísticos a una mejora de la transferencia y adquisición de conocimiento apelando a cualquiera de nuestros sentidos para hacernos llegar el contenido semántico y el sentido perceptivo que encierran.

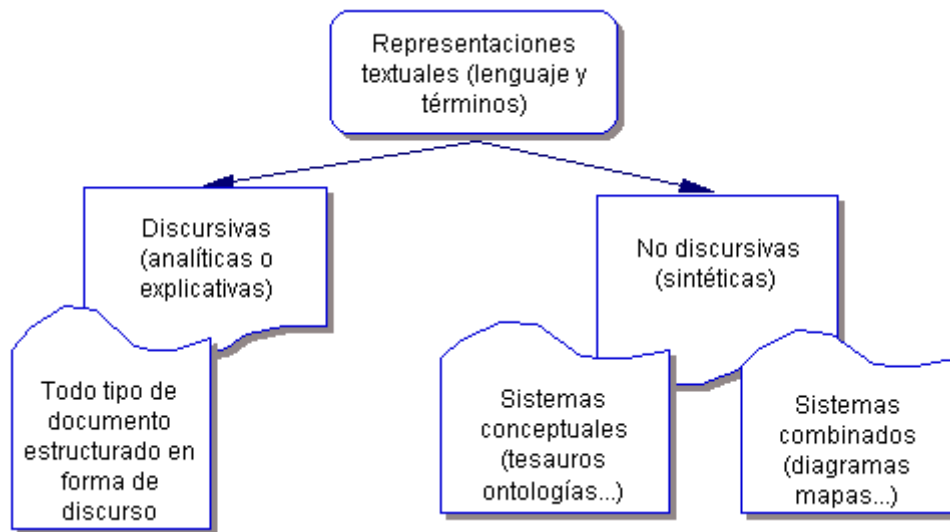
No obstante, como ya habíamos mencionado, el lenguaje sigue siendo básico en el desarrollo y transferencia de contenidos. Ambas tareas se engloban en los procesos implicados en la gestión de conocimiento explícito. Las tareas vinculadas a la gestión del conocimiento están estrechamente relacionadas con los procesos cognitivo-lingüísticos que tienen lugar en el cerebro y se materializan en forma de conocimiento representado a través

del lenguaje (Kandel, 1995; Jorba y Sanmarti, 1996). En este sentido, el lenguaje es el reflejo del modelo de procesamiento del pensamiento humano.

Aun cuando el refrán rece, *una imagen vale más que mil palabras*, el lenguaje natural, la palabra y el texto, siguen siendo la forma más potente de representación del conocimiento y, por tanto, de gestionar la información (Codina, 2001: 201); es más, la generación, representación y transferencia del conocimiento a través del lenguaje es uno de los factores determinantes para la supervivencia de nuestra especie sobre otros homínidos (Junyent, 2003: 40-41).

El lenguaje es la materia prima de los recursos lingüísticos, ya bien en su manifestación espontánea —lenguaje natural— como en la controlada —lenguaje controlado con fines específicos como la indización (lenguaje documental), la redacción automatizada, por citar algunos ejemplos—.

La representación lingüística discursiva de los textos técnicos aglutina el lenguaje natural y los términos que conjuntamente dan forma a cualquier tipo de documento presentado en forma de texto, como por ejemplo en un manual de instrucciones. Este tipo de representación discursiva la consideramos de carácter analítico o explicativo, frente a las representaciones estructurales, que son la síntesis de un análisis, como es el caso de las representaciones lingüística no discursivas. Éstas se abastecen para la producción de sus estructuras, ante todo, de términos ordenados alfabéticamente y/o conceptualmente en estructuras más o menos sofisticadas y que representan sistemas conceptuales muy diversos (tesauros, ontologías, bases de conocimiento, glosarios, opciones de software).



**Fig. 8: Formas textuales para la representación del conocimiento técnico**

Los términos son unidades de naturaleza lingüística que representan conocimiento de un determinado dominio y son consensualizaciones para denominar conceptos a fin de poderlos transmitir. Si el lenguaje es trascendental como fenómeno determinante para el desarrollo de conocimiento y los términos representan unidades de conocimiento consensuadas para su transferencia, podemos afirmar que no existe conocimiento explícito especializado sin terminología.

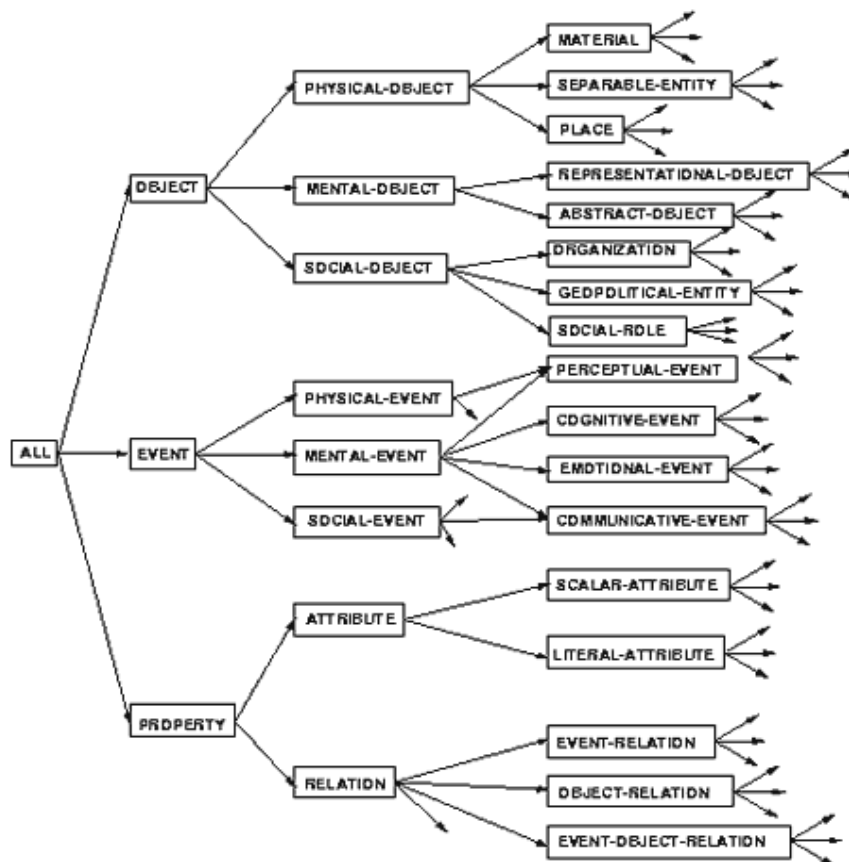
La trascendencia de estas unidades de conocimiento lingüísticas para la gestión de contenido técnico se deriva en la aplicación de estos mismos términos en diferentes entornos. Y esto es así, porque los términos:

- componen la matriz cognitiva o esqueleto conceptual (Cabré, 2002: 94) sobre la que se construye un texto (documento), como los índices analíticos;
- sirven de descriptores para identificar el contenido de un documento, como la información contenida en las propiedades semánticas del documento que mostramos a continuación:



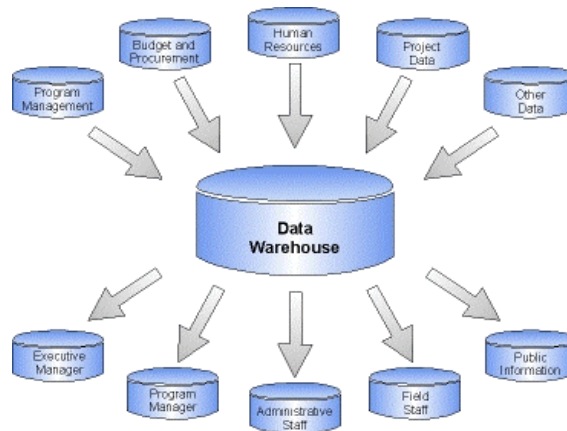
**Fig. 9: Identificación para la catalogación de documentos**

- componen sistemas conceptuales como los glosarios, diccionarios, tesauros, ontologías, como se ejemplifica en el siguiente diagrama:



**Fig. 10: Ontología Mikrokosmos**

- ayudan a identificar las representaciones de las ilustraciones como gráficos, esquemas, mapas mentales, como en el esquema siguiente:



**Fig. 11: Esquema**

- integran los nodos, vértices de los diferentes esquemas de representación del conocimiento (mapas conceptuales, marcos, redes semánticas...).

Si, como hemos visto, los contenidos técnicos requieren incorporar nuevos recursos de representación para optimizar la transferencia de conocimiento, la creación de esos contenidos se beneficia de las técnicas aportadas en el ámbito de la publicación digital, como veremos a continuación.

#### **2.2.4. Técnicas y tecnología de desarrollo de contenido técnico (multilingüe)**

Como se puede inferir de lo expuesto hasta ahora, la ampliación de la diversidad en la representación de la información en el entorno de la documentación técnica está relacionada con el desarrollo experimentado en las técnicas de publicación electrónica. Los documentos electrónicos han revolucionado la gestión del conocimiento gracias a las técnicas de las que ha hecho uso.

Una de estas técnicas es la del *hipertexto*. Ésta supera la secuencialidad del texto organizándolo en una serie de bloques que se conectan entre sí mediante nodos, ofreciendo al usuario la posibilidad de decidir el orden de adquisición de contenido. Otros autores identifican esta característica de los textos electrónicos bajo el término multidireccional (Pinto, 2000a) en el sentido de que las conexiones que realiza el usuario marca la dirección del conocimiento. Inseparablemente relacionada con el método hipertextual se ha desarrollado la técnica hipermedia. Ésta recoge y expande el concepto de hipertexto al incluir no solamente recursos textuales, sino también visuales y auditivos.

La elaboración sistemática de *bases textuales de conocimiento* con vistas a la recuperación de la información es otra de las técnicas aplicadas en la gestión de contenido. Estas bases son la superación de las bases de datos tradicionales de codificación textual como, por ejemplo, las bases de datos bibliográficas, ya que están orientadas al acceso y recuperación de información no exclusivamente atendiendo a su ubicación física, sino a la obtención de documentos totales o parciales y a las que se llega desde múltiples accesos. Las ayudas en línea son un buen ejemplo de estas bases de conocimiento. Se puede buscar por palabra clave, por título, por unidades terminológicas compuestas, estableciendo entre los elementos de la microestructura un campo semántico que facilita al usuario el autoaprendizaje.

La creación de *entornos amigables* o ergonómicos es también una tendencia clara en las técnicas de publicación (Pinto, 2000a). Estos entornos intentan ofrecer al usuario un grado de satisfacción predisponiéndoles a adquirir conocimiento. Esto se logra gracias a una representación cada vez más clara de los contenidos con la ayuda de *interfaces gráficas* fáciles de usar, una integración de las ayudas en línea al producto y una mejora en el acceso de la información que convierte el proceso paulatinamente en más intuitivo e interactivo. Las interfaces gráficas persona-máquina son el componente interactivo de una aplicación informática, pues a través de ésta el usuario

visualiza la aplicación y opera con ella. Las interfaces persona-ordenador, en general, incorporan todos aquellos recursos por los cuales se permite la comunicación entre el usuario y la máquina. Para ello, deben tener un enfoque cognitivo, es decir, deben considerar las actividades mentales como percepción, memoria, aprendizaje y resolución de problemas (Lewis y Rieman, 1993: 1). El desarrollo de interfaces utiliza diferentes recursos, ya que pueden incluir iconos y términos como en los menús en cascada o en los cuadros de diálogo, así como sonidos.

Estas interfaces son interactivas, pues permiten a los usuarios tocar o hacer clic en los iconos y otras áreas sensibles de la pantalla para seleccionar la información o proceso que deseen. Es una tendencia que se da no sólo en los programas de software o en los juegos relacionados con el ocio electrónico, sino en artefactos como por ejemplo, las máquinas de uso industrial, los electrodomésticos o los cajeros automáticos dotadas de interfaces operadas por pantalla táctil o mediante botones, como en el siguiente ejemplo:



**Fig. 12: Interfaz gráfica de cajero automático**

El desarrollo de mejores interfaces que logran un grado de satisfacción para el usuario estriba en su facilidad de uso. Éstas se activan mediante el ratón, botón, tecla o pantalla táctil y tienen como objetivo simplificar y facilitar la tarea de los usuarios en interacción con los sistemas. Por ello, son instrumentos de comunicación.

La aplicación de estrategias de reutilización de textos es otra de las características de las técnicas aplicadas en la producción de documentación técnica. Esta técnica acorta los ciclos de producción, además de los costes de la documentación técnica multilingüe (Ahmad *et al.*, 1992: 7-15). Las razones para poder aplicar técnicas de reutilización residen en las características propias de los contenidos técnicos. Esto es:

- ❑ un gran número de repeticiones en un mismo documento
- ❑ documentos con la misma estructura textual
- ❑ versiones anteriores de un documento

La tendencia de reutilización ha relegado la reformulación de los textos ya elaborados acerca de un producto concreto para trabajar con módulos de información combinables (Höge, 2001a: 23). Para lograrlo es necesaria, por una parte, la estandarización de los elementos textuales y, por otra, la codificación de estos elementos, por ejemplo, mediante metadatos. De esta manera, es previsible que, por ejemplo, las ayudas en línea —uno de los módulos textuales que abordamos en el siguiente apartado— se generen de forma automatizada a partir de los textos explicativos de la documentación general del producto, como es el caso de un manual de instrucciones.

Las técnicas y tecnologías a las que hemos aludido —hipertexto, hipermedia, bases textuales de conocimiento, desarrollo de interfaces persona-máquina amigables y ergonómicas—, están revolucionando la forma de (re)presentar los contenidos. A la par de estas técnicas, la evolución de los soportes electrónicos que almacenan memoria como los CD-ROM ofrecen hoy en día posibilidades de presentación eliminando, por una parte, el soporte en papel y facilitando la aplicación de las técnicas de recuperación y reutilización, por otra.

Asimismo, si los productos y su documentación técnica tienen una proyección para mercados internacionales, los contenidos tienen que tener en cuenta los aspectos de internacionalización y localización, lo que afecta a

la forma de estructurar, redactar, traducir y procesar los diferentes elementos textuales.

La localización es el proceso de adaptación de un producto de software, hardware o cualquier ingenio que incorpore interfaces de usuario a un mercado teniendo en cuenta aspectos lingüísticos, culturales y técnicos. El término estandarizado en inglés es *localization*, abreviado como L10N. Y según los ámbitos de aplicación se distingue entre:

- localización técnica: abarca aspectos tales como la conversión de doble byte, la compatibilidad del sistema operativo, etc.;
- localización nacional: cubre la adaptación de los alfabetos, convenciones de símbolos y formatos y no se limita únicamente a la traducción de la interfaz de usuario;
- localización cultural: adapta el software y la documentación a las características culturales de un lugar (costumbres, historia, convenciones lingüísticas (véase Esselink, 2000: 20-27). Y siguiendo con esta línea de pensamiento podríamos afirmar que la traducción de cualquier texto es una tarea de localización, en cuanto a adaptación a un entorno lingüístico y cultural concreto.

El término *internacionalización* representa un concepto que define el proceso de diseño o rediseño de un producto a fin de que pueda adecuarse a múltiples idiomas y convenciones culturales, es decir, a fin de que pueda localizarse. El término estandarizado en inglés es *internationalization*, abreviado como i18n.

No atender los aspectos de localización e internacionalización puede suponer una pérdida millonaria para la empresa que intenta introducir un producto en un nuevo mercado. Este es el caso por ejemplo de la versión de Windows 95 para el mercado de India. Tras el lanzamiento del sistema a ese mercado, tuvo que retirarse ante las protestas de las autoridades. El sistema operativo contenía un mapa de 80.000 píxeles coloreados de color verde, de los cuales

ocho píxeles identificaban la región de Cachemira —la zona india del conflicto entre India y Pakistán— en un verde mucho más oscuro. La retirada de más de 200.000 versiones del sistema operativo costó a la empresa varios millones de dólares de pérdidas (Der Spiegel, 26.08.2004).

Las pautas de internacionalización para la publicación de contenidos están relacionadas con:

- ❑ las herramientas de software empleadas
- ❑ las pautas de redacción e ilustración
- ❑ las pautas de maquetación
- ❑ la aplicación de terminología

Las herramientas de software que asisten a la redacción y publicación utilizadas juegan un papel determinante en el desarrollo de contenidos con previsión a la internacionalización. Si las herramientas utilizadas dan soporte a metadatos estandarizados como XML/SGML será posible la separación completa de la información textual de la información de diseño o maquetación. Todas las herramientas utilizadas para el desarrollo de contenido (procesadores de texto, programas de DTP) deben prever la compatibilidad con todas las lenguas de destino (codificación de caracteres) a las que previsiblemente se traducirá.

Las pautas de redacción deben aglutinarse en una guía de estilo que describa reglas comunes para todos los redactores. La guía debe contemplar pautas básicas y generales a todas las lenguas, así como otras específicas relativas a una lengua determinada. Algunas de las pautas básicas y generales para elaborar contenido técnico preparado para su adaptación en otros mercados son:

- ❑ redacción de frases cortas (para el inglés se habla de 10 palabras por frase por término medio);
- ❑ aplicación de reglas de puntuación adecuadas al final de la frase (. ! ?);

- elaboración de una lista de todos los acrónimos y abreviaturas, expandirlos, explicarlos y definirlos y, si fuera necesario, se deben definir las reglas de generación de estas abreviaturas para otros mercados. Por ejemplo, la abreviatura ABS viene de *antilock brake system* y en otros idiomas se mantiene esa abreviatura, a pesar de que las iniciales de las palabras que constituyen el término no corresponden con las inglesas. Así para *sistema antibloqueo* o *Antiblockiersystem* en castellano y alemán respectivamente se utiliza ABS;
- exclusión de abreviaturas que acaben en punto. Esto será determinante para que después, en la fase de traducción, se obtenga una segmentación adecuada de los textos para su posterior elaboración con sistemas de traducción asistida;
- exclusión de elipsis o pronombres personales para referirse a otras oraciones o sintagmas;
- redacción de frases completas, prescindiendo en lo posible del estilo telegráfico;
- reutilización al máximo del texto redactado, esto es copiar y pegar antes de volver a redactar;
- revisión de los textos desde el punto de vista lingüístico y técnico antes de someterlo al proceso de traducción.

En cuanto a los elementos ilustrativos que están anidados en la documentación técnica se debe tener en cuenta lo siguiente:

- utilizar números con leyendas o cuadros de texto, en vez de incluir los elementos ilustrativos en el texto;
- si se utilizan cuadros de texto, se debe utilizar para cada elemento un cuadro y dejar espacio para la expansión del texto;



- ❑ si el documento contiene gráficos específicos para cada lengua, no se deben copiar en el documento, sino enlazarlos.

Las pautas de maquetación deben planificarse previamente ya que su aplicación repercute positivamente en la creación de contenidos para su internacionalización. En este sentido, también debe existir una guía para la maquetación de los documentos asociadas a la guía de estilo que hemos visto más arriba. La guía de maquetación debe contemplar pautas de maquetación básicas, generales y comunes y otras más específicas relativas a los programas de publicación (programas de DTP) empleados. Algunas de las pautas básicas y generales de maquetación para elaborar contenido técnico preparado para su adaptación en otros mercados son:

- ❑ creación de una plantilla antes de empezar a redactar que contenga estilos;
- ❑ empleo de la plantilla por todos los redactores;
- ❑ explotación de todas las funciones automáticas contenidas en los programas de redacción como, por ejemplo, la elaboración automática de índice de contenidos, de índices alfabéticos de palabras clave, de tablas mediante la función de insertar tablas, etc.

La composición terminológica del contenido técnico obliga también a crear procedimientos para su generación y aplicación en la fase de la redacción con vistas a una óptima internacionalización. En este sentido se debe tener en cuenta lo siguiente:

- ❑ utilización coherente de la terminología;
- ❑ creación de una base terminológica cuyas entradas contengan el término, la definición, la abreviatura si fuera el caso y las equivalencias a otros idiomas, siempre y cuando el redactor posea esta información;

- utilización de un gestor de terminología en red para dar acceso a la base terminológica para que todos los redactores cuenten con la información terminológica que precisen;
- utilización de herramientas de gestión de terminología de acceso remoto en el caso de que los redactores no estén físicamente en un mismo lugar;
- obligación de introducir nuevas entradas en la base terminológica, siempre que los redactores empleen un nuevo término en el texto;
- acceso a la base terminológica inicial creada por los redactores para que los traductores obtengan información suficiente de cada concepto.

Resumiendo, el hipertexto, los sistemas de ayuda en línea, las interfaces persona-ordenador y, en general, todo lo relacionado con la tecnología web, en cuanto a su soporte electrónico o digital en el que se crea, ofrece unas ventajas de redacción, edición y publicación inimaginables en papel. Al verse superado el papel como soporte del contenido de la documentación y utilizarse el soporte electrónico con tecnología web, las posibilidades de representación de contenidos técnicos mediante recursos lingüísticos y de imagen se han sofisticado. A su vez, la aplicación de técnicas de internacionalización y localización, es decir, la creación de contenido con una previsión internacional pero sin perder de vista la visión local repercute en la producción de documentación técnica. Y no en último lugar, la organización de la información con vistas a su reutilización y recuperación mediante las bases textuales de conocimiento es, por una parte también posible, gracias al soporte y, por otra, gracias al modelado que prevé el almacenamiento y posterior recuperación. Por ello, la tendencia a incorporar en la documentación técnica de un producto no sólo un manual impreso, sino sistemas de ayuda en línea digital para el acceso a la información es una ventaja no sólo en sintonía con las exigencias del usuario, sino en sintonía con los ciclos de producción de la documentación técnica (Knopp, 2000).

### 2.2.5. Los tipos / módulos textuales de la documentación técnica

Tanto desde la perspectiva del redactor técnico, del informático, como del traductor la tipologización de la documentación técnica en módulos textuales obedece a diferentes razones, a las que nos referiremos en este apartado, no sin antes aludir a la cuestión terminológica empleada.

Scholand (1996: 78-91) en su propuesta de tipologización de los tipos de texto (*Texttypen*) que aparecen en un videojuego o programa interactivo — aplicable, en cuanto a sus elementos compositivos, a cualquier programa informático, así como a cualquier artefacto con códigos de programación, procesamientos e interfaces persona-máquina asociadas con las operaciones del artefacto en sí— tipifica los textos en *tipos textuales sencillos* que combinados puede conformar *tipos textuales compuestos*. Esta concepción coincide con la definición de género textual empleada por Gamero (2001: 65-66) en su estudio acerca de los textos técnicos. Así, la combinatoria de los diferentes tipos textuales pueden conformar un género textual, y a la vez un mismo tipo textual puede considerarse en algunos casos como género en sí mismo. Y esto es así, porque dependiendo del fabricante la documentación puede estar estructurada en diferentes documentos.

En otro estudio de Höge (2001b) se utiliza el término *text modules* para identificar los tipos textuales simples que determinan el producto documental final. El concepto que engloba el término *modular* en diferentes ámbitos y concretamente en el ámbito del diseño de sistemas de software, puede darnos las pautas de adecuación e idoneidad del uso del término *módulo textual* en el contexto de la documentación técnica multilingüe.

El modelo de Baldwin y Clark para el diseño de sistemas de software modulares (2000: 63-92) define esta estructura como un sistema compuesto de elementos relacionados entre sí, los módulos. Y en las características de

diseño de estos módulos estriban las ventajas de todo el sistema modular. En este sentido, cada elemento es un módulo caracterizado por ser:

- independiente
- sustituible
- exclusivo
- escalable
- de relevancia modificable
- adaptable

Siguiendo este modelo de Baldwin y Clark y si además tomamos en cuenta la tendencia de aplicación de estrategias de reutilización expuesta para acortar los ciclos de producción de documentación técnica —como consecuencia de esto, también se acortan los ciclos de producción de la redacción y la traducción— recogida en el apartado anterior, el término módulo textual sería aplicable para el diseño de documentación técnica.

Por tanto, nos decantamos por el término módulo textual, porque son éstos los que conforman después géneros textuales de textos técnicos, ya bien sean expositivos o exhortativos, tal y como los tipifica Gamero (1998: 34-38) en su estudio acerca de la función de los textos técnicos.

Del producto que se documente se podrán prever los módulos textuales que deben gestionarse, además de la secuencia de su elaboración y procesamiento. Por ejemplo, si se gestiona un proyecto de contenido multilingüe para un cepillo de dientes, previsiblemente contará con menos módulos textuales que si gestionamos el contenido multilingüe de cualquier programa de software. Y esto no depende únicamente de la complejidad del artefacto descrito, sino de la inclusión o no de elementos de software como, por ejemplo, una interfaz gráfica de usuario, con requisitos textuales adicionales.

Como ya hemos mencionado al final del apartado 2.1. A propósito del término “documentación”, el término *documentación* designa el material explicativo que acompaña a un producto. En el caso de los sistemas o programas de software existe un tipo de material específico de interacción entre el sistema y el usuario que debe tenerse en cuenta en el proceso de diseño, desarrollo y traducción del producto. Este material se agrupa bajo el término de *recursos de software* para diferenciarlo de otra clase de material que se conoce también como *documentación* y al que dedicamos el siguiente apartado.

### **2.2.5.1. Documentación**

Los documentos que acompañan a un producto tienen como objetivo comunicar los requisitos técnicos, los aspectos legales y sobre todo las posibilidades de explotación de un ingenio para que, ante todo, el usuario adquiera conocimiento del producto, o lo que es lo mismo, competencias funcionales y tecnológicas para explotar las posibilidades del ingenio en cuestión, es decir, para la transferencia de conocimiento técnico. Ese material está constituido por los manuales de usuario también conocidos como manuales de instrucciones, los manuales técnicos o de reparaciones y las guías de referencia rápida. Sin embargo y, a pesar del papel fundamental para la que se concibe la documentación de un producto y que cumplen estos documentos enunciados, existen una serie de módulos textuales adicionales que también dan información acerca del producto, como es el caso de las condiciones de garantía, información de embalaje y, en el caso de productos informáticos, la información legal de licencias, así como las actualizaciones de información anexa, como los archivos *léame (readme)*.

La mayoría de los manuales, tanto para los programas informáticos como para los artefactos, se estructuran de forma similar. Lo que puede cambiar es la forma de organización, el almacenamiento de los datos en archivos, así como el soporte de almacenamiento (papel, digital). En este sentido, es posible que la guía de instalación esté incorporada al manual de

instrucciones general o que se presente de forma independiente, como también sucede con las condiciones de garantía. Sin embargo, podemos decir que, en conjunto, la documentación que acompaña a un producto consta de los siguientes módulos textuales:

- ❑ introducción a las funciones del producto o descripción genérica del producto
- ❑ instrucciones de uso del propio manual, identificación de siglas, símbolos, etc.
- ❑ guía de instalación
- ❑ instrucciones de uso de las funciones del producto, contenidas o no en las ayudas en línea
- ❑ guías didácticas o tutoriales
- ❑ información sobre el mantenimiento del producto
- ❑ catálogo con diagnóstico de posibles errores
- ❑ condiciones de garantía
- ❑ contrato de licencias y otras informaciones legales como la información del copyright (aplicable a los programas de software)
- ❑ índice analítico (palabras clave)
- ❑ índice de contenidos
- ❑ información de embalaje
- ❑ folletos publicitarios adicionales
- ❑ archivos *léame* (*readme*) o actualizaciones con información anexa (aplicable a los programas de software)

Asimismo y, dependiendo nuevamente de la planificación realizada para la gestión de estos contenidos, es posible encontrar subcomponentes como archivos gráficos anidados en el propio documento de texto o con marcas de hiperenlaces que nos llevan a los documentos gráficos. Desde el punto de

vista de la traducción, este tipo de archivos son relevantes si contienen términos que deben adaptarse a la lengua de llegada.

Las *ayudas en línea* son instrucciones de uso de las funciones del producto y, junto con los *tutoriales*, son documentos para el autoaprendizaje. Estas ayudas constituyen también la documentación de un producto, especialmente si se trata de un sistema o programa de software. Estos tutoriales son guías didácticas que describen paso a paso las funciones de los artefactos y sistemas para su demostración mediante, por ejemplo, capturas de pantalla. La tendencia es que los documentos tipo instrucciones funcionales o los archivos de ayuda tengan la función didáctica de un tutorial, por eso la delimitación de este tipo textual es difusa.

Las ayudas en línea aunque contengan instrucciones funcionales como en los manuales, se diferencian de estos últimos por utilizar el principio del hipertexto e interfaces de búsqueda por donde se accede a la información a través de palabras clave o tema de los capítulos. Esta información relacionada con las palabras clave y los títulos puede generarse automáticamente mediante los programas de maquetación y edición de contenidos conformando el *índice de contenidos* y el *índice analítico*. Estas informaciones se guardan en archivos aparte con el índice de contenidos que sirven de hipervínculos conocidos comúnmente como los *helpcontext* y los archivos *helpk* que contienen el índice de las palabras clave.

Las ayudas en línea se activan directamente en el entorno informático de la aplicación y contienen instrucciones y descripciones relacionadas con el contexto actual de la aplicación. Sin embargo, desde el punto de vista traductológico, casi no existen diferencias entre la documentación y las ayudas en línea y esto es porque la documentación cada vez más tiende a elaborarse con información codificada, haciendo posible la generación de, por ejemplo, textos explicativos de la documentación general del producto. Por eso, en muchos casos sólo se traduce el manual de instrucciones que se

utiliza para la documentación impresa, para la documentación en línea y para las ayudas en línea.

Los *archivos "léame"* son textos con informaciones actualizadas del producto. Éstos se abren en la pantalla con un simple editor de textos. Al igual que el caso de las ayudas en línea, si estos archivos son textos de pantalla, se consideran parte de la documentación de un programa, puesto que son archivos con información generada tras la impresión del manual con informaciones generalmente de funciones adicionales del sistema.

### **2.2.5.2. Recursos de software**

Cualquier software, tanto sistema operativo como aplicación, consta de un código de programa que controla los procesos y contiene las instrucciones necesarias para activar las funciones. Los archivos dll (Dynamic Library Link) son un ejemplo componente de un programa que ejecuta alguna función.

Parte de estas instrucciones son específicas del control del programa: formato de fecha y hora, sistema de unidades de pesos y medidas, moneda, combinación de teclas, teclado, etc. Éstos son los recursos del sistema o archivos de recursos. En el entorno de la localización de software, la aplicación de la regla de oro de la separación del bloque de código del bloque de datos se ha generalizado hasta definir el estándar para la externalización de los datos localizables: los archivos de recursos o archivos \*.rc (Scholand, 2004).

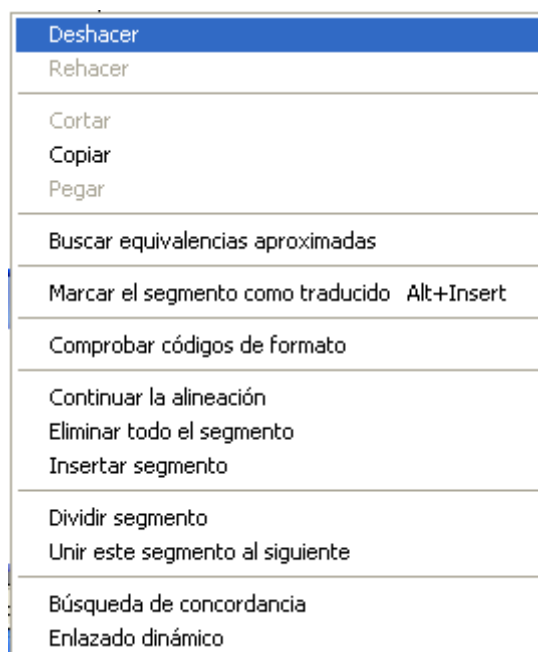
Sin embargo, desde el punto de vista de la explotación de las funcionalidades de un producto que cuente con algún elemento textual en el software, son otros los tipos que nos ocupan y que denominamos: recursos interactivos de software de interacción y los ejemplos de uso de la aplicación. Los recursos interactivos de software cuentan con información textual y pueden ser de los siguientes tipos:

- interfaz de usuario



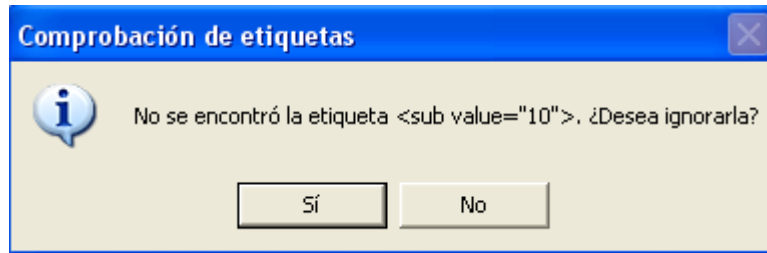
- ❑ cuadros de diálogo
- ❑ asistentes
- ❑ secuencias de imagen con sonido

La *interfaz de usuario* se (re)presenta mediante recursos textuales y visuales permitiendo la operatividad entre usuario y sistema, como es el caso de la información de los botones de las funciones, los menús generales y los contextuales, como se ejemplifica con la siguiente ilustración:



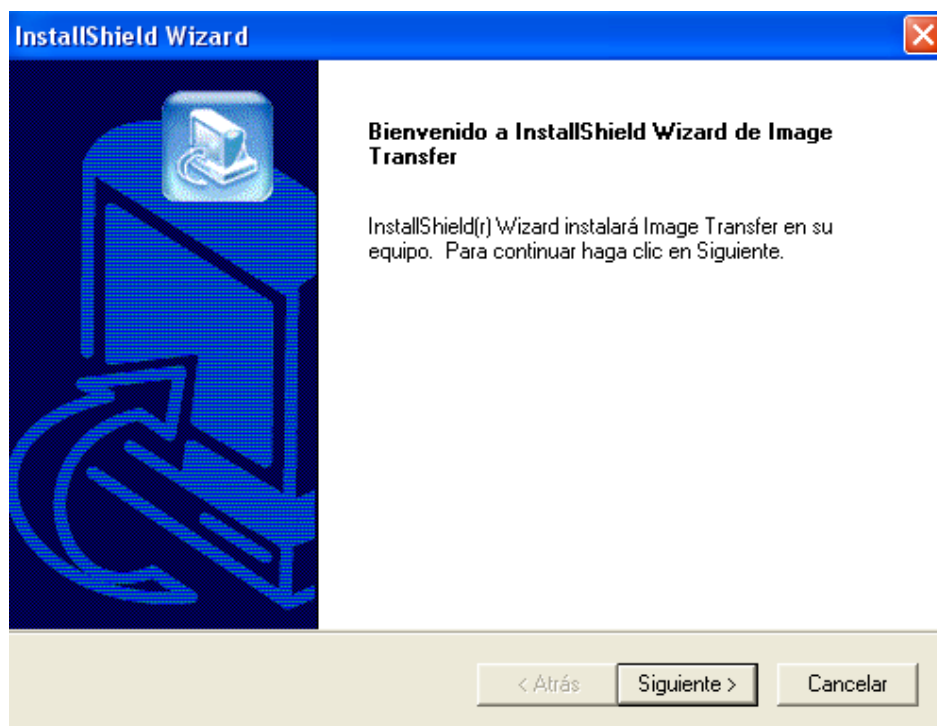
**Fig. 13: Menú contextual**

Los *cuadros de diálogo* son ventanas emergentes con cadenas textuales que contienen mensajes de error, estado o advertencias, explicaciones adicionales o sugerencias funcionales. Con la siguiente ilustración mostramos un ejemplo de cuadro de diálogo:



**Fig. 14: Cuadro de diálogo con advertencia**

Los *asistentes* son aplicaciones sencillas dentro de otro programa que ayudan al usuario a manejarse con funciones del programa y a ejecutarlas por pasos. Este acompañamiento didáctico se realiza mediante interfaces que se asemejan a los cuadros de diálogo y que requieren de intervención del usuario hasta completar la tarea. Existen asistentes para todo tipo de tareas como, por ejemplo, los de instalación de software o hardware en la computadora que mostramos aquí:



**Fig. 15: Ventana de un asistente de instalación**

Las *secuencias de imagen con sonido* son diálogos con figuras de juego asistidas por ordenador que se componen de las expresiones lingüísticas recogidas en un guión presentadas no sólo como texto en la pantalla, sino

como lengua oral reproducida. Este tipo de secuencias son características del software interactivo como los videojuegos, así como de los cursos asistidos por ordenador. Asimismo, las explicaciones narrativas o informaciones de fondo pueden ser parte de los recursos de software si interactúan en función de las opciones seleccionadas por el usuario. Ahora bien, si no existen variables seleccionables, entonces estas locuciones de fondo se pueden catalogar como documentación propiamente dicha, como es el caso de los CDs de presentación de nuevos modelos de coches de cualquier empresa de automoción, por citar solo un ejemplo.

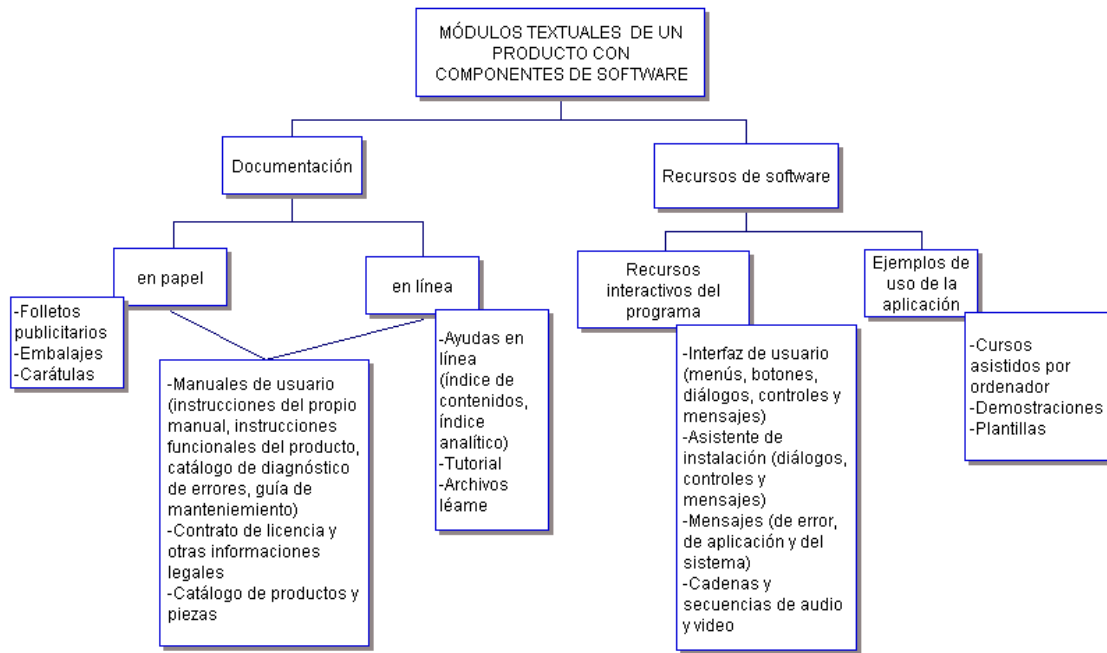
La tendencia general es la de preparar el software desde su diseño contemplando los aspectos de internacionalización, por ejemplo, separando los componentes de recursos interactivos del software en archivos independientes del código fuente, modificando la forma de diseñar y organizar el código de un programa (Wahle, 2000: 34-41).

Como mencionábamos al principio de la descripción de los recursos de software, además de los tipos textuales interactivos, también existen diferentes tipos de ejemplos del uso del software, como pueden ser:

- cursos asistidos por ordenador
- demostraciones
- plantillas

Estos ejemplos de uso pueden estar incluidos en capturas de pantalla de los tutoriales en línea o impresos, así como también pueden simular las acciones principales del programa sin intervención del usuario; los componentes llamados demos (demostraciones o ejemplos de uso), o aquellos que precisan la intervención del usuario, como los cursos asistidos por ordenador.

En el siguiente cuadro mostramos a modo de síntesis los módulos textuales que hemos identificado y descrito:



**Fig. 16: Tipología de los módulos textuales posibles en la documentación de un producto**

La tipificación de los módulos textuales del diagrama es adaptable a la mayoría de los productos técnicos que tengan incorporado algún elemento de software, aun cuando el producto documentado no sea en sí una aplicación informática. Y esto es porque existen cada vez más artefactos que tienen incorporados en sus mecanismos sistemas electrónicos de activación mediante una interfaz de usuario, como puede ser, por ejemplo, el caso de un móvil, una máquina de impresión a cuya configuración se accede mediante una interfaz de usuario o las modernas máquinas que cortan metal dotados de asistentes para el usuario donde éste introduce los valores deseados para que la máquina lleve a cabo los cálculos y las secciones deseadas del metal.

Resumiendo, la documentación técnica es el conjunto de documentos que describen requisitos de procedimientos, especificaciones, explicaciones funcionales y/o de uso de un ingenio tecnológico, tanto artefacto como sistema o en combinación de ambos. Estos documentos representan conocimiento tácito de tipo técnico y son un instrumento esencial para la comunicación técnica, cuyo objetivo principal es facilitar la gestión del

conocimiento al usuario para que pueda explotar su uso. Para ello, estos documentos se valen de diferentes recursos de representación, sobre todo los lingüísticos en combinación o no con recursos visuales y auditivos. A su vez, estos documentos son directos usuarios y beneficiarios de las más modernas técnicas de publicación como el hipertexto, hipermedia y recursos interactivos. A los elementos textuales que componen el contenido total de un producto los hemos llamado módulos haciendo una simetría con el modelo de módulos para el diseño de sistemas de software. Las ventajas del módulo estriban en su condición de independiente, sustituible, exclusivo, escalable, de relevancia modificable y adaptable.

### ***2.3. Gestión y traducción de la documentación técnica multilingüe***

La producción de documentación técnica multilingüe está marcada por la expansión del sector industrial que la produce: la industria GILT. Este sector es al mismo tiempo dependiente del sorprendente desarrollo tecnológico experimentado en los últimos años, así como de la increíble difusión de este conocimiento tecnológico traspasando las barreras lingüísticas.

El ciclo de producción de contenido técnico, en concreto de la documentación técnica, debe entenderse como un proceso industrial compuesto de diferentes fases. Es un proceso complejo de cuya buena planificación y ejecución se derivan después consecuencias que pueden repercutir sensiblemente en la producción y gestión de esos contenidos en diferentes lenguas. En este sentido, la gestión de los proyectos de traducción representa una de las fases de la cadena de producción documental multilingüe.

La gestión de los recursos lingüísticos empleados en la fase de traducción es, sin duda alguna, una de las tareas estratégicas más relevantes para controlar la calidad y reducir los costes de producción. Los recursos

lingüísticos del tipo memorias de traducción y diccionarios terminológicos se componen de información textual que se analiza, planifica, traduce, manipula y explota mediante sistemas profesionales integrados de traducción asistida por ordenador. Las herramientas que componen estos sistemas integrados son imprescindibles, junto con otras anexas que dan apoyo a la gestión de los recursos lingüísticos multilingües.

### **2.3.1. La documentación técnica en su ámbito global**

En los últimos 15 años, la globalización de la economía, el consumo masivo de software en entornos empresariales y domésticos fomentado también por la veloz popularización de Internet ha cambiado el espectro del sector industrial dedicado a la gestión de la documentación técnica multilingüe.

En un artículo publicado hace unos años se estimaban en 20 millones de palabras registradas por día en concepto de información técnica (Murray, 1996). Estas cifras, aun cuando parezcan estar hinchadas, no dejan de ser muy significativas porque reflejan el índice de crecimiento que experimenta el sector y que entendemos mejor, si las ponemos en relación con épocas pasadas.

En ese sentido, el informe de Hubbard (1997: 219) basado también en un análisis estadístico afirma que en los últimos 30 años se ha generado más información que en los 5.000 anteriores. Los antecedentes de la gestión de la información técnica están vinculados a la industria de la localización que emergió impulsada por el gran aumento de las necesidades de traducción y adaptación técnica y cultural a otros mercados. Este crecimiento ha sido desde sus comienzos notable debido a:

- las disposiciones legales de cada vez más países para que los productos estén documentados y localizados para el mercado de destino;

- la ventaja que obtienen los productos localizados con respecto a la competencia;
- la necesidad de que la tecnología cada vez más sofisticada sea descrita para poder explotar mejor sus posibilidades de uso y funcionamiento.

Si antes se hablaba de *industria de la localización* ahora este sector empresarial ha aumentado su radio de acción ajustándose a las necesidades de sus clientes potenciales ofreciendo una paleta de servicios ampliada que va desde la gestión de contenidos (creación, edición, publicación), es decir, gestión de documentación, aplicando técnicas de internacionalización y localización, hasta el diseño y desarrollo de herramientas que gestionan contenidos en diferentes lenguas (sistemas de traducción asistida, gestores terminológicos, convertidores, alineadores, extractores, correctores, bancos terminológicos, corpus en línea, etc.).

Por tanto, de la denominación original *industria de la localización* se ha pasado a la denominación *industria GILT*, como así lo muestra el texto introductorio actual de la prestigiosa asociación LISA (*Localisation Industry Standards Association*):

Organization for the GILT (Globalization, Internationalization, Localization, and Translation) business community. Over 400 leading IT manufacturers and services providers, along with industry professionals representing corporations with an international business focus, have helped establish LISA's best practice guidelines and language-technology standards for enterprise globalization (LISA, 2004).

El ajuste denominativo está relacionado con la propia proyección internacional de cualquier producto. Esto obliga a la planificación estratégica para:

- la gestión de los contenidos multilingües,

- la asesoría a los clientes en el proceso de planificación de estrategias de globalización, es decir, de internacionalización y de localización.

En este sentido, globalización se entiende como el proceso de adaptación de productos para el mercado internacional (IBM, 2004 y Microsoft, 2004). Debido a la necesaria implicación de estrategias de internacionalización, de localización y traducción en el proceso de la globalización, se entiende que en estos últimos años se esté popularizando la identificación del ampliado sector industrial de la localización también como industria GILT. Este sector industrial se ha convertido en el motor que aviva y dirige la sociedad de la información multilingüe donde los recursos lingüísticos y su gestión se han convertido en el factor clave impulsador de estrategias, formatos y tecnologías.

Las cifras mundiales en cuanto a ingresos de este sector se estimaban en los 2.800 millones de dólares en 1997 con una previsión de 6.200 millones en el 2000 (Loughman, 1997). Las cifras más actualizadas muestran un crecimiento aún mayor en los años sucesivos. De 10.400 millones de dólares cifrados en el 2000, se preveían unos ingresos de 17.200 millones para el 2003 (Austermühl, 2001: 4), aunque otras estimaciones apuntaban los 20.000 millones de dólares (Allied Business Intelligence, 1998). Según el último análisis consultado, las previsiones anteriores ya han sido superadas, ya que en el 2004 ya se ha traspasado la cifra de los 20.000 millones de dólares (Schäler, 2004: 2).

Esto quiere decir que la industria GILT sigue en expansión, y esto es porque la producción de documentación técnica multilingüe va en aumento. Sin embargo, este aumento progresivo del volumen de negocio en cifras absolutas no debe entenderse en relación proporcional a la producción (creación y traducción) de documentación multilingüe. La documentación gestionada aumenta aún más como consecuencia del esfuerzo que están llevando a cabo los implicados en el sector para optimizar los costes.



A la par, son también reflejo de la expansión de este sector la fundación de asociaciones como LISA en 1990 o de la asociación GALA (*Globalization and Localization Association*).

GALA se funda en 2002 con el objetivo de promover la colaboración entre las empresas proveedoras de productos y servicios vinculados a la industria GILT y con ello aumentar y optimizar el conocimiento industrial del sector con el objetivo de reducir costes de producción (GALA, 2002). Asimismo, el espectro de los miembros asociados da cuenta de la interdisciplinariedad, multisectorialidad y globalidad de los aspectos abordados.

A su vez, las asociaciones de comunicación técnica que se han ido creando en los últimos años, como la STC (*Society for Technical Communication*) o la pionera Tekom (*Deutscher Fachverband für Technische Kommunikation, Dokumentation und Informationsentwicklung*) es una prueba de la expansión y trascendencia de la comunicación técnica.

En estas asociaciones se tratan las necesidades del sector de documentación técnica multilingüe vehiculados por los agentes implicados.

Las pautas de trabajo desarrolladas por estas asociaciones son instrumentos para redactores técnicos, editores, autores de contenidos, documentalistas, ilustradores técnicos, diseñadores, profesores, arquitectos de la información, diseñadores gráficos y de web. Estas pautas o recomendaciones son herramientas para todos aquellos implicados en crear y convertir la documentación técnica en información y conocimiento accesible a todos los que lo necesiten.

Para lograr sus objetivos, estas asociaciones fomentan la promoción y el reconocimiento de estas actividades para que adquieran el estado de profesión reconocida, informan acerca de las tecnologías, métodos y aplicaciones de la comunicación técnica multilingüe y ofrecen formación profesional, entre otras iniciativas (STC, 2004 y Tekom, 1978).

### **2.3.2. Ciclo de producción de la documentación técnica (multilingüe)**

Producir documentación técnica de un producto presenta en la actualidad grandes retos, insospechados hace dos décadas. Y esto es así porque:

- ❑ los plazos son cada vez más apretados,
- ❑ el volumen de información que se debe gestionar crece espectacularmente y,
- ❑ el carácter innovador de su contenido dificulta aún más el proceso de redacción por no existir referentes, problema que se extiende también en la fase de traducción.

Cada vez más la gestión corporativa de una empresa entiende que la producción de documentación es crucial para hacer su lanzamiento al mercado. Por tanto, hay que entender su creación como parte de la planificación del producto completo y tratar esa información de forma óptima.

El tratamiento de la información supone considerar los recursos informacionales como activos susceptibles a ser reutilizados como, por ejemplo:

- ❑ los módulos textuales
- ❑ otros recursos no lingüísticos

Éstos aparecen en diferentes tipos de documentación y publicados en diferentes formatos y entornos (Hofmann y Mehnert, 2000: 59-60).

La creación de la documentación técnica se compone de diferentes pasos. Es un proceso complejo de cuya buena planificación y ejecución se derivan después consecuencias que pueden repercutir sensiblemente en la producción y gestión de esos contenidos en diferentes lenguas. En este contexto cada vez más se concientia tanto a redactores como traductores a

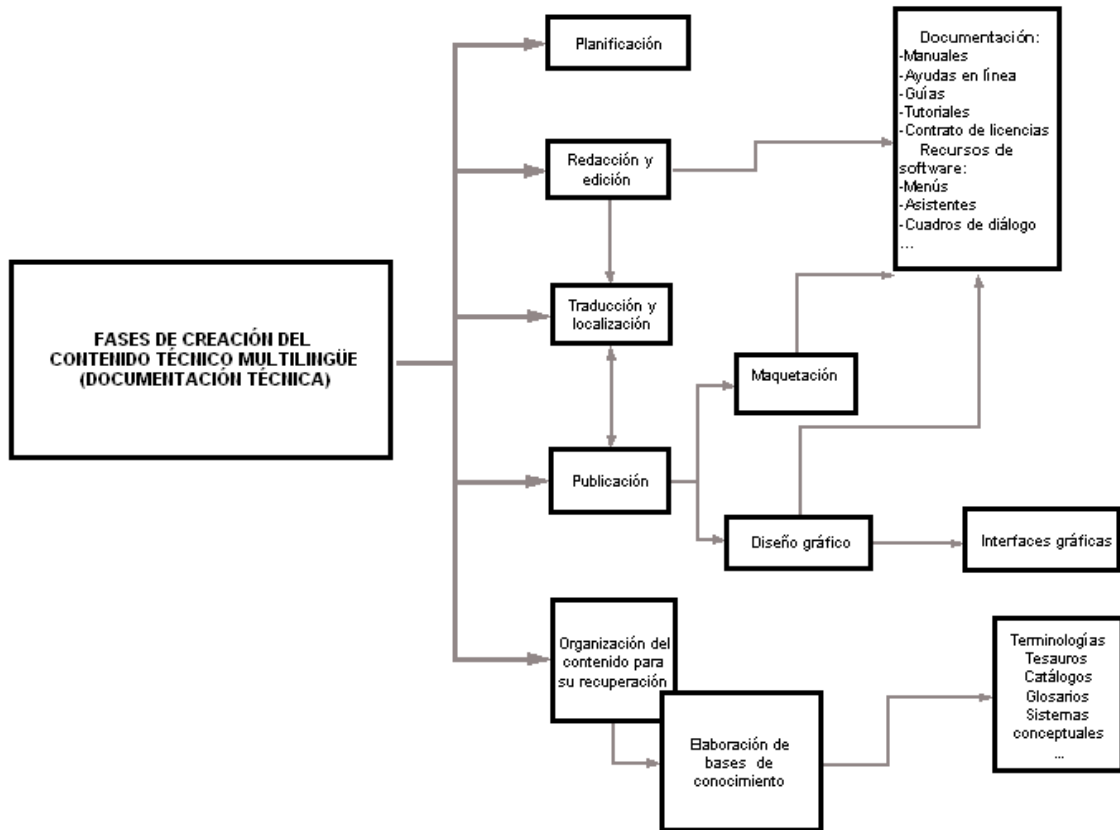
simplificar sus tareas para poder aumentar la velocidad de producción, aplicando estrategias de internacionalización, como la estandarización y la reutilización de textos.

El ciclo de producción del contenido técnico es comparable con la cadena documental de un documento entendido éste como objeto manipulable y susceptible a ser reutilizado (véase Méndez, 2002: 46-47).

Las fases de nuestra propuesta que identifica el ciclo de creación de contenidos multilingües son:

- ❑ la planificación
- ❑ la redacción y edición
- ❑ la traducción y localización
- ❑ la publicación
- ❑ la organización del contenido para su recuperación

En el siguiente gráfico mostramos las fases a las que nos hemos referido:



**Fig. 17: Creación de contenido técnico**

Este gráfico presenta escuetamente las fases de creación del contenido técnico, coincidiendo, de arriba abajo, con el orden del proceso. Hemos proporcionado algunos ejemplos a cada una de las fases. Como se puede ver en el gráfico, hemos vinculado las fases de redacción, edición y publicación, no así la de planificación del proyecto y de la organización del contenido. Esto se justifica porque la planificación de un proyecto de documentación técnica abarca todas las fases. De una buena planificación de los objetivos, de los recursos humanos empleados, de la tecnología aplicada y de la gestión de los recursos de representación empleados dependerá que el producto final sea de calidad con un coste de producción asumible. La etapa de planificación tiene que incorporar en su estrategia una organización sistemática del contenido. Y esto se lleva a cabo mediante una serie de procedimientos documentales como la catalogación, indización y el diseño de bases de datos, que desemboquen en la creación de bases de conocimiento abocadas a la recuperación y extracción.

Asimismo, la recuperación del contenido fragmentado facilita la adquisición del conocimiento, además de servir para la reutilización de esos contenidos para la creación de posteriores versiones de contenidos técnicos (Hofmann y Mehnert, 2000: 69-71), o para hacer productos adicionales, como diccionarios terminológicos.

Si la fase de planificación ha previsto la organización sistemática (conceptual y tecnológica) de los contenidos, previsiblemente será posible su reutilización. Pensar en reciclar es imprescindible debido al corto ciclo de vida de los documentos (Arevalillo, 2004: 467). Si además el producto que se documenta está pensado para otros mercados con variantes lingüísticas y culturales, se debe tener en cuenta también en la fase de planificación, la incorporación de técnicas de internacionalización que haga después más sencilla la fase de localización en cualquiera de los niveles a los que ya hemos aludido y que comentaremos con más detalle en el apartado 2.5. Formatos de codificación e intercambio.

La creación automatizada de contenidos es un ejemplo de que, sin planificación previa, difícilmente podría llevarse a cabo. Para ello, se planifica cómo se va a emplear el lenguaje para poderlo controlar en la fase de la redacción de documentación. El lenguaje controlado que se preestablece en la fase de planificación se compone de un conjunto de recursos lingüísticos acotados, dotados de una gramática reducida y un vocabulario restringido que permite ser entendido y procesado sin mucho margen de error por sistemas, por ejemplo, de traducción automática (Budin y Wright, 2001: 872). En este sentido, la fase de planificación está íntimamente relacionada desde un principio con la organización del contenido. Asimismo, la fase de planificación debe prever estrategias de control de calidad como la corrección ortográfica y de estilo de los elementos lingüísticos o la revisión de la posición de los recursos extralingüísticos, así como el control de funcionamiento de los componentes que precisan la intervención del usuario.

Así, una vez se han redactado los contenidos, se pasa a la fase de edición en la que se debe contemplar la corrección lingüística como paso previo a la publicación. En la fase de publicación es donde se maqueta y termina de organizar los elementos lingüísticos y extralingüísticos maquetando e implementando el diseño gráfico de, por ejemplo, las interfaces gráficas de usuario.

Con referencia a los procedimientos de control de calidad en servicios profesionales de producción de contenidos multilingües, cabe destacar la existencia de estándares. Un estándar es un documento definido y consensado por un grupo de trabajo reconocido. Un estándar define normas, guías y criterios de evaluación para una actividad. Las normas internacionales generales y específicas aplicables a la producción técnica multilingües son, por ejemplo:

- ISO 9000, 9001, 9002, contemplan normas para garantizar la calidad en el proceso industrial. De estas normas se desprende que hay ciertos elementos que todo sistema de calidad debe tener bajo control, con el fin de garantizar que los productos y servicios de calidad se fabriquen de forma consistente y a tiempo. La serie ISO 9000 fue creada entre el 2000 y el 2002 por comités integrados por representantes de 27 países quienes se encargan de las actualizaciones necesarias. Ha sido adoptada por más de 70 países alrededor del mundo como la norma de mayor aceptación que establece requisitos para los sistemas de calidad.
- ISO 2384, norma publicada en 1977, establece las reglas para presentar de forma normalizada las traducciones con el objetivo de facilitar su uso por las distintas categorías de usuarios. Son reglas de documentación que se aplican a las traducciones (completa, parcial o abreviada) de cualquier tipo de documento.
- LISA QA 3.0, modelo normativo publicado en 1995, establece un conjunto de normas propuestas por la asociación LISA para la gestión

y control de la calidad en el proceso de localización de todos los componentes de un producto.

- UNI 10574, norma italiana publicada en 1995, documenta por primera vez en el entorno europeo las actividades relacionadas con la traducción e interpretación para el control de la calidad específicamente para los servicios de traducción. El organismo responsable de su publicación es UNITER (*Organismo di Normazione e di Certificazione di Sistemi di qualità italiano*).
- DIN 2345, publicada en 1998 por el Instituto alemán para la normalización (*Das Deutsche Institut für Normung*), propone normas específicas para la gestión de proyectos de traducción y localización. Se centra en el proceso de traducción y en la gestión empresarial de esos proyectos. Atiende a aspectos organizativos como presupuestos, contratos, formas de entrega, acuerdos de confidencialidad, entre otros. Esta norma recoge también aspectos de contenido relacionados con el control de las traducciones y la interacción con el cliente.
- ÖNORM D 1200 y 1201, realizado en el 2000 por el Centro de servicios austriaco para la normalización (*Österreichische Dienstleistungszentrum für Normen*), conforman un paquete de pautas específicas para la gestión de proyectos de traducción. Estas normas abordan las mismas actividades que la norma DIN 2345. La diferencia de la norma austriaca reside en la tipificación de los servicios ofrecidos por las empresas del sector, la amplia descripción de las competencias de los agentes implicados, así como la regulación de responsabilidades del proveedor de servicios.
- CEN/BTTF 138, norma iniciada en 2002 y que previsiblemente deberá terminarse en el 2005 por el Comité Europeo de Normalización, que pretende aglutinar propuestas europeas para conseguir el establecimiento de un estándar europeo común aplicable a todos los proveedores de servicios de traducción, tanto empresas de traducción, como traductores autónomos. La iniciativa parte de la Asociación de la

Unión Europea de empresas de traducción —EUATC (*European Union of Associations of Translation Companies*)—. AEONOR, la Asociación española de normalización y estandarización, asume la dirección del proyecto repartido en varias tareas ejecutadas por diferentes países. Así, el grupo de trabajo de Alemania se encarga de establecer los términos y las definiciones del sector; el equipo de Austria establece los requisitos básicos tanto técnicos como humanos para ofrecer servicios de traducción; el finlandés define la relación entre el cliente y el proveedor de servicios (contrato marco, obligaciones y derechos de las partes); el grupo español está elaborando los procedimientos administrativos, lingüísticos, técnicos, terminológicos de los servicios de traducción; y por último, el equipo inglés define los servicios de valor añadido para el entorno de la traducción. Todas estas propuestas abocadas al control de calidad muestran también algunos de los flujos de trabajo relacionados con el proceso industrial de la producción de contenido multilingüe y que ha inspirado el siguiente apartado relativo a las tareas relacionadas con la fase de traducción.

### **2.3.3. Fase de traducción**

La gestión de los proyectos de traducción representa uno de los engranajes que integra todo el mecanismo del flujo de trabajo de la gestión de contenido multilingüe. Desde un punto de vista de creación documental, los proyectos de traducción técnica pasan prácticamente por las mismas fases de creación a las que están sometidos los originales. Ciertamente hay algunas diferencias, por ejemplo, durante la fase de redacción de traducciones, los documentos originales sirven de plantilla para estructurar el documento equivalente en otro idioma, lo que simplifica esa tarea de organización del contenido.

Si se toman estos paralelismos en cuenta, es lógico afirmar que de la buena gestión de la documentación técnica original dependerá que la parte de



adaptación tanto lingüística como técnica a entornos multilingües sea más sencilla.

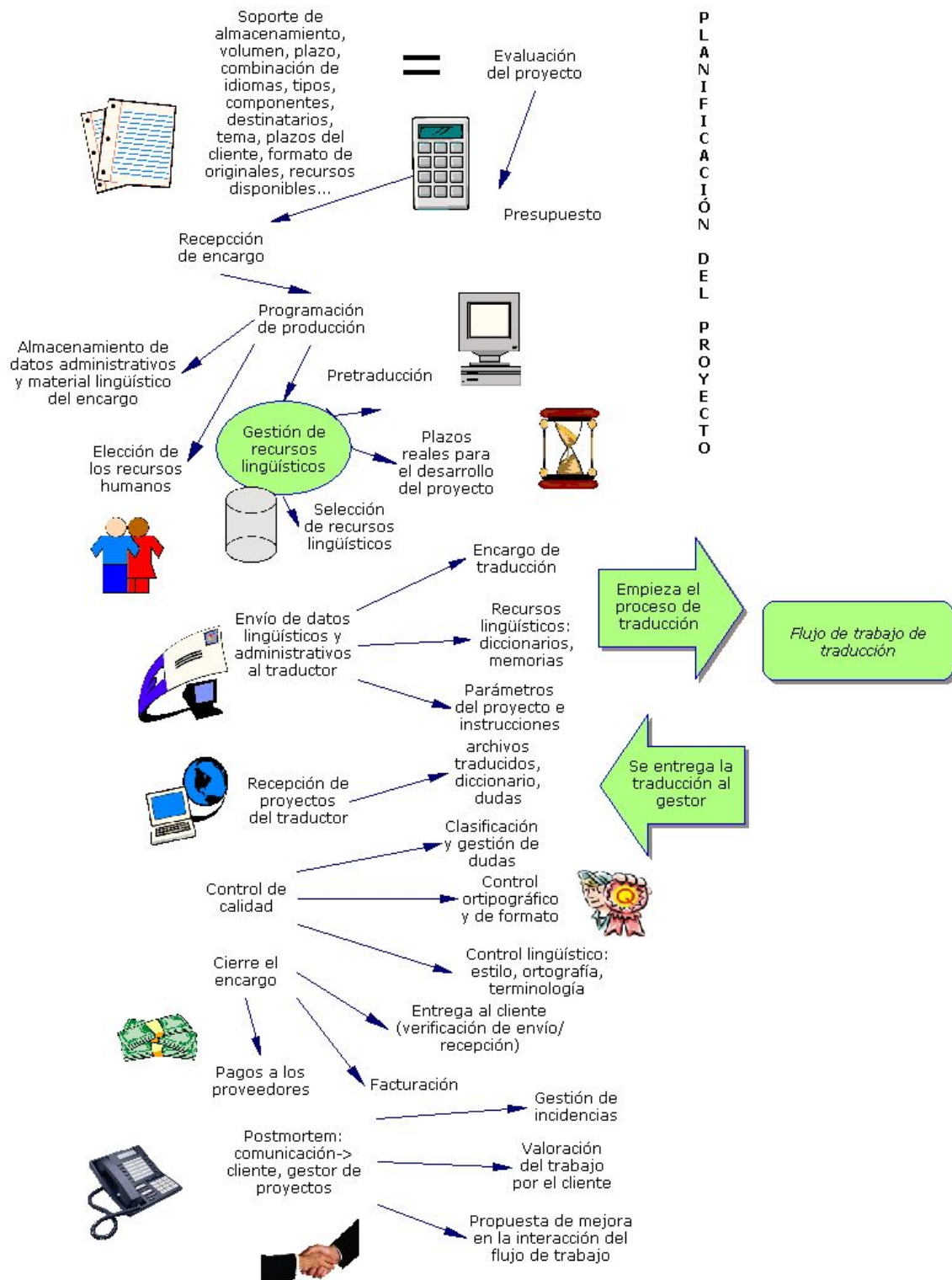
En ese sentido, y tal como hemos explicado, se han de tomar en cuenta criterios de internacionalización no sólo para el desarrollo del producto en sí, sino para la producción de documentación. A la par, también son determinantes las tareas específicas de control de calidad que deben llevarse a cabo en el proceso de redacción y edición: control ortográfico, de estilo, armonización terminológica; en la fase de maquetación: control ortotipográfico y formato, verificación del funcionamiento (*betatesting*), por ejemplo, de una aplicación de software o de los hiperenlaces de un sitio web. La verificación del funcionamiento o más conocido como *betatesting* es el término dado al proceso de probar la versión beta de un programa. La industria del software utiliza el término versión beta para identificar aquellas versiones de un programa o aplicación que no están completamente depuradas. Se trata pues de versiones de prueba de un software determinado, desarrolladas antes de la versión definitiva.

Puesto que usualmente y concretamente los programas de software antes de sacar la versión final, se lanzan en etapas (versiones alfa y beta) para probarlos, y precisamente durante las fases previas a la final se lleva a cabo la localización en diferentes idiomas, es frecuente que los implicados en la revisión final tanto traductológica como técnica en proyectos de traducción y localización sirvan de *betatesters*. Haciendo las pruebas se detectan los posibles errores no sólo relativos a la localización, sino al funcionamiento del programa. Como norma, los *betatesters* no pertenecen a la empresa desarrolladora del software en cuestión, por tanto, que la empresa encargada de la localización del producto asuma esta tarea de realización de las pruebas del programa en un entorno real, es muy beneficioso para ambas partes.

Por tanto, la calidad de la redacción de originales, la estructuración de la información contenida y la calidad de las traducciones y localizaciones influye

decisivamente en el cumplimiento del principal objetivo de transferencia de conocimiento que debe cumplir la documentación técnica.

En el siguiente gráfico hemos representado el flujo de gestión de proyectos de traducción de una empresa proveedora de servicios lingüísticos con perfiles de los colaboradores internos y externos diversificados para optimizar el flujo de trabajo. En el caso de un traductor autónomo que trabaja para un cliente directo, coinciden muchos de los pasos que mostramos, ya que, en realidad, el autónomo constituye una empresa unipersonal, pero, sin la estructura de una empresa real, lo que no le permite diversificar las tareas implicadas en el proceso de gestión de un proyecto de traducción.



**Fig. 18: Flujo de gestión de los proyectos de traducción**

A diferencia del gráfico sobre la creación de contenido técnico, en el anterior se contempla la interacción existente con el cliente y con los recursos

humanos implicados. La parte específica de realización de la traducción no la hemos desarrollado, por dos razones: la primera, porque este trabajo no aborda los problemas cognitivos específicos de resolución de translación semántica, sintáctica y morfológica, y, en segundo lugar, porque el diagrama de flujo propuesto atiende la perspectiva de la empresa de traducción y localización. Esto quiere decir que parte de la traducción se gestiona con traductores externos en condición de autónomos, mientras que las tareas de preparación del proyecto se llevan a cabo por gestores de proyectos y las de control de calidad por revisores, que son traductores con experiencia demostrada.

### **2.3.3.1. Planificación del proyecto**

Tras la solicitud de un presupuesto, se procede a la comprobación del encargo para poder planificar el proyecto de traducción. Se analizan los datos que componen la totalidad del material. Las variables que determinan este análisis son de tipo textual, pero igualmente relevantes son las de tipo extratextual como se explica a continuación:

- El soporte de almacenamiento. Si el cliente entrega los datos de forma electrónica o digital, se puede automatizar la planificación y presupuestación del proyecto. Obtener los datos en papel ralentiza el proceso de evaluación, planificación y, por ende, de gestión del proyecto. De todas maneras, cabe mencionar que la tendencia creciente es la de entregar todos los documentos para traducir en formato electrónico.
- El componente textual y formato de codificación. Vinculado al soporte de almacenamiento debemos tener en cuenta qué módulos textuales se deben gestionar y en qué formato de codificación se entregan.
- El volumen del documento y el plazo de entrega (el nivel de urgencia). La combinación de la información del número de palabras o de líneas de las que consta el proyecto de traducción, junto con el plazo de

entrega ofrece una relación de las personas que debemos utilizar en el proyecto. Utilizando herramientas adecuadas, como por ejemplo, las de traducción asistida, también es posible saber cuánta repetitividad interna tienen los textos. Esto quiere decir, cuántas veces se repite una unidad de traducción.

- El tipo de texto y grupo destinatario de los textos. La información de a quién va dirigido el texto que se debe traducir da pautas acerca del grado de complejidad conceptual que entraña la traducción. Un manual de instrucciones destinado a un usuario no necesariamente especializado, previsiblemente no tiene la misma dificultad que un manual de reparaciones destinado a técnicos especializados.
- El tema de la traducción. Si se tiene esta información se puede resolver mejor la selección de los recursos humanos que deben emplearse para la realización.
- La combinación de idiomas. El idioma original en que está escrito y al que se desea traducir es muy relevante ya que dependiendo de esto se contarán con más o menos recursos. Asimismo, la ausencia de estos recursos en las bases de datos disponibles, obliga a buscarlos, seleccionarlos y evaluarlos, lo que previsiblemente afecta a los plazos de realización del trabajo.
- El material adicional de referencia (recursos lingüísticos y visuales) que aporta el cliente en forma de textos paralelos, glosarios específicos y publicaciones anteriores, en general. Los textos paralelos son textos equivalentes, bien porque uno es el resultado de la traducción de un original, o bien porque ambos se han generado semiautomáticamente siguiendo parámetros fijos de control lingüístico. Contar con textos paralelos simplifica el trabajo de documentación y traducción, lo que supone poder afinar los presupuestos.

### **2.3.3.2. Programación de producción**

Fruto de toda la fase anterior, se presupuesta el trabajo al cliente. Si se acepta, el proyecto entra en la fase de programación de producción, asignado a un gestor. El gestor de proyecto es el punto de contacto con el cliente para solventar dudas, informar sobre el desarrollo del trabajo con respecto a la programación de producción acordada. El término *gestor de proyectos* puede variar y denominarse *jefe de proyectos* (Arevalillo, 2003: 14 y 20) identificando con ello el mismo concepto. Dependiendo de la estructura empresarial pueden coexistir el término *gestor de proyectos* y *jefe de proyectos*, siendo éste último el que coordina todo el trabajo de los gestores.

El gestor de proyectos se ocupa también de decidir qué personas realizarán la traducción, adaptación y revisión. La disponibilidad de los traductores, la especialización y las tarifas que solicitan son las variables que influyen en la selección de los recursos humanos, ya que el gestor debe responsabilizarse de que los plazos y presupuestos acordados se cumplan.

Además de la selección de los recursos humanos implicados, el gestor debe ocuparse de gestionar (seleccionar y manipular) los recursos lingüísticos disponibles para:

- llevar a cabo la pretraducción automática basada en las memorias de traducción seleccionadas y, en casos determinados, basados en bases terminológicas;
- enviar a los traductores recursos lingüísticos (memorias y glosarios) vinculados al proyecto de traducción para contribuir en la producción de material de calidad.

### **2.3.3.3. Envío de los datos al traductor**

El traductor a quien se le asigna el proyecto de traducción recibe del gestor el encargo de la traducción, con información administrativa: tarifas, palabras, número de encargo, etc., tal y como se muestra en el siguiente ejemplo:

Nº encargo int. <sup>1</sup>	9918_DAI	Título	Alfing
Fecha encargo	05-10-04	Componente	STRINGS DEU/ENG-ESP
Fecha entrega	07-10-04	Gestor	FG3
Traductor/a	DAI	Idiomas	DEU->ESP
Tipo trabajo	TR	Precio Pal LS	0,051
Palabras LS	943	Precio Pal Parc. Prctr.	
Palabras LD		Precio hora	
Tiempo estim.		<b>Total suma encargo</b>	<b>48,09 €</b>

**Fig. 19: Ejemplo parcial de hoja de encargo para el traductor**

Una vez que se han tomado las decisiones relacionadas con las herramientas que se van a utilizar para la traducción —esta cuestión la abordamos con más detalle en el apartado 2.6. Herramientas para la gestión de recursos lingüísticos multilingües—, el gestor debe enviar al traductor:

- los parámetros que definen el proyecto (formato, combinación de lenguas, memorias y diccionarios asociados, etc.);
- los documentos que se deben traducir importados al entorno técnico de trabajo previsto, si fuera necesario;
- el material de referencia, es decir, memorias y diccionarios, si existen para el proyecto;
- los documentos en formato \*.pdf para poder visualizar el texto maquetado en sus formatos de archivo originales, en el caso de que los archivos que se deben traducir ya estén importados a una herramienta de trabajo de la traducción.

Las instrucciones específicas del proyecto, así como las guías de estilo específicas del cliente o el producto, si es que ha lugar, conforman todo el paquete de información para el traductor que el gestor envía a los traductores, o por lo menos, eso es lo que, en una situación ideal, debería recibir el traductor para poder optimizar la calidad del trabajo.

### **2.3.3.3. Traducción**

El análisis del proceso traductor encierra una gran complejidad. Además de la dificultad intrínseca que comporta cualquier proceso cognitivo por no ser directamente observable, se añade la dificultad que genera el análisis de un proceso, como es el traductor. Este proceso se desarrolla en diversas fases e intervienen no sólo numerosos conocimientos y habilidades, sino también numerosas destrezas que plantean retos y exigencias al traductor técnico y, en general, a cualquier traductor sea cual sea la especialidad que tenga.

Los procesos mentales que desarrolla el traductor en el momento de empezar su actividad traductora se aglutinan bajo el término de proceso traductor, tema abordado desde diversas perspectivas por diversos autores (Krings, 1986; Lörscher, 1991 y 1996; Kiraly, 2000; Hurtado, 2001). Sin embargo, carecemos de estudios empíricos que demuestren la incidencia de la aplicación sistemática de herramientas profesionales de traducción en este proceso. La razón que alega Lörscher (1991: 5) para explicar esta situación es que la mayoría de los estudios que analizan el proceso de la traducción lo hacen desde una perspectiva teórica y especulativa, sin contemplar la perspectiva pragmática.

Sin embargo, se puede intuir que la participación sistemática de herramientas que automatizan partes del proceso repercute en el proceso traductor. En ese sentido, el flujo de trabajo de la traducción con herramientas de traducción asistida se asemeja más a un proceso complejo e interactivo humano-máquina. Este proceso se aleja cada vez más de modelos lineales en el sentido secuencial, acercándose a propuestas que identifican la traducción como un modelo de análisis de retorno que vuelve sobre las decisiones o propuestas terminológicas o de traducción ya realizadas, generando un conocimiento traductológico basado en procedimientos de rutinas (Nord, 1991: 30-34; Wilss, 1992: 115).

De todas maneras, ya que analizar el proceso de la traducción propiamente dicha no es el objetivo de este trabajo, lo hemos simplificado en el diagrama



*Fig. 18: Flujo de gestión de los proyectos de traducción mediante una flecha de salida y otra de entrada.*

#### **2.3.3.4. Recepción del proyecto traducido → Control de calidad**

El papel del gestor de proyectos también es fundamental en esta fase, ya que recae sobre su responsabilidad que se cumplan los procedimientos de calidad y de que el proyecto se lleve a cabo respetando los plazos. Y esto es, aun cuando el gestor delegue la parte del control de calidad lingüística a otra persona. En este sentido, el gestor debe comprobar que los datos enviados por el traductor estén completos, todos los archivos traducidos, en el formato solicitado. Una vez verificada la recepción de los datos entra en juego el revisor.

El revisor se encarga de comprobar que la utilización de la terminología técnica es correcta, en función de las instrucciones establecidas, así como la corrección lingüística (morfológica y sintáctica) y de estilo. Una vez hecho el control de calidad se debe confeccionar un informe evaluativo para el traductor con el fin de mejorar su trabajo si fuera preciso. Este informe puede componerse de recomendaciones, lista de problemas y dificultades y/o catálogo de errores detectados. Esta tarea ayuda a la formación continua del traductor, aumentando la calidad de trabajo.

Dependiendo de la diversificación de los perfiles de una empresa de traducción y del tipo de trabajo solicitado por el cliente, el control de maquetación, que es la fase de cohesión de los recursos lingüísticos con los recursos ilustrativos y de formato, puede llevarla a cabo un especialista en el campo o, a veces, el mismo gestor. Dependiendo de la organización de la empresa, esto lo asume el gestor o bien, la persona específica con estas competencias en maquetación.

Dependiendo del proyecto puede ser que se precise compilar los datos, esto es, convertir los datos de los recursos textuales de un producto de software, es decir, el conjunto de instrucciones en forma de texto, al código máquina

para que los datos de la traducción se puedan ejecutar por la aplicación que ha sido traducida y localizada. En este caso, también dependiendo de la diversificación de las tareas de la empresa y de la formación del gestor, esta tarea puede llevarla a cabo el gestor o un técnico. Una vez realizado, se debe hacer la comprobación del funcionamiento (betatesting).

#### **2.3.3.5. Cierre del encargo**

El proyecto se cierra cuando se entrega al cliente según las condiciones establecidas en el encargo. Asimismo, como parte de las tareas para cerrar el encargo, se almacenan adecuadamente los datos enviados y los recibidos de los proyectos, se eliminan los archivos temporales de los directorios del proyecto y se verifican y confirman las facturas de honorarios. Los gestores de proyectos responsables de un proyecto tienen que comprobar las facturas emitidas por los colaboradores que han participado en dicho proyecto.

#### **2.3.3.6. Postmortem**

Pero más allá de la entrega de trabajo realizado, la fase postmortem del proyecto es básica. En esta fase se reestablece la comunicación entre el cliente y el gestor de proyecto y puede repercutir positivamente en un futuro en el flujo de gestión de proyectos de traducción con ese cliente.

La forma de hacer la evaluación final del proyecto puede variar. Un método común es la realización de cuestionarios que sirvan al cliente de guía para evaluar su propio proceso de producción de la documentación técnica original y el de la empresa proveedora. La entrevista puede cumplir el mismo fin. Las listas de verificaciones o de procedimientos de trabajo, las recomendaciones, informes del desarrollo del proyecto, listas de problemas, dificultades e incidencias, así como recuperación de errores son instrumentos que pueden aplicarse para la mejora. Esta mejora puede derivar en nuevos procedimientos de trabajo para el cliente, con el fin de que optimice los resultados. Asimismo, esta evaluación puede ser de aplicación interna, para

gestionar mejor las posibles dificultades con proyectos similares o del mismo cliente.

## **2.4. Gestión y tipos de recursos lingüísticos multilingües**

La gestión de los recursos lingüísticos, desde su creación, almacenamiento, recuperación y manipulación representa una de las partes más importantes en la realización de cualquier proyecto de traducción.

El principio de reutilización ha marcado una tendencia en el desarrollo de tecnologías de procesamiento del lenguaje natural y de codificación (Arrarte, 1995).

En el marco del proyecto EAGLES (*Expert Advisory Group on Language Engineering Standards*) donde se elaboraron recomendaciones comunes para cinco áreas de la ingeniería lingüística —corpus textuales, diccionarios, formalismos gramaticales, evaluación de productos y lengua hablada— y cuyos resultados han marcado las tendencias actuales de la ingeniería lingüística en Europa, nace la Asociación Europea de Recursos Lingüísticos (ELRA). Esta asociación está dedicada a crear pautas de gestión de los recursos lingüísticos para fomentar su desarrollo, codificación común e intercambio, igual que lo hace la asociación LISA, tal y como hemos visto en el apartado 2.3.1. La documentación técnica multilingüe en su ámbito global al describir los objetivos de esta asociación para el entorno de la traducción y localización. La necesidad de reutilizar y compartir los recursos lingüísticos ha llevado por una parte a identificar los tipos de recursos lingüísticos relevantes para cada entorno específico y la búsqueda de un consenso para establecer estándares en este campo.

Según el catálogo realizado por ELDA (*Evaluations and Language Resources Distribution Agency*), el organismo de desarrollo y ejecución de las propuestas comerciales de la asociación ELRA, los recursos lingüísticos se distribuyen en cuatro categorías generales (ELDA, 2003):

- recursos orales: corpora orales, lexicones fonéticos
- recursos escritos: corpora escritos, lexicones monolingües y multilingües
- recursos terminológicos: bases terminológicas monolingües y multilingües
- recursos multimodales o multimedia: combinación de diferentes recursos en diferentes soportes digitales

Por tanto, se puede decir que un recurso lingüístico es cualquier documento digital que contiene conocimiento lingüístico o que es susceptible de utilizarse para crear conocimiento lingüístico (Moreno, Ares y Marcos 1999).

La aplicabilidad de éstos dependerá por una parte del sector productor de los recursos y, por otra, del sector consumidor de estos mismos recursos. En este sentido, según el estudio de Gonzalo (2004: 300-306) sobre las fuentes de recursos lingüísticos en línea con orientación al traductor, éstos se distribuyen de la siguiente forma:

- recursos lexicográficos: diccionarios, glosarios, vocabularios, léxicos, corpus lingüísticos, siglas y acrónimos, abreviaturas y símbolos
- recursos terminológicos: bancos de datos terminológicos, neologismos, tesauros, nomenclaturas oficiales
- recursos gramaticales: gramáticas, ortografías, servicios de consultas

Desde el punto de vista de la gestión de proyectos de traducción y localización, la selección de los recursos lingüísticos de referencia para el traductor está vinculada a los tipos de recursos susceptibles de ser seleccionados y definidos en los parámetros de un proyecto de traducción para su gestión automatizada con sistemas de traducción asistida.

Además de los recursos gestionados con herramientas de traducción asistida, es decir, las memorias de traducción y los diccionarios

terminológicos, existen muchos más recursos consultables, tal y como hemos visto en la tipología de Gonzalo. De todas maneras, en este estudio, tenemos sólo en cuenta aquéllos que, en un entorno de traducción asistida, como ya hemos mencionado, se utilizan como material de referencia específico de cada proyecto.

Una de las funciones del gestor consiste en la preparación de los recursos lingüísticos pertinentes para el traductor, tal y como hemos visto más arriba. Estos recursos lingüísticos se pueden considerar internos por estar almacenados en el entorno de trabajo donde se preparan. Existen recursos externos y se caracterizan por poderse consultar fuera del entorno de trabajo donde se preparan los proyectos de traducción, no están almacenados en la empresa y pueden ser consultados en Internet.

De todas las fuentes de búsqueda son las internas, especialmente las de tipo texto paralelo (las memorias de traducción) y las bases terminológicas (diccionarios terminológicos específicos) las que ofrecen muchas más ventajas con respecto a otras posibles fuentes de consulta. Y esto es, porque estos recursos lingüísticos almacenados permiten:

- ser explotados de forma mucho más rápida por estar en formato digital, relegando a un segundo plano el uso de las fuentes en soporte papel. La sustitución de los contenidos publicados en papel por la utilización de los contenidos digitales publicados en Internet es una tendencia ya constatada (Hutchenson, 1988: 11);
- facilitar su ubicación en un espacio finito, como el de una intranet, extranet, o PC, con respecto a la ubicación en el espacio infinito de Internet;
- prever su fiabilidad, ya que en un entorno limitado donde sistemáticamente se han almacenado los recursos lingüísticos tras su revisión, se presupone que se puede recuperar un material fiable a diferencia de los documentos y recursos publicados en Internet. En el

entorno de Internet conviven los recursos publicados de calidad con aquellos que no lo son, concepto que ha sido identificado como *Multi-Media Mediocrity* (Ciolek, 1997) y que se pretende paliar mediante una metodología de evaluación de los recursos digitales como así lo muestran los trabajos de Codina (2000b y 2000c ) o de Merlo (2003), entre otros;

- realizar la traducción en menos tiempo que el que se emplea sin este tipo de recursos; de esta manera, tanto la empresa de traducción y localización como los traductores pueden ofrecer precios competitivos, sin perder el margen de beneficio previsto por la realización de las tareas encargadas.

La utilización de un almacén propio de recursos lingüísticos ofrece en el entorno de ese depósito más o menos restringido soluciones a las cuestiones lingüísticas, terminológicas y traductológicas fiables que pueden plantear tanto el traductor como el gestor durante el desarrollo de los proyectos de traducción. Con el fin de agilizar el flujo de trabajo, por tanto, se reconoce la validez, utilidad y operatividad de los recursos lingüísticos ya almacenados en entornos controlables, finitos y locales. En el entorno de la documentación técnica multilingüe y profesional, este almacenamiento debe realizarse de forma sistemática mediante el uso de herramientas de traducción asistida basadas en el principio de la reutilización, como abordamos en el apartado 2.6. Herramientas para la gestión de recurso lingüísticos multilingües.

La selección de estos recursos por parte del gestor de proyectos —en el caso de un entorno de trabajo donde estén diferenciados los perfiles de los recursos humanos que gestionan la traducción— y su entrega a los traductores implicados en el proyecto en cuestión, tiene ante todo la finalidad de controlar la calidad desde un principio, ya que, de la entrega de un trabajo de calidad depende no sólo satisfacer las necesidades del cliente, sino fidelizarlo. De la misma forma, la disponibilidad de buenos recursos lingüísticos incide en la calidad del trabajo entregado. Y esto es así porque

facilita el proceso cognitivo del traductor encontrando respuestas a sus preguntas, pudiendo éste concentrarse en las cuestiones traductológicas o terminológicas no solucionadas con los recursos ofrecidos. La disponibilidad de recursos lingüísticos previos garantiza la rentabilidad del trabajo, porque de esta manera se pueden obtener mayores resultados en menos tiempo, por ello la relevancia de crear mecanismos adecuados para su selección.

### **2.4.1. Las memorias de traducción**

Las memorias de traducción son almacenes compuestos de textos originales en una lengua alineados con su traducción en otra(s). Esta definición de memorias de traducción coincide literalmente con una de las definiciones más aceptadas de corpus paralelo (Baker, 1995: 230). Por esto se puede decir que las memorias de traducción son corpus paralelos.

Los corpus paralelos se componen de textos originales o lengua original y el texto traducido en la lengua de llegada. Estos textos además están alineados por unidades de traducción o segmentos. Las unidades de traducción que se almacenan junto con sus equivalentes se definen de forma variable siendo la segmentación tras un signo de puntuación que marca el final de la frase (., ?, !, :, ...) o un salto de párrafo las más frecuentes ofrecidas por defecto en el entorno de los sistemas de traducción asistida.

En la literatura especializada existe otra denominación que encaja con el concepto de texto paralelo: *bitexto*. El término inglés *bitext*, acuñado por Harris (1988a: 41y b) —en la literatura escrita en español encontramos *bitexto* como cognado equivalente— se define como:

Text stored in such a way that each retrievable segment consists of a segment in one language linked to a segment in the other language which has the same meaning.

Algunos autores, sin embargo, establecen diferencias entre la memoria de traducción y el bitexto, arguyendo que en el entorno de las memorias se

pierde el contexto de las unidades originales (Arrouart y Bédard, 2001). La pérdida contextual está vinculada a la forma de almacenar las memorias, que si bien es cierto, en la mayoría de los casos, se almacenan en bases de datos, existe también otra técnica de almacenamiento: la de gestión de archivos y directorios anotados mediante metadatos utilizada por algunos sistemas como STAR Transit o por el sistema Ecco (véase la forma de almacenamiento en el catálogo de herramientas de traducción asistida de Sánchez, 2001). En estos sistemas, la recuperación no implica pérdida contextual porque las propuestas de traducción no se almacenan descontextualizadas, sino que permanecen en sus respectivos archivos de trabajo, tema que abordaremos más adelante.

Mediante almacenamiento sistemático de unidades de traducción dentro del entorno de la traducción asistida por ordenador, se va alimentando uno de los componentes más importantes de los que disponen las herramientas de traducción asistida, el almacén de las memorias de traducción.

El almacén de traducciones, es decir, el corpus paralelo, va creciendo en función del volumen y frecuencia de alimentación de las memorias de traducción y éstas se crean durante la validación de las unidades de traducción durante el proceso de traducción en el entorno de la traducción asistida.

Otra de las técnicas para la generación de memorias de traducción es mediante la alineación de textos traducidos y por tanto equivalentes. La alineación se lleva a cabo con herramientas específicas para ello, conocidas como alineadores. Éstos son programas que a partir de material textual en dos lenguas diferentes enlazan segmentos más pequeños de una lengua con la otra, con el fin de crear o alimentar las memorias de traducción. Estas herramientas normalmente se hallan integradas en los sistemas de traducción asistida y a pesar de las diferencias específicas de cada producto en particular, los pasos que sigue el alineador son los siguientes:



- ❑ separación de los elementos textuales de los no textuales de los textos que se deben emparejar;
- ❑ segmentación del texto en sus distintas frases o segmentos definidos como unidades de traducción;
- ❑ emparejamiento o alineación propiamente dicho de las frases o los segmentos de origen y de destino;
- ❑ creación de la memoria de traducción a partir de las frases correlacionadas.

La utilización de una herramienta de traducción para generar memorias de traducción mediante la traducción manual o automatizada o mediante la alineación de textos equivalentes es decisivo para la reutilización de recursos lingüísticos.

Cuando las memorias de traducción se crean en el entorno de un sistema de traducción asistida, el sistema las anota de forma automatizada, conformando un corpus paralelo anotado, ahora bien las informaciones son de tipo extralingüístico como: lengua, fecha de creación y de modificación, longitud del texto (en caracteres, palabras, oraciones, párrafos), autor, traductor, edición y anotaciones estructurales de códigos de correspondencia que asocian las unidades de traducción con sus equivalencias alineadas. No existen informaciones adicionales de tipo lingüístico.

La principal función de las memorias de traducción es extraer sugerencias traductológicas totales o parciales de una frase y concordancias para términos. El contenido de estos recursos lingüísticos paralelos es fundamental, pero también es crucial el motor de búsqueda que permite explorar una gran cantidad de texto e identificar patrones lingüísticos y terminológicos comunes. En el apartado 2.6. Herramientas para la gestión de recursos lingüísticos multilingües abordamos con más detalle estas cuestiones técnicas relacionadas con los mecanismos de funcionamiento de los sistemas de traducción asistida. Por tanto, si el contenido de las

memorias y si su sistema de indización ofrece buenos resultados, estas memorias se convierten en el mejor instrumento de trabajo del mediador lingüístico. En esta línea Abaitua (2002) afirma que los corpus paralelos son una alternativa a los diccionarios como fuente única de información y, como ya hemos dicho, por definición las memorias de traducción constituyen textos paralelos.

Asimismo, esos corpus generados por los sistemas de traducción asistida también tienen relevancia en el área de la traducción automática basada en ejemplos obtenidos de corpus paralelos y no la traducción automática regida por reglas de análisis lingüístico morfológico y sintáctico (Hutchins y Somers, 1995: 127-149 y 184-185).

#### **2.4.2. Las bases terminológicas: diccionarios específicos y especializados**

Las bases terminológicas contienen recursos lingüísticos conocidos también como diccionarios terminológicos. Éstos, a diferencia de los diccionarios generales, están orientados mayoritariamente al concepto y no a la palabra. Esto es, para el registro *ratón* en un diccionario general se incorporarían los diferentes significados que se atribuye a esa palabra (enfoque semasiológico), mientras que en un diccionario terminológico, por ejemplo, acerca de experimentos de laboratorio con animales, *ratón* aparecería en dos registros diferentes: uno para identificar *ratón* como *cobaya* y otro para identificar *ratón* como *periférico del ordenador*.

Estos recursos están representados de forma digital y se gestionan mediante herramientas concretas, los gestores terminológicos, que pueden estar integrados o no en sistemas de traducción asistida. Mediante estas herramientas se facilitan algunos de los procesos implicados en la gestión, como la elaboración, ordenación, numeración, búsqueda, actualización e incluso edición de las fichas terminológicas. En el apartado 2.6. Herramientas para la gestión de recursos lingüísticos multilingües recogemos

específicamente cuestiones técnicas involucradas en la gestión de las bases terminológicas.

Las bases terminológicas son recursos lingüísticos absolutamente necesarias para la traducción especializada en general. Las razones abordadas por Cabré (1999: 191-193 y 2004: 104) están vinculadas a las características del texto técnico o especializado y su repercusiones en la traducción como recordamos a continuación:

- los elementos terminológicos constituyen el aparato conceptual imprescindible sobre el que se apoya el texto especializado;
- la calidad dependerá del uso sistemático de términos, adecuados al nivel de especialización del texto;
- los términos representan el principal conocimiento que ofrece el producto final traductológico.

Asimismo, la terminología tiene una gran repercusión para la organización de la información y del conocimiento. Los mapas de conocimiento, gráficos, bases de conocimiento se componen de términos. Las técnicas de publicación también precisan de bases terminológicas bien gestionadas. Por una parte, la técnica del hipertexto obliga aún más no sólo al uso, sino a la coherencia terminológica (Moreno-Torres, 1996: 99). Y, en segundo lugar, el principio de reutilización obliga a armonizar el uso coherente de los recursos terminológicos. Las bases terminológicas pueden reutilizarse en cualquiera de las técnicas documentales susceptibles a ser desarrolladas y diseñadas para explotar mejor la documentación técnica como, por ejemplo, mediante la creación de índices de diferente índole aplicables en los motores de búsqueda o en las bases de conocimiento.

Además del carácter simbiótico que se establece entre la gestión de la terminología y el proceso de traducción, en otros entornos también es imprescindible gestionar (producir y organizar) de forma automatizada los recursos terminológicos.

Las bases terminológicas pueden servir de instrumento de control empleado en la elaboración de una imagen corporativa de la empresa. En el mercado global actual es importante crear una imagen de una empresa con actividad internacional que dé idea de uniformidad. La documentación técnica de un producto debe transmitir consecuentemente la misma imagen y esto no solo se consigue con recursos extralingüísticos como logos, tipografía, colores de fuentes, maquetación, sino con el uso consecuente de una terminología específica común para los departamentos de marketing, de desarrollo y de finanzas.

Así, los gestores de proyectos, los creadores de contenido, los editores, los redactores técnicos, los arquitectos de la información, los expertos en marketing necesitan gestionar recursos terminológicos que sirvan de soporte para desarrollar calidad y eficiencia en sus tareas, especialmente si están trabajando en entornos de proyección internacional y multilingüe. Ninguna entidad con aspiración internacional y multilingüe puede optimizar la gestión del conocimiento dentro de su entidad sin un sistema de organización de la información de bases terminológicas. El fundamento para la gestión del conocimiento es la comunicación. Si la comunicación se establece de forma digital se deben utilizar bases terminológicas para poder gestionar la organización hipertextual (sitio web, intranet, extranet) y para organizar y gestionar los documentos de una empresa.

Los recursos terminológicos bien gestionados participan en el desarrollo de los fundamentos y técnicas de la planificación, construcción, gestión, uso y evaluación de sistemas de descripción, catalogación, ordenación, clasificación, almacenamiento, comunicación y recuperación de los documentos creados por las personas.

Por tanto, más allá del soporte imprescindible para el ámbito de la traducción, las bases terminológicas bien gestionadas posibilitan la gestión eficiente de la información, contribuyendo a gestionar el conocimiento de una entidad. Y esto es así, porque se transfiere mejor el saber (know-how) dentro de

cualquier empresa si se unifican los conceptos utilizados en una corporación, y porque mejora la comunicación entre diferentes centros de la misma entidad, especialmente si pertenecen a diferentes comunidades lingüísticas (Cámara y Cuchí, 2003a).

También será imprescindible el uso de bases terminológicas si además los productos que se desarrollan y documentan se fundamentan en estándares internacionales con terminología específica.

Y por último, desde el punto de vista de la ingeniería lingüística, las bases terminológicas conforman además una imprescindible base de conocimiento lingüístico para el desarrollo de verificadores ortográficos, sistemas de traducción automática, sistemas de traducción o sistemas expertos, sistemas de redacción asistida cuyo rendimiento y eficacia, sin las bases de terminología, se reduciría.

Las bases terminológicas se componen de fichas con informaciones específicas para cada concepto, es decir, una microestructura. Estas informaciones específicas de cada concepto irán en función de las necesidades del entorno de aplicación. En este sentido, desde el punto de vista del traductor y más allá de las recomendaciones metodológicas para los trabajos terminográficos definidas por la obra de referencia de Cabré (1993: 281-285), el manual de trabajo del Termcat (1990: 69-83) y las categorías abordadas por la norma ISO 12616 acerca de la terminografía orientada a la traducción, las categorías de la ficha terminológica que gestiona el traductor se reducen a las mencionadas a continuación:

- término
- categoría gramatical
- definición
- remisiones: abreviaturas, sinónimos y referencias cruzadas
- equivalencias en otros idiomas del término

- contextos y ejemplos
- gráfico

Las categorías apuntadas deben seguir unas pautas comunes de trabajo para poder controlar mejor la calidad de los términos registrados.

En este sentido, el término registrado se debe introducir en orden secuencial. Es decir, el término *fuentes en línea* no se debe introducir de esta manera: *en línea, fuentes*. El término debe aparecer en minúsculas, a no ser que se trate de un nombre propio. En alemán, como la regla de la gramática alemana estipula, las iniciales de todo tipo de sustantivos se deberán escribir en mayúscula. Las siglas deben ponerse también en mayúsculas. Se debe introducir la forma canónica de los términos: sustantivos sin flexión de número o de género, verbos en infinitivo, adjetivos con su forma simple.

La categoría gramatical de un término puede o no ser obligatoria, depende de las necesidades concretas de cada proyecto. Lo importante es que los indicativos de la categoría siempre sean los mismos, por eso se debe definir una lista de valores comunes para entornos de trabajo concretos. Por ejemplo:

<b>Sigla</b>	<b>Información gramatical</b>
m	Nombre masculino
f	Nombre femenino
m /f	Nombre masculino y femenino
m pl	Nombre masculino plural
f pl	Nombre femenino plural
v	Verbo
sin	Sintagmas y locuciones

**Fig. 20: Lista de valores con información gramatical**

La información lingüística de tipo morfológico y sintáctico en principio no se crea ni se utiliza (Arevalillo, 2004: 468). Ésta no es relevante en este ámbito de trabajo del traductor por dos razones. La primera razón está relacionada con la competencia lingüística del traductor en su lengua de destino que, en general, coincide con la lengua materna. Esto significa que el conocimiento

acerca de la flexión morfológica del término no es necesaria, porque al ser el término de naturaleza lingüística, se puede derivar su flexión. Sin embargo, los casos puntuales que pueden salirse de los patrones reconocibles deben registrarse o documentarse. Por ejemplo, el término *macro* en el sector informático es femenino. Asimismo la información sintáctica atribuible al término en cuestión no es necesaria por lo previamente argumentado. La segunda razón para no atender a la inclusión de las especificaciones lingüísticas de tipo morfosintáctico está relacionada con el tipo de herramientas que gestionan estos recursos lingüísticos, nos referimos a las herramientas de traducción asistida. Éstas, al no estar dotadas de conocimiento lingüístico, no explotan ni mejor ni peor los recursos lingüísticos, simplemente trabajan con cadenas de caracteres (términos sencillos o compuestos) cuyos rasgos lingüísticos asociados no son explotados.

La definición de un término se puede añadir, especialmente si se trata de algún término conflictivo, inusual o ambiguo. Siempre que sea posible la definición estará encabezada por un término más genérico para contextualizar con rapidez el significado del término que definimos. Por ej. *ESP: Termómetro: Instrumento que sirve para medir la temperatura. Se debe evitar en la definición informaciones metalingüísticas como Nombre masculino que designa... Nombre dado a... Calificativo aplicado a...* Para definir locuciones se debe intentar introducir en la definición un segmento de frase que sea sustituible por la locución que definimos. Por ejemplo, *a bordo: Situado dentro de una embarcación / a pulmón libre: Manera de bucear sin escafandra.*

En una situación ideal, la definición del término no es tarea del traductor, sino del redactor autor del documento original. La gestión terminológica desde el origen del documento es una de las estrategias de internacionalización de contenidos que deben aplicarse en el entorno de la producción técnica multilingüe.

Según las recomendaciones en las que se inspira nuestra descripción acerca de la microestructura para el registro de un concepto, las informaciones conceptuales específicas del término más allá de la definición son la atribución del campo de especialidad. Sin embargo, el hecho de trabajar con bases de datos ya especializadas hace innecesario la asignación adicional de esta categoría. Mucho más relevante es la inclusión de remisiones conceptuales que enmarquen el término en función de relaciones de asociación a modo de tesaurus. Esto posibilita la contextualización del término junto con su desambiguación conceptual.

Las remisiones pueden ser: formas abreviadas, sinónimos y referencias cruzadas. Todas son indicadores de relaciones semánticas entre los términos. Estas remisiones pueden ser de equivalencia (abreviaturas, siglas y sinónimos), de oposición (antónimos) o de meronimia (hipónimos, hiperónimos, etc.). Las remisiones que se deben registrar durante el proceso de traducción son las de equivalencia. Otras relaciones semánticas entre los términos se agrupan en un solo campo *Referencia cruzada* y su inclusión es optativa.

Las formas abreviadas del término como *acrónimos*, *siglas* y *abreviaturas* se deben registrar en un solo campo y deben estar vinculadas con su término expandido. Si se trata de una abreviatura lexicalizada o en proceso de lexicalización, como por ejemplo, XML, SGML o ADSL, el término puede ocupar el campo de término, en vez del campo de abreviatura. El término expandido se introduce entonces en el campo de sinónimos para que, por defecto, aparezca indizado en el diccionario como entrada adicional.

Los sinónimos, en cuanto a formas conceptuales equivalentes de un término, deben registrarse en el campo de sinónimos. En casos de introducción de sinónimos se debe introducir un ejemplo contextual para identificar el uso que se hace de la equivalencia semántica aportada. Aunque en la traducción técnica se tiende a relacionar un solo concepto con una unidad de representación, también es cierto que los redactores técnicos no siempre se



atienden a esta pauta, lo que supone un trabajo adicional para el traductor que debe reconocer si el término en cuestión es una variante denominativa o conceptual.

Las especificaciones lingüísticas de los términos como las siglas o las variantes denominativas son especialmente útiles para el traductor porque ofrecen las posibles formas alternativas o de variación denominativa del tipo geolinguístico, formas abreviadas, denominaciones específicas de un cliente para un concepto ya existente, etc.

Los antónimos, hipónimos, hiperónimos, se registran en campos específicos o en un solo campo, el de referencias cruzadas. Su inclusión es recomendable siempre y cuando puedan esclarecer el significado del término al que se refiere. Éstas deben incluirse en el mismo idioma que el término al que se refiere.

Las equivalencias en otros idiomas del término se deben marcar mediante siglas internacionales, que se predefinen en el sistema de gestión terminológico usado. Estas siglas están normalizadas como estándares y se recogen en la norma ISO-639, definidos para el uso en entornos de biblioteconomía, gestión de la información, lingüística, terminología y lexicografía. Esta norma propone una lista de códigos de dos y tres letras para la representación de las lenguas como se muestra en la siguiente tabla:

<b>Lengua</b>	<b>Código 2 letras</b>	<b>Código 3 letras</b>
Español	ES	ESL / SPA
Catalán	CA	CAT
Vasco	EU	BAQ / EUS
Gallego	GL	GAL
Inglés	EN	ENG
Francés	FR	FRA / FRE
Alemán	DE	DEU / GER

**Fig. 21: Códigos ISO para la representación de idiomas**

La inclusión de las equivalencias conceptuales de los términos es obligatoria y definitiva para el proceso de la traducción y para la reutilización de las bases terminológicas para traducciones posteriores.

Los contextos de uso y los ejemplos pueden registrarse en campos separados o aglutinarse en un solo campo para contextos y ejemplos por lengua. El ejemplo o contexto que se introduzca debe intentar ser ilustrativo para su comprensión. La inclusión en el campo de contexto en la sección de la lengua de partida de una unidad de traducción con su equivalencia traductológica en el campo de contexto, pero en la sección de la lengua de llegada son especificaciones pragmáticas que ofrecen mucha información acerca del uso del término.

La información sobre las fuentes, en general, no la incluye el traductor durante el proceso de la traducción. Una de las formas para la obtención de esa información en el entorno de la traducción asistida es ver resaltados los términos en el texto. Éstos están resaltados, si están registrados en los diccionarios definidos para ese proyecto. La fuente de cada término es pues el diccionario correspondiente.

La información administrativa de las bases terminológicas se suelen incluir de forma automática mediante las herramientas de gestión terminológica con las que se gestionan las bases de datos. Estas informaciones pueden registrar de forma automática el nombre del creador del registro y las fechas de creación o modificación de la ficha terminológica.

La macroestructura de las bases terminológicas dependerá del modelo de datos de almacenamiento. Como norma, al ser bases de datos se organizan por estructuras que contienen diccionarios y que corresponden con la columna dentro de la base de datos. De esta manera se posibilita la organización adicional de los datos terminológicos dentro de la base de datos. Las organizaciones predecibles en el ámbito de la traducción técnica

—y probablemente dentro de la traducción especializada, en general— son por área temática, productos o proyectos.

En el caso de las memorias de traducción, como hemos visto, existen dos maneras de generarlas. Una mediante la validación de las unidades de traducción ya traducidas, y otra, mediante la alineación de textos equivalentes. El caso de las bases terminológicas es similar, en el sentido de que, por una parte, son de creación propia y, por otra, son el resultado de un cambio de formato para reutilizarlo en el entorno de trabajo deseado.

La situación ideal es contar con bases terminológicas ya elaboradas y validadas por el cliente que solicita la traducción. Puede ser que el cliente proporcione la terminología, pero sin las equivalencias; otras veces puede aportar diccionarios específicos de referencia pero sin haber controlado la calidad. Este material puede provenir de Internet, de bases de datos de otros clientes del sector o de un esbozo provisional de equivalencias terminológicas durante el proceso de redacción de la documentación del producto. Este material que proporciona el cliente para su reutilización, ya bien sea validado o no, debe tener el formato al que dé soporte al entorno de trabajo específico.

Otra de las formas de generar bases de datos es la elaboración durante la traducción. Esto es, se pueden ir alimentando los registros que constituyen las bases terminológicas, mientras se está traduciendo en un entorno de traducción asistida o no. Incluir las nuevas ocurrencias terminológicas como hábito instaurado de trabajo, repercute positivamente en el flujo de trabajo y en la calidad final. En el caso de trabajar con un equipo de traductores para un proyecto común, la práctica de introducir los términos puede contribuir significativamente a unificar la terminología empleada. Si además estas mismas bases de datos se pueden actualizar en red o de forma remota para entornos virtuales, las nuevas propuestas terminológicas pueden ser vistas por todos para aprovecharse de ese conocimiento o para proponer una equivalencia más acertada.

## **2.5. Formatos de codificación e intercambio**

Los formatos destinados a codificar, etiquetar, catalogar y describir información con el fin de almacenar, conservar, manipular e intercambiar los recursos de información en general son el paradigma de la sociedad de la información imperantes en la actualidad. Estos formatos de intercambio se conocen también bajo el término metadatos. Acerca de la nomenclatura de este concepto se recomienda la lectura de *Metadatos: concepto, fundamentos, aplicación y tipología* de Méndez (2002, 29-71).

La eclosión de Internet con sus millones de máquinas interconectadas con arquitecturas de hardware diferentes funcionan con distintos sistemas operativos y, sin embargo, logran comunicarse e interoperar gracias a una serie de protocolos estandarizados sencillos, abiertos y públicamente disponibles como:

- TCP/IP (*Transmisión Control Protocol / Internet Protocol*): Protocolo en el que está basado la red de Internet y la comunicación;
- HTTP (*HyperText Transport Protocol*): Protocolo de comunicación empleado por los servidores de WWW que permite de forma descentralizada la comunicación cliente-servidor;
- FTP (*File Transfer Protocol*): Protocolo usado en Internet para transferir archivos desde una ubicación remota a una máquina local, o viceversa;
- SMTP (*Simple Mail Transfer Protocol*): Protocolo para el envío de mensajería a través de Internet ;
- POP3 (*Post Office Protocol*): Protocolo de recepción de mensajes de correo electrónico a través de Internet ;
- ODBC (*Open DataBase Connectivity*): Protocolo para el intercambio de bases de datos de diferentes plataformas;

- HTML (*Hyper Text Markup Language*): Lenguaje de metadatos basado en SGML usado para dar forma y crear documentos en hipertexto para el entorno web;
- SGML (*Standard Generalized Markup Language*): Metalenguaje internacional estándar surgido por las necesidades de la industria editorial a fin de facilitar intercambio de datos ante todo tipográficos — originalmente describía una anotación u otros símbolos para indicar a un maquetador de imprenta como diseñar una determinada página: tamaño de letra, negritas, centrados, etc. (Berrocal, Figuerola y Zazo, 2000: 2)—. Sin embargo, este lenguaje de etiquetas ha traspasado y evolucionado fuera de ese ámbito editorial para ampliar el intercambio de información mediante la definición de lenguajes adicionales como el caso de HTML y más tarde XML. SGML provee un marco para describir la estructura lógica de los documentos identificando elementos estructurales y de contenido y éste codifica las relaciones jerárquicas entre sus elementos. Desde 1986 se utiliza este lenguaje de codificación para el desarrollo de lenguajes de etiquetas (Elosua y Plágaro, 2000; Budin y Wright, 2001: 863-866).

En los siguientes subapartados dedicamos especial atención a la codificación Unicode y al formato XML, ambos desarrollos a favor de la estandarización para facilitar el intercambio de datos. Tanto la codificación mediante códigos comunes de caracteres, como es el caso de Unicode, como el tratamiento de la documentación en soporte electrónico mediante metadatos (XML) son inherentes a la *internetización* mundial. Esta tendencia universalista de *todos accedemos a todo*, es por tanto, posible gracias a la desfronterización de la información mediante la superación de las barreras lingüísticas, que se posibilita por la publicación en diferentes idiomas y por la sustitución de formatos propietarios por estándares universales. Esto supone poder garantizar la perpetuidad de los datos, por almacenarse y transferirse en un formato cuyo acceso no depende de una sola persona, organización o aplicación de software. La tendencia de estandarización se muestra también

en el desarrollo de metadatos para entornos o dominios específicos con necesidades concretas (Kashyap y Shet, 2000: 20; Velluci, 1998: 204), como así lo demuestra, por ejemplo, el proyecto del consorcio internacional para la codificación de textos TEI (*Text Encoding Initiative*) cuyo objetivo es el fomento del uso de etiquetas rigurosas y productivas para cualquier clase de texto para su codificación e intercambio (Burnard, 2003).

En el entorno de la gestión de los recursos lingüísticos multilingües también se busca la manera de codificarlos, almacenarlos y transferirlos en un formato estándar (TMX, Martif, OLIF y XTL → TBX) que no ostente limitaciones de uso o patentes. Esto hace posible cambiar de proveedor de software si el anterior desaparece o intenta imponer sus condiciones al acceso de sus recursos almacenados en su formato propietario o patentado.

### **2.5.1. Unicode**

Unicode es un acrónimo resultante de la contracción de *universal* y *codification*. Es un sistema de codificación de caracteres que, por ejemplo, actualmente posibilita que en el ciberespacio de Internet circulen millones de documentos en miles de lenguas con sistemas de escritura diferentes, por ejemplo, el alfabeto cirílico, el latino o el árabe o el hebreo, con codificaciones distintas. Esta diversidad supone un problema de gestión, por ello la solución ideada para resolver esta multitud de conjuntos de caracteres incompatibles entre sí fue el desarrollo y aplicación del estándar del consorcio Unicode. Este consorcio está integrado por una amplia gama de corporaciones y organizaciones de la industria de la computación y del procesamiento de la información cuyo objetivo es desarrollar, difundir y promover el uso de la norma Unicode<sup>1</sup>.

Unicode define un único conjunto de caracteres donde caben casi todos los sistemas de escritura actualmente en uso, al igual que muchos símbolos

---

<sup>1</sup> Para más información véase <http://www.unicode.com>

matemáticos y científicos y algunas escrituras históricas. Cada carácter está representado por 16 bits. Mediante este formato de codificación de caracteres un documento en texto sencillo puede contener texto en varias lenguas con sistemas de escritura diferentes (Budin y Wright, 2001: 861-862). Unicode simplifica considerablemente el trabajo de la documentación en múltiples idiomas. La relevancia de Unicode para el entorno de la documentación técnica multilingüe, tanto durante la creación de documentación como en la traducción es obvia, ya que por definición, se trabaja con múltiples lenguas a la vez.

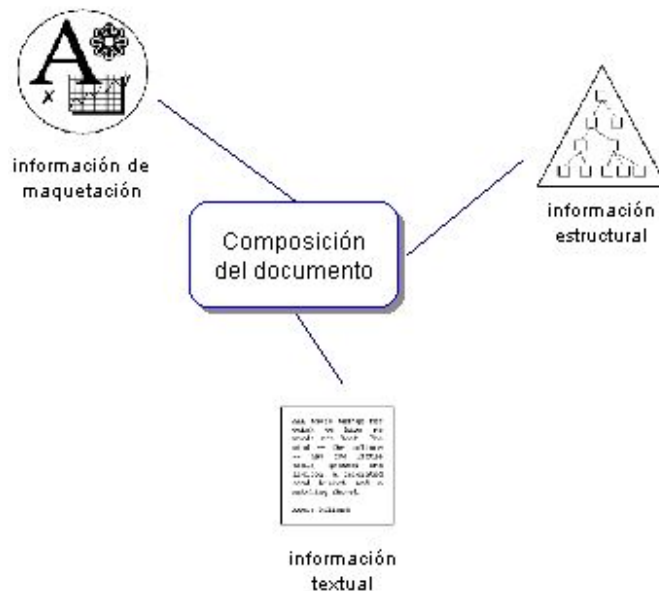
### 2.5.2. XML

El lenguaje extensible de etiquetas (XML) es una versión simplificada y modernizada del SGML (*Standard Generalized Markup Language*). Este subconjunto de SGML se describe en la recomendación del W3C sobre XML<sup>1</sup>. El formato XML reduce la complejidad de SGML, sin perder capacidad de descripción y definición. Sirve principalmente como lenguaje de descripción de datos. Su característica principal es la estructuración de documentos en forma de árbol, con un único elemento raíz y con los elementos descendientes propiamente anidados. Su objetivo es permitir que las aplicaciones puedan intercambiar datos descritos en un formato bien definido, fácil de procesar, legible tanto para las personas como para las máquinas. Este lenguaje es cada vez más utilizado para una multitud de fines y existen muchas herramientas que permiten visualizar, procesar y transformar datos descritos en XML. La sintaxis y el significado de los elementos y atributos de un tipo de documento XML específico se suele describir en un documento DTD (*Document Type Definition*) u otro formato equivalente. Para el almacenamiento y tratamiento de los recursos lingüísticos se utilizan DTD específicos. Estos DTD determinan los elementos permitidos en un tipo de documento y las relaciones entre los elementos.

---

<sup>1</sup> Para más información véase <http://www.w3c.com>

Estos DTD se rigen por las reglas descritas en XML, de fácil manejo. Los DTD componen un lenguaje de etiquetas determinado desde el inicio. Las etiquetas no tienen un significado predeterminado, sino que se asignan. Las etiquetas se entienden muy bien porque consisten en texto común que transmiten información de contenido y estructura, la información de maquetación de los documentos no está contenida en los DTD, por eso es necesario asociar hojas de estilo tipo XSL (*eXtensible Stylesheet Language*) o CSS (*Cascading Style Sheet*) para poder ofrecer toda la información de la que se compone un documento —información de maquetación, estructural y textual—, si es que se precisa, como mostramos esquemáticamente a continuación:



**Fig. 22: Tipos de información que componen el documento**

Los DTD más importantes en el entorno de la documentación técnica son aquellos lenguajes específicos que codifican los documentos que contienen recursos lingüísticos como las memorias de traducción (TMX, *Translation Memory Exchange*) y las bases terminológicas (Martif, *Machine-Readable Terminology Interchange Format*) para facilitar su intercambio.



### 2.5.3. TMX

TMX es el estándar para el intercambio de memorias de traducción. Es un DTD creado por el comité OSCAR (*Open Standards for Container/Content Allowing Re-use*), grupo de interés de la asociación LISA. El objetivo de este comité es facilitar el intercambio de recursos lingüísticos para entornos multilingües entre diferentes aplicaciones sin pérdida de información vital. En este sentido, se ocupan del diseño e implementación de formatos de intercambio de memorias de traducción.

Mediante la aplicación del formato TMX es más viable la colaboración en proyectos de traducción de personas o empresas que usan herramientas diferentes seleccionadas en función de sus necesidades y preferencias. El formato TMX también hace más fácil la migración de un sistema de traducción asistida a otro, lo que favorece la competitividad entre las tecnologías ofertadas y el desarrollo constante de las mismas para marcar diferencias con respecto a sus competidores. Como otros estándares abiertos, este formato se desarrolla con vistas a reducir los problemas de compatibilidad, impulsar la reutilización de los recursos lingüísticos, simplificar el intercambio de datos y estimular, de esta manera, la innovación tecnológica (Gómez, 2001).

Según la guía del TMX publicada por LISA (2004) las especificaciones de implementación del formato TMX presenta dos niveles. En el nivel uno sólo se da soporte al texto sencillo y en el nivel dos se da también soporte a códigos de formato. Este nivel es el adecuado para la documentación ya maquetada. Según la especificación, una aplicación con soporte de nivel dos debería poder recrear la versión traducida del documento original disponiendo únicamente del archivo TMX.

Además de facilitar el intercambio de datos entre herramientas de traducción asistida, un archivo TMX puede tener otros usos, ya que se trata de una representación de un documento en múltiples versiones lingüísticas

marcadas con un lenguaje relativamente fácil de procesar como es el XML. Esto significa que se puede acceder a la valiosa información contenida en un proyecto de traducción desde aplicaciones o extraerla y procesarla desde la mayoría de lenguajes de programación, como se muestra a continuación:

```
<tu>
<prop type="FileFormatGUID">{2E8E31BC-1690-4BCB-B2F0-5D2C6BC52373}</prop>
<prop type="Filename">0060mxs</prop>

<tuv xml:lang="en-gb" changedate="20040922T082407Z" changeid="Transit">
<seg><ph x="1">|Index: "Side doors", "Automatic locking and unlocking functions"##</ph></seg>
</tuv>

<tuv xml:lang="es-mx" changedate="20040922T082655Z" changeid="Christiane Gut">
<seg><ph x="2">|Index: "Puertas laterales", "Funciones de bloqueo y desbloqueo automático de las puertas"##</ph></seg>
</tuv>
</tu>

<tu>
<prop type="FileFormatGUID">{2E8E31BC-1690-4BCB-B2F0-5D2C6BC52373}</prop>
<prop type="Filename">0060mxs</prop>

<tuv xml:lang="en-gb" changedate="20040922T082407Z" changeid="Transit">
<seg><ph x="1">|Index: "Side door locks"##</ph></seg>
</tuv>

<tuv xml:lang="es-mx" changedate="20040922T082655Z" changeid="Christiane Gut">
<seg><ph x="2">|Index: "Cerraduras de las puertas laterales"##</ph></seg>
</tuv>
</tu>
```

**Fig. 23: Ejemplo de codificación de memorias en TMX**

Si tenemos en cuenta lo que ya hemos apuntado acerca de las memorias de traducción con la definición de corpus paralelo, cabe mencionar que existe un lenguaje de estándar de codificación para corpus, el CES (*Corpus Encoding Standard*), desarrollado siguiendo las directrices de codificación del TEI (EAGLES, 2000) y susceptible de ser utilizado como formato de codificación de las memorias de traducción. El CES es un DTD de la familia SGML que contiene las categorías y especificaciones para la codificación de anotaciones lingüísticas en los corpus. El formato de intercambio de corpus para XML es el XCES (*Corpus Encoding Standard for XML*) (EAGLES, 2002). La industria GILT no utiliza estos formatos en ninguna de las herramientas desarrolladas; tampoco proporciona soporte ni para el formato CES ni para el XCES mediante filtros o convertidores específicos. Las anotaciones de tipo lingüístico que aborda el formato CES no se contemplan en el entorno de las

herramientas de traducción asistida cuya estrategia, tal y como ya hemos visto, no se basa en el conocimiento lingüístico del que no están dotadas.

#### **2.5.4. Martif**

Martif es el formato estándar ISO 12200 para el intercambio de bases terminológicas tanto monolingües como multilingües. Es un DTD estándar basado en XML, desarrollado con ayuda del TEI y de la asociación LISA. Este formato permite la importación, exportación e intercambio de las bases terminológicas orientadas a conceptos, sin tener que especificar o asociar los campos o entidades de cada base de datos. Las categorías definidas en este lenguaje de intercambio se basan en la norma ISO 12620. Esta norma describe la normativa aplicable a las categorías de datos usadas en las bases terminológicas con la recomendación de usar todas o una parte en función de los entornos de aplicación.

Desde el punto de vista de la traducción, es común el envío de datos terminológicos. En esta situación un formato de intercambio común contribuye a solventar los problemas de incompatibilidad que pueden generarse si los diferentes usuarios no usan una plataforma común. Por este motivo, es conveniente usar archivos XML con datos descritos en el estándar Martif. Este formato sólo incluye los datos necesarios en vez de incluir toda la base de datos, lo que hace *más ligeros* los paquetes de datos. Por ejemplo, en un proyecto multilingüe el traductor al que se le envían los datos recibe sólo los términos de las dos lenguas con las que está trabajando o en otra adicional si se requiere como referencia para introducir las equivalencias terminológicas en la lengua de llegada que faltan por cumplimentar en la base de datos terminológica que se envía. Utilizando el mismo formato de intercambio, tras haber terminado la traducción y, posiblemente, editado la terminología, se envían al gestor del proyecto sólo los registros modificados, que en el destino se integrarán en el diccionario original.

Mediante la estructuración de los datos con ayuda de este formato, hace que las posibles conversiones de los datos, por ejemplo, para su publicación en Internet con el formato HTML y, obviamente a XML sea posible de una manera sencilla. Así, a los archivos Martif se les pueden aplicar hojas de estilo para convertirlos en otros formatos (HTML, PDF). Los datos se pueden filtrar y transformar de muchas maneras usando el lenguaje especializado de transformación para XML, conocido como XSLT o cualquier otro lenguaje con soporte a XML.

A continuación mostramos un ejemplo de información terminológica codificada con metadatos inscritos en Martif:

```

<ntig><termGrp>
  <term>steering gear box</term>
  <termNote type='termType'>synonym</termNote>
  <termNote type='TS_CreateId'>SCM</termNote>
  <date type='origination'>19980703T000000Z</date>
  <termNote type='TS_UpdateId'>WIT</termNote>
  <date type='modification'>20010217T010000Z</date>
</termGrp></ntig>

</langSet>

<langSet lang='fra-fr'>
<ntig><termGrp>
  <term>engrenage de direction</term>
  <termNote type='termType'>full form</termNote>
  <termNote type='TS_CreateId'>SCM</termNote>
  <date type='origination'>19980703T000000Z</date>
  <termNote type='TS_UpdateId'>SCM</termNote>
  <date type='modification'>19980706T010000Z</date>
  <termNote type='TS_Gender'>f</termNote>
</termGrp></ntig>

</langSet>

<langSet lang='spa-es'>
<ntig><termGrp>
  <term>engranaje de la dirección</term>
  <termNote type='termType'>full form</termNote>
  <termNote type='TS_CreateId'>AB1</termNote>
  <date type='origination'>20040304T000000Z</date>
  <termNote type='TS_UpdateId'>AB1</termNote>
  <date type='modification'>20040304T010000Z</date>
  <termNote type='TS_Gender'>m</termNote>
</termGrp></ntig>

```

**Fig. 24: Ejemplo de codificación de terminología en Martif**

### **2.5.5. TBX**

El formato TBX (*TermBase eXchange*), también como Martif, es un formato de intercambio de bases terminológicas pero con operatividad ampliada. Este formato representa un conjunto de metadatos inscritos en XML que describe el marco de trabajo para la gestión de metadatos en el entorno de la gestión de bases terminológicas.

Incorpora el formato ISO 12200 (Martif), que ya hemos visto, y el formato TMF (*Terminological Markup Framework*) para la descripción del marco para la gestión de metadatos en el ámbito terminológico.

TBX es el formato al que previsiblemente se tenderá, puesto que ya constituye la plataforma de gestión de bases terminológicas seleccionada por la industria GILT para incorporarla en sus metodología y desarrollo de tecnologías de la lengua (Melby, 2003).

### **2.5.6. OLIF**

El formato OLIF (*Open Lexicon Interchange Format*) es un conjunto de metadatos específicos, DTD, inscrito en XML que permite el intercambio de bases lexicográficas orientadas a la traducción automática o, lo que es lo mismo, lexicones computacionales. Los metadatos aportados codifican conocimiento fonológico, morfológico, sintáctico, semántico y pragmático. En el ámbito de la documentación técnica multilingüe no se utiliza este formato, de ahí que las herramientas empleadas en este entorno de trabajo no den soporte a este formato de intercambio.

### **2.5.7. XLT**

La sigla XLT (*eXchange format for Lex/Term-data*) hace referencia al formato de intercambio de bases terminológicas y lexicográficas. XLT se basa en

XML. El formato denominado *Default XLT* es equivalente al formato TBX que ya hemos descrito más arriba.

Las propuestas de estos formatos de intercambio surgen en el marco de un proyecto integrado en el V Programa Marco (2000-2001), el proyecto SALT (*Standards-based Access to Lexicons and Terminologies*), cuyas entidades participantes avalan la relevancia de la iniciativa. En este sentido, el Comité Técnico de la ISO, el grupo de trabajo de LISA (OSCAR), el consorcio OLIF2, el TEI, entre otras entidades, se afanan en conseguir representar y establecer el intercambio de metadatos que estructuran las bases terminológicas y lexicones computacionales.

Es lógico preguntarse por las razones que inducen al desarrollo de un formato de estas características, si se consideran los formatos de intercambio existentes tanto para las bases terminológicas (Martif), como para los lexicones computacionales (OLIF) que componen el nuevo formato XLT común e integrador propuesto. La respuesta está vinculada a la tendencia de integración de las herramientas de traducción asistida con las herramientas de traducción automática basada en ejemplos, es decir, basada en memorias de traducción. Esta posible comunión de herramientas de trabajo que no gestionan recursos lingüísticos estrictamente iguales en el entorno de la traducción requiere de un formato que pueda dar respuesta a las posibles exigencias. El hecho de que estos diferentes recursos lingüísticos compartan la misma estructura lingüística es para cualquier sistema automático o automatizado un problema de identificación diferenciada. Los sistemas identifican cadenas de caracteres en relación con otras cadenas de caracteres y en ambos casos se trata de caracteres alfanuméricos. Contar con estrategias de codificación que aborde esta problemática, supone allanar el camino para la integración de los sistemas de traducción asistida y los sistemas de traducción automática. En este sentido Melby (2001b) afirma que:

The OLIF and MARTIF formats are good enough and the need is growing. Let's get them widely enough known in their integrated form as XLT (with various user-group-specific subsets such as TBX) so that no fragmentation can take place. The language industries need one format that is good enough, not multiple competing formats. And they need it now.

## ***2.6. Herramientas de gestión de recursos lingüísticos para la documentación técnica multilingüe***

El mercado que produce documentación técnica multilingüe debe afrontar la gestión de grandes volúmenes de contenidos digitales que deben ser publicados cumpliendo criterios de calidad aceptables y precios de producción razonables. Esto es posible con la ayuda de recursos lingüísticos como las memorias de traducción (corpus paralelos) y las bases terminológicas. Ahora bien, para gestionar estos recursos, se precisan además herramientas diseñadas para su almacenamiento, recuperación, extracción, recopilación y organización. Las principales tecnologías de la lengua que procuran cubrir las necesidades de gestión de los recursos lingüísticos multilingües son, en primer lugar, los gestores de memorias de traducción y, en segundo lugar, los gestores terminológicos. Estas herramientas logran la optimización del uso y generación de los recursos lingüísticos y se utilizan no sólo en entornos de traducción, sino de gestión de la información multilingüe, en general.

Correspondientemente a las dos herramientas de software mencionadas, se ha desarrollado una serie de herramientas anexas que las complementan, como es el caso de los alineadores, buscadores de concordancias, correctores ortográficos, extractores de terminología, editores de texto, entre otros. Los sistemas integrados de traducción asistida son una muestra de la necesidad y complementariedad de las herramientas principales con las anexas. Sin embargo, no todas las herramientas utilizadas en la gestión de

recursos multilingües están integradas en los sistemas de traducción asistida, aunque sean igualmente necesarias para gestionar los recursos lingüísticos.

Los recursos lingüísticos en línea como los corpus monolingües y paralelos, los diccionarios, los bancos terminológicos son también herramientas de trabajo del traductor. Éstos no se tratarán en este trabajo, aunque son recursos susceptibles de convertirse en recursos lingüísticos gestionados por un sistema de traducción asistida, por ejemplo, importando un diccionario recuperado de Internet al gestor terminológico o alimentando las memorias de traducción mediante la alineación de textos paralelos. En los apartados 2.4.1. Las memorias de traducción y 2.4.2. Las bases terminológicas: diccionarios específicos, ya hemos abordado ampliamente la creación de estos dos tipos de recursos lingüísticos.

Antes de adentrarnos en el tema de las herramientas de gestión de los recursos lingüísticos, valga recordar que éstas son sólo una parte de la tecnología utilizada para la gestión de información en una empresa de traducción, como muestra la siguiente lista no exhaustiva:

- ❑ herramientas ofimáticas (hoja de cálculo, procesador de textos, etc.)
- ❑ herramientas audiovisuales
- ❑ protocolos de comunicación de sistemas
- ❑ redes internas (intranet)
- ❑ redes externas (Internet / extranet)
- ❑ mensajería electrónica
- ❑ sistemas de gestión de bases de datos
- ❑ gestores documentales
- ❑ programas de seguridad de datos (antivirus, gestor de respaldos)

Estas herramientas desempeñan diferentes tareas facilitando la transmisión y conservación de la información. Sin la utilización de muchas de estas



herramientas las posibilidades de explotación de las herramientas de traducción asistida se vería seriamente dañada y, en consecuencia, la gestión de los proyectos de traducción técnica.

### **2.6.1. Sistemas integrados de traducción asistida**

Los *sistemas de traducción asistida* son sistemas de creación, gestión y aprovechamiento de memorias de traducción.

En el número monográfico de la revista *Tradumática* dedicado a los gestores de traducción asistida, los autores de la publicación utilizan diferentes denominaciones para referirse al mismo concepto definido en el párrafo anterior. Así se pueden encontrar: *memorias de traducción* (Strandvik, 2001), *sistemas de traducción asistida por ordenador*, *sistemas TAO*, *herramientas de traducción asistida*, *herramientas TAO* (Editorial, 2001), *herramientas CAT* —acrónimo de *computer assisted translation*— (Gómez, 2001) o gestores de memorias de traducción (Sánchez, 2001).

Esta diversidad denominativa donde un concepto se asocia a varios términos es un fenómeno muy frecuente y fácilmente constatado por varios autores mediante la aplicación de métodos de lingüística de corpus (Nkwenti-Azeh, 1994: 61-68; Daile *et. al.*, 1996: 197-258 y Bowker, 1997:275).

La variación terminológica existente para identificar a los sistemas de traducción asistida da cuenta de la evolución de la tecnología. Las primeras herramientas que empezaron siendo almacenes de información en bases de datos sencillas con un motor de búsqueda, las *memorias de traducción*, han pasado a denominarse *sistemas de traducción asistida*. A estos sistemas se les ha ido dotando e integrando una serie de herramientas anexas que complementan las funciones originales para mejorar su rendimiento convirtiéndose en un sistema integrado de módulos.

En este trabajo utilizamos el término sistema de traducción asistida así como herramientas de traducción asistida indistintamente. Ahora bien, tal y como

hemos visto en el apartado 2.4.1. Las memorias de traducción, la utilización de *memorias de traducción* lo restringimos para identificar un tipo de recurso lingüístico compuesto de textos originales en una lengua alineados con su traducción en otra(s).

La idea original de almacenar las traducciones en un formato bilingüe y, por tanto, la base de los sistemas de traducción asistida, es de Melby (1982: 541-543). Éste manifiesta la necesidad de implementar una herramienta de concordancia entre dos lenguas para agilizar el proceso de traducción. Melby lanzó la hipótesis de que una buena concordancia bilingüe podría servir de material de referencia fiable para los traductores, además de reconocer que la correspondencia a escala inferior es muy difícil de establecer y, por lo tanto, casi imposible diseñar un sistema de equivalencia bilingüe absoluta de todos los elementos. No obstante, afirmaba que sí debía ser posible gestionar concordancias bilingües de sentido. En consecuencia, en su propuesta se decanta por la división entre segmentos de traducción, para establecer la correspondencia traductológica de, por ejemplo, un párrafo o una frase en el ámbito conceptual y no necesariamente de sintagmas y mucho menos de palabras.

Las múltiples ventajas de las herramientas de traducción asistida ya han sido descritas en la literatura especializada (Cámara, 2001; Freigang, 2001; Gómez, 2001) pudiéndose resumir en dos aspectos. Por una parte, automatizan el flujo de producción. Esto se consigue reutilizando los pares bilingües ya traducidos, reduciendo los costes de producción y acortando los plazos de entrega de los proyectos. Y por otra parte, mejoran la calidad de la producción de documentación técnica. Esto se logra debido a que se puede recuperar lo que se ha hecho y en caso necesario modificarlo para armonizar contenidos. La estrategia principal y genérica de los sistemas TAO es doble y común:

- reciclan unidades de traducción traducidas y revisadas por un humano —en el entorno de la traducción asistida, las unidades de traducción se

definen de forma variable siendo la segmentación tras un signo de puntuación que marca el final de la frase (., ?, !, :, ...) o un salto de párrafo las más frecuentes ofrecidas por defecto por los sistemas TAO—;

- recuperan unidades de traducción con un índice de coincidencia variable mediante la técnica de lógica difusa.

Por eso, el entorno ideal de aplicación son aquellos textos que se caracterizan por su carácter redundante, repetitivo y de alto contenido terminológico, como es el caso de los textos que conforman cualquier tipo de documentación técnica.

Sin embargo, las diferencias específicas de estas herramientas dependerán de los componentes que tengan integrados y de la funcionalidad de esos componentes. Los componentes generales de casi todas las herramientas profesionales de traducción asistida son los siguientes:

- motor de búsqueda que interactúa con el almacén de las traducciones denominado memoria de traducción. En el mercado existen tres tipos de gestores de memorias de traducción: El estático, que sólo identifica las equivalencias coincidentes totales, es decir de un 100%. El interactivo, que identifica equivalencias parciales con coincidencia ajustable en porcentajes y la combinación de estático e interactivo. Es obvio que el modelo combinado es el que ofrece mejores resultados (Cámara, 2001);
- herramienta de alineación que contribuye a generar material reciclable por los sistemas de traducción automática a partir de documentos ya traducidos, como ya se ha mencionado en el apartado 2.4. Gestión y tipo de recursos lingüísticos multilingüe;
- gestor de terminología que interactúa con las memoria de traducción, como veremos en un apartado especial dedicado a estas herramientas;

- gestores de concordancias que contribuyen a mejorar la explotación del corpus bilingüe y que interactúa con los corpus almacenados y con los gestores de terminología;
- correctores ortográficos que contribuyen al control de calidad;
- editor de traducción que facilita la creación de textos con funciones específicas para el traductor;
- convertidores o filtros de formatos que posibilitan la importación y exportación en diferentes formatos;
- herramienta de análisis estadístico que automatizan el análisis para el recuento de las unidades de facturación —por carácter, palabra, línea, párrafo, página— las equivalencias aproximadas y las repeticiones internas.

Los sistemas más sofisticados de traducción asistida como Transit de STAR (Díez, 2004: 491-506) o las herramientas de Trados (Carbajo, 2004: 473-490), además de los componentes mencionados, cuentan con:

- herramientas de gestión administrativa de proyectos que gestionan tarifas, monedas y parámetros para calcular el precio de un proyecto de traducción. Basados en los análisis estadísticos, se puede automatizar el proceso de facturación incluyendo el precio por la unidad de facturación y la moneda. Así mismo es posible incorporar otras variables para el cálculo, como las unidades con posible pretraducción automática, las equivalencias aproximadas y las repeticiones internas;
- API (*computer program interface*<sup>1</sup>) que permite definir interfaces entre la memoria de traducción y el gestor terminológico, entre la memoria de

---

<sup>1</sup> Un API es una serie de rutinas, protocolos y herramientas para desarrollar una aplicación de software compatible con el resto de los programas de su entorno de aplicación. Es una especie de *programa puente* (Webopedia, 2005).

traducción y una herramienta de localización y otras aplicaciones de traducción automática.

Todos estos módulos que conforman el sistema constituyen la estación del traductor (*translation workstation*) cuya aceptación entre los profesionales de la traducción ha evolucionado hasta hacerse familiar (Vasconcellos, 1988: 3-5). La integración de diferentes herramientas ha ido mejorando el modelo que ya propusiera Melby.

Un sistema de traducción asistida aligera las tareas implicadas en la traducción ya que contribuye a facilitarlas en diferentes aspectos que apuntamos a continuación:

- recuperar bitextos ya traducidos basándose en la equivalencia parcial o total de la cadena textual que debe traducirse; esta función se lleva a cabo mediante la comparación del corpus paralelo almacenado ejecutado por el motor de búsqueda. Algunos motores de búsqueda recuperan solo coincidencias exactas o totales mientras que otras emplean un algoritmo de lógica difusa para recuperar también las coincidencias parciales. La flexibilidad y estabilidad de este algoritmo influye decisivamente en el funcionamiento del sistema. Una equivalencia exacta es una coincidencia del 100% de cada uno de los caracteres tipográficos entre el segmento de partida actual y el segmento de partida almacenado. Una equivalencia aproximada o parcial es toda aquella que está por debajo de la coincidencia total del 100%. Algunos sistemas asignan además porcentajes a sus equivalencias aproximadas permitiendo al usuario el porcentaje mínimo que debe tener en cuenta el motor de búsqueda a la hora de identificar equivalencias parciales;
- traducir automáticamente. Si las equivalencias son exactas, el sistema contribuye a recuperar esas partes del texto de las memorias y sustituirlas automáticamente en el editor de la traducción sin intervención del traductor;

- controlar los elementos extraídos mediante informaciones extralingüísticas, como el porcentaje de correspondencia, fecha, nombre de archivo recuperado. Esta información contribuye a evaluar las equivalencias recuperadas por el sistema;
- generar más memorias de traducción a partir de textos equivalentes mediante un alineador. Esta herramienta es un programa que a partir de dos textos equivalentes genera pares bilingües o segmentos estructurados de unidad frasal en correspondencia con sus equivalentes también frasales. Esta idea tan sencilla la utilizan, en principio, todos los alineadores del mercado con independencia de la precisión con la que trabajen. Asimismo, la alineación se puede realizar de forma totalmente automática o se puede hacer confirmando algunas decisiones propias que toma el alineador, como unir segmentos, separarlos, moverlos o eliminarlos. En general, los alineadores profesionales como los de STAR o Trados ofrecen técnicas basadas en diferentes coeficientes regulables que trabajan con información no lingüística de tipo ortotipográfica y de maquetación ofreciendo una fiabilidad formidable. Martín Kay fue el primer científico que propuso un método para enlazar las correspondencias entre el texto de partida y el de salida para generar pares bilingües, aunque no se concretó un formalismo hasta 1988 (Kay y Röscheisen, 1988). Mediante este algoritmo sus creadores no pretenden hallar todas las correspondencias posibles entre todas las unidades que componen los textos, sino que apuntan a encontrar una correspondencia a nivel de frase, como es el caso de los alineadores integrados en los sistemas de traducción asistida que ven mejorado su rendimiento mediante el empleo de los diccionarios internos y específicos para cada tarea de alineación. Esto es, el alineador utiliza los diccionarios generales en las lenguas definidas y, si existen, también glosarios o diccionarios terminológicos específicos;

- encontrar términos en diferentes frases contribuyendo al control de estilo de la traducción y el respeto de las convenciones fijadas. Esta función se realiza con ayuda de un gestor de concordancias. Las herramientas de concordancias explotan corpus textuales en búsqueda de concordancias. Las concordancias son líneas de texto que contiene la cadena de caracteres y que tras la búsqueda se presentan en el medio de la línea del texto que ocupan. Las otras palabras a la derecha e izquierda son el contexto parcial en el que se encuentran. Cada concordancia se compone de la palabra clave de búsqueda y el texto que acompaña en esa línea de texto a la palabra clave. Los gestores de concordancias generan concordancias automáticas, siempre en función de la entrada de datos del usuario. Cuando el buscador de concordancias se halla integrado en los sistemas de traducción, las concordancias se extraen del corpus que son las memorias de traducción. En las herramientas profesionales de traducción asistida como en Trados, Transit y Déjàvù existe un motor de búsqueda de concordancias. Las concordancias son una fuente de gran valor porque contribuyen a homogeneizar la terminología, así como la fraseología, es decir, la combinación frecuente de los términos con otras piezas lingüísticas específicas en los textos especializados (Budin, 1990; Desmet, 1994) y a esta agrupación de palabras en una frase se conoce como *collocations*. Concretamente Sinclair define el término como “the occurrence of two or more words within a short space of each other in a text” (1991:170). El punto de partida de esta idea se basa en la concepción de entender la dimensión del significado lingüístico en contexto, tal y como lo entiende Firth y Cabré en su teoría lingüística contextual y su teoría comunicativa de la terminología respectivamente —parte de los principios teóricos en los que nos basamos en este trabajo ya mencionados en el apartado 1.6. Marco teórico— en oposición al concepto de la lengua como constructo mental. Más allá de la traducción asistida los gestores de concordancias también se comercializan como herramientas independientes. Pueden aplicarse en

otros entornos extrayendo las concordancias de corpus monolingües o de más lenguas también, como es el caso de la *Biblia Paralela* de Heartlight (2004) que gestiona corpus paralelos con textos traducidos del inglés, árabe y español y que mostramos a continuación:

**Búsqueda de Palabra/Frase: Búsqueda nueva**

Resultados de Búsqueda: **324 Versos** conteniendo 'altar'

Búsqueda  usando

en  con

Dentro de los resultados  Búsqueda nueva |  Incluye Recursos de Estudio

[AYUDA](#)

---

**Versos 1 - 20 de 324**

Páginas: [\[1\]](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) >> [Siguiente 20 >>](#)

<b>The New American Standard Bible</b>	<b>La Biblia de la Américas</b>
<p><b>Ge 8:20</b> Then Noah built an <b>altar</b> to the LORD, and took of every clean animal and of every clean bird and offered burnt offerings on the <b>altar</b>.</p>	<p><b>Gn 8:20</b> Y edificó Noé un <b>altar</b> al SEÑOR , y tomó de todo animal limpio y de toda ave limpia , y ofreció holocaustos en el <b>altar</b>.</p>
<p><b>Ge 12:7</b> The LORD appeared to Abram and said, "To your descendants I will give this land." So he built an <b>altar</b> there to the LORD who had appeared to him.</p>	<p><b>Gn 12:7</b> Y el SEÑOR se apareció a Abram , y le dijo: A tu descendencia daré esta tierra . Entonces él edificó allí un <b>altar</b> al SEÑOR que se le había aparecido.</p>
<p><b>Ge 12:8</b> Then he proceeded from there to the mountain on the east of Bethel, and pitched his tent, with Bethel on the west and Ai on the east; and there he built an <b>altar</b> to the LORD and called upon the</p>	<p><b>Gn 12:8</b> De allí se trasladó hacia el monte al oriente de Betel, y plantó su tienda, teniendo a Betel al occidente y Hai al oriente; y edificó allí un <b>altar</b> al SEÑOR, e invocó el nombre del SEÑOR .</p>

**Fig. 25: Concordancias resultantes de *Altar* en un corpus paralelo**

Esta herramienta compuesta de corpus paralelo con un gestor de concordancias es un potente instrumento de gestión de conocimiento no solo para el ámbito de la traducción, sino aplicable a otros contextos, como también se indica en la descripción de aplicación de la *Biblia Paralela*:



Although the Parallel Bible can be used for text comparison between two different translations in the same language, the main intent is to allow the use of the Bible as a study tool for learning another language (Heartlight, 2004).

Estas herramientas ejemplifican, utilizando las denominaciones de Cabré (1999), las diferencias existentes, entre los términos *in vitro* —terminología registrada en un diccionario— y los términos *in vivo* —terminología contextualizada y extraída de los textos—. Los términos contextualizados dan cuenta de los aspectos comunicativos de su uso, así como de la confirmación de unidades lingüísticas y extralingüísticas con la misma función de representación del conocimiento especializado;

- generar estadísticas sobre el texto. Mediante las herramientas estadísticas se puede estimar el volumen de trabajo que va a implicar un proyecto de traducción. Esta información es necesaria para planificar el calendario y para su presupuestación. El cómputo de palabras y la repetición interna en los textos son dos de las funciones típicas de las herramientas estadísticas de los sistemas de traducción asistida. El porcentaje de repeticiones internas que busque la herramienta estadística depende de la configuración que se haga de la unidad de medida para detectar repeticiones, es decir, que previsiblemente el porcentaje de repeticiones por párrafo será menor que por segmento, palabras o caracteres. En el entorno de la traducción asistida el recuento se hace tomando como unidad el segmento o frase;
- importar los textos originales al formato de la memoria de traducción, así como la exportación de los textos del formato de la memoria al formato original. Esto se lleva a cabo mediante los convertidores o filtros de formatos que son pequeños programas que gestionan la información textual y de diseño;

- controlar la ortografía de las palabras, optimizando la calidad del producto final. Esto se realiza mediante las herramientas llamadas correctores ortográficos. Estos programas de ayuda al usuario se componen de un diccionario y de un motor de búsqueda que funciona comparando las palabras del texto con el diccionario almacenado. Si la palabra en cuestión no aparece en dicho diccionario se la toma como incorrecta, pudiendo corregirla si así fuera necesario o incluirla al diccionario referente. El diccionario referente para que el motor de búsqueda pueda comparar las palabras del texto puede sustituirse por un corpus referente. Los correctores pueden también contar con una pequeña gramática en forma de reglas de derivación morfológica y de formación de palabras para la identificación de cadenas de caracteres como correctos a pesar de no estar incluidos en el diccionario. Los correctores ortográficos son herramientas muy habituales que se integran no solo en los sistemas de traducción asistida profesionales, sino en casi todos los procesadores de texto y en los sistemas de traducción automática;
- identificar los términos específicos mediante la interacción con un gestor terminológico. Si el gestor terminológico se halla totalmente integrado en el editor de traducción, el sistema está constantemente buscando términos de forma sincronizada con la unidad de traducción en la que se encuentra el traductor. El traductor no sólo puede incorporar al editor de traducción los términos encontrados en los diccionarios especializados tan sólo pulsando una tecla, sino que además puede introducir nueva terminología desde el editor de traducción;
- procesar la traducción utilizando las funciones de un editor de traducciones específico. El editor de traducción puede ser independiente de cualquier programa de procesamiento de textos y estar integrado en el sistema de traducción asistida; puede estar también integrado en un procesador de textos estándar, como es el

caso del Workbench de Trados que cuenta con una barra de herramientas específica, como la que mostramos aquí:



**Fig. 26: Barra de estado integrada en el editor específica para la traducción**

Todas estas ventajas que acabamos de enunciar ayudan ciertamente a mejorar las condiciones de trabajo del traductor. Asimismo, estas ventajas expuestas contribuyen a la reducción de costes de producción. El siguiente gráfico extraído de un análisis de Schwartz y Toon (2004) muestra la reducción de costes en la producción de contenido multilingüe utilizando herramientas de traducción asistida.

Publication type	Pages per year	Target countries	Updates per year	% Updates per year	Total pages per year	Costs for localization
Sales material	5,000	20	1	20%	19,000	\$1,520,000
Websites	100	2	2	30%	60	\$4,800
Press releases	300	2	2	80%	480	\$38,400
Contracts	1,000	20	1	20%	3,800	\$304,000
Operating manual	10,000	18	1	20%	34,000	\$2,720,000
Training manuals	10,000	4	1	20%	6,000	\$480,000
Service manuals	10,000	4	2	20%	12,000	\$960,000
Other catalogues	400	4	2	10%	240	\$19,200
R&D documents	10,000	4	2	5%	3,000	\$240,000
Control documents	500	4	1	5%	75	\$6,000
Diagnostics	1,000	4	1	5%	150	\$12,000
Company-wide communication	10,000	4	2	50%	30,000	\$2,400,000
					1 Year Total	\$6,544,400
					1 Year Potential Savings (@60%)	\$2,617,760

**Fig. 27: Ahorro potencial de costes con herramientas de traducción asistida**

Sin embargo y, a pesar de las evidentes ventajas de estas herramientas, no dejan de tener críticas de usuarios e incluso detractores, debido a las desventajas y límites que se observan.

Una de las desventajas que se mencionan con mayor frecuencia es la carente garantía de calidad de las propuestas traductológicas almacenadas. Es decir, si la traducción para un determinado segmento es incorrecta es posible que se hereden los mismos errores en el segmento que se repita con posterioridad y de esta manera perpetuar el error. Los sistemas de traducción

asistida gestionan trabajo humano previo, pero no sustituyen la parte cognitiva implicada en el proceso traductor. Esto es, el correcto proceso de decodificación del mensaje en la lengua de salida y el de recodificación en la lengua de llegada es una tarea cognitiva que debe llevar a cabo el traductor.

Los procedimientos para la resolución de problemas están vinculados a la forma en la que gestionamos la información que almacenamos en nuestra memoria. Siguiendo a Dörner (1979: 27), en general estos procedimientos pueden ser de dos tipos: el epistémico y el heurístico. La parte epistémica de la memoria o base de conocimiento determina la habilidad de resolver problemas en función del conocimiento acumulado con anterioridad. Es un tipo de conocimiento basado especialmente en la experiencia. Mientras que el procedimiento heurístico es teórico o metodológico confiado en buscar nuevo conocimiento para resolver problemas. En principio, las memorias de traducción son soluciones epistemológicas. Se trata de reutilizar el conocimiento aplicándolo a problemas ya resueltos. Wilss (1988: 86) argumenta que cuanto mayor sea la base de conocimiento, es decir, la memoria epistémica del traductor, menos tiempo empleará desarrollando estrategias heurísticas para resolver problemas traductológicos de los que no tiene referencia. Siguiendo este razonamiento, se puede afirmar que, cuanto mayor sea la memoria de traducción almacenada, menos tiempo utilizará el traductor en buscar las soluciones que no le oferta el almacén de memorias del sistema.

Sin embargo, cualquier problema que para un traductor se resuelve mediante una inferencia lógica es para el sistema de memorias de traducción un problema insalvable que requiere necesariamente intervención humana. Las soluciones que presenta el sistema y que simulan una estrategia heurística no dejan de ser soluciones estadísticas de coincidencia parcial. Y esto es, porque el sistema no *metaboliza* la experiencia, aplicando los datos acumulados en la resolución de los problemas; es decir, el conocimiento no se asimila para generar nuevo conocimiento, sino que busca, extrae y

presenta información en función del grado de coincidencia de los caracteres que se incluyen en una unidad de traducción.

Según lo recogido en el informe ESPRIT (Brace, 1992) dedicado a la evaluación de las herramientas de traducción asistida, los límites cognitivos expuestos no son las únicas limitaciones y carencias de estos sistemas. Según el citado informe las desventajas técnicas de los gestores de memorias de traducción son:

1. La falta de convertidores de formato para los programas de creación de documentación —también conocidos como filtros de importación y exportación—.
2. La falta de integración en los sistemas de traducción asistida de herramientas de localización para los módulos textuales del software.
3. Los grandes costes involucrados en la generación de memorias de traducción utilizando traducciones ya hechas.
4. La dificultad y los costes de mantenimiento de las herramientas de traducción.
5. La gran inversión para el traductor autónomo. Si se subcontrata el trabajo de traducción, los traductores autónomos necesitan de estas herramientas para poder llevar a cabo la traducción de los textos ya preparados mediante los sistemas de traducción asistida.
6. La dificultad de aprendizaje de las herramientas. Se tarda mucho tiempo en adaptarse al nuevo entorno.

Si bien las desventajas apuntadas podrían resultar un impedimento para su aplicación, la evolución de las herramientas de traducción asistida en estos últimos años se han centrado en poder resolver estos límites. En este sentido, a continuación mostramos la situación actual de todos los problemas apuntados más arriba, haciendo coincidir la identificación numérica de los problemas con las soluciones si existen:

1. Las herramientas de traducción asistida comerciales han ido desarrollando los filtros de importación y exportación en función de las necesidades de los clientes potenciales y reales de estas herramientas, aumentando su gama de convertidores cubriendo prácticamente todas las necesidades, tal y como muestra la tabla con los formatos a los que dan soporte dos de las herramientas profesionales del mercado:

<b>Tipo de formato</b>	<b>STAR Transit</b>	<b>Trados</b>
Texto ANSI / ASCII / Unicode para Windows	✓	✓
Texto para Apple Macintosh	✓	
Corel WordPerfect	✓	
HTML	✓	✓
XML (también derivados como ASP.NET, ASP, JSP, XSL)	✓	✓
SGML	✓	✓
SVG (Scalable Vector Graphics)	✓	
MS Word para Windows	✓	✓
MS Excel	✓	✓
MS PowerPoint	✓	✓
RTF y RTF para WinHelp	✓	✓
RC (Archivos de recursos )	✓	✓
QuarkXPress	✓	
Adobe FrameMaker	✓	✓
Adobe PageMaker	✓	✓
Interleaf /Quicksilver	✓	✓
Adobe InDesign	✓	✓
XGate para QuarkXPress	✓	✓
AutoCAD	✓	

**Fig. 28: Formatos de conversión de STAR Transit y Trados**

2. La falta de integración de las herramientas de localización en los sistemas de traducción asistida incluida en el catálogo de desventajas del informe ESPRIT puede decirse que doce años más tarde se ha solventado en gran parte. Las herramientas de localización son programas especializados en la gestión de formatos de archivo específicos de los programas de software como los ejecutables, los

archivos de recursos y los archivos basados en XML. Si bien es cierto que, como ya hemos visto, los sistemas de traducción asistida entre tanto y como respuesta a esa antigua carencia, dan soporte a formatos de archivos específicos de software como formatos Windows Help Files, Windows Resource Files (RC), código fuente C y CC++, código fuente Java, entre otros lenguajes de programación, también es cierto que algunas tareas se facilitan con las funciones que ofrecen estas herramientas específicas y no los sistemas de traducción asistida.

Las herramientas profesionales de localización como *Catalyst* de Alchemy Software Development<sup>1</sup> o *Passolo* de Pass Engineering GmbH<sup>2</sup> permiten traducir las cadenas de caracteres de los recursos textuales e interactivos de software en diferentes idiomas incluso en las lenguas asiáticas por dar soporte al formato Unicode y la bireccionalidad para lenguas como el árabe y el hebreo, característica que también encontramos en los sistemas profesionales de traducción asistida. Asimismo ofrecen la funcionalidad de editar en el formato real o *WYSIWYG* utilizado para *What you see is what you get*. Esto significa que, por ejemplo, el texto de un cuadro de diálogo o elemento de un menú en inglés se puede sobrescribir con el texto en español. Esto marca una diferencia considerable con respecto al funcionamiento de las herramientas de traducción asistida al traducir los módulos textuales de los recursos interactivos de software, ya que se trabaja en modo texto. La ventaja del método de edición *WYSIWYG* es la facilidad de adaptación de los campos textuales a la longitud del texto de llegada.

En cualquier proyecto de traducción con módulos textuales de recursos de software, primero se traduce el software, después las ayudas en línea y después los manuales de usuario. La traducción de software

---

<sup>1</sup> Para más información véase <http://www.alchemysoftware.ie>

<sup>2</sup> Para más información véase <http://www.passolo.com>



genera terminología en la lengua de llegada relevante para el resto de la documentación, por ello existen APIS, interfaces entre los programas profesionales más populares de traducción asistida —STAR Transit y STAR TermStar y TRADOS Translator's Workbench, MultiTerm y TagEditor—. Estas interfaces de conexión permiten el intercambio de datos y su reutilización facilitando el flujo de trabajo. Mediante el formato estándar TMX se intercambian las memorias de traducción, así como las bases terminológicas. La evolución de estas herramientas por sus funciones y por su integración en los sistemas de traducción asistida resuelve, pues, parte del flujo mecánico de los proyectos de localización.

3. A propósito de los costes vinculados a la generación de memorias de traducción, hay que mencionar que las herramientas de alineación han experimentado también una evolución. Sus algoritmos cada vez más sofisticados permiten la generación prácticamente de forma automática de memorias de traducción a partir de textos previamente traducidos. La inversión en la generación de memorias de traducción se reduce justificando así la conversión de los datos textuales.
4. El problema acerca de los costes de mantenimiento de las memorias de traducción está estrechamente vinculado a la forma de almacenamiento que utilizan los sistemas de traducción asistida. En la actualidad existen dos formas de almacenar: el modelo basado en base de datos, utilizado por la mayoría de los sistemas de traducción asistida y el basado en archivos XML, usado por Transit, que además no utiliza un formato interno propio para sus archivos, sino el estándar XML con Unicode para el formato de texto. La forma de almacenamiento utilizada afecta al mantenimiento de las memorias de traducción. La posibilidad de que las bases de datos se vean alteradas, con el consiguiente riesgo de pérdida de datos es frecuente, mientras que los archivos XML no corren ese peligro. El modelo XML ofrece datos independientes de cualquier plataforma, lo que hace viable

cualquier manipulación de estos datos para otros fines. Este lenguaje permite, a escala metasemántica, incluir datos lógicos de clasificación de los conocimientos en el flujo de datos, por ejemplo, para utilizar las memorias de traducción codificadas en XML integrándolas en otros sistemas de objetivos lingüísticos diferentes al ámbito de la traducción. Además XML es el lenguaje del futuro inmediato en Internet, como hemos explicado antes. Por tanto, ofrece mayor libertad para anidar en archivos de texto paquetes de datos como pistas de audio o de vídeo, datos gráficos, etc. o vincularlos para facilitar los efectos multimedia. La administración y gestión de las bases de datos requiere un perfil especial, ya que, por una parte, la estructuración de las bases de datos puede ser bastante compleja para un usuario y, por la otra, se complica la actualización en caso de querer descartar las malas traducciones de las memorias de traducción almacenadas en bases de datos. Otra de las desventajas del modelo de bases de datos está relacionada con la imposibilidad de acceder al contexto original de las memorias de traducción almacenadas en las mismas, mientras que el modelo XML permite el acceso directo a las memorias de traducción contextualizadas. La velocidad del flujo de trabajo es notablemente mayor con el modelo XML que con el modelo de bases de datos. Al tratarse de archivos abiertos independientes de la plataforma, pueden integrarse en otras de las tareas que repercuten directamente en todo el proceso de trabajo, como la administración y gestión. Por tanto, la dificultad y los costes de mantenimiento pueden reducirse considerablemente mediante la selección de un sistema TAO cuyo medio de almacenamiento no sea una base de datos.

5. Los costes de la tecnología para la traducción se han minimizado considerablemente. La razón de esta evolución estriba por una parte en la disminución de los precios de toda la producción de software experimentada en general en los últimos años. La otra razón estriba en la tendencia de los grandes fabricantes de tecnología para la

traducción de ofrecer versiones limitadas de sus productos que se pueden descargar gratuitamente desde el sitio web correspondiente, instalarse en el ordenador y realizar la traducción e inclusión de la terminología. Esta estrategia empresarial cubre de esta manera las necesidades de traductores autónomos que tienen las empresas de traducción sin renunciar a la aplicación de sistemas de traducción asistida. Por otra parte, mediante esta estrategia los fabricantes de estas herramientas se siguen garantizando presencia en el mercado.

Cierto es que las versiones gratuitas de los sistemas profesionales de traducción asistida tienen funcionalidades limitadas. También es cierto que permiten al traductor autónomo llevar a cabo su trabajo con una inversión cero en tecnología de la lengua.

6. El problema relacionado con el aprendizaje de estas herramientas también se ha minimizado. La experiencia acumulada durante los últimos seis años por la autora de esta investigación como docente de informática aplicada a la traducción muestra una evolución en el comportamiento y adquisición de conocimiento de los usuarios. Las razones apuntan a la popularización del PC, Internet y programas de toda índole, lo que han generado progresivamente una naturalidad en la percepción de los sistemas teniendo cada vez menos temores y limitaciones para adentrarse a las funcionalidades de cualquier programa de software. La otra razón estriba en la paulatina evolución de los entornos de traducción asistida, los cuales son cada vez más amigables y cuentan con documentación técnica mejorada, lo que reduce la dificultad de aprendizaje.

A pesar de las carencias apuntadas, no se debe olvidar que la motivación para el desarrollo de las primeras propuestas tecnológicas de traducción asistida se originan por la necesidad de automatización y por los insuficientes resultados obtenidos en el ámbito de la traducción automática. En este sentido, la tecnología de la traducción asistida ha representado un avance insustituible en la producción de documentación técnica multilingüe. Es más,

de un avance tecnológico *imperfecto*, como pretenciosamente afirman Biau y Pym (2003: 287-301), han nacido soluciones que se han ido sofisticando y que seguirán en esa línea. La prueba de que la generación sistemática de bitextos indizados y recuperables de forma total y parcial es el uso que se está dando en el ámbito de la traducción automática cuya estrategia se basa en ejemplos obtenidos de un corpus de ejemplos ya traducidos, es decir, de memorias de traducción (Somers y Fernández, 2003: 167-192).

### **2.6.1.1. Gestores terminológicos**

Los *gestores terminológicos* también llamados *gestores de bases terminológicas* son programas de software compuestos de una base de datos extensible que permiten la gestión —creación, extracción y modificación— de los datos por parte de los usuarios.

Los gestores terminológicos profesionales mayoritariamente dan soporte a un modelo de categorías con una orientación onomasiológica a diferencia de los diccionarios electrónicos de léxico común cuyas unidades se organizan con una orientación semasiológica, esto es, en cada entrada o registro del diccionario se apuntan los diferentes valores semánticos. Así, mientras la lexicografía parte de la palabra para llegar al concepto, la terminología parte del concepto para llegar a su representación: el término (Cabré, 1997). Esta orientación onomasiológica, es decir, del término al concepto puede derivar en un modelo orientado al término o al concepto en sí. Siguiendo la metodología ampliamente establecida en la gestión terminológica de entornos multilingües, se reconocen las ventajas de gestionar la terminología con una orientación al concepto (Melby *et al.*, 2001: 623); esto es, centrado en el significado y no en la representación terminográfica del concepto en cada lengua. De esta manera cada concepto será un nuevo registro que puede contar con diferentes términos en cada una de las lenguas. Así un concepto puede contener múltiples términos y variantes lingüísticas para un sólo objeto, característica o acción. En el siguiente ejemplo se muestra la

información relativa al concepto con independencia de su representación en cada lengua:



**Fig. 29: Información general de un concepto**

Por tanto, cada término en la lengua correspondiente tiene información específica, como también se muestra en la siguiente captura de pantalla:



**Fig. 30: Información específica de un término**

El modelo de las categorías que subyace al gestor de terminología ha sido motivo de debate durante años. Tradicionalmente, las categorías discutidas se agrupan entorno a cuatro grandes grupos: información administrativa, lingüística, pragmática o de uso y de representación conceptual (Pérez, 2002: 5-5). Los escenarios de aplicación de la gestión terminológica son determinantes para seleccionar el modelo de datos categorial. Un reflejo de esta situación es el desarrollo de estándares diferentes para el trabajo terminológico propiamente dicho como la norma ISO 12620 y el terminológico orientado a la traducción como la ISO 12616 que, como ya hemos visto, se trata de un modelo conceptual categorial para la terminografía orientada a la traducción.

En la mayoría de los casos, los gestores terminológicos han sido concebidos como herramientas de ayuda a la traducción que facilitan tanto la consulta de diccionarios temáticos como su gestión (Terminómetro, 1996). Así, la motivación de su diseño se basa en las necesidades del sector que consecuentemente desarrolla y aplica.

Los gestores de terminología profesionales como Multiterm o TermStar — ambos productos trabajan integrados en los sistemas de traducción asistida de Trados y STAR respectivamente, aunque también trabajan como módulos autónomos— ofrecen funciones para:

- importar datos permitiendo volcar el contenido de otros diccionarios electrónicos, bases de datos y archivos disponibles en diferentes formatos para convertir los datos a formatos de un entorno común para reutilizarlos a mayor velocidad;
- personalizar el perfil de usuario y la configuración de cada diccionario posibilitando la presentación de los datos según sus necesidades, sin pérdida de información. También cuentan con un modelo de categorías subyacente orientado al concepto y predefinido y/o con posibilidades de personalización en función de las necesidades;

- buscar y clasificar los datos terminológicos simplificando la gestión de la información que figura en el diccionario, según las preferencias del usuario;
- filtrar la información contenida en los registros de la base de datos posibilitando la visualización, de las partes del registro de datos que se deseen mostrar, por ejemplo, todas las lenguas o sólo las lenguas de partida y de llegada, con o sin información lingüística, con o sin encabezado.

Más allá de las funciones específicas, un gestor terminológico ayuda al traductor a utilizar la terminología de forma coherente. Al crear un nuevo concepto, se asigna un término preferente, que se debe utilizar de forma homogénea. Los sinónimos y otras formas del término se almacenan en campos separados. Sin embargo, estas herramientas permiten realizar una búsqueda y mostrar el concepto apropiado informando del término preferente. De esta manera, un gestor terminológico favorece el uso coherente de la terminología en cualquier entorno.

Un gestor terminológico tanto en una empresa como entidad pública permite gestionar la terminología de manera profesional para que todo el personal utilice los mismos términos de manera homogénea en todas las lenguas. Si los gestores además ofrecen además una solución web, se puede publicar la terminología en la intranet con o sin restricciones de acceso a través de Internet. De esta manera se facilita la visualización y gestión de los recursos terminológicos de una entidad de manera remota, a través de Internet desde cualquier punto. Estos gestores también pueden permitir enviar peticiones de nuevas entradas al gestor de terminología de forma pública o privada, como por ejemplo, Multiterm iX de Trados (2004) o WebTerm de STAR (2004) que permite descargar a través de Internet la terminología actualizada e integrarla automáticamente en el gestor terminológico TermStar, que está incorporado en Transit, el sistema de traducción asistida de STAR. Por una parte, un gestor terminológico puede ser uno de los módulos principales que

componen un sistema de traducción asistida; por otra parte, un gestor terminológico puede beneficiarse de otras herramientas anexas, como los correctores ortográficos que ya hemos abordado o los extractores terminológicos, a los que nos dedicamos en el siguiente subapartado.

#### **2.6.1.1.1. Extractores terminológicos**

Estas herramientas permiten la identificación y extracción de candidatos a términos de los textos explorados. Estas herramientas están abocadas a generar material para las bases terminológicas y que requieren del análisis y evaluación del usuario para la inclusión definitiva en la base de datos. Este tipo de herramientas son muy útiles en el entorno de la gestión de la información y del conocimiento —contribuyen a la extracción de palabras claves como posibles candidatos a unidades de indización—, así como en el ámbito de la traducción especializada —poder extraer automáticamente los términos de un texto contribuye a elaborar glosarios específicos—.

Sin embargo, la automatización de la extracción de los términos se enfrenta a problemas consustanciales a la propia naturaleza lingüística de los términos, así como a su variación denominativa. Estos problemas estriban ante todo en identificar y separar los términos de los no términos, tanto simples como compuestos. Para que un proceso sea automatizable, debe existir un diseño conceptual que represente las fases y estrategias que den la solución a un problema. De esta manera se podrá hacer legible por la máquina. La formalización conceptual puede ser de tipo lingüístico, por ejemplo, mediante el etiquetado previo de los textos para dotarlos de conocimiento morfológico. Asimismo se puede dotar de conocimiento sintáctico etiquetando los textos para basar una extracción en algoritmos formales de composición gramatical. Esta aplicación, sin embargo, no resuelve los problemas de delimitación de los sintagmas de significación, ya bien sean términos compuestos, ya bien simples (Ahmad, *et al.* 1996: 168).

Con un corpus etiquetado con alguna información acerca de los formantes de los términos se puede simular razonamientos para extraer, por ejemplo,



candidatos a términos en el ámbito biomédico mediante la identificación de los formantes grecolatinos (Estopà, 2001: 72). La aplicación de este nivel de conocimiento semántico y morfológico alcanza un buen rendimiento en la extracción de términos de las ciencias de la salud en general debido a la rica generación de términos con prefijos, sufijos y afijos de origen culto (grecolatino). Esta estrategia, si bien también se podría aplicar, por ejemplo, en el área temática de la química, difícilmente se podría aplicar en el lenguaje específico, como es el lenguaje técnico propio de la documentación técnica que nos ocupa.

Los métodos estadísticos basados en cálculos de frecuencias con o sin diccionarios de palabras de exclusión son otra de las estrategias utilizadas en las herramientas de extracción de términos. Ciertamente es un método no exento de problemas, pero que ofrece resultados satisfactorios a partir de una estrategia más fácil y económica de aplicar que la dotación de conocimiento lingüístico a los textos.

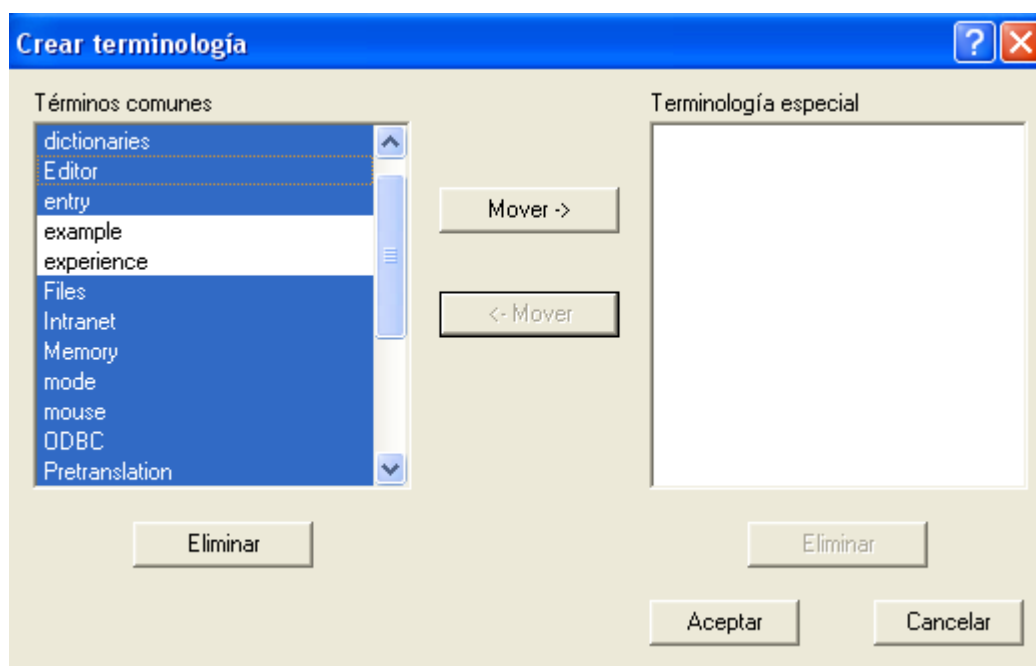
Los métodos híbridos que combinan el conocimiento lingüístico con el análisis estadístico parece que ofrecen mejores resultados, sin embargo, cualquiera de los métodos usados tienen limitaciones y éstas se evidencian mediante la necesidad de intervención humana para decidir si un candidato a término seleccionado por la herramienta ciertamente lo es o no (Lauriston, 1995).

La solución de la empresa Tamis (2003) subsidiaria de Xerox ha desarrollado una herramienta de extracción terminológica basada en conocimiento lingüístico y estadístico. Esta herramienta permite la extracción inteligente mediante un análisis superficial morfosintáctico de los textos. Los resultados se mejoran cuando la extracción se lleva a cabo a partir de corpus bilingües. La intervención humana para confirmar las propuestas señaladas por la herramienta es absolutamente necesaria. Este extractor permite la integración en sistemas de traducción asistida, siempre y cuando éstos den

soporte al formato estándar de intercambio de bases terminológicas, Martif, y al formato estándar de intercambio de memorias de traducción, TMX.

Las herramientas de extracción de términos integradas en las herramientas de traducción asistida por ordenador son también reflejo de las limitaciones de su aplicación (véase la documentación técnica de *Extraterm* de Trados y de *STAR Transit*). En este sentido, los extractores fraccionan el problema asistiendo a la solución sólo semiautomática de los vaciados dependiendo siempre de las memorias de traducción y de la intervención humana.

En el caso de la función de extracción en *Transit*, la herramienta busca en el texto de la lengua de salida las palabras que ni están en los diccionarios definidos para el proyecto de traducción en cuestión, si es que tiene, ni en el diccionario de exclusión compuesto de una lista de palabras comunes que se va alimentando por cada extracción.



**Fig. 31: Extracción y creación de terminología asistida**

En este caso, aquellos elementos considerados como términos y que deben por tanto volcarse en una bases de datos específica se seleccionan y mueven al bloque derecho. Los que permanecen en el bloque izquierdo se

guardarán en el diccionario específico de *términos comunes* que, en realidad, son elementos léxicos no portadores de conocimiento especializado. La próxima vez que se utilice la función de extracción de término para esta lengua, se excluirán los elementos ya guardados y considerados como no específicos. Los elementos que no deseemos añadir al diccionario terminológico ni a la lista de términos comunes para su exclusión se eliminan.

### **2.6.2. Herramientas no integradas en los sistema de traducción asistida**

En este apartado contemplamos dos tipos de herramientas que participan o deberían participar en el proceso de gestión de los proyectos de traducción. Estas herramientas no se hallan integradas en los sistemas de traducción asistida, sin embargo, su aplicación puede repercutir favorablemente en la mejora de la gestión de recursos lingüísticos. Nos referimos a los gestores documentales, a los gestores de tesauros y a las herramientas de gestión de proyectos específicos de traducción.

Así pues, mostraremos las ventajas de estas herramientas en los diferentes escenarios en los que pueden intervenir durante el ciclo de producción y de gestión de los recursos lingüísticos almacenados.

Concretamente, las herramientas de gestión de proyectos, requerirían más que un apartado, un trabajo de investigación por sí mismo debido a la carencia de literatura especializada. Si bien, la formación en herramientas de gestión de proyectos se está iniciando modestamente en el marco de algunos cursos especializados en el ámbito universitario español —en el *Máster de tradumática* de la Universidad Autónoma de Barcelona o en el *Máster de localización* de la Universidad Europea de Madrid, por citar los ejemplos pioneros—, también es cierto que la formación acerca de la función y el uso de estas herramientas sigue siendo una asignatura pendiente todavía por abordar en toda su dimensión.

### **2.6.2.1. Gestores de tesauros multilingües**

Los gestores de tesauros son herramientas que ayudan a crear, editar, fusionar tesauros y a buscar términos y sus relaciones. La necesidad de gestores de tesauros nace de la relevancia creciente de los tesauros para gestionar los volúmenes ingentes de información. Estas herramientas han empezado a utilizarse tímidamente en el entorno de la traducción asistida (Höge, 1991a: 29), aunque no se han explotado sus posibilidades como muestra la no integración de un gestor de tesauros en un sistema de traducción asistida con miras a la indización y recuperación de los documentos donde se incluyen los recursos lingüísticos.

La relevancia de los tesauros se refleja en la proliferación de las herramientas que los gestionan, como muestra el estudio comparativo de Ganzmann (1990) acerca de los gestores de tesauros para PC existentes en el mercado. Aun cuando las prestaciones que ofrecen las diferentes herramientas sean diversas (véase la evaluación de software de gestión de tesauros de Moya y Gil, 2001), existen funciones genéricas que deben contemplar todas las herramientas de gestión de tesauros (Milstead, 1991: 165-175 ) para proporcionar asistencia en:

- la creación de tesauros mediante el diseño de las tablas relacionales que sirvan de base para el tesoro;
- la definición de niveles de usuarios mediante códigos de acceso y mediante la definición de las funciones asignadas a cada usuario dentro de la dinámica de flujo de trabajo;
- la gestión y validación de términos así como sus relaciones. Esta función se lleva a cabo por el usuario que tiene asignada estas competencias pudiendo crear nuevos términos, relaciones o emparejamientos entre descriptores y recursos de Internet o bien validar los que han propuesto los usuarios que tienen asignadas las funciones correspondientes para proponer términos, relaciones o indización;

- la gestión de propuestas de términos, relaciones o indización. Esta función está vinculada a la definición de niveles y perfiles de usuarios. En este sentido es necesario separar el perfil del gestor de términos, relaciones o indización, con respecto al perfil de los usuarios que proponen la creación, modificación o supresión de dichos elementos. El gestor del tesoro debe validar o rechazar tras su evaluación para que los cambios se incorporen al tesoro. La definición de los términos permitirá la inclusión de notas explicativas y/o definición de los mismos, si se considera necesario;
- el control de fuentes mediante la creación de fuentes de información que se han utilizado para la obtención de la terminología del tesoro. Así, posteriormente se pueden asociar dichas fuentes a los diferentes términos y realizar revisiones o actualizaciones;
- la consulta del tesoro a través de la presentación alfabética y de la presentación sistemática.

Un tesoro es un sistema de organización conceptual sistemático del conocimiento, compuesto de términos descriptores estructurados en un sistema de índices entrelazados mediante relaciones semánticas. Los tesoros contienen notas de uso y algunos contienen definiciones del término. Si los tesoros son además multilingües tienen relación con el término equivalente en cada lengua contemplada por el tesoro.

Estos sistemas conceptuales son un instrumento de categorización semántica. Las personas categorizamos para procesar la información y metabolizarla, tal y como se desprende de la siguiente cita de Jackendoff (1998: 165):

Un aspecto esencial de la cognición humana (quizá el esencial) es la capacidad de categorizar (...) Esto es lo que hace posible emplear la experiencia previa para guiar la interpretación de la experiencia nueva, porque sin la categorización la memoria es virtualmente inútil. Así explicar la capacidad de un organismo para la categorización no es sólo un asunto de la

semántica de las operaciones predicativas; es central para toda la psicología cognitiva.

De ahí se puede afirmar que un tesoro es una herramienta que asiste al humano en la gestión de conocimiento. A su vez,

(...) las clases y relaciones conceptuales son dos de los elementos esenciales de nuestro esquema cognitivo, tanto en lo que concierne a la realidad en general, como a la especializada (Feliu, 2001: 143).

Por eso, hay que tomar en consideración una tipificación de las mismas en la que basarnos y a la que referirnos para poder avanzar en la creación de asistentes informáticos que faciliten el acceso a la información y su organización según principios semánticos.

La aplicación de tesauros en línea permiten una contextualización conceptual del término, lo que significa que mediante sus términos y sus relaciones podemos entender mejor el concepto, como se muestra en la siguiente imagen con la captura del registro *dispositivos de almacenamiento* del tesoro multilingüe *Spines*, citado con anterioridad y que mostramos a continuación:

```

DISPOSITIVOS DE ALMACENAMIENTO
EN  STORAGE DEVICES
FR  MEMOIRES D'ORDINATEUR
LT  18 Proceso de datos. Informática
=   DISPOSITIVOS DE MEMORIA
=   MEMORIAS DE ORDENADOR
<   SOPORTE FISICO DE COMPUTADORES
.<  EQUIPO DE PROCESO DE DATOS
>   ACCESO ALEATORIO
>   ACCESO SECUENCIAL
>   CAPACIDAD DE ALMACENAMIENTO
>   CINTAS MAGNETICAS
>   CINTAS PERFORADAS
>   DISCOS MAGNETICOS
>   FICHAS MAGNETICAS
>   FICHAS PERFORADAS
>   LINEAS DE RETARDO (COMPUTADORES)
>   MEMORIA DE NUCLEOS
>   MEMORIA DE PELICULA DELGADA
>   MEMORIA DE PELICULA GRUESA
>   REGISTROS DE COMPUTADORES
>   TAMBORES MAGNETICOS
-   ALMACENAMIENTO DE DATOS
-   EQUIPO EN LINEA
-   EQUIPO FUERA DE LINEA
-   GRABACION DE DATOS
-   GRABACIONES LEGIBLES POR COMPUTADOR
-   GRABADORAS DE DATOS
-   NUCLEOS MAGNETICOS
-   PROCESADORES
-   PROCESO DE DATOS

```

**Fig. 32: Extracto de captura del tesoro SPINES**

Los tesauros ofrecen a los especialistas un conjunto estructurado de términos sobre la base de un sistema de conceptos aptos para la organización del conocimiento en las áreas de trabajo aplicadas.

*Google Scholar* (2004), la herramienta especializada en búsquedas académicas, como tesis doctorales o informes técnicos posibilita la búsqueda de literatura especializada acerca de cualquier tema. La captura de cualquiera de los artículos viene acompañada de información de indización correspondiente a los tesauros utilizados por *Google Scholar*, tal y como mostramos en la siguiente imagen:

↑ INDEX TERMS

**Primary Classification:**

H. [Information Systems](#)

↳ H.3 [INFORMATION STORAGE AND RETRIEVAL](#)

↳ H.3.1 [Content Analysis and Indexing](#)

↳ **Subjects:** [Thesauruses](#)

**Additional Classification:**

K. [Computing Milieux](#)

↳ K.1 [THE COMPUTER INDUSTRY](#)

↳ **Subjects:** [Standards](#)

**General Terms:**

[Design](#), [Management](#), [Reliability](#), [Standardization](#), [Theory](#)

**Fig. 33: Captura del tesoro utilizado por *Google Scholar***

Tal y como se puede comprobar, los términos de indización relacionados entre sí son una forma de traducir el conocimiento en elementos sintetizados creando un mapa conceptual que contribuye a la contextualización del conocimiento. Las relaciones semánticas proporcionan términos más genéricos, más específicos, así como una clasificación adicional que indizan el contenido de los textos buscados. De esta manera se intenta conectar la ecuación de búsqueda del usuario con los elementos de indización que identifican a cualquier documento recuperado. En este escenario los tesauros también ayudan en la normalización de los conceptos.

Y por último, son una fuente muy útil y confiable para la indización. Trabajar con un tesoro facilita una indización homogénea permitiendo mejorar la pertinencia y exhaustividad en la recuperación de documentos como, por ejemplo, se hace en centros de información o bibliotecas (Willpower Information, 1998).



### **2.6.2.2. Gestores documentales → Gestores de contenido**

Los gestores documentales son herramientas que ayudan a crear, almacenar, indizar, editar y recuperar documentos. Los gestores documentales proporcionan además una serie de funcionalidades que enunciamos a continuación:

- ❑ clasificación de documentos (clasificación e indización);
- ❑ bloqueo de documentos (niveles de acceso a los documentos mediante perfiles de usuarios predefinibles);
- ❑ control y acceso a las diferentes versiones de los documentos, mediante acceso al historial de versiones y mediante la marcación del estado del proyecto;
- ❑ búsqueda, indización y consulta de documentos de forma eficiente;
- ❑ control de entrada / salida en el proceso de flujo de trabajo: los documentos son reservados para actualizar por usuarios con permiso, evitando su modificación por otros hasta que el documento es de nuevo accesible en el espacio de trabajo;
- ❑ discusiones de documentos en línea que permiten a los usuarios comentar dudas, problemas o diversos aspectos de las tareas o documentos, en un ambiente colaborativo;
- ❑ acceso a documentos por varias personas en distintos entornos;
- ❑ recuperación organizada de la información del sistema;
- ❑ publicación de contenidos a través de una plataforma web;
- ❑ comunicación interactiva para el usuario: acceso restringido, personalización de la interfaz, publicación de comentarios, notificación de alertas por correo electrónico, SMS o Web, a la llegada de un documento, etc.;
- ❑ estadísticas de acceso e históricos de administración (trazabilidad).

Si además los gestores documentales están dotados de una base de conocimiento para gestionar los contenidos de los documentos, esto es, no solo para recuperar información, sino para extraer información de los textos, entonces a estos sistemas se les denomina gestores de contenido.

En el entorno de un sistema de gestión de contenidos, la aplicación de sistemas conceptuales como, por ejemplo, los tesauros es absolutamente necesaria para indizar el contenido a fin de poderlo recuperar y extraer.

Estos sistemas representan las infraestructuras intelectuales (conceptuales) del contenido para hacer posible el acceso y uso colectivo del contenido (Budin, 2004). Y esto es así, porque no se puede llevar a cabo la gestión de los contenidos de los documentos sin una organización previa y sistemática de los contenidos basadas en diversos criterios de clasificación como el dominio o materia, el grado de especialización del documento, los receptores del documento, recursos de representación (texto, imagen, etc.), canal de difusión (sitio web, correo electrónico, intranet, etc.) y el tipo textual.

La relevancia de estas funcionalidades se aprecia tanto o más en entornos de comunicación técnica, ya que la gestión de contenido basada en el nivel semántico se ha convertido en un proceso que complementa la comunicación.

Tal y como se puede deducir de la lista de funciones atribuidas a los gestores documentales, el flujo de los proyectos de traducción podría beneficiarse de estas bondades; es más, la gestión documental basada en el contenido podría repercutir considerablemente en la explotación de los recursos lingüísticos almacenados.

### **2.6.2.3. Gestores de proyectos de traducción**

Las funciones de las herramientas de gestión de proyectos de traducción pueden identificarse con la gestión de los parámetros que definen un proyecto —ubicación física, organización de carpetas de trabajo, recursos lingüísticos (memorias de traducción y diccionarios) y el conteo de las palabras, las equivalencias totales y aproximadas del material que se debe traducir—. Estas funciones se hallan integradas en los sistemas de traducción asistida de mayor difusión.

Sin embargo, hay una serie de tareas vinculadas a la gestión de proyectos de traducción a las que no nos asisten estos sistemas y que son críticas para reducir costes en un proyecto de traducción, tal y como enumeramos a continuación:

- ❑ preparación de un proyecto —presupuestos, costes del trabajo de traducción: tarifas de autónomos, tarifas para clientes, costes de gestión y oferta competitiva al cliente—;
- ❑ planificación del calendario del proyecto para gestionar la disponibilidad de traductores internos y autónomos, así como su perfil (combinación de lenguas, especialidad, experiencia);
- ❑ gestión de los recursos terminológicos y memorias de traducción propios y de los clientes;
- ❑ explotación de los recursos lingüísticos almacenados;
- ❑ coordinación de los equipos (internos y externos) de traducción y de control de calidad;
- ❑ control de calidad adicional: control ortotipográfico, maquetación, betatesting;
- ❑ comunicación con el cliente para la aceptación del presupuesto, el envío/recepción del material para traducir y traducido y el envío de la factura;

- comunicación de los equipos (internos y externos) de traducción y de control de calidad, envío / recepción del material para traducir y traducido recepción de la factura, comprobación de datos de factura.

En el entorno de la empresa de traducción, muchas de estas tareas se gestionan en bases de datos independientes; sin embargo, también existen en el mercado sistemas que ofrecen la gestión integrada de todos estos datos vinculados a cada proyecto, como es el caso de *LTC Organizer*.

Este sistema ofrece un servicio integral a casi todas las tareas mencionadas arriba. *LTC Organiser* es un sistema diseñado, modificado y mejorado desde su lanzamiento en 1998 con la primera versión, hasta la versión 5 de la actualidad.

Tiene una estructura modular específica para la gestión de clientes, para la gestión de los recursos humanos internos y externos, de gestión de proyectos (gestión de directorios, trayectoria de los proyectos de traducción y su estado mediante la gestión de directorios y no de documentos), de gestión de presupuestos y por último un módulo de facturación. Es decir, existen módulos para prácticamente todas las tareas enunciadas arriba. Además *LTC Organiser* cuenta con un módulo para generar informes de todo tipo y otro para gestionar y controlar el diario de tareas. En la siguiente captura de imagen mostramos la barra de estado de *LTC Organizer*.



**Fig. 34: Componentes de LTC Organizer**

Como se puede deducir por la diversidad de módulos, es una herramienta muy completa y a saber por las fuentes (Wassmer, 2003; Guzmán, 2003; Iverson, 2003), su empleo supone la reducción de los costes de gestión de los proyectos de traducción.

Sin embargo, esta herramienta no da ninguna solución para la gestión de los recursos terminológicos y memorias de traducción propios y de los clientes. Y mucho menos para su explotación. Hasta ahora esta función se viene desempeñando con ayuda de las herramientas de traducción asistida. Sin embargo, éstas gestionan los recursos lingüísticos por proyecto y carecen de las funciones básicas de gestión documental como la capacidad para subir / bajar archivos, control de versiones —algunas herramientas de gestión documental permiten hacer accesible nuevas versiones pero lo hacen siempre tratando esta nueva versión como un archivo más; se hace accesible de forma completa sin controlar los cambios—, capacidades para distribuir los documentos en múltiples repositorios, opciones de sincronización entre diferentes repositorios, posibilidad de búsquedas no sólo en los metadatos del archivo sino en el contenido del propio archivo.

### **2.6.3. Tendencias para la gestión integral del flujo total de la traducción**

Si bien los sistemas de traducción asistida han logrado un grado de sofisticación y automatización que difícilmente puede incrementar aún más el rendimiento del traductor, no sucede lo mismo con las tareas de gestión de proyectos de traducción, tal y como hemos explicado más arriba. Y eso, aun cuando el nivel de automatización se podría mejorar mucho más, debido a la repetitividad de las tareas de bajo nivel que el gestor tiene que llevar a cabo. Así, las herramientas de traducción asistida, las de logística de gestión de proyectos, de gestión (remota) de terminología y los sistemas de información abocados a la gestión del conocimiento deben sistematizar el seguimiento eficiente de proyectos (gestión de contenido), el control de calidad continuo y la gestión de las bases de datos de clientes y proveedores. En el siguiente diagrama apuntamos de forma gráfica el sistema de flujo integral de trabajo de traducción ideal:



**Fig. 35: Integración de la tecnología en flujos de trabajo de la traducción técnica**

Este entorno nos crea la necesidad de tratar de manera global y sistemática los distintos tipos de información generados durante la producción de traducción. A saber:

- información interna que se produce en el ejercicio de la gestión de traducción;
- información que proviene de los clientes y de otras fuentes externas;
- información pública que la organización quiere transmitir a su entorno: sitio web de Internet, intranet.

Según un estudio de las tendencias en la gestión de la información y el conocimiento en las organizaciones un sistema integral de flujo de trabajo:

debe basarse en las aportaciones de la archivística, la biblioteconomía y la documentación, convenientemente adecuadas a los nuevos entornos tecnológicos, y marcando un camino de integración imparable de todo tipo de información (Bustelo y García-Morales, 2001).

Por tanto, siguiendo estas tendencias, un sistema integral del flujo de traducción requiere un gestor de contenidos y de bases de datos para gestionar la información tipificada arriba. Para la gestión de la información y

conocimiento precisamos una herramienta de gestión de documentos y otra herramienta de gestión de bases de datos interconectadas dotadas de una base de conocimiento, es decir, sistema de indización, que permita el acceso a todos los documentos de los proyectos de traducción (archivos originales, archivos traducidos, diccionarios, guías de estilo, hoja de preguntas, datos administrativos, etc.). Tanto el sistema conceptual de indización como el diseño de las interfaces de búsqueda deben tener en cuenta al usuario. Estos dos elementos deben representar el conocimiento documentado para satisfacer las necesidades de los usuarios.

A vista de lo expuesto, prácticamente podemos concluir que la tendencia para la mejora de la gestión de proyectos de traducción es la del camino hacia la integración de las diferentes herramientas de trabajo en uno único que ofrezca todas las funcionalidades ya detalladas a través de un mismo software.

En este capítulo que concluimos hemos descrito el sector industrial GILT — contexto de nuestro trabajo—, identificando el ciclo de producción de documentación técnica multilingüe y adentrándonos en la complejidad de la fase de producción de proyectos de traducción.

Puesto que los objetivos de este trabajo es la identificación de las necesidades de gestión de los recursos lingüísticos, hemos dedicado sendos apartados a la descripción del tipo de recursos lingüísticos principales implicados en este entorno, su codificación, así como de las herramientas que asisten al usuario para llevar a cabo las tareas.

En este sentido, se ha visto que en el ámbito concreto de la traducción técnica se vienen obteniendo cada vez más beneficios de las herramientas de traducción asistida. Sin embargo, el incremento del rendimiento del traductor por encima de niveles actuales mediante estos sistemas que son, en realidad, motores de búsqueda muy sofisticados que capturan elementos para su posterior reutilización, se supone casi imposible. No obstante, la

producción de las traducciones puede aumentarse mediante una solución de apoyo a las tareas de gestión de proyectos de traducción, en general y, concretamente, a las tareas de gestión de documentos.

La aplicación de la gestión documental y específicamente la catalogación semántica pueden contribuir en la realización de importantes avances en la gestión de contenidos y en la sistematización de conocimiento de experiencias en función de las necesidades de los usuarios del sistema implementado.

En este capítulo teórico descriptivo, hemos alcanzado, pues, algunos de los objetivos que ya se habían descrito en el apartado 1.2. Objetivos y que recordamos aquí:

- qué tipo de recursos se almacenan, se reciclan y consultan y;
- qué herramientas nos asisten en el proceso de gestión de los recursos lingüísticos.

La aplicación del método auditor de análisis de los recursos lingüísticos abordará el resto de las cuestiones planteadas en el apartado mencionado, es decir, pretende detectar los flujos de almacenamiento, creación, retención, transferencia y (re)utilización de los recursos lingüísticos ya descritos. Asimismo, mediante este método se deben detectar las dificultades y cuellos de botella en los procesos identificados, así como las necesidades de los usuarios en un entorno concreto de trabajo.



### **3. Auditoría de los recursos lingüísticos multilingües en una empresa de traducción: estudio de caso**

En el presente capítulo se lleva a cabo un estudio de caso de una empresa de traducción, STAR SL. Con este fin, en primer lugar, atendemos a:

- los flujo de producción
- los agentes implicados
- el volumen de recursos lingüísticos gestionados en base a los proyectos de traducción

Para diagnosticar la utilización de los recursos lingüísticos, aplicamos una técnica de análisis de los datos recogidos denominada auditoría de la información. Tratamos pues los recursos lingüísticos como recursos de información a los que se les puede asociar conocimiento adicional a fin de optimizar su explotación.

Por tanto, se aborda la gestión de la información como identificativo de gestión de activos —tratados éstos también como activos de la entidad propietaria— y se define y describe la auditoría como un método de análisis de la gestión de recursos lingüísticos.

Aplicaremos la auditoría por fases siguiendo así las etapas que prevé la metodología auditora. Mediante la aplicación de una metodología auditora obtenemos respuestas acerca de:

- las necesidades e implicación de implantar una política de la información,
- cómo evaluar el proyecto de auditoría en cuanto a coste-beneficio,
- el desarrollo planificador de la auditoría de la información,
- cómo y con qué instrumentos se debe realizar el análisis e interpretación de los datos,
- qué propuestas conceptuales se deben modelar y qué propuestas técnicas se deben diseñar e implementar,
- cómo se debe hacer el seguimiento y el mantenimiento del sistema propuesto.

### **3.1. Estudio de caso: STAR Servicios Lingüísticos**

La empresa STAR SL creada en enero de 1999 y dedicada a la producción de documentación técnica multilingüe cuenta con una cartera de clientes saneada con un nivel de fidelización muy elevada.

La empresa está constituida por un equipo fluctuante de entre 20 y 24 empleados en plantilla. Éstos, a excepción de la ingeniera informática y la responsable de administración, están formados en diferentes universidades y países con titulaciones superiores en traducción y filología. Asimismo, en muchos casos, cuentan además con estudios de postgrado y máster en

diferentes disciplinas relacionadas con el sector como la traducción técnica, la tecnología aplicada a la traducción y la localización. Todos los integrantes del equipo son productores, consumidores, distribuidores y proveedores de información relacionada con la producción de documentación técnica multilingüe, ya bien en cuanto a los contenidos descritos, ya bien en cuanto a los procedimientos de trabajo para llevar a cabo las tareas.

Adicionalmente, la empresa cuenta con traductores autónomos probados a los que se envían los encargos de traducción. A diferencia de un traductor autónomo, la empresa de traducción en cuestión actúa de forma bidireccional. Esto significa que la empresa:

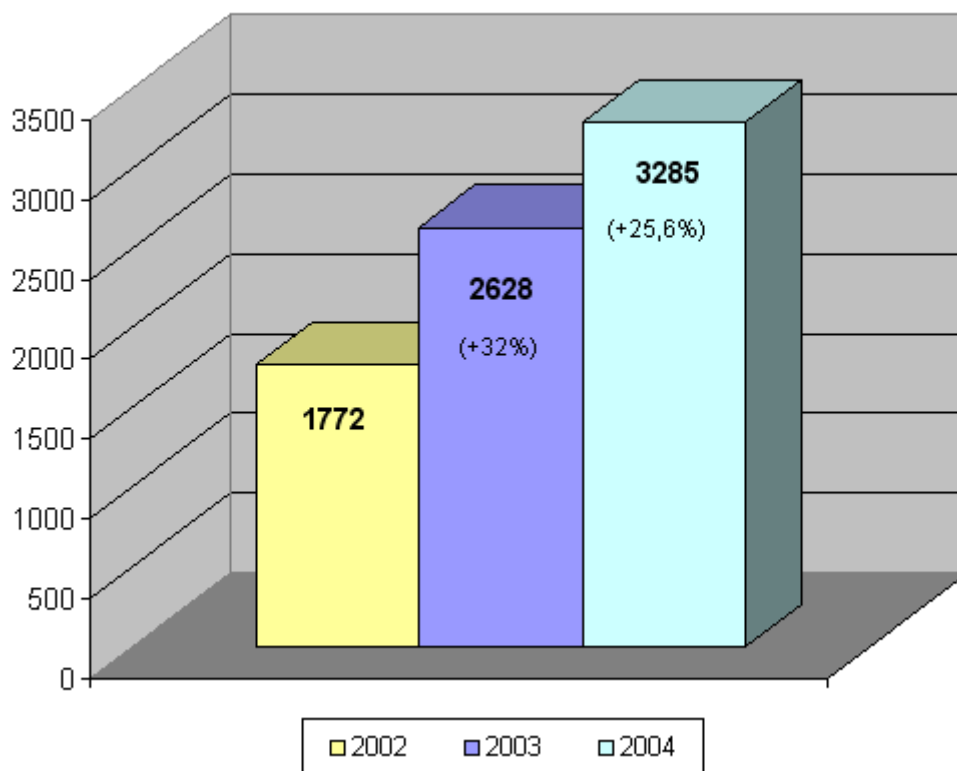
- recibe documentos de clientes,
- distribuye los documentos entre los traductores autónomos, volviéndolos a recibir una vez traducidos.

Las tareas de gestión de proyectos y de control de calidad se hacen dentro de la empresa.

El escenario compuesto por el equipo en plantilla y los traductores externos, es decir, los autónomos, responde a las exigencias del mercado. Las fluctuaciones a las que está sometido el mercado de la traducción obligan, en general, a las empresas de traducción a trabajar con un número límite de recursos humanos para poder afrontar los periodos donde disminuye el volumen de trabajo. La subcontratación de autónomos aporta flexibilidad a la empresa porque, por un lado, permite afrontar los periodos de carencia y, por otro, permite resolverse en la ejecución de un gran proyecto de traducción. La subcontratación de autónomos garantiza la reducción de los costes de infraestructura de la empresa, pues, de esa manera, se necesita menos espacio, menos estaciones de trabajo, equipos, licencias de software, etc. Y, en algunos casos, puede que incluso reduzca costes de la producción de la traducción.

Los flujos generales de trabajo que marcan la producción se definen a través de la identificación de la trayectoria de los proyectos de traducción, el almacenamiento de los documentos originales y los traducidos en cualquier formato, la gestión de las bases de datos de los clientes, por una parte, y de los traductores autónomos, por otra.

El número de clientes y de los proyectos gestionados en STAR SL va en aumento, tal y como muestra el siguiente gráfico:



**Fig. 36: Aumento progresivo del volumen de proyectos gestionados (2002-2004)**

La gestión de los proyectos de traducción se lleva a cabo principalmente mediante el sistema *Transit*, uno de los sistemas integrados para la traducción profesional, apuntados en el apartado 2.6. Herramientas de gestión de recursos lingüísticos para la documentación técnica multilingüe. Mediante esta herramienta se consigue profesionalizar, en un buen grado, la gestión de una gran parte de la información relevante e involucrada en la

gestión de los proyectos de documentación técnica multilingüe. Más adelante, al hacer el mapeo de los flujos de trabajo de los recursos lingüísticos, abordaremos con detalle el flujo de producción de las traducciones con *Transit*.

STAR SL conforma un ecosistema informacional —definido éste como un sistema de personas, prácticas, valores y tecnologías en una determinada localización física (Nardi y Vicki O'Day, 1998, citado en Cornellá, 2000a)—. En este ecosistema informacional específico todos los agentes que producen documentación multilingüe, esto es, los redactores, traductores, maquetadores, localizadores, gestores de proyectos de traducción y localización se convierten en gestores de información. Estas personas dedican cada día varias horas a crear y explotar inteligentemente toda la información que gestionan. En este sentido la sentencia de Cornellá (2000b: 34) en el sector que nos ocupa adquiere actualidad:

El mundo al que vamos es uno en el que todos seremos gestores de información; esta información será un componente esencial de cualquier trabajo y debemos aprender a manejarla de forma eficiente.

En STAR SL se produce información, se registra información, se comercia con información, se recicla información y además con una complejidad adicional, ya que se gestiona información que, al mismo tiempo, es susceptible de convertirse en el producto final que se entrega al cliente (Cámara, 2005c).

El producto final que se entrega al cliente es información resultante del conocimiento procesado en la misma lengua o en otras. La información textual que se usa, genera, recicla y modifica se compone, ante todo, de recursos lingüísticos del tipo memorias de traducción y diccionarios terminológicos. Estos recursos lingüísticos son gran parte de los activos informacionales constitutivos del capital de la empresa y por ello se debe planificar y organizar su gestión para mejorar su explotación.

La gestión de la información, en general, sigue obteniendo progresivamente un amplio reconocimiento como activo fundamental dentro del ámbito empresarial, como se desprende de la siguiente cita:

Las organizaciones que sobrevivirán serán aquellas que empleen los recursos de información como instrumentos para el ahorro de la fuerza de trabajo y la reducción de los costos. Solo competirán, de forma efectiva y exitosa, las que utilicen dichos recursos de manera estratégica para elaborar mejores productos, diseminarlos y acceder a ellos de forma creativa (Perdomo, citado en Valdés y Urra, 2000).

Boon (1990: 320) define la gestión de información como la gestión de un recurso cuyas funciones, así como sus flujos, se gestionan para mejorar los procesos de trabajo. Así, la gestión de la información se encarga de los recursos humanos, la información multimedia y los sistemas electrónicos que administran información. Es decir, gestiona la información asociada a todos los recursos que componen el entorno informacional de una organización.

Mason (citado en Roets y Boon, 1992: 138) identifica la gestión de la información como un proceso de planificación, organización, formación de empleados, coordinación y control de las diversas actividades incluyendo la descripción de los servicios, sistemas y formatos de información. Y atribuye como principal función de la gestión de la información lograr la mejor gestión empresarial.

La gestión de la información debe entenderse como el elemento clave para la coordinación económica, eficiente y efectiva de la producción. La gestión de la información es la respuesta para buscar procedimientos que mejoren las vías de acceso y los medios de control. Las pautas de gestión deben prever el almacenamiento, el control de acceso, la recuperación de la información y la diseminación desde fuentes internas y externas para la mejor marcha de la empresa (Best, citado en Fairer-Wessels, 1997: 95).

Los recursos de información son elementos claves para la organización empresarial. Asimismo, la gestión de la información utilizada en un contexto y atendiendo a las necesidades de las personas que usan esa información está abocada a la gestión del conocimiento (Harris, 1996).

Todas las tareas asociadas a la gestión de la información, así como las funciones atribuidas a la gestión de la información de los diferentes autores coinciden conceptualmente al tratar la información como un recurso estratégico que contribuye a lograr los objetivos empresariales. Esto significa desarrollar estrategias que maximizan el uso inteligente de la información — en otras palabras, con conocimiento— y minimizan el coste de producción, sin ir en detrimento de la calidad de lo producido, sino todo lo contrario.

Las colecciones de información almacenadas susceptibles a convertirse en activos para la empresa con el objetivo de crear conocimiento explotable y así mejorar los flujos de trabajo, dependen, sin embargo, de la metodología aplicada para su gestión.

### ***3.2. Auditoría de la información como método de análisis***

La información representa un recurso que debe gestionarse mediante una metodología con el fin de explotar el potencial estratégico de la propia información. Esta metodología se conoce como auditoría de la información (Buchanan y Gibb, 1998: 33) y la aplicamos aquí para analizar únicamente la explotación de los recursos lingüísticos en el proceso de producción de documentación técnica multilingüe. De esta manera pretendemos obtener las respuestas y también soluciones vinculadas a los objetivos del presente trabajo expuestos en el capítulo 1.Introducción y que recordamos a continuación:

- qué tipo de recursos lingüísticos se almacenan, se reciclan y consultan;

- de qué manera se llevan a cabo las tareas de gestión de los recursos lingüísticos (flujos de trabajo);
- de qué manera se organizan a nivel individual y corporativo (perspectiva del usuario y de la organización);
- qué herramientas nos asisten en el proceso de su almacenamiento y selección, cuáles son sus límites y posibilidades de refinamiento.

Si bien, al contextualizar el sector industrial, el tipo de producción, la relevancia de los recursos empleados y las herramientas de gestión de recursos lingüísticos —en el capítulo 2. La documentación técnica multilingüe. Recursos lingüísticos y técnicos implicados—, ya hemos abordado de forma genérica algunas de estas cuestiones; ahora, mediante la aplicación de instrumentos para auditar la información, pretendemos evaluar más detalladamente los elementos de información que inciden en la gestión de los recursos lingüísticos con el fin de poder derivar las propuestas de optimización.

La metodología que empleamos para el análisis de los recursos informacionales, como se ha dicho, se denomina auditoría de la información y se define como:

Proceso de identificación y evaluación de los recursos de la información necesarios para cumplir con los objetivos de la empresa (Soy, 2003b: 10).

Una auditoría de la información asegura que los recursos de información se gestionan y explotan convenientemente, esto es, se analizan en función de su utilidad, funcionamiento y valor. Por tanto, el papel de una auditoría de la información es proporcionar un método para el proceso de monitorización y evaluación de los flujos de trabajo de una entidad con el fin de trazar propuestas de mejora.

Las técnicas de auditoría de la información son una excelente herramienta para evaluar la gestión de la información, y así poder establecer los



requisitos conceptuales y técnicos para el desarrollo de sistemas de información abocados a la gestión del conocimiento (Martín, 1998).

Así, la aplicación de auditorías de la información está, por una parte, vinculada a la tendencia de aplicar procedimiento de control de calidad y, por otra parte, al desarrollo de instrumentos para la gestión del conocimiento — analizar, desarrollar y aplicar mecanismos para fomentar el conocimiento generado a nivel individual y de grupo—, tal y como se deduce de la siguiente afirmación de Eíto (2003: 302):

La gestión del conocimiento ha permitido recuperar unos conceptos y prácticas procedentes de la gestión de la información tradicional, entre ellas las auditorías de la información.

Por tanto, si las auditorías de la información son parte del entramado necesario para diseñar un programa de gestión de conocimiento (Bukowitz y Williams, 1999: 6), de la aplicación de esta metodología a la gestión de los recursos lingüísticos y la información asociada a los recursos lingüísticos, se pueden derivar soluciones para la automatización de esa información gestionada de forma inteligente. Esto significa diseñar y desarrollar propuestas para automatizar la gestión de conocimiento.

### **3.2.1. Fases del método auditor**

Soy (59-100), en su monografía acerca de la auditoría de la información ya comentada en el apartado 1.7. Metodología, identifica las fases y los instrumentos y técnicas para llevar a cabo una auditoría de la información, realizando una síntesis genérica e ideal de las fases de una auditoría de la información en base a propuestas anteriores de otros autores. En los siguientes subapartados se explican la función de cada una de estas fases y de las tareas implicadas correspondientes.

### ***3.2.1.1. Implicación y decisión por parte de la dirección de la entidad en implantar una política de la información***

Cualquier organización que desee mejorar la gestión de sus procesos de trabajo, necesita de una política para su gestión. Vinculada a la inversión tecnológica en procesos de informatización, los orígenes de las políticas de la información datan de hace más de dos décadas. Si bien una auditoría de este tipo, de alguna manera puede considerarse como una política de la información dentro de una organización, existen diferencias con respecto a las políticas de la información propiamente dichas.

La gran diferencia estriba en que la política de la informatización está avalada, sobre todo, por la inversión de tecnología y técnicos informáticos, mientras que la aplicación de una política de la información supone considerar de antemano la gestión de los recursos informacionales como factor de cambio, competitividad y mejora permanente para la entidad.

Por tanto, los recursos informacionales que se gestionan se convierten en un activo estratégico para la empresa y, como tal, merece ser gestionada como cualquier otro activo e integrarse en su proceso de planificación (Cornellá, 1994: 71). De esta manera, si se entienden los recursos lingüísticos como activos informacionales, es decir, como parte del capital de la empresa, de la mejora de su gestión, se puede mejorar su explotación.

La decisión de aplicar una política de la información debe tomarla la dirección de la empresa fomentada por la necesidad de explotar la información almacenada, mejorar los procesos de trabajo analizando los procesos estructurados establecidos.

### **3.2.1.2. Identificación de una necesidad y evaluación de la viabilidad del proyecto**

En esta fase se debe identificar una necesidad genérica y definir los objetivos y beneficios a corto y largo plazo de la aplicación e implementación de un proceso auditor de la información. Asimismo, se deben prever los costes y los beneficios que previsiblemente supondrá la implementación de nuevas soluciones. Esto quiere decir, que se debe prever el retorno de la inversión, abreviado ROI, del concepto anglosajón *return of investment*. Tradicionalmente, este concepto está basado en la recuperación de costes, desde un punto de vista centrado en el ámbito financiero.

Las áreas tradicionales en las que se ha desarrollado esta actividad relacionada con el cálculo del ROI han sido los recursos tangibles, es decir, aquellos que se pueden medir, como las cifras de ventas, reducción de costes, etc.

La ecuación: *inversión X, supone Y beneficios* no puede, sin embargo, aplicarse numéricamente en la previsión de coste-beneficio en un proyecto de auditoría de la información, debido a su intangibilidad. No obstante, algunos estudios realizados entre ejecutivos y empleados de algunas empresas y analizados por Ward (2003) sobre qué beneficios aporta la aplicación de una auditoría relacionada con la productividad y los recursos intangibles ofrecen muy buenos resultados. Los intangibles son aquellos recursos, experiencia y capacidades que existen en una entidad y que no están valorados desde un punto de vista contable y no aparecen ni en el balance, ni en la cuenta de resultados de la organización. Sin embargo, tras el análisis de los intangibles como la experiencia, la comunicación y la gestión de contenidos se pueden sistematizar algunas informaciones estandarizadas para registrar este conocimiento para su reutilización.

Los sistemas de información deben prever canales de información donde se facilite a los usuarios la mayor amigabilidad y agilidad para la comunicación,

el ahorro de tiempo, la mayor consistencia y fiabilidad de la información, la reducción de procesos manuales o las mejoras en la accesibilidad, disponibilidad, publicación e intercambio de información. Esto es, aparentemente positivo, pero la cuestión es de qué instrumentos disponemos para poder medir el beneficio de la inversión.

Está demostrado que el análisis ROI no es una ciencia exacta, por ello la falta de herramientas de medición y la carencia de unidades de medida. No existe un algoritmo para estimar el valor tangible de un sistema de gestión de intangibles. De hecho, es imposible medir el valor exacto de los resultados obtenidos tras auditar la información e implementar las propuestas derivadas de la auditoría. Algunos incluso opinan que hay que olvidarse de este tipo de cálculos ya que su estimación no conduce a nada (Computerworld, 2003). Sin embargo, si necesitamos hacerlo, ¿cómo podemos justificar desde un punto de vista económico la inversión en un sistema de información fruto de una auditoría de la información?

La valoración para tomar la decisión de invertir en la implementación de un sistema, debe estar motivada por criterios de la viabilidad de su desarrollo, el ahorro de costes y el incremento de beneficios tras su aplicación. En este caso, los beneficios son difíciles de determinar numéricamente, por eso, es fundamental considerar, a priori, como un valor empresarial los siguientes aspectos:

- la comunicación colaborativa,
- la gestión de activos intelectuales basados en la experiencia y
- la gestión de conocimiento basada en la toma de decisiones.

Según un artículo de Surmacz (2003), la mayoría de directivos responsables de la implantación de auditorías de la información no realizan un análisis formal del ROI que justifique la realización de inversiones. Y esto es, porque reconocen que uno de los objetivos clave perseguidos es lograr los beneficios de la gestión de intangibles. Según este artículo, los propios

directivos ofrecen prioridad a la gestión de la información y del conocimiento y están convencidos de la necesidad de invertir en sistemas basados en la comunicación y gestión de recursos de conocimiento para que el sistema resultante tras la auditoría de la información, logre la optimización de cualquier flujo de trabajo.

### **3.2.1.3. Planificación de la auditoría de la información**

La planificación, así como la evaluación de todas las fases de la auditoría, la deben llevar a cabo profesionales de la información que sean capaces de realizar las siguientes tareas:

- ❑ analizar los flujos de trabajo,
- ❑ detectar los cuellos de botella en los flujos de trabajo,
- ❑ recoger los problemas desde diferentes ópticas de los usuarios,
- ❑ canalizar sugerencias de los usuarios para la optimización de los sistemas de información,
- ❑ trazar propuestas de mejora,
- ❑ fomentar y sistematizar la gestión del conocimiento,
- ❑ identificar el conocimiento tácito y representarlo para hacerlo explícito mediante su registro en los sistemas de información,
- ❑ diseñar sistemas de información de acuerdo a las necesidades del entorno,
- ❑ definir las entidades básicas de los sistemas de información.

Ciertamente, para llevar a cabo las tareas que acabamos de mencionar, se requiere un profesional de la información *no improvisado* que diseñe, supervise y mantenga una política de información. El perfil del profesional ideal requerido coincide con las destrezas tipificadas por la SLA (*Special Libraries Association*), traducido por Doc6 (2004) acerca del profesional de la información y que recordamos a continuación:

### El profesional de la información...

- ❑ tiene un profundo conocimiento del contenido de los recursos de información, incluyendo la capacidad de evaluarlos y filtrarlos críticamente;
- ❑ conoce los temas en los que trabaja su organización o su cliente;
- ❑ desarrolla y gestiona servicios de información adecuados, accesibles y eficientes en consonancia con la dirección estratégica de la organización;
- ❑ facilita formación y soporte a los usuarios del servicio de información;
- ❑ evalúa las necesidades y diseña y promociona servicios y productos de información con valor añadido para satisfacer las necesidades;
- ❑ utiliza la información tecnológica apropiada para adquirir, organizar y difundir la información;
- ❑ utiliza los medios y las técnicas de gestión apropiadas para transmitir a la dirección la importancia de los servicios de información;
- ❑ desarrolla productos de información especializados para el uso interno o externo de la organización o para usuarios individuales;
- ❑ evalúa los resultados del uso de la información e investiga para solucionar los problemas de gestión de la información;
- ❑ mejora continuamente los servicios de información en respuesta a las necesidades cambiantes;
- ❑ es un miembro efectivo del grupo directivo y un consultor para la organización en temas de información.

El profesional de la información que planifica una auditoría en una entidad debe decidir qué métodos y técnicas empleará para la recogida de información y para el análisis de esa información recogida.

La delimitación a un fenómeno concreto y su selección por áreas supone minimizar los posibles riesgos vinculados a la aplicación de un método auditor. Asimismo, se debe planificar la aplicación de los resultados de la auditoría por fases para reducir riesgos como, por ejemplo, el impacto que produce un cambio de procesos, flujos de trabajo y sistemas para su gestión entre los integrantes del equipo de la entidad.

De la fase de planificación y de la gestión del profesional de la información depende en gran parte el éxito del progreso de la auditoría de la información que continúa con la fase que describimos a continuación.

#### ***3.2.1.4. Análisis e interpretación de los datos para modelar un catálogo de propuestas técnicas y conceptuales***

En esta etapa comienza el proceso auditor propiamente dicho. Se aplican los instrumentos para la recogida de toda la información precisada. Los instrumentos utilizados para recoger información en el marco de un proyecto de auditoría de la información son:

- ❑ catálogo o inventario de los recursos de información almacenados;
- ❑ diagramas o mapas de flujos de información aplicados a procesos de trabajo;
- ❑ encuestas entre los usuarios acerca de los sistemas de información existentes y las necesidades reales de información de los usuarios. Esto significa que los usuarios son las personas clave para obtener esta perspectiva<sup>1</sup>;
- ❑ observación de los cuellos de botella en los flujos de información y del uso y utilidad de la información.

---

<sup>1</sup> Para más información sobre la metodología utilizada para la recogida de información relacionada con las necesidades de los usuarios se puede consultar Sanz, 1996: 195-197

A partir del análisis de los datos recogidos se debe confeccionar un protocolo de problemas e incidencias en el que basar después el protocolo de propuestas.

### **3.2.1.5. Implementación de las propuestas técnicas y conceptuales**

Tras acabar el análisis e interpretación de los datos comienza la fase de implementación de las propuestas conceptuales en las que se basa el sistema de gestión de los recursos de información basada en conocimiento.

La implementación de las propuestas envuelve:

- ❑ la redacción de un protocolo incorporando las necesidades del usuario;
- ❑ la documentación de los nuevos procedimientos de trabajo propuestos, si fuera el caso;
- ❑ la diferenciación por niveles de los usuarios de la documentación propuesta, por ejemplo, orientada a los desarrolladores del sistema y orientada al usuario del sistema propuesto;
- ❑ la publicación y difusión de los nuevos procedimientos de trabajo;
- ❑ las especificaciones técnicas del sistema propuesto: modelo de almacenamiento de datos, lenguajes de programación, lenguajes de interrogación y de diseño de las interfaces de interacción persona-ordenador.



### **3.2.1.6. Seguimiento y mantenimiento**

La etapa de seguimiento comienza tras el desarrollo de las propuestas conceptuales que se traducen en protocolos de trabajo y de las propuestas técnicas de las que se derivan los nuevos sistemas de información. Estos desarrollos deben ser probados por grupos reducidos de usuarios que detecten la usabilidad de lo propuesto. La usabilidad es la percepción de que el proceso de llevar a cabo una tarea dentro de un sistema es consistente, organizado, eficiente, productivo, fácil de usar, intuitivo y específico para su entorno de aplicación. La usabilidad es uno de los criterios definidos en el capítulo 11 de la norma ISO 9241 (1998: 11) para el diseño de interfaces ergonómicas y eficientes.

Los sistemas propuestos deben seguir un modelo basado en las necesidades del usuario y, más aún, se debe hacer un seguimiento de la experiencia de su uso. Y esto es, porque el verdadero resultado de la usabilidad del sistema se conoce cuando el usuario lo emplea. En esta fase también se determina quiénes deben hacer el mantenimiento del sistema y qué tareas pueden realizar en función de su perfil (niveles de usuario).

### **3.3. Auditoría de los recursos lingüísticos multilingües: almacenamiento y explotación**

El modelo de auditoría dividido en fases que se han expuesto en el apartado anterior, lo adaptamos y flexibilizamos en función de nuestro entorno de aplicación y los objetivos marcados para este proyecto de auditoría de la información. Con esta auditoría perseguimos exponer:

- ❑ las necesidades en las rutinas de trabajo y
- ❑ las carencias de los sistemas de información vinculados a la gestión de los recursos lingüísticos.

Así, el componente empírico del modelo auditor es la aplicación de alcance parcial de una metodología auditora preestablecida. Y es parcial porque se define y aplica solo en la gestión de los recursos lingüísticos y las informaciones asociadas a la creación, retención, transferencia y reutilización de esos recursos lingüísticos.

Se persigue que, al concluir el proyecto auditor, se hayan detectado las razones que impiden una óptima gestión de los recursos lingüísticos y trazar propuestas basadas en el análisis de los datos recogidos para proponer mejoras de los flujos de trabajo.

### **3.3.1. Implementación y aplicación de una política de la información**

La implicación por parte de la dirección de la entidad en implantar una política de la información está, según nuestra interpretación en el caso que nos ocupa, estrechamente vinculada a la identificación previa de una necesidad de mejora de algún procedimiento de trabajo y la evaluación de la viabilidad para implementar un proyecto auditor con objetivos específicos.

Desde la dirección se apoya el objetivo del proyecto auditor de controlar el ciclo de las memorias de traducción y diccionarios terminológicos para la mejora de su explotación. Se prevé que la aplicación de los resultados de la auditoría revocará en beneficios que, si bien no son tangibles, por no ser contables, son activos empresariales que redundan en la mejora de la comunicación de la información de la experiencia personal, con repercusiones directas en la gestión del conocimiento.

La aplicación de esta política de la información pretende asegurar el diseño y desarrollo de un modelo de gestión total del entorno preconcebido con la perspectiva de:

- reducir el almacenamiento de recursos repetidos,

- mejorar la recuperación de los recursos para su posterior explotación,
- avanzar en la automatización del trabajo de mantenimiento y actualización de los recursos,
- optimizar el proceso industrial de gestión de documentación técnica.

### **3.3.2. Planificación del proyecto auditor**

Quien suscribe este trabajo es la responsable del proyecto de la auditoría que nos ocupa. Además de compactas habilidades asociadas a las actividades del profesional de la información, contamos con conocimientos teóricos y aplicados de lingüística, traducción, documentación e informática aplicada. Esta interdisciplinaridad nos permite:

- obtener una panorámica general del origen de los problemas relacionados con los flujos de información existentes;
- determinar las áreas de análisis relevantes relacionadas con la gestión de los recursos lingüísticos del entorno de la producción de documentación técnica multilingüe;
- delimitar el fenómeno concreto vinculado a la explotación de los recursos lingüísticos con el objetivo de concretar una propuesta conceptual que tenga una implementación técnica;
- interpretar las necesidades de los usuarios en su contexto.

El fundamento de partida es la recogida de las informaciones relacionadas con la gestión de los recursos lingüísticos utilizando inventarios, diagramas y mapas de flujos, catálogo de incidencias, cuestionarios para detectar necesidades o mejoras de información, así como, mediante la observación de la explotación de la información utilizada. Estas tareas se realizan en la fase de análisis e interpretación de los datos.

### 3.3.3. Análisis e interpretación de los datos

Durante esta fase el objetivo es modelar un catálogo de propuestas técnicas y conceptuales basados en la observación e interpretación de los datos que previamente vamos a recoger. La propuesta conceptual debe estar avalada por el análisis de la información relacionada con los recursos lingüísticos y su gestión. Para ello se deben llevar a cabo tareas utilizando instrumentos metodológicos concretos como los que proponemos a continuación:

- confeccionar un inventario de los recursos que vamos a auditar;
- representar mediante mapas y diagramas los flujos de la información;
- identificar y catalogar los problemas que afectan a los flujos de información mediante un protocolo de identificación de los cuellos de botella que afectan a esos flujos;
- recoger las necesidades de los usuarios mediante encuestas relacionadas con las herramientas de gestión de los recursos lingüísticos disponibles —los informantes son los usuarios que abordan las cuestiones desde su perspectiva— y mediante entrevistas personales;
- identificar la información utilizada, la infrautilizada o la no utilizada mediante la observación de los flujos de información.

### 3.3.3.1. Análisis de los recursos lingüísticos almacenados

En esta primera fase de la auditoría recogemos información mediante la recopilación de los datos. Según Robertson (1994:31) hasta la forma más sencilla de auditar información supone la aplicación de un proceso de recogida de información muchas veces limitada al mero inventario de los recursos informacionales. Las formas más sofisticadas de auditar incluyen diagramas y mapas de la información. Estos instrumentos deben servir para la identificación de la trayectoria de la información y la incorporación de la perspectiva del usuario, como veremos más adelante, tras hacer el inventario de los recursos lingüísticos almacenados.

En la siguiente tabla se muestra el volumen en cifras absolutas y la progresión de las memorias de traducción almacenadas:

Fecha	Bytes	Estimación palabras
2004-05-27	17185072884	38445353
2004-09-14	20850213785	46644774
2005-01-20	25570269950	57204183
2005-01-23	25706628140	57509235
2005-01-30	26053237957	58284649

**Fig. 37: Evolución de las memorias de traducción almacenadas**

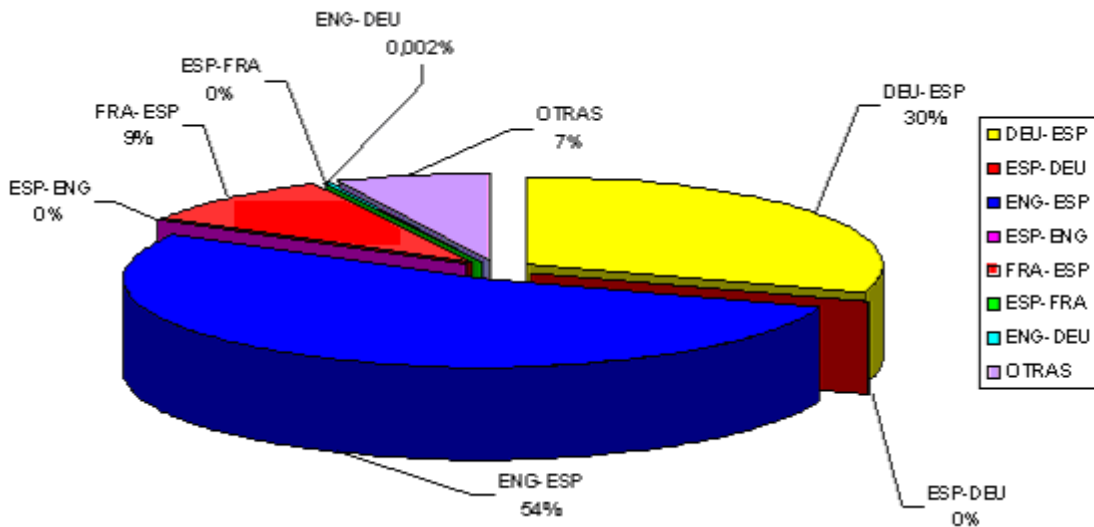
Asimismo, la siguiente tabla muestra el volumen en cifras absolutas y la progresión de los registros de los diccionarios terminológicos almacenados:

Fecha	Bytes	Estimación registros
2004-05-27	4577401959	6492769
2004-09-14	4856997673	6889358
2005-01-20	5200772871	7376983
2005-01-23	9320674486	13220815
2005-01-30	9348344004	13260062

**Fig. 38: Evolución de registros de los diccionarios almacenados**

Todos estos recursos lingüísticos se manipulan, traducen, editan, recuperan, extraen y reciclan.

El porcentaje del número total de los recursos repartidos por lenguas gestionadas se deriva del siguiente gráfico:



**Fig. 39: Porcentaje del nº total de recursos según combinaciones lingüísticas**

Como se puede deducir de los datos, todos estos recursos lingüísticos almacenados que se han inventariado crecen interrumidamente. No obstante, el análisis de los mismos muestra que, muchos de los diccionarios se hallan duplicados bajo nombres diferentes. Esto significa que

previsiblemente el volumen de los datos de los diccionarios inventariados se verá parcialmente reducido, tras eliminar los elementos duplicados.

### 3.3.3.1.1. El corpus paralelo de STAR SL

Concretamente, los recursos tipo memorias de traducción conforman un almacén de recursos lingüísticos compuestos de textos originales en una lengua alineados con su traducción en otra, es decir, se trata de un corpus paralelo, tal y como hemos expuesto en el apartado 2.4. Gestión y tipo de recursos lingüísticos multilingües. En el informe EAGLES, ya citado en dicho capítulo, se enumeran los parámetros que deben cumplir los textos almacenados electrónicamente para que puedan considerarse corpus. Según el informe los criterios de tipologización de los corpus se relacionan con:

- ❑ la distribución por combinación de lenguas y volumen,
- ❑ la especificidad representada en sus contenidos,
- ❑ la codificación y anotación de los textos almacenados,
- ❑ la documentación que acompaña a los textos almacenados.

Como ya hemos adelantado, el almacén de textos se compone de textos distribuidos en la lengua de salida con sus respectivas traducciones a una o más lenguas. La alimentación con textos y su traducción a una o más de una lengua hace que el corpus de STAR SL sea un *corpus paralelo multilingüe*.

El volumen del corpus va creciendo en función del progreso de los proyectos gestionados, es decir, que se actualiza permanentemente, por eso y según los criterios del informe EAGLES, se clasifica como un *corpus dinámico* o *corpus extenso*. La extensión del corpus no está, por tanto, delimitada. El corpus recoge un número elevado de palabras, —más de 50 millones (el British National Corpus se compone de 100 millones de palabras), como se ha visto en la tabla— sin discriminar su representatividad, ya que se parte de la base de que todo el material procesado en el marco de un proyecto de

traducción es representativo en cuanto a que es susceptible de volverse a utilizar.

Los contenidos de los textos del corpus son de tipo técnico, ya que se generan en el entorno de la documentación técnica multilingüe que se produce en STAR SL. En este sentido, y continuando con la tipología de los corpus descrita en EAGLES, nuestro corpus es un *corpus especializado*.

En cuanto a la codificación y anotación de los corpus estos pueden ser anotados o no anotados. Siguiendo la clasificación de Abaitua (2000), las anotaciones posibles pueden ser de diferentes tipos, como enunciamos a continuación:

- *extralingüístico*: lengua, fecha de creación y de modificación, longitud del texto (en caracteres, palabras, oraciones, párrafos), autor, traductor, edición, dominio y, en general, toda aquella información que pueda ser relevante para la catalogación;
- *ortotipográfico*: códigos de caracteres de las lenguas y protocolos de intercambio, etiquetas de maquetación y formato;
- *lingüístico* de tipo *textual o estructural*: elementos que componen un texto en cuanto a su macroestructura: párrafos, segmentos, niveles de títulos o encabezados, etc. En este tipo de anotaciones estructurales cabrían los códigos de correspondencia que asocian las unidades de traducción con sus equivalencias alineadas. Y la anotación *lingüística* de tipo *morfosintáctico*: categoría morfosintáctica de cada unidad léxica, otras propiedades morfológicas generalmente asociadas a la flexión (género, número, persona, caso, tiempo, asociación con forma canónica, etc.), categorías sintagmáticas de las oraciones: grupos verbales y nominales, cláusulas subordinadas, etc.

Según la tipología estándar de los corpus, la documentación que acompaña a los textos almacenados es otro de los criterios clasificadores de los corpus. Un corpus documentado es aquel en el que cada documento textual lleva



asociado un archivo del tipo asiento bibliográfico o ficha documental con informaciones en diferentes formatos como, por ejemplo, incluido en el mismo documento que se cataloga como cabecera de las propiedades que describen el documento, o en una DTD (*Definición de tipo de documento*) específico —en el apartado 2.5. Formatos de codificación e intercambio se puede consultar la morfología y aplicabilidad de este tipo de documentos—. Las informaciones asociadas a un documento coinciden con la mayoría de las anotaciones clasificadas como extralingüísticas.

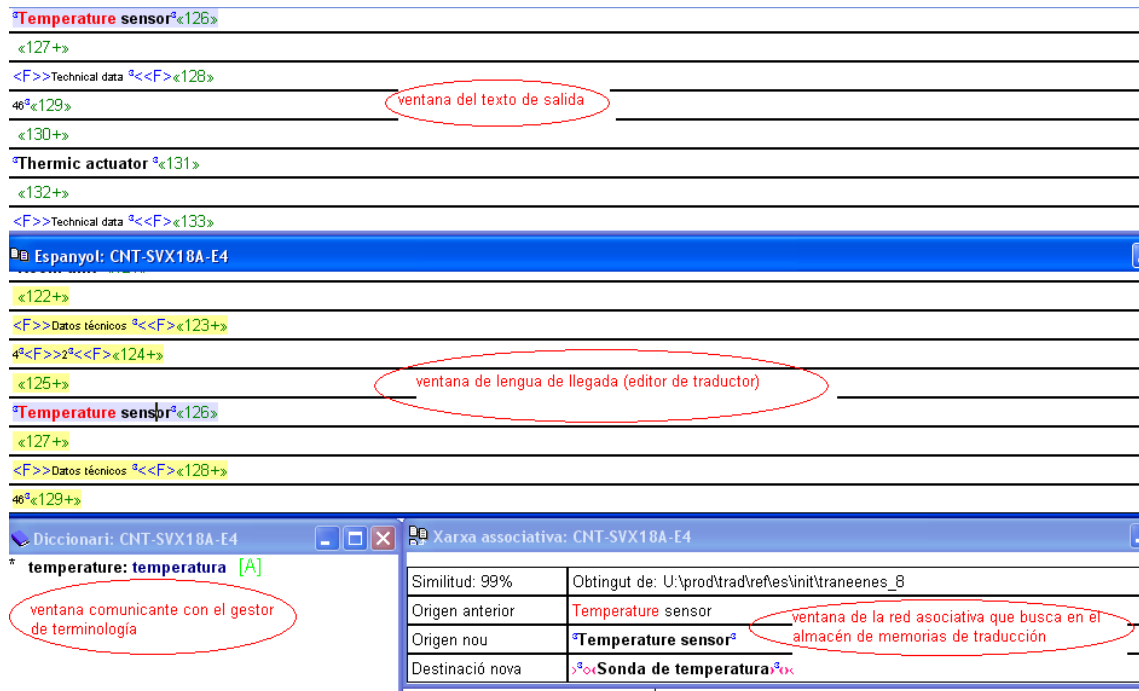
#### **3.3.3.1.2. Gestión del corpus paralelo con *Transit***

Las anotaciones de las que dispone nuestro corpus paralelo son las heredadas por el sistema de traducción asistida empleado para generar las memorias de traducción. Así, cuando las memorias de traducción se crean en el entorno de *Transit*, el sistema las anota de forma automatizada. Éstas conforman el *corpus paralelo anotado*.

En general, cuando el corpus paralelo se crea en el entorno de un sistema de traducción asistida, el sistema anota de forma automatizada las memorias de traducción. Las herramientas de traducción dotan a las memorias de traducción, es decir, a los corpus paralelos de anotaciones extralingüísticas: lengua, fecha de creación y de modificación, longitud del texto (en caracteres, palabras, líneas, párrafos), autor, traductor, edición y anotaciones estructurales de códigos de correspondencia que asocian las unidades de traducción con sus equivalencias alineadas.

Mientras estamos trabajando en el entorno de *Transit*, el sistema tiene activas cinco ventanas —la del texto de salida, la del editor de traducción donde se redacta el texto de la lengua de llegada, la del diccionario terminológico asignado, la del bloc de notas para incorporar comentarios asociados a una unidad de traducción y la ventana de la red asociativa— haciéndolas interactuar buscando en las memorias de traducción asignadas a ese proyecto sugerencias de traducción, específicas a una unidad de

traducción, como se muestra a continuación, con la captura de pantalla del entorno descrito:



**Fig. 40: Interfaz del entorno de trabajo de Transit**

En la ventana de la red asociativa —en terminología específica de *Transit* corresponde a la interfaz de búsqueda que explora las memorias de traducción—, el sistema nos proporciona anotaciones al recuperar de la memoria propuestas, como se muestra a continuación:

Similitud: 73%	Recuperado de: J:\Alden\6911\trwork\Troubleshooting
Origen anterior	Some of the common causes of a Leak alert include:
Origen nuevo	Some of the common causes of a No Movement alert include:
Destino nuevo	Algunas de las causas más comunes de las alertas de fugas son:

**Fig. 41: Interfaz de la red asociativa de Transit con información de ruta definida por defecto**

Como se puede ver, la sugerencia que nos ofrece el sistema para una nueva traducción contiene informaciones relacionadas con:

- el nombre del archivo de donde se extrae la sugerencia,

- la ubicación del archivo y
- el porcentaje de coincidencia del segmento recuperado.

Más allá de la información no lingüística que acompaña a las propuestas de traducción recuperadas, también se puede extraer información adicional de los archivos específicos donde se guarda la configuración de cada proyecto de traducción en el entorno de *Transit*.

En este entorno de *Transit* cada proyecto de traducción queda definido mediante diferentes formas de parametrización. La información de los archivos que contienen estos documentos definen el proyecto de traducción. Estos archivos específicos son para el entorno de *Transit* los \*.prj.

En la siguiente captura de pantalla se puede ver un archivo \*.prj abierto con el *Bloc de notas*, por tanto, con un formato .txt de fácil manipulación:

```

prj_bloc de notas - Bloc de notas
Archivo Edición Formato Ver Ayuda
[Languages]
Source=ENG
Destination=SPA

[Admin]
ProjectName=Tiles
Customer=UPF2
Username=lidia cámara

[FileType]
GUID={91409CA0-910D-11D2-A4DA-00104B026E97}

[Files]
File1=1|1|1|0|1|C:\Archivos de programa\Transit XV\
Projects\Tiles\Tiles2.doc|Tiles2|Tiles2|0

[References]
File1=1|1|1|0|1|C:\Archivos de programa\Transit XV\
PROJECTS\Tiles\REF\Tiles1_ENG.doc|Baldosas_ENG|Baldosas_ENG

[Dictionaries]
Dict1=Materiales//TermStar
Dict2=tiles//tiles
Active=2

```

**Fig. 42: Archivo \*.prj abierto con el *Bloc de notas***

En el apartado de las anotaciones con información sobre las lenguas [*Languages*] encontramos la combinación de lenguas asociadas a toda la

información textual que comprenden los documentos relativos a un proyecto; en las anotaciones con información administrativa [*Admin*] se encuentran las informaciones extralingüísticas relacionadas con el nombre del proyecto (*ProjectName*), cliente (*Customer*) y gestor de proyectos (*UserName*). La información ortotipográfica viene dada por el tipo de archivo o formato del documento original [*FileType*]. Las informaciones descriptivas acerca de la ubicación física de los documentos almacenados relacionados con los documentos que sirven de base para generar las nuevas memorias de traducción se encuentran en [*Files*] y los documentos asociados, que son memorias de traducción preseleccionadas para reutilizarlas en el progreso de la traducción están en [*References*]. Y por último, obtenemos información acerca del nombre de los diccionarios y su base de datos correspondiente que se han creado o utilizado para el proyecto en cuestión bajo la etiqueta [*Dictionaries*]. A su vez podemos saber qué diccionario se ha seleccionado como activo (*Active*), esto es, el diccionario que por defecto será utilizado para incorporar automáticamente nuevos términos en el proceso de traducción o creación de memorias de traducción.

Esta información asociada a los documentos que componen el corpus paralelo en el entorno de *Transit* es información relativa a dichos documentos. Lo que significa que cada documento lleva asociado un archivo o una cabecera descriptiva con las características que ya hemos apuntado. En este sentido, podemos afirmar que nuestro corpus puede calificarse como un corpus documentado.

Sin embargo, esta información de los \*.prj no se utiliza para buscar recursos lingüísticos (memorias de traducción y diccionarios) en todo el corpus de textos almacenados. Por tanto, esta *catalogación* no es tal, porque los recursos almacenados aunque cuenten con información documental en los \*.prj, no están previstos para ser explotados como unidades documentales normalizadas que especifiquen los datos relacionados con su forma y contenido para facilitar la identificación y ubicación, en una palabra, su acceso. Por el contrario, el acceso a las memorias de traducción, es decir, al

corpus paralelo, se hace a través de su almacenamiento físico, tema que abordamos en el apartado 3.3.3.2.1. Flujo de almacenamiento y acceso físico a los documentos.

Resumiendo, los recursos lingüísticos inventariados se componen de memorias de traducción y de diccionarios terminológicos creados para servir de herramienta auxiliar en el proceso de traducción de cualquier proyecto. Las memorias de traducción almacenadas en STAR SL conforman, por tanto, un corpus dinámico, paralelo, especializado, multilingüe y documentado, en cuanto al progreso constante del volumen almacenado, sus características estructurales de textos alineados, la especificidad de los contenidos técnicos que alberga, las combinaciones de lenguas con textos alineados y el tipo de información que puede ir asociada a cada documento almacenado.

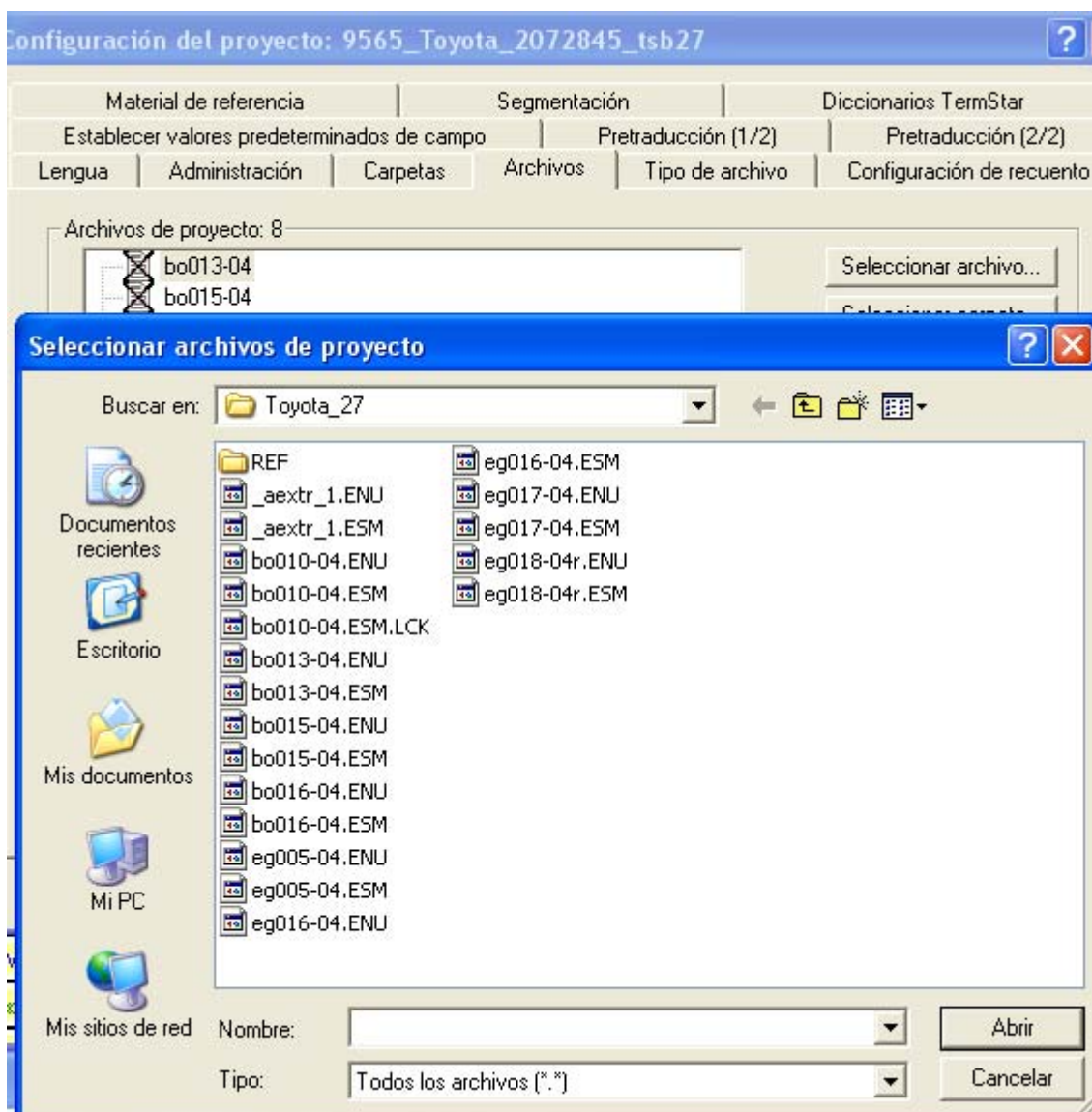
### ***3.3.3.2. Análisis de los flujos de los recursos lingüísticos (RLL)***

Mediante la descripción de tareas y la creación de diagramas de procesos o flujos determinamos los pasos y secuencias que definen el flujo de la gestión de los recursos lingüísticos en STAR SL. El examen atomizado de cada paso y su interrelación con el resto de las secuencias involucradas en el proceso o procesos debe contribuir a identificar la fuente de problemas reales y potenciales. Las ventajas de la aplicación de los diagramas de flujos se traducen no solo en la identificación y representación de las fases de cualquier proceso y sus tareas vinculadas. Del análisis de los flujos representados en los diagramas se obtendrán propuestas que presumiblemente afecten a la estandarización de los procesos y a un mayor control sobre los mismos.

En los siguientes subapartados nos centraremos en el almacenamiento y acceso físico de los documentos almacenados, en el flujo de gestión de las memorias de traducción y de los diccionarios desde el comienzo hasta el final.

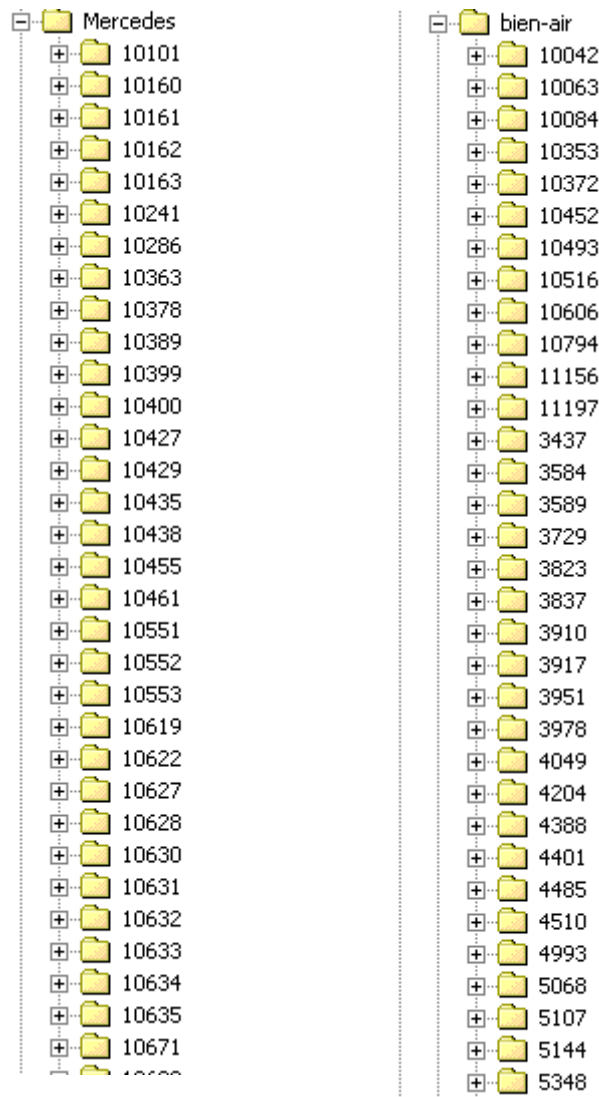
### **3.3.3.2.1. Flujo de almacenamiento y acceso físico a los documentos**

A estos documentos vinculados con los proyectos de traducción se accede de forma directa a través de las carpetas donde quedan almacenados físicamente. También se accede a ellos de forma indirecta a través de los asistentes que sirven de puente entre el usuario y la organización de las carpetas que se gestionan a través del *Explorer* que incorpora *Transit*, como se muestra en la siguiente captura de pantalla:



**Fig. 43: Acceso a los recursos lingüísticos almacenados a través de los asistentes de *Transit***

Cuando trabajamos con las carpetas gestionadas a través del *Explorer*, el nombre de las carpetas nos da cierta información asociada a los proyectos donde previsiblemente encontraremos las memorias de traducción que deseamos seleccionar para reutilizarlas en el desarrollo de otros proyectos de traducción. En el entorno de STAR SL, las carpetas se organizan por clientes y por número de proyecto.



**Fig. 44: Organización de las memorias de traducción por clientes y por nº de proyecto**

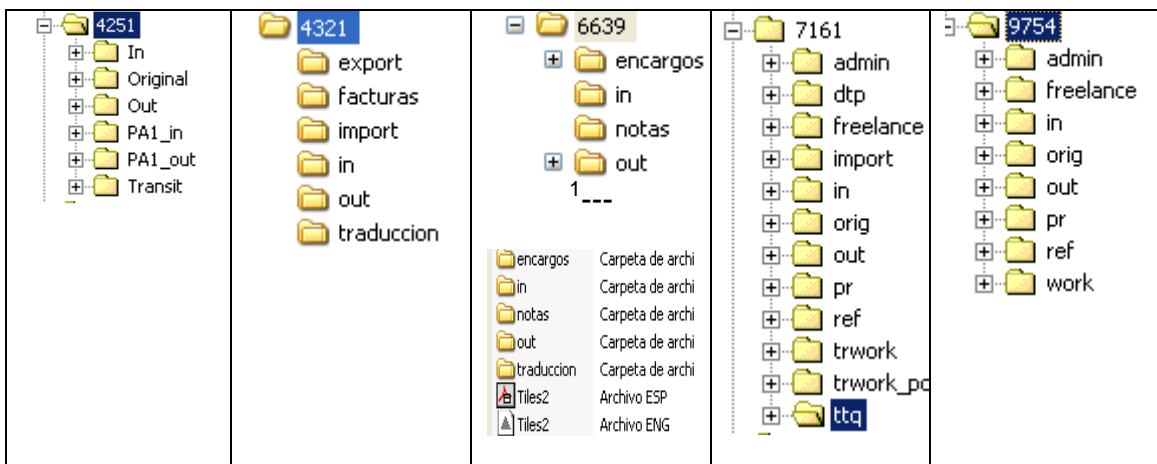
Como se ve en la figura anterior el almacenamiento físico de las memorias y sus documentos asociados se distribuye en carpetas ordenadas por clientes y por el nombre de proyecto que coincide con un número identificativo de cada proyecto que se asigna de forma secuencial y ordenada en función de la fecha de entrada de los encargos globales de traducción recibidos.

Estos criterios de organización de las carpetas pueden generar silencio y ruido durante el proceso de selección del material deseado y esto es porque la información del cliente y el número de proyecto da escasa información



acerca de los contenidos de los documentos, especialmente si desconocemos el sector industrial al que se dedican los clientes.

Dentro de esa carpeta identificativa de un proyecto se pueden encontrar carpetas adicionales para, por ejemplo, almacenar los documentos en formato original, para almacenar los documentos que sirven como material de referencia, como se muestra en los siguientes ejemplos de distribución por carpetas de los documentos asociados a un proyecto:



**Fig. 45: Clasificación de subcarpetas de proyectos**

Tal y como se puede deducir de los diferentes ejemplos de carpetas, coexisten diferentes formas de categorización representadas en carpetas jerarquizadas. Tras un análisis de los archivos que contienen las carpetas se puede constatar que, por ejemplo, las carpetas *facturas*, *encargos*, *admin* contienen información relacionada con los encargos de traducción para los autónomos, las facturas recibidas y las instrucciones generales sobre el proyecto. Esta última información del proyecto 6639 se encuentra almacenada excepcionalmente en la carpeta *notas*. En la carpetas *in* se guarda, en principio, lo que envía el cliente, sin embargo, coexisten con carpetas como la de *Impor*, *Originales*, donde se guardan innecesariamente

<sup>1</sup> Muestra del almacenamiento de los archivos de trabajo directamente en el directorio del proyecto 6639.

de forma duplicada los archivos originales antes de su importación al formato *Transit*. Los archivos de trabajo donde se hace la traducción propiamente dicha se almacenan en las carpetas *transit*, *traducción*, *trwork* y *work* o directamente sin directorio específico, como en el caso del proyecto 6639.

Esta distribución por carpetas está vinculada, por una parte, a las diferentes formas de categorización de los individuos que las generan y, por otra parte, al flujo de trabajo de *Transit* descrito en el apartado 3.3.3.1.2. Gestión del corpus paralelo con *Transit*. Estas formas de organizar siguen una estructura lógica, sin embargo, se constata la aplicación de diferentes criterios de organización en un mismo entorno.

### 3.3.3.2.2. Gestión de las memorias de traducción

Antes de entrar a especificar los pasos vinculados a la gestión de las memorias de traducción, hemos trazado un diagrama de flujo que representa de forma general las diferentes fases de su gestión:

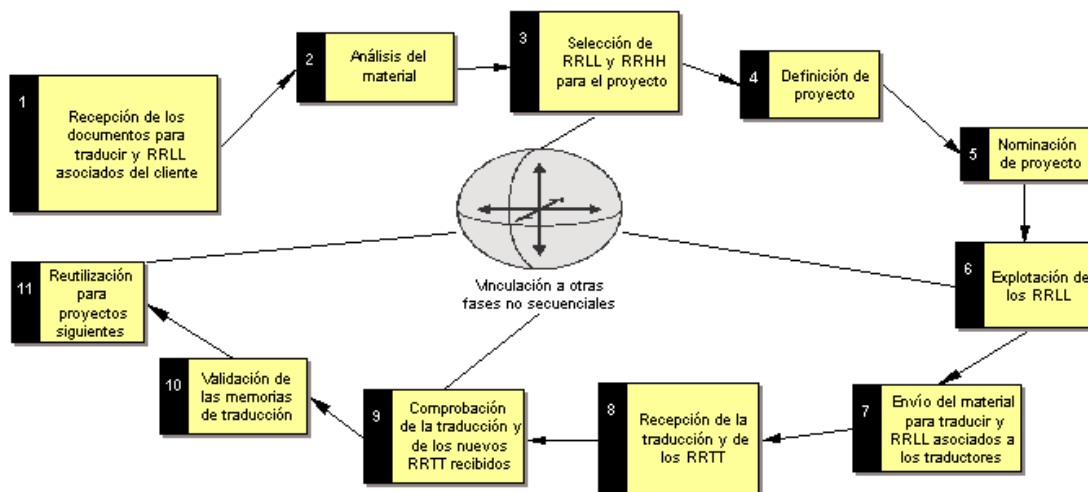


Fig. 46: Fases del flujo de gestión de las memorias de traducción

Tal y como se puede deducir del diagrama, la gestión de las memorias está vinculada a la gestión de proyectos de traducción y esto es, porque estos recursos se generan o reutilizan en el entorno de los proyectos de traducción. Asimismo, si la gestión de las memorias de traducción es inmanente a la

gestión de los proyectos de traducción, el análisis de la herramienta empleada para generar y explotar estas memorias de traducción, es otro de los aspectos unidos a la gestión, ya que la herramienta empleada, en este caso *Transit*, determinará parte del flujo analizado.

Las memorias de traducción que se generan en el entorno de un proyecto constituyen solo una parte de los documentos asociados con el proyecto de traducción. Otros documentos asociados pueden ser los documentos originales que se deben traducir, memorias de traducción —no generadas directamente en el entorno de un proyecto concreto— y diccionarios.

Durante las dos primeras fases, de recepción y análisis —fases 1 y 2 del diagrama—, puede ser que se reciban:

- archivos ya importados en Transit (documentos originales, memorias de traducción adicionales y diccionarios). Estos archivos están en formato \*.pxf, que se corresponde al archivo comprimido de *Transit* que aglutina los archivos para traducir, los diccionarios en formato *TermStar* y la configuración del proyecto —denominados archivos \*.prj que, como veremos más adelante, cobrarán una relevancia clave en el diseño de las propuestas de automatización en la gestión de las memorias de traducción—.
- archivos en formato original (FrameMaker, Interleaf, Word, Quark, HTML, XML, etc.) de versiones anteriores del documento para traducir y las traducciones de ese documento, ya bien en los formatos originales, ya bien en forma de memorias de traducción en otros formatos que no sean *Transit*, como por ejemplo, en formato TMX. Para más información relacionada con este formato, se puede consultar el apartado 2.5. Formatos de codificación y de intercambio de los recursos lingüísticos.

Si el cliente envía archivos \*.pxf, en muchos casos, antes de recibirlos en el entorno de trabajo de *Transit* —como se muestra en la siguiente captura de pantalla—, el gestor de proyectos ya ha procedido a su almacenamiento:



**Fig. 47: Recepción de proyectos de traducción en el entorno de *Transit***

Este almacenamiento está vinculado a la clasificación tanto lógica como física de los proyectos de traducción, como se ha mostrado en la *Fig. 44: Organización de las memorias de traducción por clientes y por nº de proyecto* y la *Fig. 45: Clasificación de subcarpetas de proyectos*. Es decir, los proyectos se crean y almacenan en carpetas específicas con el nombre de cliente creando, adicionalmente, otra subcarpeta con el nombre del proyecto concreto, como veremos más ampliamente cuando tratemos la forma de definir y nombrar el proyecto de traducción —fase 4 y 5 del diagrama—.

En el caso de recibir el material en otro formato, tras el análisis se procederá igualmente al almacenamiento físico, distribuido en carpetas de clientes y proyectos. Una vez almacenado, se llevará a cabo la manipulación de los documentos, como explicamos al tratar la fase de explotación de los recursos lingüísticos.

Durante la selección de los recursos humanos necesarios, así como los recursos lingüísticos pertinentes para la ejecución del proyecto de traducción —fase 3 del diagrama— se utiliza la información obtenida durante el análisis

del material recibido para realizar la traducción. Las diferentes variables que se toman en cuenta para la selección de los recursos humanos se pueden consultar en el apartado 2.3.3. Fase de traducción de la gestión de documentación multilingüe.

El tipo de búsqueda que realiza el gestor para la identificación de recursos lingüísticos potenciales es por cliente. Si ya se han efectuado trabajos para el cliente del proyecto actual, existe una carpeta con el nombre de ese cliente donde previsiblemente encontramos subcarpetas con material susceptible de ser reutilizado. Si el cliente es nuevo, el gestor ya no busca posibles recursos lingüísticos pertinentes para la traducción de ese encargo.

Se procede a la definición del proyecto —fase 4 del diagrama— en el entorno de *Transit*. Definir un proyecto implica:

- ❑ nombrar el proyecto —fase 5 del diagrama—,
- ❑ especificar la combinación de lenguas,
- ❑ especificar el filtro que se debe utilizar para importar los archivos del proyecto, si éstos los hemos recibido en un formato diferente a *Transit*,
- ❑ determinar la forma de segmentación (frases y párrafos),
- ❑ seleccionar diccionarios existentes y crear de nuevos diccionarios,
- ❑ establecer las rutas de los archivos que se deben traducir,
- ❑ establecer las rutas de las memorias de traducción que se deben reciclar para el proyecto.

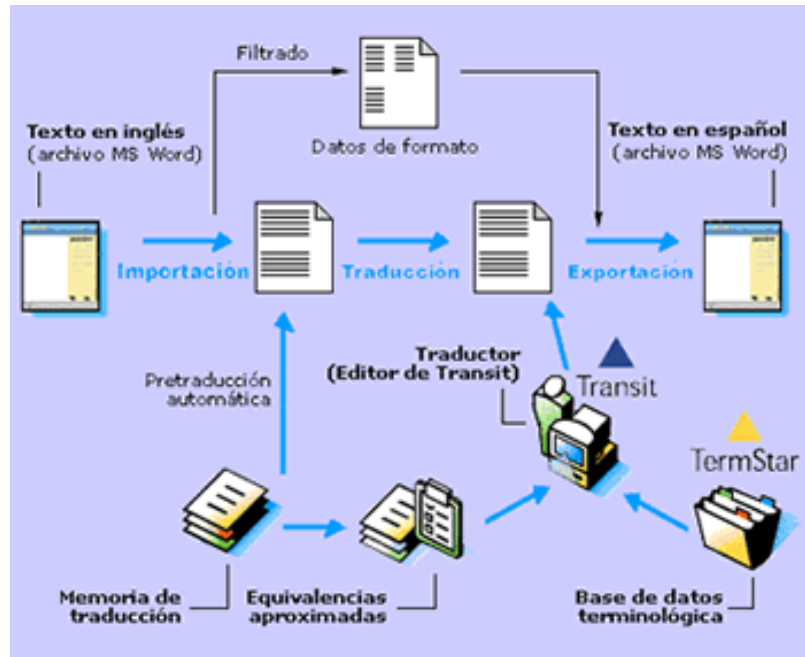
De esta definición dependerá que la explotación automatizada de los recursos lingüísticos sea efectiva —fase 6 del diagrama—. La primera etapa para la explotación de los recursos lingüísticos se inicia con la *importación* de los documentos para traducir. *Transit* extrae el texto traducible de los archivos originales y marca la estructura del documento y la información ortotipográfica con etiquetas XML. Asimismo, puede crear uno o más archivos auxiliares con la información específica del formato de archivo de

origen. Esta información se utiliza en la *exportación* para crear archivos en el formato original a partir de las traducciones. También durante la importación, el sistema segmenta el texto en unidades de traducción tales como párrafos y frases. Los detalles del marcado y de segmentación los controla el *tipo de archivo*. Si el proyecto tiene asignado material de referencia, es decir, simplemente archivos de un proyecto anterior realizado o importado a *Transit*, el siguiente paso durante la importación es la *pretraducción*. Durante la *pretraducción*, el sistema traduce automáticamente segmentos del proyecto actual que tienen una correspondencia exacta en el material de referencia que son las memorias de traducción que previamente se han seleccionado. Ahora, también se crea un índice de equivalencias aproximadas, es decir, correspondencias de unidades de traducción con respecto a las memorias de traducción almacenadas con un índice de coincidencia por debajo del 100% y con un umbral superior predefinible.

En el siguiente diagrama se puede ver de todo el flujo completo de un proyecto de traducción realizado con *Transit*<sup>1</sup>.

---

<sup>1</sup> Para más detalles relacionados con el flujo de trabajo con esta herramienta se puede consultar Cámara y Chrupala 2003b; Díez y Scholand, 2003.



**Fig. 48: Flujo de producción de las traducciones con *Transit***

En el caso de contar con material traducido pero no importado en *Transit*, se lleva a cabo la tarea de alineación que se asemeja bastante a la de importación, con la diferencia de que se activan los algoritmos definidos en la herramienta de alineación integrada en *Transit* para sincronizar las unidades de traducción identificadas por los segmentos numerados.

Los archivos resultantes de la importación de los archivos originales que se deben traducir, se almacenan en una carpeta llamada en la jerga del sistema usado *directorío de trabajo*. En este directorio, *Transit* creará los archivos resultantes de la importación de los archivos originales que se deben traducir. A cada archivo original le corresponderán al menos dos archivos: uno para la lengua de salida y otro para la lengua de llegada. Este par de archivos se denomina par bilingüe y, tras su traducción, se convertirán en las memorias de traducción. En el caso de proyectos multilingües habrá más de dos archivos: uno por cada lengua de llegada, más uno de la lengua de salida. Sin embargo, en un momento dado el traductor siempre trabajará sólo con dos lenguas, de ahí la importancia del par bilingüe.

Una vez se han terminado de preparar y explotar los recursos lingüísticos existentes para un proyecto —importación, alineación y pretraducción—, el gestor envía al traductor el material preparado para traducir, es decir, el archivo \*.pxf, —fase 7 del diagrama—. Y el traductor, tras haber realizado su trabajo, envía al gestor la traducción —fase 8 del diagrama—.

La preparación del material para enviar se lleva a cabo de forma automatizada, al igual que el envío de los archivos \*.pxf. Ahora bien, a diferencia de lo que envía el gestor al traductor, el gestor recibe también un archivo comprimido de *Transit*, un archivo \*.txf, donde no se incorporan los archivos con la información de maquetación, los archivos originales, el material de referencia y los archivos que contienen la información de los parámetros del proyecto, es decir, los archivos \*.prj. El gestor no necesita estos archivos mencionados, ya que la persona que preparó el proyecto para su envío cuenta con toda esta información. Un archivo \*.txf contiene pues los archivos con la traducción y, en algunos casos, también los nuevos registros terminológicos incorporados en los diccionarios definidos para el proyecto en cuestión.

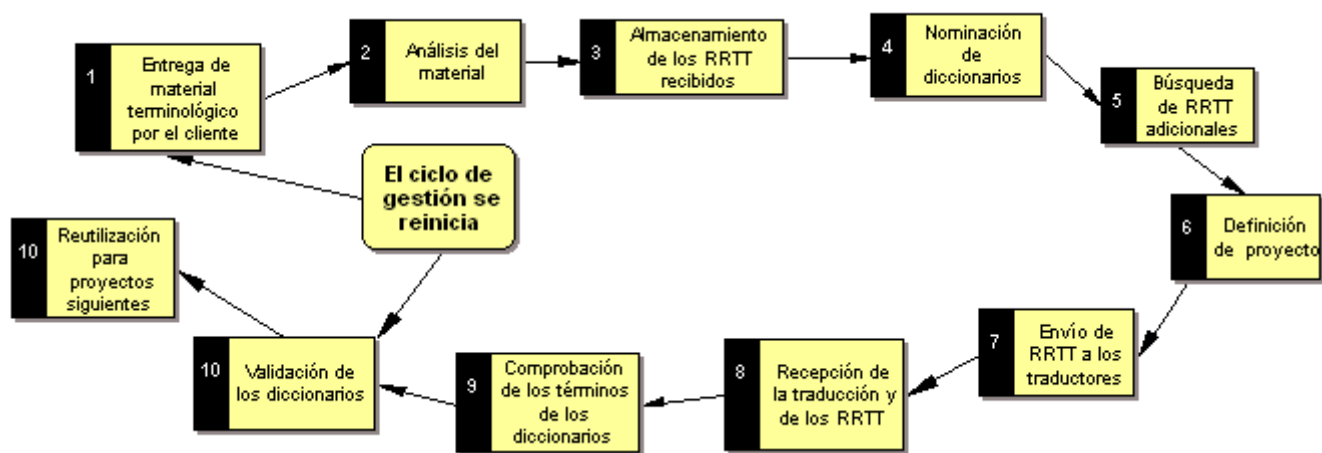
El gestor tras recibir el material traducido dispone quién debe llevar a cabo el control de calidad de la traducción. Un traductor revisor tras corregir la traducción —fase 10 del diagrama—, valida el estado de las memorias de traducción —fase 10 del diagrama— predisponiéndolas a convertirse en material para ser reutilizado de diferentes formas, como se explica a continuación:

- mediante una pretraducción de versiones posteriores a ese proyecto del mismo cliente,
- como material consultable para la búsqueda y extracción de términos y concordancias.



### 3.3.3.2.3. Gestión de los diccionarios (recursos terminológicos)

La gestión de los diccionarios es uno de los flujos de trabajo cruciales dentro de la gestión global de los recursos lingüísticos en STAR SL y, como en el caso de la gestión de las memorias de traducción, crítica para la empresa. Antes de entrar en detalle acerca de las tareas vinculadas a todos los pasos de gestión, se ha trazado un mapa de flujo que representa de forma general las diferentes fases de la gestión.



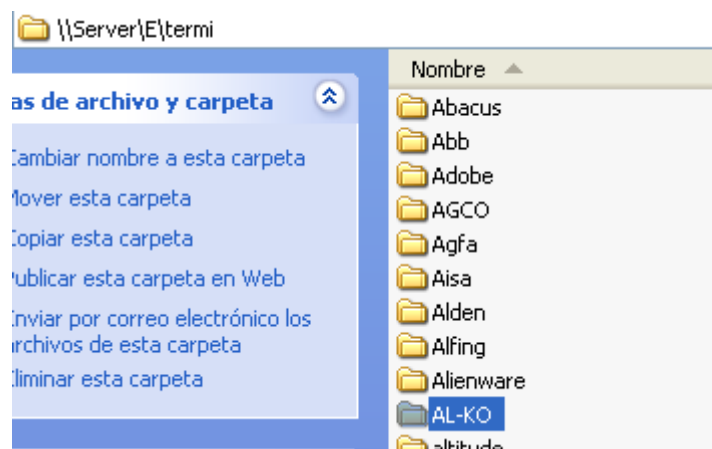
**Fig. 49: Fases del flujo de gestión de los diccionarios**

Durante las dos primeras fases, de recepción y análisis —fases 1 y 2 del diagrama— puede ser que se den tres diferentes situaciones, como tipificamos aquí:

- que se reciban diccionarios en formato *TermStar* como elementos independientes o dentro del archivo \*.pxf que, como se ha visto más arriba, este formato comprimido de Transit aglutina los archivos que se deben traducir, los diccionarios en formato *Termstar* y la configuración del proyecto;
- que se reciban glosarios o diccionarios en formato diferente a *TermStar*. En estos casos se intenta importar los diccionarios al entorno de *TermStar* para poder reutilizar mejor los recursos;

- que no se reciba ningún recurso terminológico. Cuando no se reciben glosarios o diccionarios por parte de los clientes, de la fase 1 y 2 se pasa directamente a la fase 5.

Cuando se reciben recursos terminológicos del cliente se procede a su almacenamiento —recuadro 3 del diagrama—. El almacenamiento físico se lleva a cabo siguiendo la estructura en carpetas de clientes predeterminada, en una unidad de disco asignada para este cometido (E:\termi), tal como se muestra en la siguiente captura de pantalla:



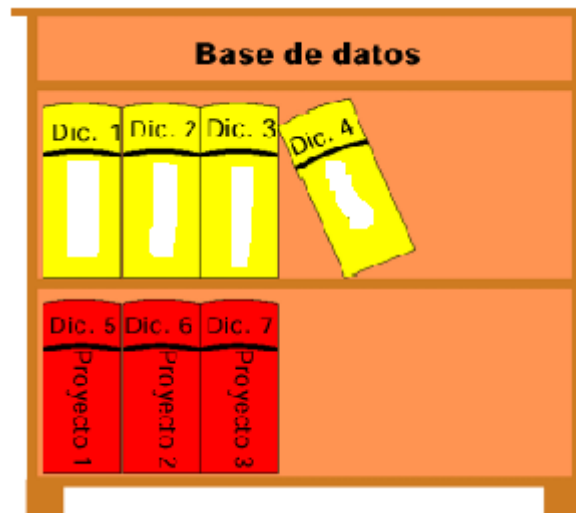
**Fig. 50: Almacenamiento de diccionarios en carpetas de clientes**

Así, cada cliente cuenta con una base de datos donde se guardan todos los diccionarios de ese cliente. Se crean, por tanto, diccionarios específicos para cada cliente. El nombre que se le asigna al diccionario —fase 4 del diagrama— puede coincidir con la nomenclatura dada por el cliente o puede renombrarse siguiendo los criterios individuales de cada gestor de proyectos. Al inventariar los diccionarios hemos podido constatar duplicados de diccionarios nombrados de forma diferente y diccionarios con el mismo nombre pero con contenidos diferentes.

Si el material está en *TermStar* se recibe y se almacena en la carpeta del cliente correspondiente. Si el material no está en formato *TermStar* se procede a su importación.

Cuando se importa un diccionario al formato *TermStar*, la herramienta terminológica almacena los datos en un diccionario que a su vez se almacena en una base de datos con el nombre del cliente.

En *TermStar*, la terminología de un cliente puede repartirse en varios diccionarios. La posibilidad de distribuir la terminología en varios diccionarios ofrece ventajas con respecto a su almacenamiento en un diccionario genérico. Por ejemplo, se puede disponer de diccionarios independientes por productos de cliente para garantizar, de esta forma, el empleo de la terminología específica de cada uno de ellos. La siguiente imagen pretende mostrar lo que acabamos de explicar:



**Fig. 51: Almacenamiento de los diccionarios por proyectos o productos en bases de datos de clientes**

El almacenamiento en bases de datos que contienen diferentes diccionarios facilita la gestión de grandes volúmenes de datos, ya que, cuando se manejan cantidades ingentes de términos, el empleo de un único diccionario puede resultar confuso y poco operativo. Al adscribir un diccionario a cada proyecto o producto del cliente, se garantiza la adecuación y la homogeneidad de la terminología específica en cada caso. Este almacenamiento en diccionarios y bases de datos no condiciona, sin embargo, el uso o el no uso de los diccionarios para un proyecto en concreto. Esto significa que pueden abrirse y consultarse varios diccionarios al mismo

tiempo. Los diccionarios elegidos se abren en una sola ventana, a modo de un solo diccionario virtual. De esta manera, las tareas de visualización, clasificación y búsqueda se realizan como si de un único diccionario se tratara.

Una vez analizado el material enviado por el cliente, ya bien haya mandado recursos terminológicos o no, el gestor de proyectos busca recursos terminológicos adicionales —fase 5 del diagrama—. La exploración se lleva a cabo en:

- la red local, en busca de los recursos terminológicos almacenados físicamente por los criterios del cliente. Si el cliente es nuevo, se pregunta a los compañeros acerca de la existencia de un cliente que corresponda al mismo sector industrial;
- Internet, pero no en busca de términos concretos extraídos del texto, sino en busca de glosarios publicados en Internet, ya bien sean monolingües con definiciones, ya bien multilingües relacionados con el tema de la traducción.

Una vez que el gestor ya ha recopilado el material que considera pertinente, define el proyecto de traducción —fase 6 del diagrama—. Una de las tareas vinculadas a la definición de proyectos es la de creación de diccionarios para que el traductor los alimente a lo largo de la traducción. El gestor crea a veces un diccionario específico para el traductor y a veces no.

El gestor envía al traductor un archivo \*.pxf, que como hemos mencionado más arriba, es un formato comprimido que utiliza *Transit* donde se guardan los archivos para traducir, los diccionarios en formato *TermStar* y la configuración del proyecto —fase 7 del diagrama—.

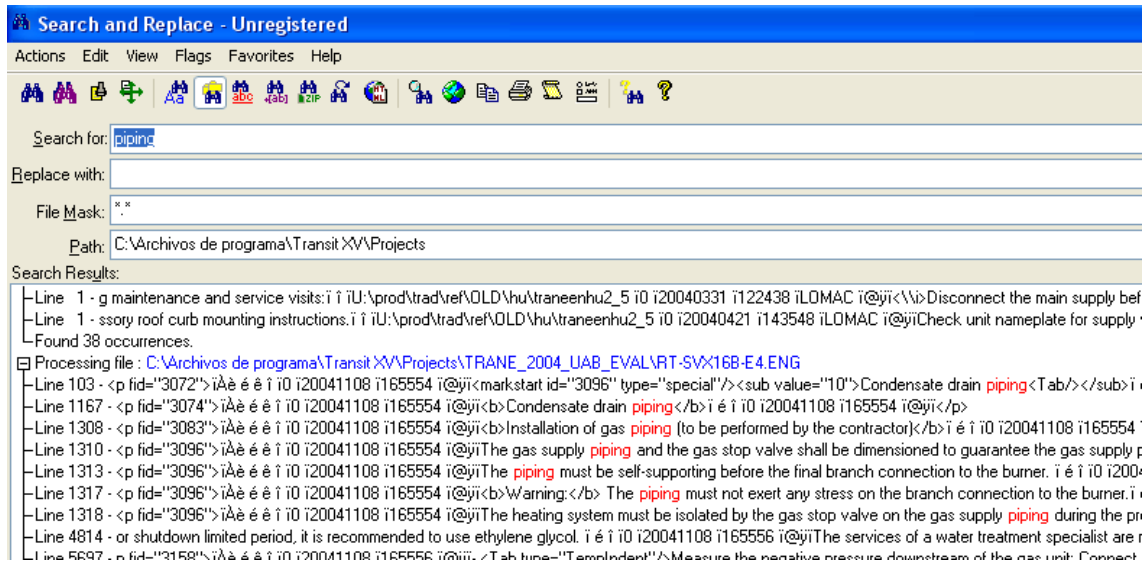
Una vez que los traductores han terminado su trabajo, envían los archivos traducidos y el diccionario con las nuevas entradas que recibe el gestor del proyecto—fase 8 del diagrama—.

El gestor de proyectos procede a la comprobación de las entradas que propone el traductor y a la edición de las mismas, según sea necesario — fase 9 del diagrama—. El gestor también puede delegar en un traductor revisor esta tarea. Para comprobar la terminología se realiza una exploración, si es pertinente, en los recursos internos almacenados mediante la herramienta *Search & Replace* de la empresa Funduc<sup>1</sup>. Mediante esta herramienta se pueden realizar operaciones de búsqueda y reemplazo en ficheros de múltiples formatos codificados en ASCII como texto plano, HTML y también en su última versión en múltiples formatos con codificación de Unicode. Esto significa que puede explorar dentro de los archivos de *Transit*, ya bien estén o no comprimidos. Se trata de un programa muy fácil de usar, con un gran rendimiento. Da soporte a la sintaxis característica de los motores de búsqueda y expresiones regulares, con palabras sueltas, frases concretas y comodines.

Tras introducir la ecuación de búsqueda, la interfaz del programa muestra los resultados obtenidos, resaltando los términos que aparecen con sus concordancias, como se muestra en la siguiente captura de pantalla:

---

<sup>1</sup> Para más información véase [http://www.funduc.com/search\\_replace.htm](http://www.funduc.com/search_replace.htm)



**Fig. 52: Resultado de las capturas de concordancias mediante *Search & Replace***

Tras analizar los resultados, se pueden abrir los documentos donde se encuentra el término. Puesto que lo que precisamos no solo es el término de búsqueda, sino su equivalencia en el otro idioma del par bilingüe, necesitamos abrir dos archivos. Es decir, que debemos abrir *Transit* previamente, extraer la ruta de los documentos deseados y propuestos por *Search & Replace*. De esta manera podremos abrir el documento de donde se extrae la concordancia del término de búsqueda y el documento en el otro idioma. Como en *Transit*, las ventanas de los dos idiomas están sincronizadas; la búsqueda de un término nos lleva a su equivalente. De esta manera, podemos visualizar el extracto de corpus paralelo deseado, como podemos ver a continuación:



**Fig. 53: Sincronización de las ventanas en *Transit***

Una vez comprobados los términos, se procede a la fusión del diccionario específico alimentado por el traductor con el diccionario general del cliente, o bien, se procede a la validación de los nuevos términos en el mismo diccionario del cliente, si es que el traductor no contaba con un diccionario específico. Puede ser que el traductor no haya incluido ningún término, en este caso, el traductor revisor procede a la inclusión de los mismos.

Una de las fases ideales es la de validación de los diccionarios por parte del cliente —fase 10 del diagrama—. El gestor debe ocuparse de que el cliente valide las entradas de los diccionarios que se envían periódicamente. Puesto que esta fase depende de la frecuencia con la que este cliente envía encargos, del número de entradas de los diccionarios y de la disponibilidad del cliente, no siempre se lleva a cabo esta tarea. Durante la fase de comprobación, los recursos terminológicos se preparan para su reutilización en proyectos posteriores —fase 11 del diagrama—.

### 3.3.3.3. Necesidades de los usuarios

Para recoger la perspectiva del usuario se ha realizado una encuesta entre los miembros del equipo de STAR SL dirigida a obtener información acerca de las necesidades relacionadas con la explotación de los recursos lingüísticos almacenados en STAR SL. Muchas de las respuestas de las diez personas encuestadas son coincidentes ya que, a través de las necesidades expresadas, abordan un problema genérico de acceso a los recursos almacenados y que se concreta en las siguientes necesidades específicas:

- poder consultar, por ejemplo, términos y concordancias en otro material de referencia que no esté definido en los parámetros del proyecto de traducción en la que el traductor está trabajando en ese momento de una forma más sencilla que la utilizada con *Search & Replace*;
- poder buscar, más allá de las búsquedas por clientes, por criterios de búsqueda adicionales como identificadores del contenido de las memorias de traducción almacenadas;
- poder buscar por criterios de búsqueda adicionales como identificadores del contenido de los diccionarios terminológicos; es decir, sin tener que abrir los diccionarios, sino que un sistema ofrezca una selección de diccionarios que se adapte a la ecuación de búsqueda del usuario;
- poder identificar los diccionarios existentes basándose en el nombre del diccionario, es decir, que los nombres sean más intuitivos;
- poder identificar los diferentes tipos de diccionarios existentes en función de su fiabilidad para la mejora de su selección;
- poder aplicar diferentes criterios de búsqueda para recuperar el material requerido, como nombre del gestor, combinación de lenguas, fechas, etc.;
- poder asociar a los diccionarios información adicional, como las recomendaciones de mantenimiento o las especificaciones concretas



para la creación y alimentación de un diccionario concreto de un cliente;

- contar con una interfaz de búsqueda similar a un buscador de Internet;
- lograr que los recursos lingüísticos estén organizados como un sistema que funcione como una biblioteca virtual.

La incorporación de las necesidades del usuario en el diseño del nuevo sistema de gestión de recursos predispone al éxito y esto es porque entenderán el funcionamiento del sistema de forma intuitiva, sin tener que analizar prolongadamente el entorno del sistema para aprender a usarlo (Baeza-Yates, *et al.*, 2004: 177-178).

Asimismo, contar a priori con las necesidades del usuario por una parte da pautas de diseño. Así se puede diseñar un nuevo sistema abocado a satisfacer las carencias y necesidades apuntadas. Por otra parte, basar los sistemas en las necesidades de los usuarios reduce el impacto que siempre provoca en un ecosistema informacional la implementación de nuevas rutinas de trabajo en el entorno de un nuevo sistema de información.

#### **3.3.3.4. Protocolo de cuellos de botella, incidencias o problemas**

A partir del análisis de los recursos inventariados, de los flujos de gestión de los RRLL, de las necesidades de los usuarios y las observaciones hemos elaborado un protocolo de los problemas que suponen un cuello de botella en la gestión de los RRLL y en consecuencia una traba para la mejora de su automatización.

En este protocolo hemos procedido a tipificar los problemas detectados en las siguientes categorías, siendo conscientes de la dificultad existente a la hora de delimitar un problema de otro:

- Problemas de almacenamiento, categorización y diversidad
- Problemas de acceso a los recursos lingüísticos

- Problemas por falta de protocolos
- Problemas de explotación de los recursos lingüísticos existentes

#### **3.3.3.4.1. Problemas de almacenamiento, categorización y diversidad**

Al mostrar las capturas de pantalla con las diferentes formas de clasificar los documentos asociados a un proyecto y, en definitiva, asociados a las memorias de traducción, se mostraban formas diferentes de categorizar las carpetas —también denominadas *directorios*—. El proceso de categorización, en general, es un proceso cognitivo vinculado a la manera en que almacenamos información para después recuperarla. Es un proceso de establecimiento de un nivel de representación mental para comprender, tal y como explicábamos en el apartado 2.6.2.1. Gestores de tesauros multilingües.

Asimismo, las diferentes formas de categorización y, por tanto, su representación mediante clases y relaciones conceptuales —en el caso que nos ocupa, las carpetas y subcarpetas representan las clases y relaciones conceptuales de tipo jerárquico y asociativo entre esas clases— son siempre válidas para el individuo que las establece, ya que se basa en su lógica personal. En el momento en que cualquier miembro del equipo deba acceder a cualquiera de las carpetas y subcarpetas de un proyecto que no haya creado él mismo, precisa de un tiempo para entender, basándose en la categorización existente, cómo se distribuyen los datos de un proyecto concreto. Y más allá del tiempo que se utiliza para entender la lógica de otra persona, que evidentemente supone un problema de gestión, la diversidad de las categorizaciones puede derivar en un problema para la automatización de los procesos. El almacenamiento de los documentos, por tanto, debe llevarse a cabo de forma estructurada y lógica, de manera que pueda ser fácil para cada uno encontrar cualquier documento en cualquier momento con independencia de quién lo haya traducido y cuándo. Se debe lograr la estandarización regulada por procedimientos preestablecidos para el

almacenamiento de los documentos que contemple el nombre y la distribución de las carpetas. Asimismo, los procedimientos deben abordar el almacenamiento de los documentos como medida de protección de los documentos para que no sean dañados o eliminados de forma inadvertida, ya que están accesibles a todas las personas de la empresa.

#### **3.3.3.4.2. Problemas de acceso a los recursos lingüísticos almacenados**

El acceso a los recursos lingüísticos se lleva a cabo a través del almacenamiento físico de las carpetas que están distribuidas por clientes. El problema que genera este tipo de acceso y criterio de búsqueda está vinculado, entre otras cosas, al hecho de que no todos los encargos o proyectos que se realizan para un cliente versan necesariamente sobre un mismo tema, por tanto la selección del material de referencia capturado tras la búsqueda por cliente puede provocar ruido; esto quiere decir que lo que encontramos no es pertinente. Igualmente, puede que existan recursos lingüísticos almacenados de otro cliente que versen sobre la misma temática que el encargo del cliente actual. La distribución física dada genera silencio; esto es que no encontramos lo que es pertinente, cuya consecuencia inmediata es la infrautilización de los recursos lingüísticos almacenados.

Si bien hemos apuntado que las memorias de traducción, —es decir, el corpus paralelo—, poseen anotaciones de tipo no lingüístico, éstas no se pueden explotar de forma global, sino solamente de forma parcial en el entorno de un proyecto concreto de traducción donde se haya definido previamente el material del corpus que, como ya hemos dicho, se selecciona en función del cliente. Asimismo, si se pudieran explotar estas anotaciones de forma global, obtendríamos los mismos problemas de ruidos y silencios apuntados más arriba. Por tanto, se precisan de puntos de acceso adicionales para identificar los posibles recursos lingüísticos pertinentes y así explotar mejor el corpus paralelo.

El problema de acceso a los diccionarios también está relacionado con las limitaciones de las posibles búsquedas condicionada por el almacenamiento distribuido en carpetas de clientes. En este sentido, puede que se dé el caso de que existan recursos terminológicos relacionados con el proyecto de traducción que se está gestionando, pero no directamente relacionados con el cliente de esa traducción. Puesto que, sin embargo, esta información no se haya registrada en ninguna parte, no se accede a estos recursos terminológicos pertinentes. Esto significa que, si se tiene un cliente para el que se ha hecho traducciones de mecánica de relojería y para eso se ha creado una base de datos del cliente y sus diccionarios específicos, previsiblemente se puede reutilizar ese material terminológico para el nuevo cliente dedicado a la fabricación de relojes. En el ecosistema informacional de STAR SL puede que se dé el caso de que esta información sí fluya de forma oral porque un integrante del equipo se lo comenta a otro y éste segundo recuerda un proyecto similar que se llevó a cabo para otro cliente. Asimismo, puede que simplemente se pierda esta información debido a las siguientes razones:

- porque no se realiza una consulta a ningún integrante del equipo en plantilla;
- porque los miembros del equipo han olvidado el contenido de las traducciones de algún cliente que podría ser aplicable a la consulta del compañero;
- porque quien gestionaba el cliente cuyos recursos terminológicos son susceptibles de ser reutilizados ya no es miembro del equipo en plantilla;
- porque quien gestionaba el cliente cuyos recursos terminológicos son susceptibles de ser reutilizados está ausente por vacaciones o baja por enfermedad o maternidad.

En estos casos apuntados, la experiencia de cada gestor con cada cliente y proyecto se pierde. Por tanto, para evitar estas situaciones se deben crear

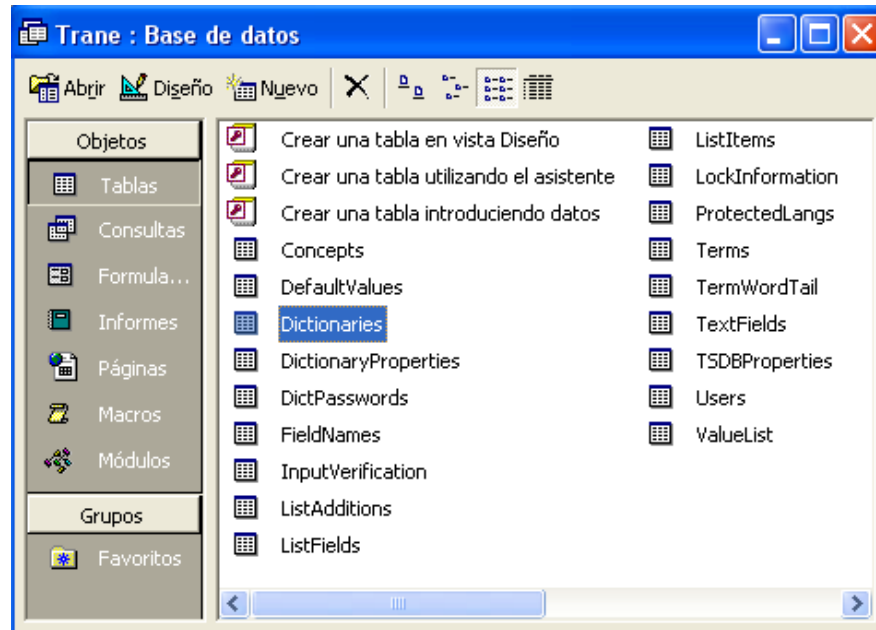
mecanismos que garanticen la mejor gestión de ese conocimiento basado en la experiencia.

Adicionalmente, el almacenamiento de los diccionarios en bases de datos distribuidos en las carpetas de los clientes no permite ver en la estructura qué diccionarios existen en un base terminológica concreta. Para saber el contenido que hay en las bases de datos existen dos posibilidades que explicamos a continuación.

Una de ellas es accediendo en el entorno de *TermStar* a la función de abrir diccionarios no vinculados a un proyecto. Mediante esta función *TermStar* mostrará todos los diccionarios de todas las bases de datos siendo esto de poca ayuda debido a la cantidad de datos mostrados sin discriminar. Si se abren los diccionarios vinculados a un proyecto, entonces se visualizan todos los diccionarios de las bases de datos enlazadas a ese proyecto con el consiguiente ruido y silencio, ya que habrá diccionarios que no interesan y otros que ni siquiera se visualizan por no tener enlazada la base de datos. Esta función no suele utilizarse debido a la poca satisfacción que obtiene el usuario.

La segunda forma para saber qué diccionarios existen en una base de datos consta de dos pasos:

1. desde el *Explorer* directamente abriendo con *Access* la base de datos del cliente que nos interesa, como se muestra en la siguiente captura de pantalla:



**Fig. 54: Tabla de diccionarios de una base de datos**

2. abriendo la tabla de los diccionarios de la base de datos del cliente, tal y como mostramos aquí:

Dictionaries : Tabla			
		DictID	DictName
	+	1	edicion-imageni
	+	2	graphics
	+	3	botones softwar
	+	4	ELSA
	+	5	C310
	+	6	GENERAL

**Fig. 55: Ejemplos de diccionarios contenidos en una base de datos**

Como se puede deducir de lo expuesto, el acceso al contenido de las bases de datos, si bien no se puede calificar de difícil, sí que resulta incómodo ya que se deben realizar dos pasos, sin lograr una satisfacción en el resultado, ya que no desaparecen los problemas de ruido y silencio utilizando otras vías.

### 3.3.3.4.3. Problemas de gestión de los recursos lingüísticos por falta de protocolos

Algunos de los procedimientos de trabajo relacionados con la gestión de los recursos lingüísticos están estructurados, como por ejemplo, el almacenamiento de documentos relacionados con un proyecto en una carpeta específica de proyecto que se incluye en otra correspondiente al cliente. En el caso de los recursos terminológicos, se almacenan todos los diccionarios específicos de un cliente en una base de datos con el nombre del cliente donde se almacenan diferentes diccionarios distribuidos por proyectos específicos o por productos.

Existen también procesos estructurados con lagunas ya que no se utilizan todas las informaciones asociadas a los recursos lingüísticos y también existen procesos no estructurados como es el caso de la *comunicación boca-oreja*, por la que una persona que ha gestionado un determinado proyecto comunica a otra, en un momento posterior, su experiencia a través de información relevante.

Muchos de los flujos existentes son el resultado de trabajo profesional ejercido por procedimientos basados en rutinas practicadas, pero no avaladas por una política de la información. Una política de la información atendería a protocolos básicos de gestión del trabajo que eviten la aplicación de diferentes criterios que, aunque lógicos, no están unificados. La aplicación de una política de la información debe proporcionar la dirección estratégica y las directrices para la creación y utilización de la información en el seno de STAR SL.

En ese sentido, al inventariar los diccionarios hemos podido constatar que, por una parte, existen duplicados de diccionarios pero que están nombrados de forma diferente y, por otra, existen diccionarios con el mismo nombre —se ha contabilizado más de cincuenta diccionarios nombrados como *General*— pero con contenidos diferentes vinculados al cliente y que están

almacenados en la carpeta física del cliente que corresponde. Y esto es, porque no existe uniformidad en la forma de nombrar diccionarios; se nombran los diccionarios almacenados con una nomenclatura libre, ya que no se ha previsto un protocolo para esta tarea.

La confección de un protocolo que ataje los problemas de duplicados de información, debe prever también la etiquetación de los diccionarios con un nombre del que se extraiga información relevante.

En este sentido, la decisión acerca de qué sintaxis debe regir el nombre de un diccionario —y, por extensión, también de un proyecto— debe incorporar elementos de los que se pueda extraer información que ayuden a identificar y seleccionar el diccionario de forma automatizada.

El protocolo para nombrar diccionarios debe abordar también la fiabilidad del contenido de los diccionarios. Esto quiere decir que el gestor o traductor debe saber en cada momento el ranking de aplicación de los diccionarios que pueden ser pertinentes para un proyecto de traducción concreto. Si en el nombre del diccionario se incorpora un elemento identificativo de su estado, el usuario podrá ponderar los diccionarios de forma rápida. Asimismo, la utilización de los términos de los diccionarios en función de su estado debe contribuir a mejorar la calidad.



#### **3.3.3.4.4. Problemas de explotación de los recursos lingüísticos existentes**

El empleo de la herramienta *Search & Replace* para explotar las memorias de traducción es una forma que si bien es útil en algunos casos, en otros, las bondades de la herramienta empleada disminuyen, por ejemplo, cuando se debe realizar una exploración en la totalidad del corpus almacenado. Igualmente, las múltiples capturas que oferta esta herramienta en función de la ecuación de búsqueda utilizada, representa un trabajo adicional de exploración y selección de los documentos que se desean. Pero si, además, se quieren abrir los documentos, entonces se debe copiar la ruta, abrir *Transit* y pegar la ruta de archivo ofertada en el entorno de *Search & Replace*. El flujo de búsquedas en las memorias almacenadas es un proceso lento e insatisfactorio, pues, aunque puede que se extraigan términos contextualizados, adicionalmente se debe abrir uno por uno el archivo donde se encuentran almacenados los términos de la búsqueda; de otra manera no se obtienen equivalencias de los términos buscados en la primera selección de las memorias.

Si se quiere obtener un mejor rendimiento de la herramienta de búsqueda, se puede hacer la exploración en subdirectorios; en este caso, se elegirán los subdirectorios del cliente en cuestión. Esto significa que previsiblemente se obtendrá el mismo ruido y silencio que se detectaba durante la búsqueda en la red local de diccionarios terminológicos almacenados físicamente por los criterios de cliente, tal y como hemos explicado más arriba.

La anotación documental de los textos que componen el corpus paralelo es relevante en el entorno de un proyecto, sin embargo, el acceso se hace a través de su almacenamiento físico reduciendo las posibilidades de búsqueda a dos: por cliente y por proyecto. Si se desconoce el nombre de cliente, se debe averiguar primero el nombre del proyecto, lo que supone tener que abrir otra aplicación adicional donde se encuentra almacenada la información relativa a la trayectoria de los proyectos. Las anotaciones

asociadas a las memorias de traducción recuperadas en el entorno de un proyecto y en los archivos \*.prj no permiten hacer una selección automática de los recursos lingüísticos en el entorno de *Transit*.

Resumiendo, en general, la ineficiencia y costes se asocian a la dispersión de la información y la carencia de buenos puntos de acceso. La existencia de multitud de repositorios de recursos lingüísticos almacenados y distribuidos por toda la red local de STAR SL con contenidos duplicados y con nomenclaturas también pueden inducir al error en la selección.

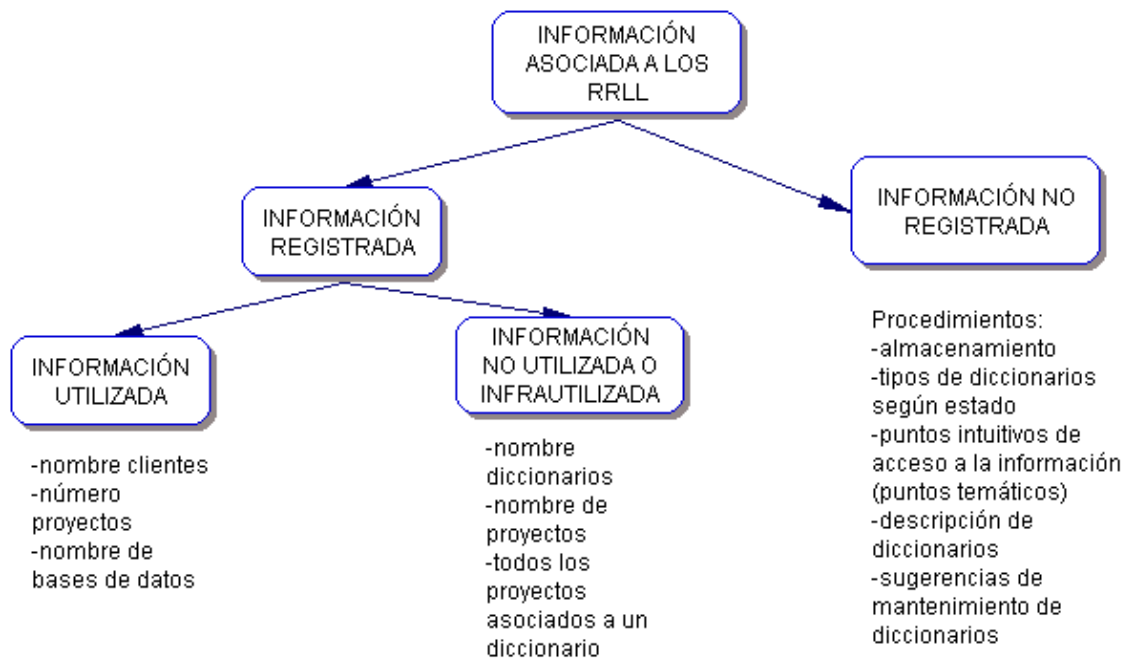
Si bien, el coste de los dispositivos electrónicos de almacenamiento de la información ha disminuido notablemente, los costes asociados a la gestión de la información y a la dificultad de encontrar la información deseada genera pérdidas para la empresa. Los costes de tiempo asociados a las variables que apuntamos a continuación son, en resumen, los puntos que se deben resolver mediante las propuestas de diseño e implementación que planteamos en el siguiente apartado:

- ❑ la búsqueda de los contenidos derivados de la dispersión,
- ❑ la ausencia de control del estado de los recursos,
- ❑ la carencia de criterios amplios de almacenamiento como puntos de acceso temáticos.

### ***3.4. Diseño e implementación de las propuestas metodológicas, conceptuales y tecnológicas***

Del análisis de los datos recogidos ya se atisba el modelado de posibles soluciones y que desarrollamos a continuación. Así, si hemos constatado una carencia de protocolos para la armonización de procesos, presumiblemente la aplicación de una política de la gestión de información garantiza el diseño e implementación de los protocolos que documentan los procesos para la mejora del flujo.

En esta línea hemos constatado los límites de explotación de los recursos lingüísticos vinculados a la discriminación solo por el nombre de los clientes y la necesidad de la gestión global de las informaciones asociadas —ya bien estén registradas o no— a esos recursos lingüísticos (RRL), tal y como se muestra en el siguiente diagrama:



**Fig. 56: Información asociada a los recursos lingüísticos**

Asimismo se han mostrado algunos de los límites de las herramientas utilizadas para explotar los recursos lingüísticos almacenados. En este sentido nos proponemos dar una solución general a los problemas expuestos mediante el diseño e implementación de:

- protocolos de trabajo con la intención de avanzar en la documentación de experiencias armonizadas para generar conocimiento explícito compatible (*corporativo o colaborativo*),
- diseño de un sistema conceptual para la gestión optimizada de los diccionarios,

- diseño de un sistema conceptual para la gestión y explotación optimizada de las memorias de traducción,
- diseño de un sistema integrado de gestión de recursos lingüísticos y de conocimiento.

### **3.4.1. Establecimiento de protocolos**

Los protocolos que proponemos dan respuesta a algunos de los problemas relacionados directamente con la gestión de los recursos lingüísticos e indirectamente con la ausencia de una política previa de información. En este sentido abordamos:

- la nomenclatura de los proyectos
- la nomenclatura de los diccionarios
- los tipos de estado de los diccionarios
- la estructura de las subcarpetas de proyectos

#### **3.4.1.1. Nomenclatura de los proyectos de traducción**

Dentro de todos los proyectos de traducción, existen tipos de proyectos que requieren de una formalización en cuanto a su nomenclatura para su rápida identificación. En este sentido, los proyectos los clasificamos según la siguiente tipología:

- Traducción: proyectos de traducción que se reciben de los clientes. La mayoría de los proyectos gestionados son de este tipo.
- Alineación: proyectos que se crean para obtener material de referencia a partir de textos que no se hayan traducido con *Transit*.
- Presupuesto: proyectos que se crean para la elaboración de un presupuesto.

- Temporal: proyectos que se crean con carácter temporal para hacer pruebas. Se eliminan cuando ya no son necesarios.
- Prueba de traducción para candidatos: traducciones preparadas para enviar a candidatos.

Aunque sea muy importante la formalización de todos los tipos de proyectos, la nomenclatura de los proyectos de traducción propiamente dichos es la de mayor trascendencia debido al número de proyectos gestionados de este tipo, tal y como veíamos en el apartado 3.3.3.1. Análisis de los recursos lingüísticos almacenados.

Para la nomenclatura de los proyectos especiales, es decir, todos los proyectos menos los de traducción propiamente dicha, obligatoriamente se incluirá el número de proyecto y el nombre del cliente, con un tercer elemento adicional que corresponde a un código identificativo de estos proyectos que no son de traducción, como sigue a continuación:

Nº proyecto\_ Nombre cliente \_ Tipo proyecto especial

**Fig. 57: Sintaxis de proyectos especiales (1)**

Los códigos identificativos de los proyectos especiales y por tanto del tercer elemento son:

<b>Tipos de proyecto</b>	<b>Códigos identificativos</b>
Presupuesto	_PRESU
Alineación	_ALIN
Temporal	_TEMP
Pruebas	_TEST

**Fig. 58: Tabla de categorías de proyectos**

Para los proyectos de alineación, se utilizará el número de proyecto relacionado con el proyecto de traducción para el que se lleva a cabo la alineación.

Puede ser que la nomenclatura de los proyectos especiales requieran de identificación adicional. Los elementos adicionales que pueden constituir un tercer elemento de carácter identificativo, se pueden derivar de las pautas que rigen la nomenclatura de los proyectos de traducción, como se muestra a continuación.

Nº proyecto_ Nombre cliente _ 3er elemento_Tipo proyecto especial
---

**Fig. 59: Sintaxis de proyectos especiales (2)**

En cuanto al nombre de los proyectos gestionados en STAR SL se compone de dos elementos obligatorios:

Nº proyecto_ Nombre cliente
-----------------------------

**Fig. 60: Sintaxis obligatoria para proyectos de traducción**

Los nombres de los proyectos de traducción, además de los dos elementos obligatorios pueden tener un elemento opcional. Un elemento opcional tiene por objeto proporcionar información adicional para discriminar la información

de los proyectos relativos a un cliente. De esta manera, dotamos al proyecto de un identificador adicional, opcional, que se aplicará según el criterio del gestor:

Nº proyecto_ Nombre cliente_ 3er elemento
---

**Fig. 61: Sintaxis para proyectos de traducción (1)**

Este tercer elemento puede identificar tres tipos de informaciones que enunciamos aquí:

- ❑ modelo del producto
- ❑ versión de la documentación
- ❑ número externo del proyecto

como se muestra en la sintaxis de los siguientes ejemplos:

8069_ Minolta_C450-COPY
7901_Alden_Menu
8901_Dremel_33572

**Fig. 62: Ejemplos de nomenclatura para proyectos de traducción (1)**

Y, por último, la nomenclatura de los proyectos puede contar con un cuarto elemento también de carácter optativo:

Nº proyecto_ Nombre cliente_ 3er elemento_4º elemento
---

**Fig. 63: Sintaxis para proyectos de traducción (2)**

Este cuarto elemento, al igual que el tercer elemento, identifica informaciones adicionales relativas al proyecto del tipo:

- ❑ módulo textual del proyecto

- subproyectos
- formatos de los documentos

En el caso relativo al tipo textual del proyecto podemos identificar los diferentes módulos textuales —acerca de la denominación de módulo textual y no tipo textual, ya nos hemos referido ampliamente en el apartado [2.2.5. Los tipos / módulos textuales de la documentación técnica](#)— mediante las siglas utilizadas en el ámbito de la localización, tal y como se muestra en los siguientes ejemplos:

<b>Modulo textual</b>	<b>Código identificativo</b>
software	_SW
readme	_READ
ayuda en línea	_HELP

**Fig. 64: Ejemplos de códigos identificativos de los módulos textuales**

En el caso de que creamos subproyectos a partir de los proyectos, sí que se debe añadir obligatoriamente el cuarto elemento que siempre es de tipo numérico. Mediante la identificación de los subproyectos podemos tener constancia de la pertenencia de un subproyecto a un proyecto concreto al que se le aplica la guía de estilo específica del cliente o las normas de trabajo establecidas para el proyecto o las tarifas. De esta manera, la coherencia en la nomenclatura dividida en subproyectos facilita el trabajo al departamento de administración, por una parte para facturar al cliente y, por otra, para efectuar los pagos a los autónomos.

Los subproyectos pueden ser de dos tipos:

- Proyectos de entrega por lotes: los clientes entregan por lotes todo el material para traducir, como se muestra en la siguiente sintaxis:



Nº proyecto_ Nombre cliente _ 3er elemento_ LoteA
Nº proyecto_ Nombre cliente _ 3er elemento_ LoteB
Nº proyecto_ Nombre cliente _ 3er elemento_ LoteC

**Fig. 65: Sintaxis para subproyectos de traducción (1)**

- Proyectos que se dividen a posteriori: los gestores dividen el material para traducir entre diferentes traductores, como se muestra en la siguiente sintaxis:

Nº proyecto_ Nombre cliente _ 3er elemento_ 01
Nº proyecto_ Nombre cliente _ 3er elemento_ 02
Nº proyecto_ Nombre cliente _ 3er elemento_ 03

**Fig. 66: Sintaxis para subproyectos de traducción (2)**

En el caso de que el cliente envíe material para traducir en diferentes formatos, entonces la extensión del formato de archivo en cuestión es el cuarto elemento de la nomenclatura y la sintaxis sería:

Nº proyecto_ Nombre cliente _ 3er elemento_ extensión formato
---

**Fig. 67: Sintaxis para proyectos de traducción (3)**

### **3.4.1.2. Nomenclatura de los diccionarios**

La nomenclatura de los diccionarios está condicionada por su organización por clientes. Son diccionarios que:

- envía el cliente,
- se crean, importan, modifican y actualizan en el proceso de la traducción para un cliente.

El nombre completo de los diccionarios almacenados por cliente se compone como mínimo de dos elementos obligatorios, esto es, el *nombre del cliente* y el *estado del diccionario* que abordaremos más adelante. De esta manera evitamos la confusión descrita en el apartado 3.3.3.4.1. Problemas de gestión de los recursos lingüísticos por falta de protocolos. El nombre del cliente del diccionario debe coincidir con el nombre del cliente que aparece en los proyectos para ese cliente concreto.

En la sintaxis de la nomenclatura de los diccionarios podemos añadir un elemento opcional que identifique, por ejemplo, un componente concreto de una aplicación de software o un producto o modelo determinado de un cliente. En este campo también se puede añadir la nomenclatura elegida por el cliente para el diccionario que envía. En caso de no tener esta información y tratarse de terminología genérica de un cliente sin especificaciones de componente, este campo tendrá la etiqueta GRAL —abreviatura de *General*— o una temática genérica como en el siguiente ejemplo:

MBW_ GRAL
MBW_ automocion

**Fig. 68: Ejemplos de nomenclatura para diccionarios**

La combinación lingüística de los diccionarios, si bien es una información imprescindible, se puede extraer de otras informaciones asociadas a ese cliente, como veremos más adelante. Por ese motivo, no es una información obligatoria que debe aparecer en la sintaxis de los nombres de los diccionarios. No obstante, si el gestor, por motivos pragmáticos, lo considera necesario, se incluye la combinación lingüística del diccionario en el nombre. Si se incluye esta información en el nombre es obligatorio utilizar los códigos de lengua estandarizados usados en *Transit* y *TermStar XV*.

Si un elemento de la nomenclatura está compuesto por más de una palabra, éstas se separan con guión [-] sin espacios, tal y como hemos formalizado en el protocolo para nombrar los proyectos de traducción.

#### **3.4.1.2.1. Clasificación de diccionarios por estado**

Tal y como hemos mencionado al iniciar este apartado, la nomenclatura de los diccionarios se compone de dos elementos obligados —*nombre del cliente* y el *estado del diccionario*—

El campo del estado de diccionario se denomina también tipo de diccionario y es fundamental para poder ponderar los diccionarios de forma rápida. En este sentido agrupamos los diccionarios en cinco categorías que describimos a continuación:

- Diccionario activo: diccionario creado por el gestor responsable del proyecto de traducción y modificado, alimentado y actualizado o por el gestor responsable del proyecto de traducción o por el traductor en plantilla. Parte de la terminología verificada por el gestor proviene de la fusión de diccionarios tipo *freelance*. En la nomenclatura de los diccionarios, este tipo se identifica con la abreviatura *ACT*.
- Diccionario freelance: diccionario creado por el gestor responsable del proyecto de traducción y alimentado y modificado por el traductor autónomo. Tiene un carácter temporal, pues su contenido es susceptible de fusionarse con los diccionarios tipo *ACT*. En la nomenclatura de los diccionarios este tipo se identifica con la abreviatura *FRL*. Asimismo se debe añadir antes la sigla del traductor. De esta manera, la codificación sería:

Nombre cliente \_ Componente \_ sigla traductor autónomo-FRL

**Fig. 69: Sintaxis para diccionarios FRL**

- Diccionario histórico: diccionario archivado y comprimido por el responsable del departamento de terminología. Este tipo de diccionarios no se utilizan para proyectos actuales del cliente. Pueden corresponder a:
  - antiguos clientes para los que desde hace más de dos años ya no se hacen más trabajos de traducción,
  - formatos ya no usados por el cliente; siempre se guardan comprimidos.

En la nomenclatura de los diccionarios este tipo se identifica con la abreviatura HIS.

- Diccionario de referencia: diccionario importado de Internet cuya temática está vinculada al contenido del proyecto de traducción para el que se usa. Los diccionarios de referencia se bajan o importan de Internet y están relacionados con áreas temáticas que previsiblemente pueden aportar terminología a proyectos de clientes relacionados con la materia. Estos glosarios pueden ser de:
  - empresas relacionados con los clientes en cuanto a materia,
  - centros de documentación específicos de un tema y no de una empresa.

Estos diccionarios los puede enviar el cliente, sin ser terminología específica del cliente o se crean e importan en el entorno de STAR SL. Los diccionarios de referencia también pueden ser todos aquellos diccionarios terminológicos en papel que hayan sido digitalizados —por ejemplo, escaneados— y que, por ello, se convierten en un recurso lingüístico que requiere de su almacenamiento estructurado. En la nomenclatura de los diccionarios este tipo se identifica con la abreviatura REF.

- Diccionario temporal: diccionario creado para hacer pruebas que lleva a cabo el responsable de la terminología. Tienen carácter temporal como los diccionarios *FRL*, pero no están creados por traductores autónomos. En la nomenclatura de los diccionarios este tipo se identifica con la abreviatura *TEM*.
- Diccionario validado: diccionario recibido y validado por el cliente. El traductor debe priorizar los términos verificados de este tipo de diccionarios. Tras el cambio de nomenclatura del diccionario, se debe registrar *obligatoriamente* la información del nombre original que el cliente ha dado al diccionario. No se permiten cambios en el nombre del diccionario, ni la modificación de sus registros. No se debe cambiar *nunca* al estado activo. En la nomenclatura de los diccionarios este tipo se identifica con la abreviatura *VAL*.

Puede ser que durante el trabajo traductológico se cuente con diferentes diccionarios catalogados por su estado. En estos casos la prioridad de la aplicación de los términos registrados se aplica de la siguiente forma:

<b>Tipos de estado</b>	<b>Sigla</b>	<b>Ranking uso</b>
Validado	VAL	1
Activo	ACT	2
Referencia	REF	3
Freelance	FRE	2
Temporal	TEM	-
Histórico	HIS	-

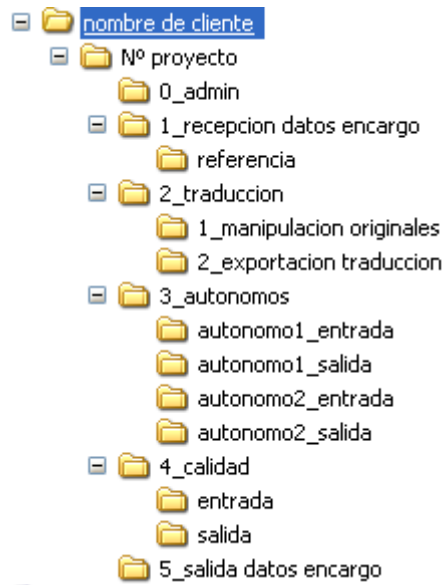
**Fig. 70: Estado de diccionarios: siglas y ponderación**

Algunos diccionarios —los temporales y los históricos— no son nunca susceptibles a ser definidos en el entorno de un proyecto de traducción, de ahí que aparezcan sin valor en el ranking de uso.

#### **3.4.1.3. Estructura de las subcarpetas de proyectos**

En respuesta a los problemas generados apuntados en el apartado 3.3.3.2.1. Flujo de almacenamiento y acceso físico a los documentos, se propone la organización sistemática de las estructuras de directorios, es decir, de carpetas y subcarpetas asociadas a la producción de los documentos de traducción.

Los archivos e instrucciones recibidos al principio y durante el transcurso de un proyecto se almacenan en la carpeta identificada con el número de proyecto que ocupa el segundo nivel jerárquico con respecto a la carpeta del cliente. La estructura de esta carpeta debe ser estándar. La estructuración de los datos sigue un modelo, como el propuesto:



**Fig. 71: Estructura de directorios estándar para un proyecto**

Nombre de carpeta	Significado y función
nombre de cliente	Contiene todos los proyectos gestionados para el cliente.
Nº proyecto	Contiene todos los directorios relacionados con el número de proyecto.
0_admin	Contiene información sobre presupuesto, hojas de cálculo para el seguimiento del proyecto, instrucciones y guías para el cliente, intercambio de correspondencia, comunicaciones con las personas de contacto del cliente, los autónomos, etc.
1_recepción datos encargo	Contiene los archivos mandados por el cliente que se deben traducir. Si el cliente además envía material traductológico adicional, se guarda en la carpeta <i>referencia</i> . El subdirectorio <i>referencia</i> contiene archivos con documentación adicional relacionada con el proyecto de traducción como, p.ej., traducciones anteriores.
2_traducción	Directorio principal de trabajo donde se crean y elaboran las traducciones, lo que es igual a las memorias de traducción. El subdirectorio <i>1_manipulación originales</i> se crea si se deben manipular archivos originales antes de importarlos a <i>Transit</i> , contiene, por tanto, los archivos manipulados. El subdirectorio

	<i>2_exportación traducción</i> se crea para almacenar los archivos exportados de Transit.
3_autonomos	Contiene archivos mandados por los autónomos en el subdirectorio <i>autonomo1_entrada</i> . Los archivos enviados a los autónomos también en su subdirectorio correspondiente <i>autonomo1_salida</i> . Estas carpetas pueden renombrarse con las siglas asignadas al autónomo.
4_calidad	Contiene archivos, aplicaciones, fotos de pantalla y todo el material utilizado para realizar el control de calidad. Los archivos en concreto para revisar se guardan en el subdirectorio <i>entrada</i> y los revisados en el directorio <i>salida</i> . El subdirectorio <i>entrada</i> contiene el material que se debe revisar; el subdirectorio <i>salida</i> contiene el material revisado.
5_salida datos encargo	Contiene los archivos finales que se mandan al cliente.

**Fig. 72: Leyenda de estructura de directorios**

No obstante, si la estructura, por requisitos específicos del proyecto, tiene que diferir de la estructura estándar, tiene que ser transparente y reconstruible para un tercero.

### 3.4.2. Modelado para la gestión de los diccionarios

El modelado que proponemos para la gestión de los diccionarios es una metabolización de lo recogido en los apartados 3.3.3.3. Necesidades de los usuarios y 3.3.3.4. Protocolo de cuellos de botella, incidencias o problemas. En función de esto hemos resuelto basar la estrategia en técnicas de gestión documental como:

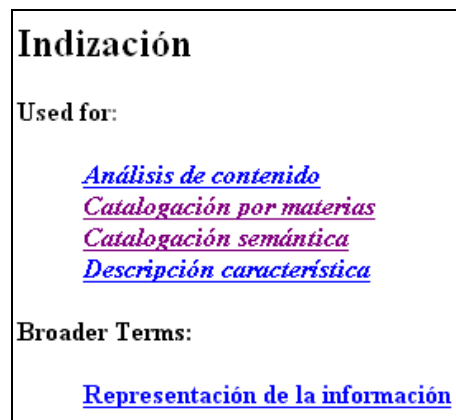
- la dotación de nuevos puntos de acceso adicionales a los diccionarios,
- la utilización de un sistema conceptual que nos sirva de lenguaje documental para la indización,



- el modelado de la estructura de un catálogo que atienda a las características de los datos que manejamos y las necesidades de recuperación.

### 3.4.2.1. Dotación de nuevos puntos de acceso

Los recursos lingüísticos deben estar dotados de más puntos de acceso para evitar su infrautilización generada por la distribución, clasificación y discriminación por cliente y proyecto. Según Martínez de Sousa (1993: 117) esta clasificación entraría en la definición de *catalogación descriptiva*, para diferenciarla de la *catalogación por materias*, también llamada *catalogación semántica*, según los autores (Bolognini y Pedrini, 1993: 25-28), así como *análisis de contenido*, *descripción característica* e *indización*, tal y como muestra el tesoro Docutes de la ciencias de la documentación de la Universidad de León, tras introducir la ecuación de búsqueda *Indización*:



**Fig. 73: Tesoro DOCUTES**

Así, la propuesta que presentamos atiende a la indización o catalogación semántica. Esta propuesta aborda la etiquetación semántica de las macroestructuras terminológicas para dotarlas de puntos de acceso temáticos y de esta manera llevar a cabo una recuperación de diccionarios que proporcione mayor satisfacción a los usuarios. Entendemos bajo *macroestructuras* los diccionarios considerados como documentos independientes. Esta definición de macroestructura amplía el concepto

descrito por otros autores (Díez, 2004: 424-425) donde la macroestructura concierne a la organización alfabética de los términos y a la organización de los conceptos en registros que aglutinan las diferentes equivalencias conceptuales, o en entradas que se corresponden a un solo concepto.

La identificación de estas macroestructuras como documentos las predispone a ser gestionadas aplicando técnicas de análisis documental. El análisis documental se define como:

Técnica formada por un conjunto de operaciones destinadas a la recuperación de información que tienen una función de intermediación en el proceso de circulación de datos entre el productor y el usuario (Garrido, 1996: 231).

Para esta *intermediación* se utilizan los productos resultantes del análisis documental, esto es, documentos secundarios que engloban dos aspectos que indicamos a continuación:

- la catalogación bibliográfica (autor, fecha de creación / publicación, formato),
- la catalogación semántica, mediante los identificadores de contenido del documento primario.

Los documentos secundarios son documentos de referencia a partir de los documentos originales también llamados primarios en la jerga especializada de biblioteconomía y documentación (Barite, 1996). Los documentos secundarios sirven para inventariar la existencia de documentos primarios y se caracterizan por ser un extracto mucho más manejable en cuanto al tamaño de su contenido y en cuanto a su sistematización en la organización de la información. El proceso de elaboración de los documentos secundarios que describen los documentos primarios se conoce bajo el término de *catalogación* y el catálogo se compone de los registros con sus respectivas informaciones de los documentos secundarios.

En nuestro caso los documentos secundarios no sólo deben identificar la ruta física u otras cuestiones no lingüísticas —aspecto conocido como *catalogación descriptiva*. La información contenida en estos documentos secundarios debe contar además con identificadores de contenido que sirvan de puntos de acceso a la información representados en forma de términos. Estos términos se denominan *descriptores*, si se trata de términos predefinidos y *palabras clave*, si se trata de términos seleccionados de forma libre. Estos identificadores semánticos son el resultado de la traducción de forma simbólica del conocimiento contenido en un documento.

Por ello, gestionar los contenidos supone gestionar el conocimiento representado en los archivos, como ya aludíamos en el apartado 2.6.2.2. Gestores documentales → Gestores de contenido.

Para poder controlar la gestión de los contenidos se debe proceder a la catalogación semántica de los diccionarios. Para ello, se prevé contar con un lenguaje de indización compuesto de un lenguaje documental predefinido, por tanto, controlado y con términos de selección libre, tema al que dedicamos el siguiente apartado.

### **3.4.2.2. Lenguaje de indización**

En el apartado 2.2.3.1. Recursos lingüísticos: Texto y estructuras conceptuales hemos abordado la representación textual del conocimiento a través de sistemas conceptuales compuestos de elementos terminológicos. El sistema conceptual propuesto para nuestros fines debe, por una parte, abarcar todo el conocimiento experto almacenado en los recursos lingüísticos y, por otra parte, debe establecer las coordenadas para una futura recuperación automatizada no solo de los diccionarios, sino también de las memorias de traducción —la forma de aplicación del sistema conceptual para indizar las memorias de traducción lo abordamos más adelante en el apartado 3.4.3. Modelado para la gestión de las memorias de traducción.

Este lenguaje de indización debe permitir el acceso a la información de una forma más intuitiva para optimizar la gestión. Los modelos conceptuales abocados a la organización del lenguaje aplicados en diversos entornos de trabajo se dividen en dos categorías (Fernández Cepedal, 1976: 47-52):

- lingüístico-epistemológica
- material-ontológica

Los sistemas lingüístico-epistemológicos son sistemas conceptuales constituidos por piezas lingüísticas de carácter léxico que representan la estructuración del conocimiento. Tienen carácter específico en contraste con el carácter universal pretendido por los modelos conceptuales ontológicos, que explicaremos después. Este carácter de especificidad se manifiesta en las posibles descripciones de estructuras aplicables a un ámbito de conocimiento experto delimitado para su estructuración. La estructuración del conocimiento científico y su representación son el objeto de estudio de la epistemología.

Estas representaciones de las estructuras del conocimiento especializado son codificaciones a partir de criterios concretos de categorización de objetos realizadas por sujetos inmersos en una realidad —experiencia individual, experiencia corporativa— que determinan los procesos epistemológicos de descripción. Las epistemologías se nutren, por un lado, de entidades discursivas y, por otro, de una entidad preconstituida, es decir, de un sistema conceptual que, al mismo tiempo, es el conocimiento sintetizado que constituyen las estructuras para poder gestionar el conocimiento representado. Por tanto, la tarea epistemológica es una operación de correlación entre la representación y la abstracción del conocimiento con la realidad.

Los modelos conceptuales lingüísticos y epistemológicos más populares son:

- vocabularios, glosarios y diccionarios especializados, que son listados de conceptos compuestos de términos vinculados con un ámbito de conocimiento especializado. Estos términos se representan alfabéticamente.
- tesauros y redes semánticas almacenadas en bases de datos léxico-conceptuales compuestas de términos y de relaciones entre los términos específicos de un ámbito de conocimiento especializado. Los términos se organizan de forma sistémica.

Ambas clases pueden servir de modelo conceptual referencial para catalogar semánticamente, es decir, para la indización. Esto significa que este conocimiento semántico es previsiblemente el idóneo para constituir la base de conocimiento de un posible sistema de gestión de información.

La segunda categoría apuntada son los sistemas conceptuales del tipo material-ontológico, constituidos también por piezas lingüísticas. Éstos pretenden reflejar las categorías de la realidad mediante la esencia de los elementos. Grossmann (1992) define la disciplina de la ontología como la rama de la filosofía que tiene como objetivo responder cuáles son las categorías del mundo y qué leyes regulan esas categorías. La paradoja de la ontología es que las categorías de la realidad deben representarse de forma lingüística, lo que equivale a una codificación humana no exenta de interpretación. El proyecto Wordnet es un ejemplo de implementación de un sistema conceptual con aspiraciones ontológicas. La ontología de Wordnet se compone de elementos ontológicos, concretamente de primitivos semánticos que, organizados jerárquicamente entre sí, agrupan significados (Díez, 1999: 7-2). El fundamento teórico del sistema tiene su origen en la idea de primitivos semánticos que conforman un vocabulario matriz en el que se basa el sistema (Miller, 1986, 1993, citado en Moreno, 2000: 2-4-2).

Nuestra propuesta de modelado se basa en un lenguaje de indización predefinido del tipo tesoro con palabras clave asociadas de selección libre.

Siguiendo a Fernández Cepedal, ya citado más arriba, este sistema sería del tipo lingüístico-epistemológico.

El lenguaje de indización —concepto conocido también como *lenguaje documental*— es un conjunto de conceptos normalizado y consensuado de términos que ayudan a identificar el contenido de documentos, catalogándolos a fin de poder recuperarlos después. Estos lenguajes pueden ser meras listas de términos o sistemas conceptuales de diferente sofisticación, dependiendo en este último caso de la estructura de las relaciones que se establecen entre los términos.

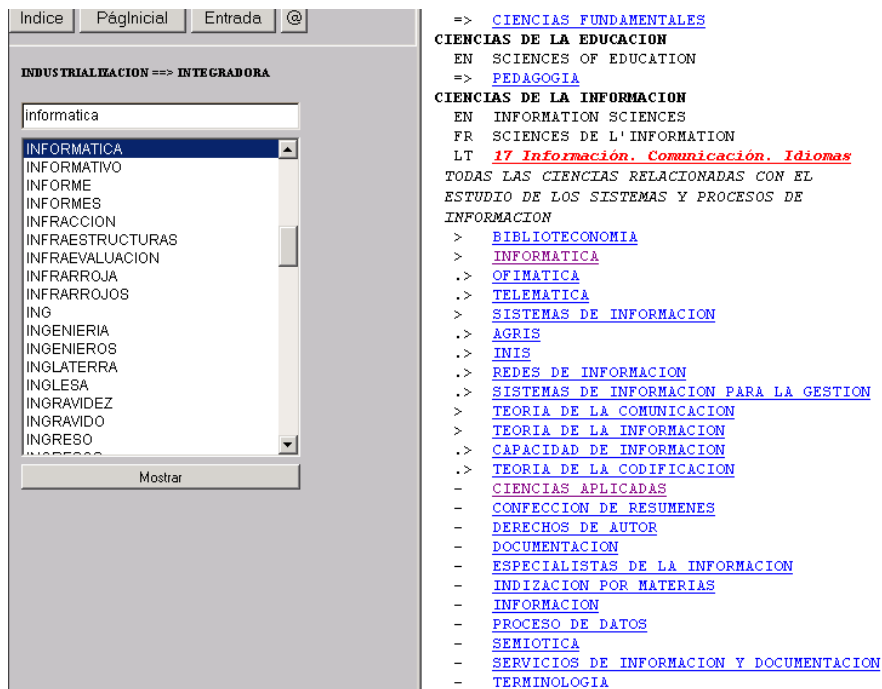
La decisión de guiarnos por un modelado conceptual híbrido —lenguaje de indización de tipo tesoro y asociación de palabras claves— está motivada por la necesidad de crear pautas comunes que armonicen sin dejar de atender las anotaciones personales de los usuarios que pueden describir de una forma más acertada los documentos mediante términos que no aparecen en el tesoro. Las anotaciones deben llevarse a cabo por dos tipos de usuarios:

- los traductores que llevan a cabo la gestión de los proyectos de traducción, es decir, gestores de información improvisados y
- el/los profesionales de la información con formación en terminología y en documentación para que a partir de los términos de selección libre, introduzca los descriptores formalizados extraídos del lenguaje de indización predefinido (Cámara, 2005a).

Asimismo la decisión de seleccionar un tesoro está inspirado en las ventajas —apuntadas en el apartado 2.6.2.1. Gestores de tesauros multilingües— de aplicar tesauros para gestionar los volúmenes ingentes de información.

El lenguaje de indización seleccionado es el tesoro Spines, creado en 1972 —la primera versión española es de 1984— por la UNESCO. Este tesoro

sigue manteniéndose en la actualidad con el objetivo de servir de instrumento para la identificación y transferencia del conocimiento (Irazazábal, 2000: 127). La elección de este tesoro está motivada por su carácter internacional, público, multilingüe, representativo para todos los ámbitos de conocimiento y por su accesibilidad en línea a través de Internet. Su estructura morfológica es terminográfica y su carácter tesauro lo diferencia de un glosario o diccionario terminológico debido a la estructura de relaciones semánticas que se establecen entre los términos. Las relaciones jerárquicas más comunes son las de hiperonimia y de hiponimia, siendo las relaciones del tipo asociativo, relaciones transversales no horizontales, las que nos posibilitan el salto de un campo semántico a otro:



**Fig. 74: Tesoro Spines de acceso público**

La catalogación semántica de los diccionarios se prevé manual. Y en el proceso debe participar primero el gestor de traducción que prepara los recursos terminológicos para un proyecto de traducción concreto. Este debería ser el encargado de proporcionar las palabras clave que, según su criterio, identifican el contenido del diccionario en cuestión. Estas palabras

clave deben servir de referente a la persona encargada para la selección de los términos descriptores predefinidos del Spines.

Tanto las etapas del proceso de catalogación semántica como el perfil humano necesario para llevar a cabo la catalogación semántica de los diccionarios, es un tema que tratamos más adelante en el apartado 4. 3. 1. Formación continua como impulsador de la gestión del conocimiento procedimental en el marco de descripción de los agentes implicados en el mantenimiento y seguimiento de COGNmática, el sistema implementado y descrito en el cuarto capítulo.

### **3.4.2.3. Diseño del gestor de diccionarios**

Una vez se decida la dotación de más puntos de acceso y el lenguaje de indización, debemos decidir la estructura del almacén donde se van a registrar los datos. Para poder tener a disposición una biblioteca virtual de los recursos lingüísticos almacenados internamente —necesidad expuesta por los usuarios—, proponemos la implementación de un catálogo, instrumento básico de gestión en el ámbito de biblioteconomía, de acceso digital.

Un catálogo es una relación ordenada de documentos secundarios en la que se incluyen o describen de forma individual libros, documentos, etc., que están relacionados entre sí. Un catálogo tiene su formalización electrónica en una base de datos, siendo el conjunto de los documentos secundarios una colección ordenada de registros estructurados en función de unos criterios concretos. Ahora bien, la estructura de un catálogo de documentos o base de datos documental no siempre es igual y eso es porque, siguiendo la subsecuente cita:

La naturaleza de la información determina la organización de los datos que se van a almacenar y procesar, así como las formas en las que éstos se van a recuperar. Y todo ello, a su vez, se encuentra condicionado por las características y necesidades del colectivo que van a hacer uso de esos datos (Moscoso, 1996: 393).



En este sentido, la estructura del catálogo de los diccionarios registrados debe prever las siguientes opciones de búsqueda derivadas también de las necesidades de los usuarios:

- ❑ nombre de diccionario
- ❑ cliente
- ❑ materia
- ❑ combinación de campos

También se prevé poder filtrar la información en función del formato de almacenamiento del diccionario.

Asimismo, se prevé expandir el resultado de las consultas accediendo a toda la información asociada a cada diccionario. Es decir, a todas las informaciones de cada registro, como si se tratara de la ficha documental de cada diccionario y obtener las siguientes informaciones:

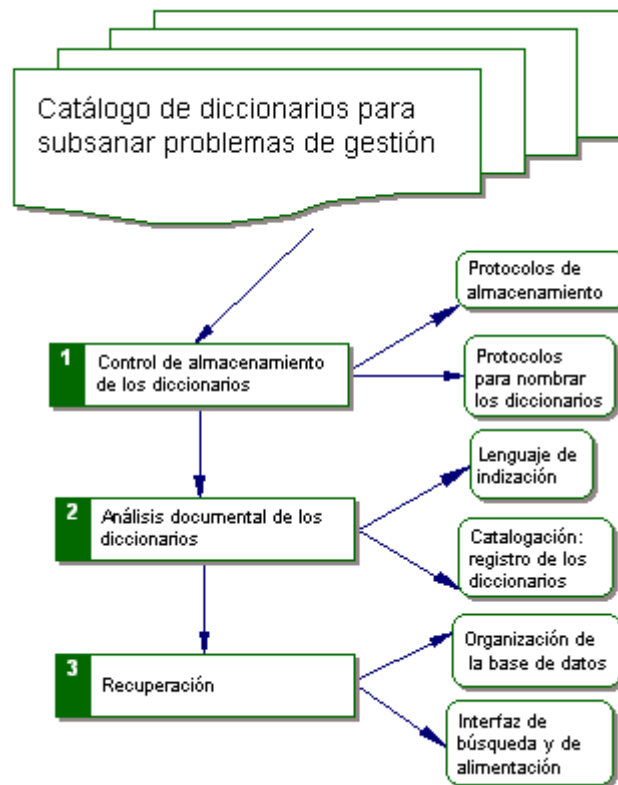
- ❑ ubicación física
- ❑ tema de traducción
- ❑ descriptores
- ❑ palabras clave
- ❑ lenguas del diccionario
- ❑ tipo de diccionarios según su estado
- ❑ comentarios y sugerencias de mantenimiento

Y por último, debe considerarse el registro (edición) de los diccionarios y sus informaciones asociadas.

Previsiblemente, los instrumentos de interacción para la gestión de búsqueda, consulta y registro de las informaciones apuntadas son una interfaz de búsqueda y una interfaz de alimentación, como veremos en el

capítulo 4. Diseño e implementación de un sistema integrado de gestión de conocimiento y de recursos lingüísticos: COGNimática.

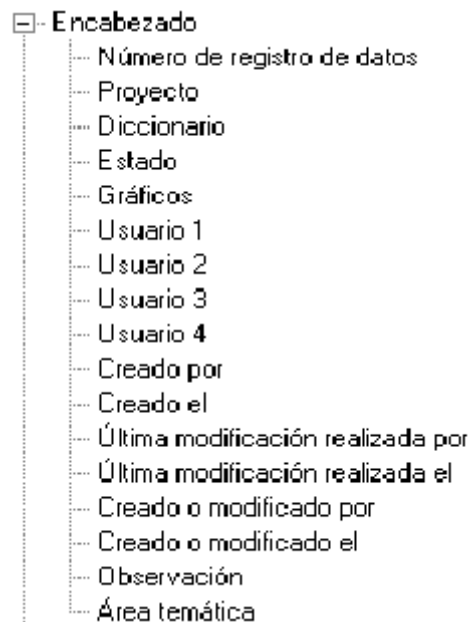
Resumiendo, las propuestas que abarcan los problemas de gestión de los diccionarios se centran en el diseño de un catálogo cuya implementación debe llevarse a cabo en tres fases con sus respectivas tareas asociadas, tal y como se deriva del siguiente diagrama de flujo:



**Fig. 75: Flujo propuesto para un sistema de gestión de diccionarios**

El sistema de gestión de diccionarios propuesto debe contribuir al almacenamiento de forma controlada de las macroestructuras de los diccionarios atendiendo especialmente a un marcado semántico para la mejora de su posterior recuperación. Planteamos llamar a esta propuesta *gestor documental de diccionarios* para diferenciarlo de los gestores de terminología como *TermStar*, el gestor terminológico integrado en *Transit*. Los gestores terminológicos gestionan, ante todo, las microestructuras de los diccionarios, es decir, las fichas de cada registro de un diccionario. Ciertamente es

que dentro de los campos de las microestructuras se encuentran, además de los campos específicos para cada término con sus equivalencias conceptuales —abreviaciones, sinónimos, acrónimos, equivalencias conceptuales en otras lenguas, lengua, variantes geográficas, tipográficas, etc.— y su información semántica —definición, contexto—, gramatical y administrativa —autor, fecha de creación, modificación, estado del término, etc.—, información genérica específica del registro aplicable a todas las representaciones terminográficas de cada concepto, como se muestra en la siguiente imagen:



**Fig. 76: Sección de la estructura del encabezado de un registro terminográfico de *TermStar***

No obstante, esta información, aunque importante, sigue atañendo a la microestructura sin la posibilidad de gestionar los diccionarios como documentos independientes.

### **3.4.3. Modelado para la gestión de las memorias de traducción**

Tomando como referencia el procedimiento propuesto para la gestión de los diccionarios, proponemos su aplicación en la gestión de las memorias de traducción con la aspiración, sin embargo, de que todo el proceso de catalogación semántica sea automático, en contraste con la catalogación semántica de los diccionarios que, como hemos visto arriba, se propone manual. El siguiente apartado, pues, lo dedicamos a abordar la dotación de nuevos puntos de acceso a las memorias de traducción para seguir con la estructura que debe tener el almacén de las memorias de traducción para su explotación.

#### ***3.4.3.1. Dotación de nuevos puntos de acceso***

Como ya se ha venido viendo, la gestión de las memorias de traducción no se puede tratar de forma independiente de la gestión de los proyectos de traducción. Y recordemos, que esto es así, porque en el marco de un proyecto de traducción se generan nuevas memorias de traducción y reutilizan otras generadas a su vez en el marco de otros proyectos de traducción, como hemos visto en el apartado 2.4.1. Las memorias de traducción. Si bien el caso de los diccionarios puede considerarse de la misma manera, el hecho de que puedan utilizarse como unidades documentales autónomas de referencia les predispone a un tratamiento independiente de un proyecto concreto, tal y como ya hemos propuesto. Ahora bien, los diccionarios en cuanto unidades referenciales de un proyecto de traducción son portadoras de información reciclable para todo el material asociado al proyecto. En este sentido, la propuesta que planteamos para la gestión de las memorias de traducción es la siguiente: si las etiquetas semánticas —descriptores y términos clave— que indizan los diccionarios son identificadores del contenido, previsiblemente los descriptores y términos clave de los diccionarios asociados a un determinado proyecto se podrán

reutilizar para la catalogación semántica de las memorias de traducción. Esto implica desarrollar una estrategia de reutilización de la catalogación semántica de los diccionarios, lo que significa que las memorias de traducción se indizarán con el mismo sistema conceptual que se ha propuesto en el apartado 3.4.2.2. Lenguaje de indización.

### **3.4.3.2. Diseño del gestor de proyectos**

La estrategia de reutilización concebida pretende asociar los nombres de los diccionarios a proyectos de traducción. Esto significa que debemos lograr que uno de los campos previstos para el catálogo de diccionarios pueda vincularse a la información externa al catálogo relacionada con los proyectos de traducción. La información correspondiente a cada uno de los proyectos de traducción gestionados con *Transit* se guardan en los archivos \*.prj, que, como recordamos, un archivo \*.prj contiene informaciones vinculadas a un proyecto de traducción —véase *Fig. 42: Archivo \*.prj abierto con el Bloc de notas*— como:

- ❑ combinación de lenguas
- ❑ nombre del proyecto
- ❑ cliente
- ❑ archivos de trabajo
- ❑ memorias de traducción de referencia
- ❑ diccionarios asociados al proyecto

A pesar de que estos archivos \*.prj contienen información relativa a los diccionarios vinculados a cada proyecto, ésta no se utiliza para buscar los diccionarios asociados, así como tampoco otros recursos lingüísticos como las memorias de traducción; por tanto está infrautilizada. Los archivos \*.prj no están previstos para ser explotados como unidades documentales normalizadas que especifiquen los datos relacionados con su forma y

contenido para facilitar la identificación y ubicación, en una palabra, su acceso. Sintetizando, pretendemos con nuestra propuesta reutilizar los contenidos de los archivos \*.prj como unidades de gestión documental.

Por tanto, haciendo un paralelismo con la descripción de documento secundario, identificamos estos archivos \*.prj como los documentos *secundarios* idóneos que nos deben servir para gestionar los documentos primarios.

Todos los documentos primarios, esto es, los relacionados con un proyecto de traducción como los documentos originales, los documentos donde se generan las memorias de traducción, las memorias de traducción ya realizadas y seleccionadas como material de recursos lingüísticos reutilizables y, por último, los diccionarios que se reutilizarán y alimentarán durante el proyecto de traducción, se almacenan físicamente en estructuras de carpetas y archivos no avaladas por una solución de gestión documental, a pesar de que la gestión ordenada del registro de las informaciones relevantes las predispone a ser gestionadas de forma automática.

Nuestra solución es el diseño de un gestor de documentos asociados a un proyecto. Este gestor de proyectos debe gestionar las memorias de traducción extrayendo información contenida en los \*.prj y con ello generar un catálogo que se cree y actualice de forma automática. Por tanto, utilizamos los \*.prj como unidades de gestión documental de donde se extraigan las informaciones pertinentes que favorezcan la alimentación de un catálogo de creación, edición y actualización automática. La información asociada a un archivo \*.prj es un registro del nuevo catálogo. Asimismo, la información que se extraiga de los \*.prj debe satisfacer las necesidades de los usuarios y atajar los problemas analizados. En este sentido, este catálogo debe reutilizar la información de catalogación semántica existente en el catálogo de diccionarios. Para ello se deben asociar ambas bases de datos. El punto de unión de esos catálogos es el nombre(s) de diccionario(s), ya que esa información se recoge en uno de los campos del registro de un

diccionario catalogado y en el archivo \*.prj donde entre todas las informaciones asociadas al proyecto se encuentran también los diccionarios asociados a un proyecto de traducción. Puesto que los proyectos de traducción no se catalogan semánticamente, se reutiliza la marcación semántica de los diccionarios asociados a cada proyecto. De esta forma podemos hacer búsquedas más intuitivas por materia del proyecto sin dotar a los datos de conocimiento semántico.

La implementación de esta propuesta debe contribuir a la mejora de la recuperación de los recursos lingüísticos almacenados porque se explotará la información contenida en el gestor de proyectos, organizada y anotada semánticamente.

Si al catálogo lo dotamos de funciones adicionales, podemos convertirlo en un gestor de corpus paralelo para explotar mejor los recursos lingüísticos almacenados. Y esto sería previsiblemente posible porque el catálogo:

- ofrece información acerca de los documentos almacenados,
- posibilita el acceso a los documentos completos correspondientes a las memorias de traducción gestionadas.

Por ello, la estructura del catálogo la abocamos a la gestión de corpus paralelos.

#### ***3.4.3.3. Diseño del gestor de corpus paralelos***

Por lo expuesto anteriormente, la estructura del gestor de proyectos de traducción se diseña para ser concebida como un gestor de corpus paralelos. Por tanto, el gestor propuesto es:

- un instrumento de catalogación automática de los archivos gestionados,

- ❑ un gestor que presenta todas las memorias de traducción de las que disponemos en forma de corpus paralelos para explotar mejor el capital lingüístico.

La mejora del acceso a las memorias de traducción, es decir, al corpus paralelo, posibilita a los usuarios explotar internamente las bases traductológicas. Las ventajas genéricas de los corpus paralelo en la producción de traducciones ya han sido descritas en anteriores estudios, a saber:

Los traductores que han tenido la oportunidad de tener como material de referencia un corpus electrónico especializado en el idioma materno mejoraron su productividad y cometieron menos errores de contenido, de selección terminológica y de estilo (Bowker, 1998 citado en Vargas, 2002: 473).

La estructura de este gestor debe prever el acceso a todas las informaciones registradas y relacionadas con los proyectos que puedan ser relevantes para los usuarios, como la búsqueda de las memorias de traducción registrados por:

- ❑ nombre de proyecto
- ❑ nombre de diccionario
- ❑ nombre de gestor
- ❑ nombre de cliente
- ❑ fechas de modificación
- ❑ categorización semántica: materia (recuperación automática de información de los diccionarios relacionada con proyectos de traducción)
- ❑ combinación de campos



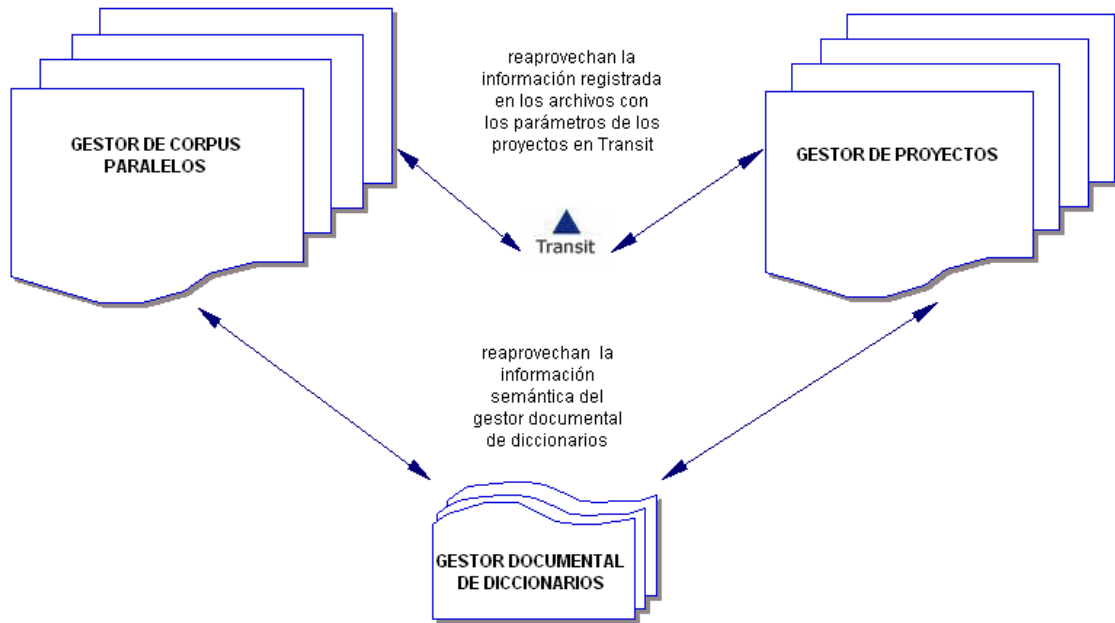
Los resultados de las búsquedas por los criterios apuntados deben constituir el paso previo para luego acceder a los documentos más pertinentes, tras haber analizado las informaciones asociadas a las memorias de traducción de una primera selección. Recordemos que uno de los problemas apuntados se refería a las herramientas de búsqueda utilizadas para la explotación de los recursos lingüísticos almacenados cuya captura por concordancias no proporciona el término equivalente en el otro idioma. Para solventar este problema proponemos dotar al gestor de las memorias de acceso completo a los textos paralelos.

Para ello, a partir de los registros del catálogo recuperados debemos contar con la posibilidad de filtrar nuevamente los registros capturados para marcar y abrir los proyectos seleccionados mediante una segunda interfaz, esta vez de selección, en contraste con la de consulta propuesta arriba.

La razón para proponer una herramienta para generar subcorpus está vinculada al aumento del rendimiento que puede ofrecer la explotación de los recursos lingüísticos con una acotación más precisa, evitando así el posible ruido resultante tras una búsqueda.

Como se puede deducir de lo expuesto, el diseño del gestor de proyectos y corpus paralelos es de evolución secuencial. No obtendremos un gestor de corpus paralelos sin la información recogida en el gestor de proyectos. Y no obtendremos información del gestor sin la ayuda de:

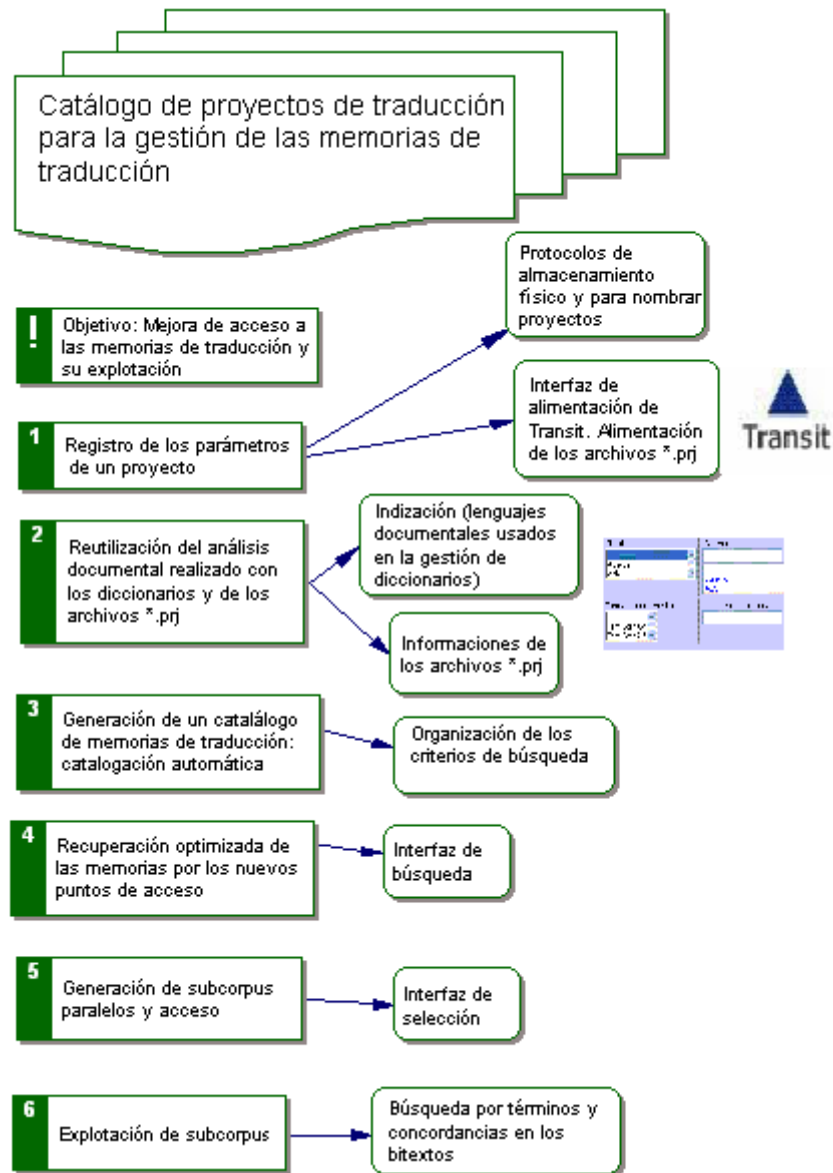
- el gestor documental de diccionarios y
- Transit



**Fig. 77: Sistema integrado de gestión de conocimiento y recursos lingüísticos**

El sistema a través de todos sus módulos debe servir para gestionar la biblioteca virtual de recursos lingüísticos almacenados en la empresa de producción de documentación técnica multilingüe analizada utilizando el conocimiento semántico registrado por los usuarios y explotado por la máquina.

En el siguiente gráfico se han identificado las etapas previstas para el desarrollo del gestor de proyectos y corpus paralelos con sus respectivas tareas asociadas:



**Fig. 78: Flujo propuesto para un sistema de gestión de proyectos y de explotación de corpus paralelos**

La estructura de un catálogo y su generación vuelve a ser de suma relevancia en nuestra propuesta, ya que al entorno de éste se derivan propuestas y soluciones para la mejora de la explotación de corpus paralelos.

Resumiendo, la explotación de los corpus paralelos se debe basar en la catalogación manual que se debe llevar a cabo con los diccionarios para poder reutilizarse *automáticamente* por el gestor de corpus paralelo. Asimismo, la propuesta de diseño contempla la presentación en forma de

corpus paralelos de todas las memorias de traducción de las que dispone STAR SL. De esta manera, se debe lograr una mejor explotación del capital lingüístico para la búsqueda de terminología o soluciones traductológicas o para conseguir información acerca de las memorias de traducción pertinentes para cada proyecto.

Para llegar a ello se ha aplicado como método de análisis una auditoría de la información para inventariar los recursos lingüísticos, determinar los flujos de gestión de los recursos lingüísticos tanto a nivel individual como colaborativo y analizar las herramientas de gestión de los recursos empleadas para asistir al usuario.

La auditoría de la información como técnica de análisis no se había aplicado hasta ahora en el análisis de los recursos lingüísticos, lo que supone indiscutiblemente una novedad en la metodología aplicada. Las razones para la aplicación de una auditoría de la información son dos: la primera se puede justificar fácilmente, ya que, esta metodología confiera a los recursos lingüísticos el valor de activo empresarial; la segunda razón estriba en la propia naturaleza de nuestro objeto de estudio, que son las necesidades de optimización en la gestión de las memorias de traducción y las bases terminológicas y, por tanto, también sus informaciones asociadas.

El siguiente capítulo muestra los resultados de nuestro sistema integrado implementado, así como la última fase de la auditoría correspondiente al mantenimiento y seguimiento tanto de una política de la información, como del sistema implementado.

#### **4. Diseño e implementación de un sistema integrado de gestión de conocimiento y recursos lingüísticos: COGNImática**

En las diferentes secciones del capítulo anterior se han expuesto varias de las fases del proceso auditor. En este sentido, se ha tratado: en primer lugar, la necesidad de implantar una política de la información y la dificultad de evaluar el proyecto de auditoría en cuanto a coste-beneficio; en segundo lugar, se ha abordado la relevancia de la planificación de la auditoría de la información para establecer los instrumentos con los que se lleva a cabo el análisis e interpretación de los datos; por último, se han desarrollado protocolos de trabajo y se ha diseñado el modelado conceptual para la gestión optimizada de los diccionarios y de las memorias de traducción a partir de las propuestas conceptuales previstas como determinantes para el diseño de un sistema integrado de gestión de conocimiento y de recursos lingüísticos.

En este cuarto capítulo se abordan las dos últimas fases de la auditoría donde, además de mostrar el diseño final del sistema propuesto, COGNImática<sup>1</sup>, y su implementación —penúltima fase de la auditoría—, se cierra el proceso auditor mediante el establecimiento de las estrategias para hacer un buen seguimiento y el mantenimiento del proceso metodológico empleado.

Así pues, especificaremos las bases técnicas y funcionales del sistema, los requerimientos de contenido, los resultados de los betatesters y las propuestas para mejorar los flujos de trabajo del sistema implementado a partir de la organización secuencial de las etapas de desarrollo de COGNImática.

A continuación mencionamos las fases de desarrollo de COGNImática establecidas a partir de las etapas de desarrollo de cualquier sistema informático y que nos han servido de referencia:

- diseño: especificaciones técnicas y funcionales
- desarrollo de la primera versión
- pruebas
- rediseño y desarrollo de la versión afinada
- seguimiento y mantenimiento

Y, finalmente, para cerrar el capítulo se tratará la última parte de la auditoría, que corresponde al seguimiento y mantenimiento de la gestión de los

---

<sup>1</sup> Al sistema lo hemos denominado COGNImática. Este neologismo es una forma de reducción léxica, morfológicamente hablando, formado a partir de dos términos identificativos del sistema: *conocimiento* y *automatización*, ambos implícitos en el neologismo propuesto. El término está inspirado en el término *tradumática* que también se acuña para identificar la inmersión de la *informática* en el proceso de la *traducción* (véase *Tradumática*, 2001); asimismo el término *informática* también nace de un proceso de reducción léxica — recordemos que apareció en Francia en 1962 uniendo las palabras *information* y *automatique*, de ahí se llevó a cabo una analogía al español a partir de la fusión de las palabras *información* y *automática*—.

recursos lingüísticos. Este tema se abordará desde la revisión y mantenimiento de una política de la (in)formación que abarca a los agentes implicados, las unidades de información tratadas y, por último, COGNlímica, el sistema de información implementado en el entorno de STAR SL.

Asimismo, se enunciarán las perspectivas de mejora a corto y medio plazo para seguir optimizando la gestión de los recursos lingüísticos, su conocimiento asociado, así como las propuestas técnicas de afinamiento del sistema implementado.

#### ***4.1. Fase de diseño: especificaciones técnicas y funcionales***

El desarrollo del sistema COGNlímica, como cualquier herramienta, no comienza con el diseño e implementación de la arquitectura de la información y sus recursos de interacción —interfaces gráficas, sistema de navegación—. Todo tipo de desarrollo de un sistema se debe iniciar con la detección de un problema, de un cuello de botella en un flujo de trabajo específico, como se planteaba en el apartado 1.1. Problema del presente trabajo. Los objetivos del sistema se trazan en función del problema principal y de las necesidades de los usuarios que integran el equipo recogidas mediante los instrumentos específicos para este cometido como la encuesta<sup>1</sup>.

Así, el sistema ha tenido como principal objetivo dar soporte a la gestión automatizada de recursos lingüísticos para los proyectos de traducción para asistir a los usuarios en el almacenamiento, la recuperación, la catalogación, la selección y la generación de subcorpus de recursos lingüísticos para su explotación como base de conocimiento traductológico. Con ello se ha pretendido además la mejora en la gestión del conocimiento corporativo a

---

<sup>1</sup> Este tema ya se ha tratado genéricamente en el apartado 3.3.3. Análisis e interpretación de los datos y concretamente en el apartado 3.2.1.4. Análisis e interpretación de los datos para modelar un catálogo de propuestas técnicas y conceptuales.

través de la realización de la diversidad de las tareas mencionadas. En este sentido, el sistema COGNImática se ha pensado como un sistema de gestión de conocimiento<sup>1</sup> que gestiona información de procedimiento y de contenido con el fin de explotar otras bases de conocimiento, las traductológicas y terminológicas.

Este gestor de conocimiento se ha concebido para fomentar la comunicación y la automatización de tareas de gestión documental y explotación de recursos lingüísticos, porque los recursos lingüísticos que se hallan contenidos en los documentos gestionados contienen conocimiento y los flujos de trabajo representados que asisten en las tareas también (base de conocimiento y flujos de procedimiento) y, por último, el producto resultante, es decir, el sistema, es un producto desarrollado que se fundamenta en el conocimiento (experiencia) del usuario y en el conocimiento del analista<sup>2</sup>.

En los siguientes apartados mostramos, por una parte, el diseño e implementación de la arquitectura de la información donde se establecen los requerimientos de contenido y las especificaciones funcionales y, por otra, las bases técnicas y funcionales que han determinado la implementación del conjunto del sistema.

#### **4.1.1. Diseño e implementación de la arquitectura de la información**

Los prolegómenos de un sistema informático se originan a partir de la necesidad de resolver un problema; asimismo, el prediseño de COGNImática

---

<sup>1</sup> En un sistema de gestión de conocimiento los esfuerzos de gestión van encaminados a cómo generar conocimiento y no perderlo.

<sup>2</sup> En este caso se trata del conocimiento de la autora de este trabajo.



se deriva como consecuencia de un análisis exhaustivo de los procesos existentes y de las necesidades de los usuarios<sup>1</sup>.

Como se ha visto en el apartado 3.4. Diseño e implementación de las propuestas metodológicas, conceptuales y tecnológicas, los análisis de los datos y las informaciones de los usuarios nos proporcionan fundamentos sobre los que reformulamos el ecosistema informacional que se pretende gestionar. Este ecosistema pronostica qué tipo de instrumentos se emplearán para llevar a cabo las funciones principales y, así, satisfacer los requisitos de los diferentes usuarios.

En este sentido, se ha previsto un sistema compuesto, en primer lugar, de un catálogo de diccionarios cuyos elementos de interacción son una interfaz de consulta, desde donde se recupera información relativa a los diccionarios, y de una interfaz de alimentación, donde se catalogan los diccionarios y se controla su almacenamiento. En segundo lugar, hemos diseñado un catálogo adicional de proyectos compuesto de una interfaz de consulta desde donde se recuperan informaciones relativas a los proyectos y a los diccionarios. Y en tercer lugar, se ha esbozado una tercera herramienta compuesta de una interfaz de selección desde donde se accede a las memorias de traducción almacenadas con el fin de extraer las concordancias deseadas por el usuario.

Igualmente, hemos decidido que todos los usuarios no deben realizar todas las funciones, por ello, se han determinado tres perfiles de usuario que corresponden a los diferentes permisos de acceso que debe proporcionar el sistema:

- primer nivel: sólo consulta
- segundo nivel: consulta y alimentación parcial de los registros

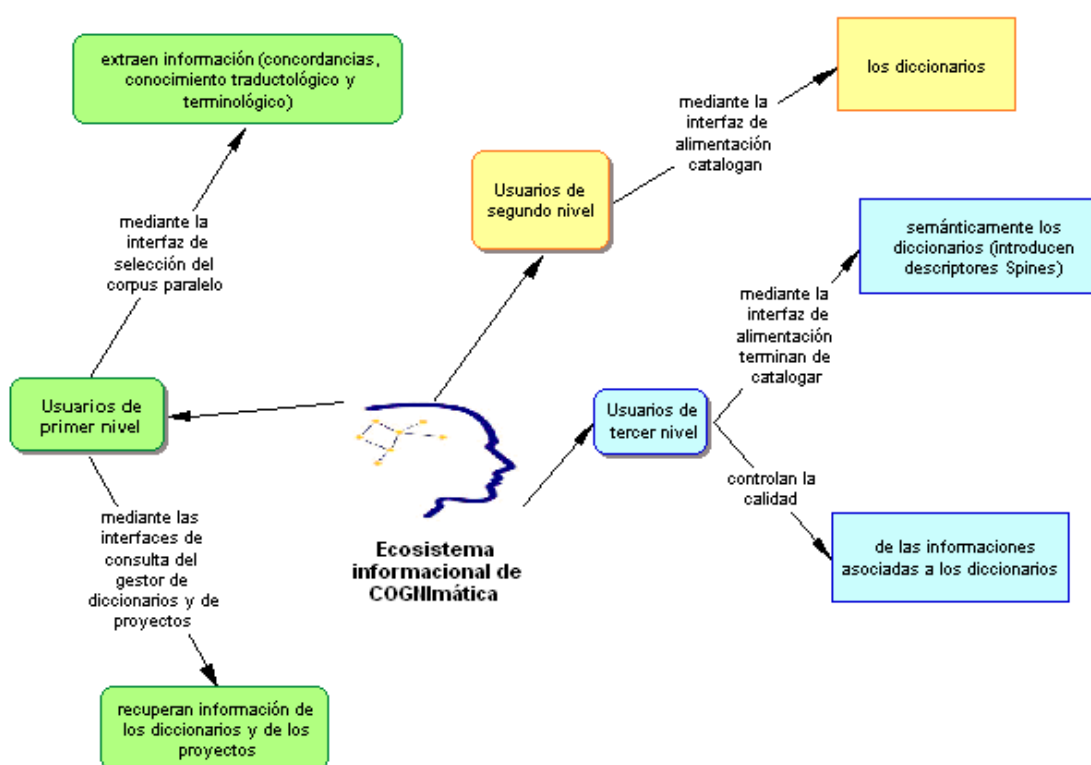
---

<sup>1</sup> Ambas informaciones se tratan específicamente en todas las secciones del apartado 3.3.3. Análisis e interpretación de los datos.

- tercer nivel: administración del sistema con permisos para introducir datos en todos los campos del registro, modificarlos o eliminarlos

El usuario de cualquier nivel debe poder predecir la respuesta del sistema ante su acción, para ello el nombre de los enlaces y componentes de interacción son significativos y precisos.

Las especificaciones de información y la tipología de los usuarios decididas, conforman el siguiente mapa conceptual donde se representa la arquitectura de la información del sistema y, más aún, el ecosistema informacional que gestiona COGNlímica:

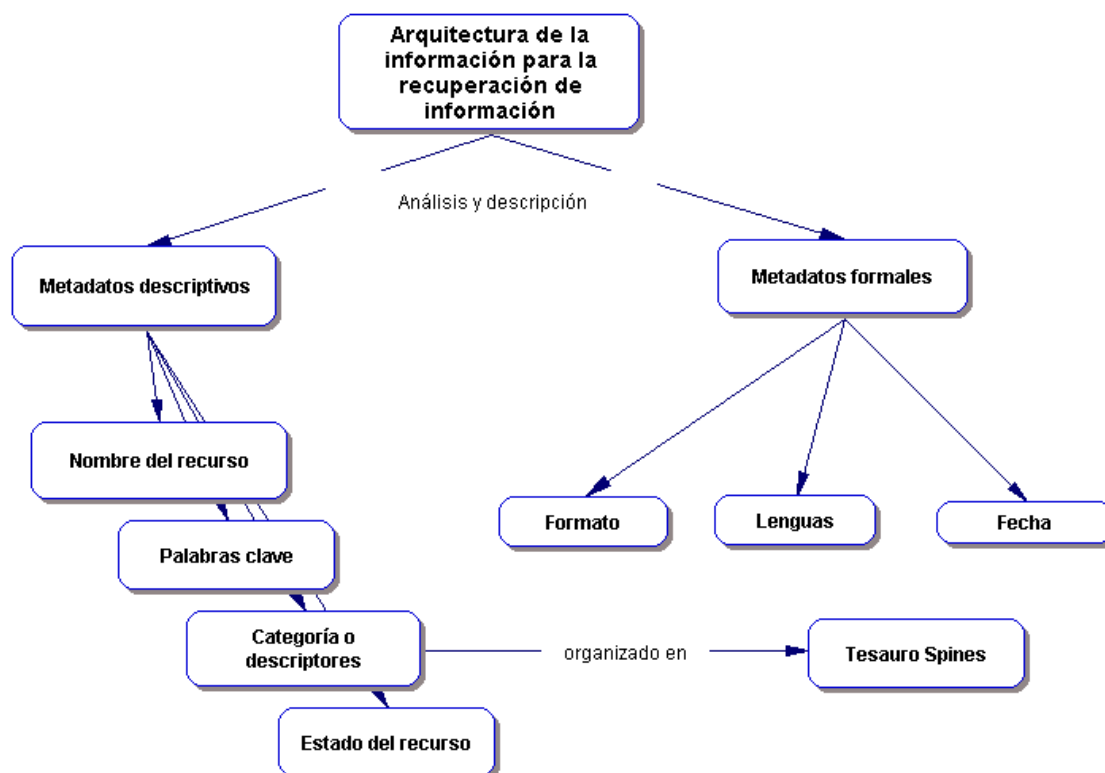


**Fig. 79: Ecosistema informacional de COGNlímica**

Siguiendo este modelo, se han implementado los elementos de interacción como interfaces, enlaces, botones que han derivado en una primera propuesta informática implementada del sistema y que tratamos en el apartado 4.2. Fase de desarrollo de la versión 1.0.

Además de la panorámica de la arquitectura informacional del sistema en general, se han definido las estructuras puntuales que deben tener los diferentes componentes del sistema.

La formalización de las estructuras de información actúan como estándares de intercambio posibilitando el almacenamiento de los puntos que sirven de acceso a los documentos con recursos lingüísticos. Ninguna de las técnicas de acceso a la información podría dar resultado sin información estructurada previa. Por ello, la información para que pueda recuperarse, se ha analizado y se ha descrito conformando así parte de la arquitectura de la información del sistema. En el apartado 3.4.2.3. Diseño del gestor de diccionarios se han previsto las opciones de búsqueda del catálogo de diccionarios derivadas también de las necesidades de los usuarios. Esta información junto con el lenguaje de indización relativos a la categorización semántica de los contenidos de los documentos conforman los datos formalizados determinantes para el funcionamiento en conjunto del sistema y que mostramos en el siguiente diagrama:



**Fig. 80: Arquitectura de la información principal**

De este modo, en el apartado 3.4.3.2. Diseño del gestor de proyectos se han formalizado las estructuras de información que se deben gestionar, éstas son, sin embargo, de generación automática. Las informaciones gestionadas por el gestor de proyectos se aprovechan, por una parte, de la información ya introducida en el gestor documental de diccionarios y, por otra, de la información ya asignada a cada proyecto de traducción introducida a través del asistente de *Transit* y que se almacena en los archivos \*.prj, tema ampliamente tratado en el apartado que acabamos de mencionar.

Y por último, en el apartado 3.4.3.3. Diseño del gestor de corpus paralelos hemos atendido a las estructuras de información pertinentes para esta última herramienta. En este apartado también se ha explicado la dependencia de la información recogida en el gestor de proyectos, el que depende igualmente del gestor documental de diccionarios y de *Transit*.

### 4.1.2. Bases técnicas y funcionales

Las bases técnicas que hemos establecido para la implementación del sistema COGNImática se han derivado principalmente de las tendencias técnicas expresadas en el apartado 2.6.2. Herramientas no integradas en los sistema de traducción asistida. Las funciones de este desarrollo técnico son, asimismo, el resultado de recomendaciones apuntadas en las diferentes secciones que componen el apartado 3.4. Diseño e implementación de las propuestas metodológicas, conceptuales y tecnológicas.

Así, los fundamentos del sistema propuesto se han basado en un diseño tecnológico que incorpora las ventajas de la estructura modular, de la tecnología web y de las técnicas de gestión documental que dan soporte a las principales funciones del sistema. Finalmente, las funciones implementadas en COGNImática como sistema de gestión de conocimiento han tenido en cuenta las enseñanzas de la ecología de la información y el diseño orientado al usuario.

#### 4.1.2.1. Estructura modular

COGNImática se ha definido como un sistema de tres módulos relacionados entre sí:

- ❑ gestor documental de diccionarios
- ❑ gestor de proyectos y
- ❑ gestor de corpus paralelos

La estructura modular que concreta la organización de COGNImática se deriva de las directrices ya explicadas en el apartado 2.6.3. Tendencias para la gestión integral del flujo total de la traducción y sintetizadas en la *Fig. 35 Integración de la tecnología en flujos de trabajo de la traducción técnica*.

La decisión de diseñar el sistema de forma modular no es casual. Ya en el apartado 2.2.5. Los tipos / módulos textuales de la documentación técnica se ha presentado el modelo de Baldwin y Clark para el diseño de sistemas de software modulares, que hemos aplicado a la tipificación de los elementos que componen el contenido total de un producto de software. Como se ha dicho, a los elementos textuales que componen el contenido total de un producto los denominamos módulos, debido a su condición de independientes, sustituibles, exclusivos, escalables, de relevancia modificable y adaptables. La estructura de módulos para el diseño de sistemas de software que hemos extrapolado para la identificación de contenidos, recupera su entorno original para el que lo conciben Baldwin y Clark. En este sentido, cada módulo de COGNImática se caracteriza porque puede:

- ❑ crearse, modificarse y eliminarse de forma independiente;
- ❑ ser intercambiable y sustituible; puede constituir una solución independiente por sí misma<sup>1</sup>;
- ❑ ser escalable, es decir, se pueden incorporar nuevos módulos a uno existente;
- ❑ tener diferentes grados de importancia dentro del sistema en función de diferentes variables;
- ❑ utilizarse y adaptarse en otros contextos más allá del entorno para el que fue diseñado.

Los módulos de COGNImática tienen diferentes funciones, que se han previsto en los apartados correspondientes del tercer capítulo<sup>2</sup> y en el

---

<sup>1</sup> A pesar de que la óptima explotación del sistema incluye la aplicación de todas las funcionalidades de los diferentes módulos, cada módulo tiene funciones específicas y alguno de ellos constituye una solución independiente por sí misma.

<sup>2</sup> 3.4.2.3. Diseño del gestor de diccionarios, 3.4.3.2. Diseño del gestor de proyectos y 3.4.3.3. Diseño del gestor de corpus paralelos

#### apartado 4.1.1. Diseño de la información e implementación de la arquitectura de la información.

En este sentido, el módulo de gestión documental de diccionarios es un catálogo organizado en una estructura de base de datos relacional, es decir, tablas, compuestas de registros, filas, y de campos de alimentación manual a través de una interfaz de alimentación. En ella se catalogan *manualmente* los diccionarios y a través de una interfaz adicional, la interfaz de búsqueda, se capturan los elementos deseados.

Este módulo es el componente principal del sistema que gestiona los contenidos recuperables mediante enlaces hipertextuales. Estos nodos constituyen la base sobre la que se apoya el catálogo como un sistema tanto independiente como integrado en un sistema mayor. El catálogo está dotado de los instrumentos necesarios para llevar a cabo las operaciones atribuibles a la gestión documental como la clasificación, la indización, el almacenamiento de documentos, la búsqueda y la captura de los documentos que los albergan. La eficiencia de los resultados de estas tareas depende del sistema conceptual aplicado para la organización del conocimiento semántico.

Asimismo, dicho módulo permite satisfacer algunas de las necesidades manifestadas por los usuarios: lograr acceder a los diccionarios terminológicos a través de identificadores del contenido sin tener que abrir los diccionarios; identificar los diccionarios existentes basándose en el nombre del diccionario porque los nombres ya son más intuitivos; identificar los diferentes tipos de diccionarios existentes en función de su fiabilidad y así mejorar su selección; y asociar a los diccionarios información adicional, como las recomendaciones de mantenimiento o las especificaciones concretas para la creación y alimentación de un diccionario de un cliente. Por tanto, con el gestor documental de diccionarios y, gracias a la catalogación semántica, se capturan los diccionarios idóneos para cada proyecto sin tener que recordar el contenido de la producción y, más aún, sin necesidad de saber lo

que ha gestionado el resto del equipo de la entidad. En ese sentido, el gestor documental de diccionarios es también un gestor de conocimiento.

El módulo de gestión de proyectos también es un catálogo donde se almacenan los datos en una estructura de base de datos relacional, pero, a diferencia del módulo de gestión de diccionarios, éste no es de alimentación manual, sino automática. Y esto es así, porque el módulo consta de una relación con la tabla del catálogo de los diccionarios que enlaza con la anotación semántica de los diccionarios y que determina la anotación de las memorias de traducción gestionadas por el módulo de gestión de proyectos. Esta anotación semántica se corresponde con *Materia* en la interfaz de búsqueda y aglutina los campos *Tema*, *Palabras clave* y *Descriptor Spines* contemplados en la interfaz de alimentación del módulo de gestión de diccionarios. Por otra parte, este módulo de gestión de proyectos consta de un script que recoge informaciones de los archivos de configuración de proyectos de *Transit* informaciones acerca del *Cliente*, *Nombre de diccionarios*, *Material de referencia asociada*, *Fecha de modificación*, *Gestor* y *Combinación de lenguas*. El script está programado para que actualice el catálogo cada día a una hora determinada.

Así como el primer módulo descrito pretende satisfacer algunas de las necesidades de los usuarios, también el gestor de proyectos contribuye a los mismos objetivos. En este sentido, este módulo posibilita poder buscar, más allá de las informaciones registradas en los archivos de configuración de cualquier proyecto en *Transit*, como las búsquedas por identificadores del contenido en las memorias de traducción almacenadas, así como la aplicación de diferentes criterios de búsqueda para recuperar el material requerido.

Y por último, el gestor de corpus paralelos no es un catálogo en sí como los dos módulos descritos, sino que se compone de un *script* y una interfaz de selección que se genera tras una búsqueda en el gestor de proyectos. Esta interfaz de selección da soporte al acceso de los documentos seleccionados

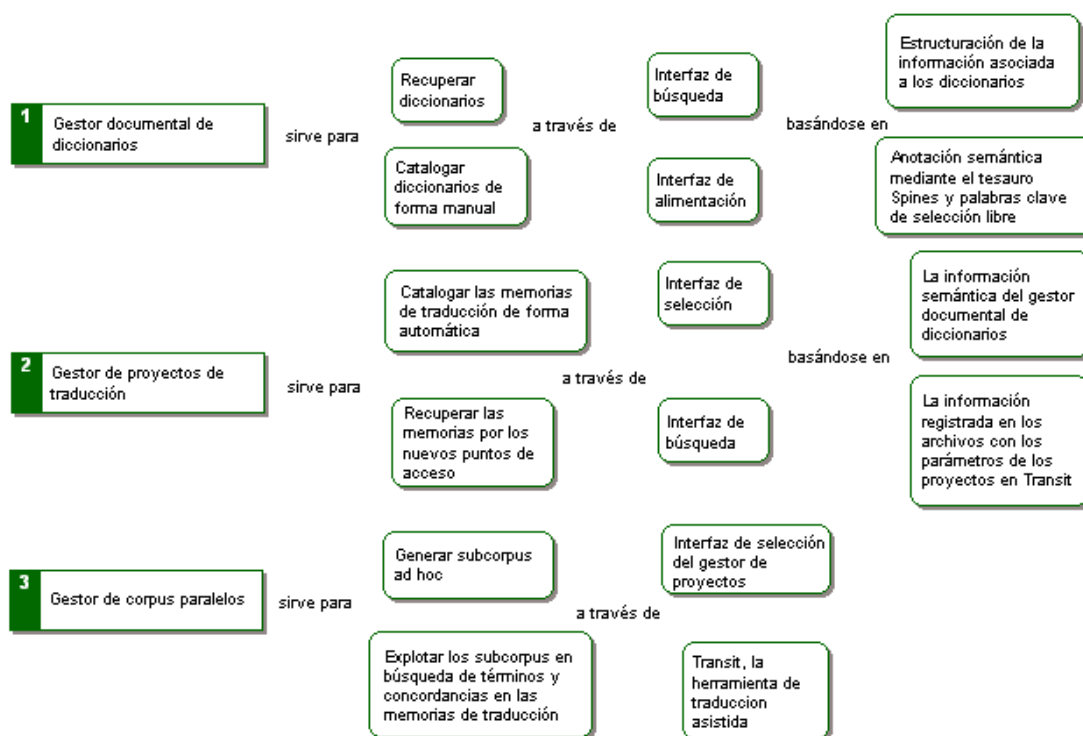


que se gestionan aprovechando la organización de las memorias basándose en directorios que ofrece *Transit*.

Este último módulo permite hacer subcorpus paralelos y poder consultar, por ejemplo, términos y concordancias en otro material de referencia que no esté definido en los parámetros del proyecto de traducción en la que el traductor está trabajando en ese momento de una forma más sencilla que la utilizada con otro de los programas de búsquedas de concordancias usadas por los traductores.

En síntesis, para satisfacer los requerimientos de contenido de los usuarios se hacen funcionar entre sí las tres herramientas que componen todo el sistema modular y se logra que los recursos lingüísticos estén organizados como si fuera una biblioteca virtual autogestionable.

A continuación mostramos en forma de mapa conceptual los módulos que componen COGNImática y sus funciones:



**Fig. 81: Funciones de los módulos de COGNlímica**

De las funciones mostradas en este mapa se puede deducir que COGNlímica actúa como una biblioteca virtual que recupera y extrae recursos lingüísticos, técnicas que se abordan en el siguiente apartado.

#### **4.1.2.2. Técnicas documentales: recuperación y extracción**

La relación de COGNlímica con las técnicas documentales es doble, ya que, por una parte, las utiliza para satisfacer las necesidades de contenido de los usuarios y, por otra, da soporte a las tareas implicadas en la gestión documental, por ejemplo, la catalogación.

Para satisfacer los requerimientos de contenido —como ya se ha perfilado en el apartado 4.1.1. Diseño de la información e implementación de la arquitectura de la información—, GOGNlímica utiliza principalmente dos técnicas documentales:

- la recuperación de información y

- la extracción de información.

Mediante la técnica de recuperación de información se obtienen documentos pertinentes a la búsqueda, mientras que, mediante la extracción de información, se consiguen las cadenas de caracteres contextualizadas en el texto completo contenido en los documentos (Baeza y Riberio, 1999: 34-35). Estas cadenas de caracteres corresponden a las unidades semánticas de búsqueda solicitadas por el usuario<sup>1</sup>.

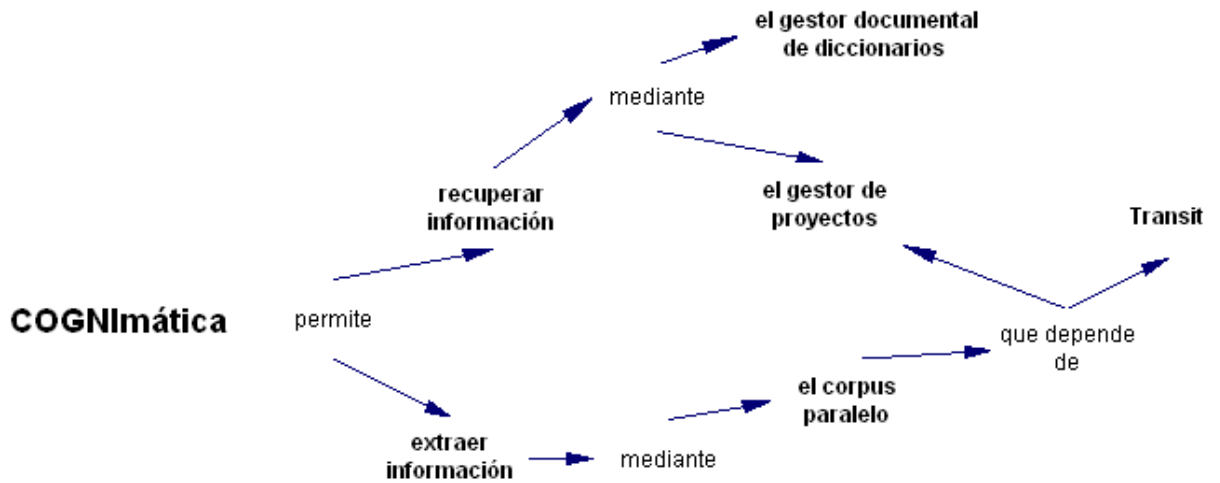
Tal y como hemos visto en el apartado 2.6.2.2. Gestores documentales → Gestores de contenido, los sistemas de gestión documental tratan documentos, es decir, crean, almacenan, indizan, editan, recuperan documentos; los sistemas de bases de datos, gestionan registros. No obstante, la tendencia tecnológica se inclina al desarrollo de aplicaciones mixtas. Esto significa que los gestores de bases de datos pueden ser gestores documentales en potencia ya que a través de los registros ofrecen acceso a los contenidos de los documentos (Tramullas, 1995). Estos sistemas sirven para registrar por tanto contenidos para que puedan ser compartidos (Bustelo y García-Morales, 2001).

GOGNImática, pues, es de alguna manera un gestor documental y de bases de datos que ofrece acceso a los recursos almacenados. Las bases de datos que contienen los datos de las referencias a los documentos se tratan mediante el gestor documental de diccionarios y mediante el gestor de proyectos y que el usuario recupera. Las bases de datos textuales y de conocimiento traductológico y terminológico se tratan a través del corpus paralelo de donde el usuario extrae las cadenas de caracteres deseadas.

---

<sup>1</sup> En otro tipo de sistemas de información, la extracción corresponde a eventos, hechos descritos o explicados en los documentos o a unidades conceptuales. La extracción de la información es una técnica que se recoge en la disciplina del procesamiento del lenguaje natural y que se halla vinculada a la inteligencia artificial.

De esta manera los requerimientos de contenidos de los usuarios especificados en el apartado 3.3.3.3. Necesidades de los usuarios se abastecen mediante las técnicas documentales expuestas y en interacción con *Transit* tal y como ilustramos en el siguiente gráfico:



**Fig. 82: Técnicas documentales de COGNImática**

En el sentido conceptual de proceso de gestión de información, tanto documentalistas como traductores comparten técnicas, debido a que, paulatinamente, los traductores, al igual que los documentalistas, se confrontan con volúmenes de información ingentes no sólo en cuanto a tamaño, sino también en cuanto a su diversidad. *Por eso, los traductores se han convertido en verdaderos expertos en la búsqueda y selección de información* (Tradumatica, 2002), compartiendo estos dos colectivos profesionales técnicas propias de acceso a la información como la recuperación y la extracción de información.

### 4.1.2.3. Tecnología web

El sistema COGNImática se basa en una arquitectura técnica compatible con el entorno web e integrada en un red de área local. En consecuencia, se han empleado los avances técnicos fundamentales que se aplican en la implementación de las tecnologías web para Internet.

El sistema corre sobre la plataforma<sup>1</sup> *Linux* y *Windows*, utilizando así los dos sistemas operativos empleados en STAR SL. Así, sobre la plataforma *Linux* corre el servidor web *Apache* que se utiliza para la implementación de COGNImática. Este servidor web de código abierto<sup>2</sup> funciona con cualquier sistema operativo y se basa en un protocolo HTTP<sup>3</sup> diseñado para transferir elementos hipermedia.

La red local utilizada en STAR SL da soporte al acceso y la distribución con celeridad de las tareas solicitadas para múltiples usuarios (arquitectura cliente-servidor), permitiendo realizar operaciones de forma descentralizada. La tecnología web ejerce una influencia positiva en la creación, uso e intercambio de recursos lingüísticos mediante su codificación en formatos estándar. Este intercambio puede ser público (Internet), restringido (extranet) o privado (intranet).

---

<sup>1</sup> El término *plataforma* se utiliza como sinónimo del término *sistema operativo* que se define como el conjunto de programas de software destinado a gestionar el hardware y a permitir la comunicación del usuario con el ordenador.

<sup>2</sup> El software de código abierto —también llamado *código fuente abierto*, en inglés *Open Source*— puede ejecutarse con cualquier propósito y por quien lo desee, puede distribuirse libremente y puede adaptarse a las necesidades del entorno modificando el código fuente (Free Software Foundation, 2002-2005). Puesto que este software goza de un sistema de distribución que no se basa en las licencias del software propietario garantiza los requisitos de interoperabilidad entre los sistemas, además de tener potencialmente la capacidad de impulsar y popularizar el desarrollo de otras tecnologías para la comunicación de conocimiento, (véase Castells, M. *et al.* 2004).

<sup>3</sup> Para más información sobre http véase el apartado 2.5. Formatos de codificación e intercambio.

Creemos que la tecnología web y, por tanto, las aplicaciones web afectan al trabajo de organización de la información, además de fomentar la comunicación, porque el medio no es un simple proveedor de datos, sino que se diseña en respuesta a una auténtica necesidad de comunicación en la que la información adquiere valor en función de las posibilidades de acceder a ella. La posibilidad de acceso a través de un navegador y la posibilidad de compartir información que ofrece la tecnología web la hace insustituible.

En síntesis, se puede afirmar que la popularización de la tecnología web se basa por una parte en su usabilidad y por otra en su aplicabilidad en los más diversos entornos de trabajo. La aplicación de una tecnología tan extendida y popular para el desarrollo de nuestro sistema de información presumiblemente minimizará el impacto entre las personas que componen el entorno de implementación.

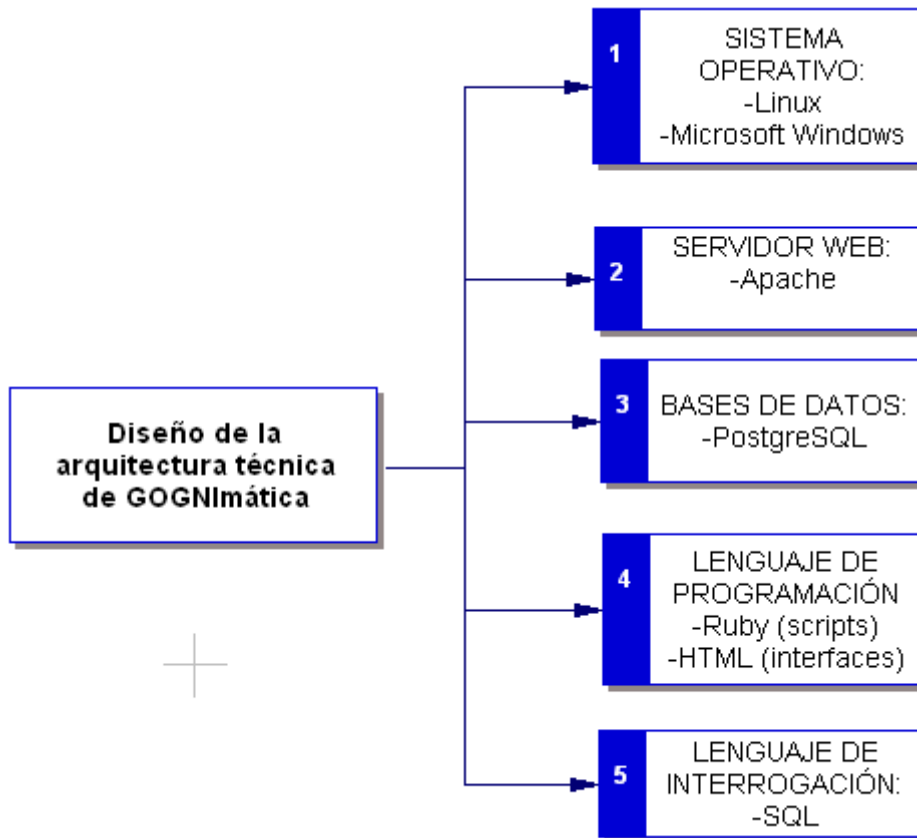
La base de datos utilizada en COGNImática es *PostgreSQL*. Se trata de una base de datos relacional orientada a objetos de código abierto que puede utilizarse como un servidor de datos de aplicación o como uno de base de datos en web. La herramienta que gestiona *PostgreSQL* es *PgAdmin*, también de código abierto.

*Ruby on Rails*, el lenguaje de programación de scripts seleccionado, es el marco (*framework*) para el desarrollo de una arquitectura de la información de aplicaciones web. Este lenguaje permite la manipulación de cadenas de texto y números, como la agrupación de cadenas de texto y la utilización de expresiones regulares. Son patrones o plantillas que se comparan con cadenas de texto. Mediante expresiones regulares se consigue la manipulación automática de, por ejemplo, grafías alternativas, variantes, etc.

Las líneas de código de *Ruby on Rails* se incluyen en el código HTML o XML y se programan en el lado del servidor, ejecutándose técnicamente como un CGI (*Common Gateway Interface*). El término *CGI* identifica un programa que toma información del servidor web y con ella realiza alguna operación de

gestión de entrada y salida de información. El segundo término que identifica esta miniaplicación mencionada, CGI, se encuentra dentro de un tipo de programa mayor muy popular en entornos de web, nos referimos a las *pasarelas de información*. Una pasarela de información es una aplicación que lleva a cabo la conversión de protocolos entre diferentes tipos de redes (Internet, intranet, etc.) y aplicaciones cuyo objetivo es proporcionar información al usuario que la solicita.

El lenguaje de consultas utilizado en COGNImática es SQL (*Structured Query Language*). Se trata de un estándar implementado por ANSI (*American National Standards Institute*) para los principales motores o sistemas de gestión de bases de datos relacionales. Este lenguaje permite recuperar y extraer información almacenada por medio de consultas (ecuaciones de búsqueda).



**Fig. 83: Tecnologías aplicadas en COGNimática**

Sobre la base de esta tecnología se debe garantizar el flujo técnico general de COGNimática para que el usuario pueda llevar a cabo las siguientes tareas:

- solicitar datos o documentos,
- interrogar al sistema y capturar datos almacenados mediante interfaces de búsqueda,
- transmitir las peticiones de información para que el sistema las reconozca en forma de sentencia de lenguaje de interrogación estándar (SQL),
- solicitar cualquier recurso para que el servidor definido para el sistema lo prepare y ofrezca la respuesta en un formato estándar independiente (HTML).



Todo este flujo debe estar fundado no sólo en el diseño de la arquitectura técnica, sino en un diseño que también dé soporte a la arquitectura de la información.

#### **4.1.2.4. Ecología de la información y diseño centrado en el usuario**

El término *ecología de la información*, acuñado por Davenport (1997), aborda la gestión de la información desde una perspectiva holística de la información; esto es, no considera relevante las unidades informacionales independientes entre sí, sino que las entiende solo dentro de un contexto en una organización.

Esta visión otorga a las fuentes donde se genera y utiliza la información, es decir, a los flujos de trabajo, así como a los usuarios implicados un papel preponderante e insustituible para poder desarrollar sistemas informáticos que den respuesta a las necesidades<sup>1</sup>.

La integración de flujos y usuarios en una estrategia de la gestión de información crea condiciones idóneas para gestionar el conocimiento de una organización, así como para el desarrollo de instrumentos asociados (Davenport, 1997: 28-45).

En este sentido GOGNImática se entiende como sistema en permanente cambio basado en el diálogo constante entre el equipo diseñador y los usuarios del sistema de información. Mediante la orientación al usuario — necesidades, percepción del sistema de información, cultura informacional de los empleados— sabemos:

- qué aplicación real hace el usuario del sistema,

---

<sup>1</sup> El auge de disciplinas como la arquitectura de la información y usabilidad web se encuadran en el mismo contexto de la ecología de la información cuyo principal denominador común es la visión centrada en el usuario. Esta orientación está viviendo una gran expansión coincidiendo con la proliferación de aplicaciones para Internet.

- los grados de dificultad en el manejo de las funciones del sistema,
- qué posibilidades de mejora existen.

En el apartado 2.2.4. Técnicas y tecnología de desarrollo de contenido técnico (multilingüe), hemos abordado el tema de las interfaces gráficas de usuario como una técnica de gestión de contenidos en el ámbito de la documentación técnica. Ahora volvemos a este tema ya que la interacción persona-ordenador es uno de los elementos centrales de gestión de contenidos que incorpora el sistema GOGNImática.

El diseño de la interfaz de usuario, tanto de búsqueda como de alimentación es el resultado del diseño centrado en el usuario. Desde hace años, la disciplina que se ocupa del diseño de las interfaces de usuario en los sistemas de información se llama *interacción persona-ordenador* o *interacción persona-máquina* (abreviado IPO —en inglés *human-computer interaction*, abreviado *HCI*—). Siguiendo a Marcos (2004: 36-38), esta disciplina se centra en la creación de productos informáticos que contribuyan a incrementar la satisfacción de los usuario en cuanto a su interacción con los sistemas. Para lograr sus objetivos debe abarcar por un lado los aspectos humanos, por otro los tecnológicos y, por último, la comunicación entre ambos. La satisfacción de usuario está vinculada a la percepción de que el proceso de llevar a cabo una tarea dentro de un sistema sea consistente, organizado, eficiente, productivo, fácil de usar, intuitivo y específico para su entorno de aplicación, es decir, su usabilidad. Por tanto, según los criterios de satisfacción de un usuario, el éxito de un sistema se define por su usabilidad, tal y como aludíamos en el apartado 3.2.1.6. Seguimiento y mantenimiento al abordar el concepto de *usabilidad* definido en el marco de la norma ISO 9241.

El diseño de la navegación está estrechamente relacionado con el diseño de la información. Si se diseña una arquitectura de contenidos hipertextual se deben prever los elementos de interacción entre el usuario y el sistema. Pero no sólo son las interfaces los elementos que condicionan la navegación, los

elementos como las referencias cruzadas entre los nodos relacionados entre sí, el tipo de relación de estas asociaciones, las acciones permitidas al usuario y las respuestas del sistema ante las entradas del usuario orientadas a la gestión semántica de la información —por ejemplo, agrupación de los nodos por categorías o propiedades en función de la indización llevada a cabo— repercuten también en la navegación (Garret, 2000).

Así, podemos concluir que las bases técnicas de GOGNImática establecen una sinergia con las bases funcionales tomando a los usuarios como centro de atención y considerando este enfoque en todas las etapas del ciclo de vida del sistema que recordamos a continuación:

- ❑ fase de diseño: especificaciones técnicas y funcionales
- ❑ fase de desarrollo de la versión 1.0.
- ❑ fase de pruebas
- ❑ fase de rediseño de desarrollo de la versión 1.1.
- ❑ fase de seguimiento y mantenimiento

## ***4.2. Fase de desarrollo de la versión 1.0.***

Este apartado pretende mostrar la primera versión del sistema ya implementado que hemos dividido en tres subapartados coincidiendo con los módulos del sistema propuesto. El orden de los apartados coincide con el orden de diseño e implementación de los módulos que hemos intentando documentar con capturas de pantalla de la aplicación real.

### **4.2.1. El gestor documental de diccionarios**

El módulo para la gestión documental de diccionarios se compone de dos interfaces: la interfaz de búsqueda y la interfaz de alimentación.

La interfaz de búsqueda es uno de los elementos interactivos del gestor documental de diccionarios y que mostramos a continuación:

**Gestor documental de diccionarios: catálogo**

Introducir nuevos datos terminológicos

[Criterios de clasificación](#)

Buscar

<b>Cliente</b> Todos Abacus Abb	<b>Materia</b>  Tecnología I INSTRUMENTOS DE MEDIDA DEL TIEMPO;DISPOSITIVO Tecnología I. ACEITES LUBRICANTES. PRODUCTOS DEL PETROLEO Tecnología I. APARATOS ELECTRICOS. TECNOLOGIA ELECTRICA. A
<b>Formato contenedor</b> -- 2.7 TermStar XV	<b>Nombre del diccionario</b>  

Buscar Borrar

**Fig. 84: Interfaz de búsqueda del gestor documental de diccionarios (1ª propuesta)**

Como se ve en la captura de pantalla, en el entorno de esta interfaz de búsqueda se accede mediante hiperenlaces a la documentación adicional, esto es, a [Criterios de clasificación](#) que abordan sólo los procedimientos de clasificación por estado de diccionarios, como se muestra en la siguiente captura de pantalla:

Código	Tipo de diccionario	Descripción
ACT	ACTIVO	Creado por el gestor responsable del proyecto de traducción y modificado, alimentado y actualizado por el gestor responsable del proyecto de traducción y por el traductor en plantilla. Parte de la terminología verificada por el gestor proviene de la fusión de diccionarios tipo FRL.
FRL	FREELANCE	Creado por el gestor y alimentado y modificado por el traductor autónomo. Tiene un carácter temporal, pues su contenido es susceptible fusionarse con los diccionarios tipo ACT, por tanto no se deben registrar en STARTERM.
HIS	HISTÓRICO	Archivado y comprimido. Este tipo de diccionarios no se utilizan para proyectos actuales del cliente. Pueden corresponder a clientes no activos o a formatos ya no usados por el cliente. Siempre se guardan comprimidos.
REF	REFERENCIA	Importado de Internet u otras fuentes cuya temática está vinculada al contenido del proyecto de traducción para el que se usa. La diferencia entre diccionarios REF enviados por un cliente de STAR SL y los que importamos nosotros después de seleccionarlo en Internet se identifican en la nomenclatura. Para la nomenclatura del tipo de diccionario temático de referencia aparece siempre la etiqueta semántica y

**Fig. 85: Descripción de tipos de diccionarios**

Estos criterios los habíamos trazado en el apartado 3.4.1.2.1. Clasificación de diccionarios por estado derivados de los problemas detectados y explicados en el apartado 3.3.3.4.2. Problemas de acceso a los recursos lingüísticos almacenados.

Asimismo, en la interfaz de búsqueda se intuyen a través de los campos representados los posibles tipos de búsquedas que permite el gestor documental de diccionarios y que especificamos a continuación:

T  
I  
P  
O  
S

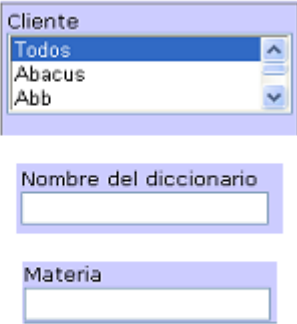
**Cliente** (seleccionable en tabla predefinida)

D  
E

**Nombre de diccionario** (introducción de texto libre, parte o cadenas completas)

B  
Ú  
S  
Q  
U  
E  
D  
A

**Materia** (introducción de texto libre, parte o cadenas completas)



**Fig. 86: Tipos de búsqueda en la interfaz de consulta**

En el campo *Materia*, el motor de búsqueda explora en los tres campos que dan información semántica acerca de los contenidos de los diccionarios: tema de traducción, palabras clave y descriptores *Spines*. Todos estos datos

están almacenados y se pueden visualizar en una lista con las secuencias totales de los descriptores relacionados y con las unidades independientes.

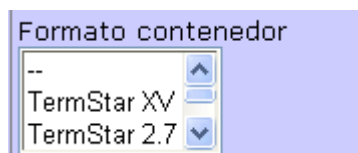
El término o frase que describe el tema de traducción, se deriva de lo que aporta el cliente, admitiendo el uso de palabras en español y en el idioma de los proyectos originales. Las palabras clave son los términos propuestos por el gestor que identifican el área temática del diccionario. Se pueden introducir tantos términos como sean necesarios. Se recomienda el uso de palabras clave en español. También se pueden introducir términos en el idioma de la traducción, siempre y cuando se introduzca adicionalmente la equivalencia en español. Los descriptores son los términos de indización extraídos del tesoro *Spines*.

Así, si realizamos una búsqueda por el campo *Materia* introduciendo, por ejemplo, *resina* el gestor de diccionarios nos ofrece la siguiente captura:

Nombre	Cliente	Tema	Palabras clave	Idiomas	Formato
BYK_quimica_ACT	BYK-Chemie URL <a href="http://www.byk-chemie.com/englisch/start.cfm">http://www.byk-chemie.com/englisch/start.cfm</a>	Aditivos químicos	Química; aditivos; resinas; pinturas; lacas; plásticos	deu esp	XV

**Fig. 87: Captura de diccionarios tras *resina***

Asimismo, las búsquedas de los diccionarios se pueden afinar adicionalmente seleccionando el formato del diccionario deseado, como se muestra en la siguiente captura parcial:



**Fig. 88: Menú desplegable de formatos de los diccionarios**

Una vez obtenida la primera captura utilizando las diferentes opciones de búsqueda, el usuario tiene la posibilidad de seleccionar los elementos que desee para obtener informaciones asociadas a cada diccionario, como

mostramos en la siguiente captura de pantalla resultante de filtrar la información por *Materia* “impresora” y *formato* “TermStar XV”:

Nombre	Cliente	Tema	Palabras clave	Idiomas	Formato
Minolta_C350_VAL	<a href="#">Minolta</a> URL	Software de impresión	copiadora; impresora; scanner; software; controlador	esp eng	XV
Minolta_Di1610f_VAL	<a href="#">Minolta</a> URL	Software de impresión	copiadora; impresora; scanner; software; controlador	esp eng	XV
Heidelberg_impresoras_ACT	<a href="#">Heidelberg</a> URL	Impresoras rotativas	impresora; papel; impresión; rodillo	deu esp	XV
Minolta_PageScopeDirectPrint_VAL	<a href="#">Minolta</a> URL	Software de impresión	copiadora; impresora; scanner; software; controlador	esp eng	XV
Hitachi_Key_VAL	<a href="#">Hitachi</a> URL	Impresión	Impresión; copia; documento; papel; página	esp eng	XV
Xerox_DMT_REF	<a href="#">Xerox</a> URL <a href="http://www.xerox.com">http://www.xerox.com</a>	Impresoras	Tóner; bandeja; tipo de letra;	esp enu	XV
Xerox_DMT1_REF	<a href="#">Xerox</a>	Impresoras	separador;	esp enu	XV

**Fig. 89: Captura de diccionarios tras *impresora***

De esta manera, si el usuario está interesado por uno de los diccionarios mostrados, por ejemplo *Heidelberg\_impresoras\_ACT* mediante un clic de ratón se puede desplegar la información correspondiente a ese diccionario, como se puede ver a continuación:

**Heidelberg\_impresoras\_ACT**

**Ubicación**  
E:\termi\Heidelberg\Actuales\Heidelberg.mdb

**Descriptor SPINES**  
Proceso de datos; Informática; IMPRESORAS;

**Palabras clave**  
impresora; papel; impresión; rodillo

**Descripción del diccionario**  
Diccionario activo. Se puede cargar en los proyectos que vienen a través del TWS. Se ha fusionado con el diccionario "Heidelberg\_impresoras\_REF" y el par DEU-SPA del diccionario impresoras\_eng-fra-deu-esp\_ACT a día de 11/11/2004. Queda con 4.547 entradas.

**Tipo de diccionario**  
ACT

**Sugerencias de mantenimiento**  
Ir alimentándolo con las traducciones.

Contraseña de administrador:

**Fig. 90: Ficha de un diccionario**

Las informaciones que obtenemos de las fichas de los diccionarios se han introducido previamente a través de la interfaz de alimentación.

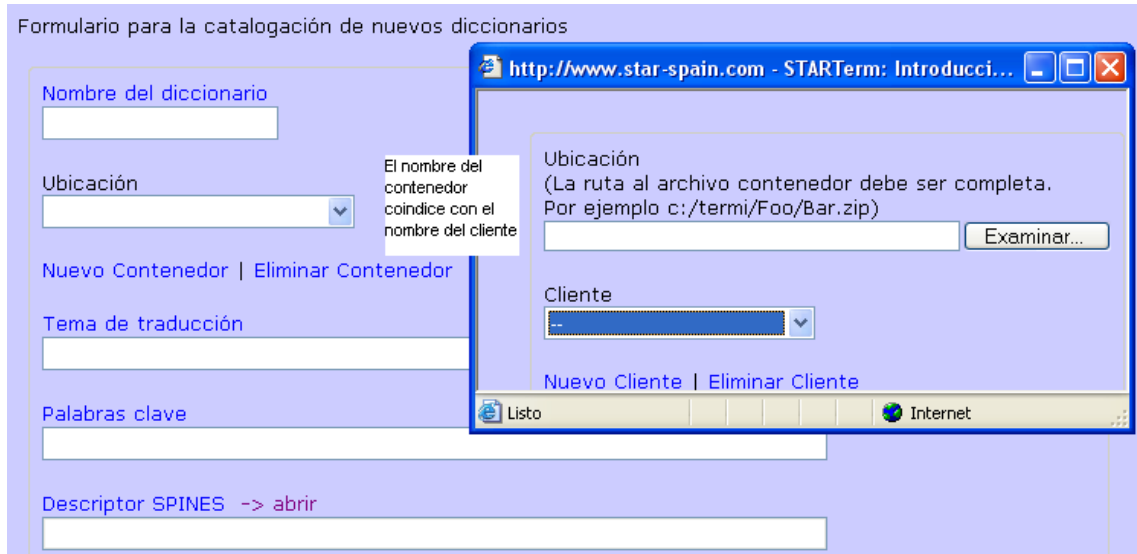
La interfaz de alimentación es el segundo de los elementos del módulo que tiene como fin la recogida sistemática de los datos. A ella se accede a través de un elemento hipertextual, el botón de función:

integrado en la interfaz de consulta, como se ha visto en la *Fig. 83: Interfaz de búsqueda del gestor documental de diccionarios (1ª propuesta)* y también a través del botón *Editar* que aparece al final de las fichas desplegadas de los diccionarios, como se ha visto en la figura anterior.

A través de la interfaz de alimentación se recoge información indizable y no indizable. Los campos con información indizable son los que sirven de puntos



de anclaje para hacer las consultas posteriores. A continuación mostramos una parte de la interfaz de alimentación:



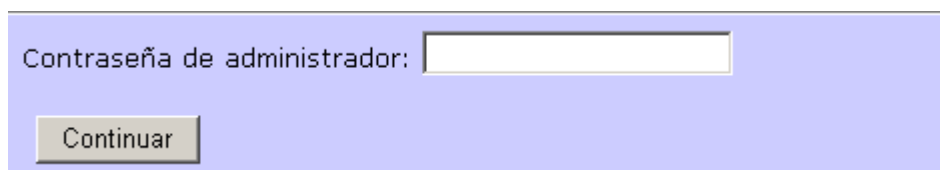
**Fig. 91: Sección (1) de la interfaz de alimentación del gestor documental de diccionarios**

Menos el campo de ubicación donde se introduce la ruta física para acceder al contenedor (base de datos) donde está almacenado el diccionario, todos son campos indizables. Esto quiere decir que a través de la interfaz de búsqueda se puede buscar por *Nombre del diccionario* y también por *Cliente*. Las búsquedas por *Tema de traducción*, *Palabras clave* y *Descriptor Spines* quedan agrupados en el campo *Materia* que aparece en la interfaz de búsqueda.

Los gestores de proyectos, es decir los usuarios del segundo nivel de acceso, introducen todos los datos del diccionario, a excepción del campo *Descriptor Spines*. Este campo está protegido contra escritura, a no ser que se entre con la clave de administrador.

El administrador del sistema, usuario del tercer nivel, tiene todos los permisos para activar todas las funciones que proporciona la herramienta. Para activarlos debe introducir su contraseña en el campo correspondiente que mostramos a continuación y que se activa en la interfaz de alimentación

tras activar la función de introducir nuevos registros de diccionarios en el diccionario:



Contraseña de administrador:

Continuar

**Fig. 92: Acceso restringido al administrador de COGNImática**

En este sentido, puede introducir los términos descriptores basándose en las palabras clave seleccionadas por los gestores.

A continuación, en la sección de captura de la siguiente pantalla, mostramos el resto de los campos pertenecientes a la interfaz de alimentación:

The image shows a software interface for managing dictionaries. It features several sections:

- Idiomas:** A list box containing the following languages: Afrikaans, Albanés, Alemán, Bielorruso, Búlgaro, Catalán, Checo, and Chino. Below the list is a small text input field containing the value 'afk'.
- Tipo de diccionario:** A dropdown menu currently showing '--'.
- Descripción del diccionario:** A large, empty text area for entering descriptive notes.
- Sugerencias de mantenimiento:** Another large, empty text area for entering maintenance suggestions.

**Fig. 93: Sección (2) de la interfaz de alimentación del gestor documental de diccionarios**

Como se ve, en la interfaz de alimentación también existen campos no indizables como los *Idiomas* de los diccionarios, el *Tipo de diccionario* con información de tablas predefinidas y dos campos adicionales para texto libre. Así, en *Descripción del diccionario* y *Sugerencias de mantenimiento* los gestores y el administrador de la herramienta pueden introducir notas descriptivas del diccionario, así como sugerencias de mantenimiento. La anotación de esta información es manual.

El gestor documental de diccionarios goza de independencia con respecto a la herramienta que se describe a continuación y que es parte del sistema diseñado y desarrollado.

### 4.2.2. El gestor de proyectos

Los elementos de interacción implementados tal y como se ha propuesto en la *Fig. 76: Sistema integrado de gestión de conocimiento y recursos lingüísticos* son:

- la interfaz de búsqueda
- la interfaz de selección

Mediante esta herramienta de gestión de proyectos el sistema puede hacer búsquedas por *Nombre de proyecto*. La cadena alfanumérica que constituye el nombre del proyecto contiene el nombre del cliente y el número de entrada del proyecto. La búsqueda no se hace por cadena exacta, es decir, no es necesario introducir el nombre completo, para obtener una captura completa de resultados, como podemos ver en el siguiente ejemplo:

8582_AGFA_Help_Update
8069_AGFA_Update
8366_AGFA
7900_AGFA
6937 Agfa
6938 Agfa
4586 Agfa

**Fig. 94: Resultado parcial de la captura de los proyectos relacionados con el cliente de búsqueda (AGFA)**

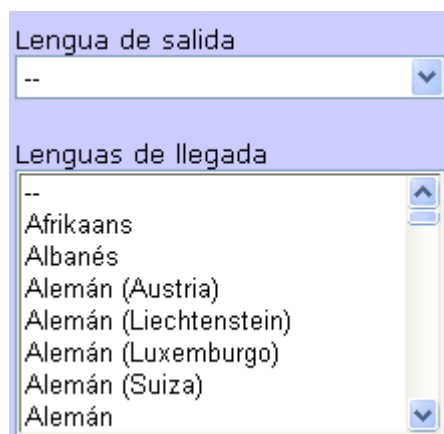
Si buscamos por *Nombre del diccionario*, el sistema nos ofrece todos los proyectos a los que se le ha asociado el diccionario de nuestra búsqueda. Así, por ejemplo, si queremos saber a qué proyectos se le ha asociado el diccionario con terminología del sistema operativo de Microsoft, y no sólo los proyectos concretos que se han realizado para este cliente, se introduce el nombre del diccionario. Así, el sistema nos ofrece los siguientes resultados:

8213_Instron04-230-01_text
8253_BMW_SAM2_DIS
6760_BridgeQuest_Install
6484_BridgeQuest_RC
8067_Arsenal_DOC
7225_UI
7144_Alden_TREK_SWIN_SW_ESP

**Fig. 95: Resultado parcial de la captura de proyectos relacionados con el diccionario de búsqueda (Microsoft\_2000)**

Del mismo modo, el sistema también permite la búsqueda por gestor. Este campo aglutina a los miembros del equipo encargados de preparar los proyectos de traducción, decidiendo qué traductores realizarán el trabajo, configurando los proyectos —formato de importación y exportación, lengua de salida, lengua de llegada, selección del material de referencia (memorias y diccionarios)— y controlando la calidad final —etiquetas de formato, control ortográfico, terminológico y traductológico—. Dependiendo de los proyectos, el gestor no se encarga del control de calidad. Las razones pueden estar vinculadas al tema de la traducción, a la combinación de lenguas y a aspectos meramente pragmáticos relacionados con picos de trabajo que hace necesaria la subcontratación del control de calidad. Si se realizan búsquedas de este tipo, las capturas pueden dar información acerca de, por ejemplo, la tipología de clientes que administra cada gestor.

Asimismo, el sistema permite la búsqueda por lengua de salida y de llegada de los proyectos de traducción, como también mostramos a continuación:



**Fig. 96: Sección (lenguas) de la interfaz de búsqueda del catálogo de proyectos**

Todas las informaciones de los campos recién descritos son anotaciones contenidas en los archivos \*.prj correspondientes a la configuración de proyectos en *Transit*. Estos archivos son los que se emplean como documentos secundarios que genera el catálogo automáticamente, tal y como habíamos explicado ampliamente en el apartado [3.4.3.2. Diseño del gestor de proyectos](#).

Pero más allá de las informaciones deterministas que podemos recuperar de la configuración de cualquier proyecto, el gestor de proyectos, al igual que el gestor documental de diccionarios, nos permite hacer otro tipo de búsquedas: por *Materia*.

Esta información no está contenida en los parámetros que configuran un proyecto, sin embargo, se reaprovecha de la marcación semántica que se ha introducido en el gestor documental de diccionarios y recupera los proyectos de traducción automáticamente mediante la información asociada a cualquiera de los diccionarios ya catalogados.

Además de estas informaciones, nuestra herramienta de gestión adicionalmente anota semánticamente y de forma automática las memorias de traducción de la misma manera que quedan marcados los proyectos, mediante la utilización de la información semántica del gestor documental de

diccionarios. De esta manera recogemos la idea formulada en el apartado 3.4.3.1. Dotación de nuevos puntos de acceso de que las informaciones semánticas de las macroestructuras de los diccionarios utilizados en un proyecto concreto de traducción son previsiblemente aplicables a las memorias de traducción que se generan en el ámbito de ese proyecto.

Si introducimos un término de búsqueda en el campo *Materia* del gestor de proyectos, el motor de búsqueda explora en los campos asociados a la materia de los diccionarios o en la información semántica contenida en el gestor documental de diccionarios —*Tema de traducción, Palabras clave y Descriptor Spines*—, como mostramos en la siguiente captura de pantalla a modo de ejemplo:

**Fig. 97: Interfaz de búsqueda del gestor de proyectos**

Así se captura, por una parte, los diccionarios relacionados con la ecuación de búsqueda y, por otra, los proyectos a los que se asignaron estos diccionarios.

Y siguiendo la ecuación de búsqueda *motos*, propuesta en la captura anterior, en la siguiente captura se muestra parte del resultado que ofrece el sistema:

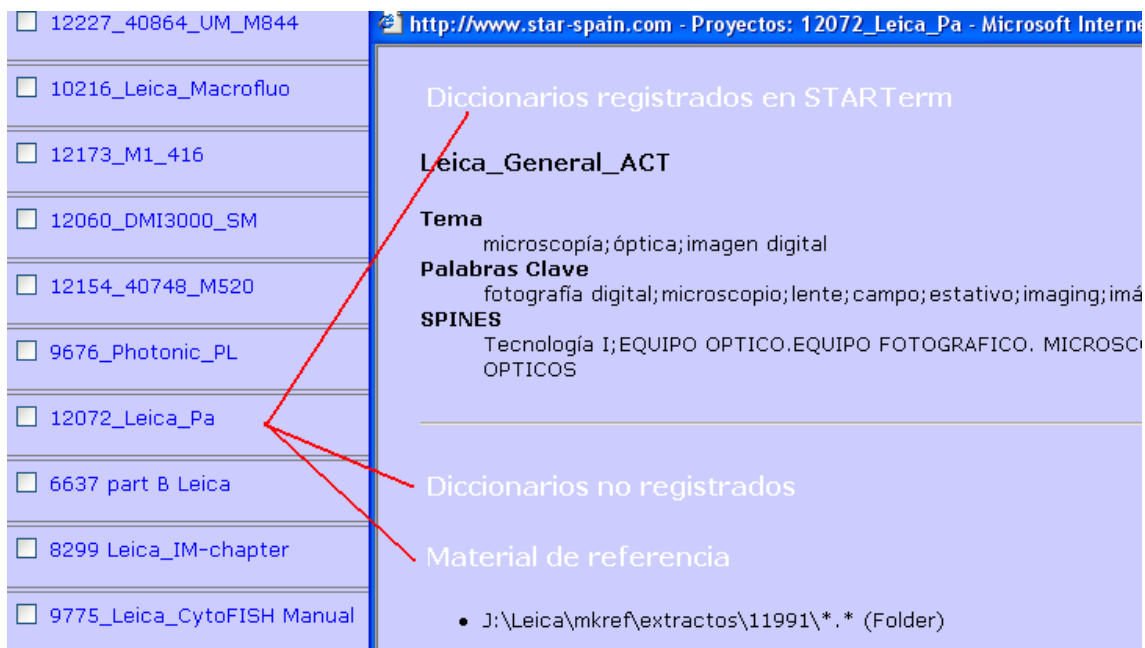
Nombre de proyecto	Gestor	Lengua de salida	Lenguas de llegada	Fecha modificación
<input type="checkbox"/> 12277_2152184_CIP-IDC	EP3	deu	esp	2005-06-15 11:11:53
<input type="checkbox"/> 12176_GRIPS_RepROM_K25HP2	FG3	deu	esp	2005-06-14 08:53:47
<input type="checkbox"/> 12069_GRIPS_SPEZWKZ_ESP	LA2	deu	esp	2005-06-13 08:23:18
<input type="checkbox"/> 12159_ETK_Teile_030605_ES	FG3	deu	esp	2005-06-07 15:20:28
<input type="checkbox"/> 12177_GRIPS_RepROM_K2X_ESP	FG3	deu	esp	2005-06-07 14:19:05
<input type="checkbox"/> 8244_BMW_GT1-Texte	monica	deu	esp	2005-05-31 07:31:56
<input type="checkbox"/> 9657_BMW_PUMA003462	PB1	deu	esp	2005-05-31 07:31:49
<input type="checkbox"/> 9223_BMW_Motorrad_EBA	MR3	deu	esp	2005-05-30 10:46:33

**Fig. 98: Interfaz de selección del gestor de proyectos**

Así, se presenta la interfaz de selección del gestor de proyectos. El usuario obtiene por una parte los diccionarios sobre esta materia, y ante las propuestas de captura podrá realizar un barrido visual, discriminando automáticamente la información que no le interesa.

Si le interesa la información sobre los recursos lingüísticos asociados con un proyecto concreto, el usuario hará clic sobre el nombre del proyecto deseado. El sistema entonces proporciona una ventana emergente con la información asociada, tal y como se muestra en la siguiente captura de pantalla:





**Fig. 99: Ventana emergente con información de recursos lingüísticos asociadas a un proyecto**

Como se puede deducir de la anterior captura, se muestra por una parte el resultado de la búsqueda y, por otra, mediante un clic, la información del documento pretendido por el usuario. Así, por medio de una ventana emergente se muestran no sólo las anotaciones semánticas que sirven de nexo para conectar el gestor documental de los diccionarios y el gestor de proyectos, sino las informaciones asociadas a los recursos lingüísticos que enunciamos a continuación:

- ❑ los nombres de los diccionarios asociados al proyecto registrados en el gestor de diccionarios,
- ❑ los nombres de los diccionarios asociados al proyecto pero que todavía no se han registrado en el gestor de diccionarios y,
- ❑ las memorias de traducción que se han predefinido en el proyecto concreto.

### 4.2.3. El gestor de corpus paralelos

La interfaz de selección del gestor de corpus paralelos depende del gestor de proyectos descrito en el apartado anterior. La dependencia estriba en que, a través de la interfaz de consulta del gestor de proyectos, podemos hacer una selección previa de los documentos que nos interesan, como se muestra en la siguiente captura de pantalla:



**Fig. 100: Interfaz de selección de las propuestas capturadas**

Una vez que el usuario ha realizado el barrido visual del material proporcionado por la herramienta de gestión de proyectos, el usuario puede afinar la selección del material que va a componer el subcorpus temporal, mediante los campos selectores de la lengua de salida y de llegada y los botones situados a la izquierda de cada proyecto, tal y como se muestran resaltados en la captura anterior.

El usuario entonces puede hacer clic en el botón  para que se active el script que genera automáticamente el subcorpus paralelo seleccionado.

La función del botón mencionado interactúa con el servidor web donde está alojado el catálogo de diccionarios y de proyectos. La activación implica el envío de un archivo con la especificación de los proyectos marcados previamente en la interfaz de selección con las propuestas capturadas. En la parte del cliente (ordenador del usuario) se crea una asociación con un script que lee los datos seleccionados y adapta la definición de proyectos especial para el gestor de corpus paralelo, el *corpus.prj*. El nombre *corpus.prj* es el que adjudica el script al activar la función de Abrir proyectos. El usuario puede cambiar el nombre del archivo.

El archivo \*.prj base es un archivo estándar que se adapta según la selección previa que haya hecho el usuario, esto es, según la combinación de lenguas y los directorios de trabajo donde se almacenan las memorias de traducción. Este archivo \*.prj base y adaptado se incluye en un proyecto temporal para ejecutar después el programa *Transit*.

Como explicamos a continuación, la activación de *Transit* permite el aprovechamiento de algunas de sus funciones claves. Estas características funcionales, que abordamos a continuación, hacen posible un buen rendimiento de explotación del corpus paralelo por disponer de:

- funciones de gestión de directorios con información etiquetada en XML —en el apartado [2.5.2. XML](#) hemos abordado ampliamente este concepto—; esta información vinculada a la estructura del almacenado de las memorias de traducción no se almacena en tablas, sino que permanece en sus archivos de creación, etiquetados en su origen con el lenguaje XML. Este modelo XML de marcación metasemántica hace viable cualquier manipulación con fines específicos, así como la integración en flujos de datos independientes del entorno de creación, como ya apuntábamos en una publicación anterior (Cámara, 2001). Y esto es así, porque, al tratarse de archivos abiertos independientes de la plataforma, pueden integrarse fácilmente en otros flujos de trabajo, así como proponemos hacer en el gestor de corpus paralelos. De esta

manera, se reutilizan las memorias de traducción —los corpus paralelos ya alineados y generados con *Transit*—;

- la función de búsqueda de concordancias, funcionalidad principal requerida por una herramienta de explotación de corpus. Así mediante la activación de *Transit* se pueden hacer búsquedas de concordancias bilingües o guardar el proyecto bajo otro nombre para usarlo en el futuro. Las ventajas de explotación de los corpus paralelos mediante las herramientas de concordancias ya se han señalado en el apartado 2.6.1. Sistemas integrados de traducción asistida. En el entorno de la traducción, estas bondades se resumen en la búsqueda contextualizada de terminología con sus equivalentes, así como en soluciones traductológicas para un proyecto de traducción. Así, mediante *Transit*, se realizan búsquedas sencillas, búsquedas de concordancias, para obtener información acerca de dónde y cómo se utiliza un término o cualquier cadena de caracteres. Los resultados siempre se visualizan en pares bilingües equivalentes, aprovechando la alineación de los datos. Estos datos se hallan importados en formatos binarios para describir armónicamente la estructura y la maquetación de los documentos. Este formato es el llamado TFF (*Transit File Format*), un DTD en XML cuya etiquetación se utiliza para la explotación de los corpus paralelos.

Una vez que el usuario determina qué documentos compondrán la colección seleccionada, que es siempre sólo una parte del contenido total del corpus paralelo, se genera el subcorpus. Y una vez se ha abierto la colección textual pretendida, desde ese subcorpus se hacen las búsquedas dentro de los textos para obtener la equivalencia del término en la otra lengua seleccionada, ya bien extraída de las concordancias ofertadas o de las búsquedas simples.

Por ejemplo, haciendo búsquedas por concordancias podemos buscar términos completos o cadenas. Nuestro resultado muestra una lista de

unidades de traducción donde aparece la ecuación de búsqueda, junto con la unidad de traducción equivalente en el otro idioma, como mostramos a continuación:

**Búsqueda de concordancia**

Buscar:  Equivalencia mín. (%):

Búsqueda de frases  
 Buscar en la lengua de llegada  
 Mayúsculas y minúsculas

Número de resultados que se muestran: 4

Similitud: 100%	Recuperado de: J:\VSR\9235\trwork\AD-Kapitel
Alemán	<i>Bei Störungen oder Schadensfällen infolge schlechter Druckluftqualität-Versorgung ist jegliche <b>Gewährleistung</b> ausgeschlossen!</i>
Español	<i>En caso de averías o daños causados por la mala calidad del abastecimiento de aire comprimido se descarta cualquier tipo de garantía.</i>
Similitud: 100%	Recuperado de: J:\Elexxon\8509\trwork\ww3\723_Allg.-Geschaeftsbedingungen1
Alemán	<b>§6-Gewährleistung</b>
Español	<b>§6-Garantía</b>
Similitud: 100%	Recuperado de: J:\Elexxon\8509\trwork\ww\Garantie-Zertifikat
Alemán	Der Käufer muss sich auf Garantie ansprüche anrechnen lassen, was er vom Verkäufer im Wege der <b>Gewährleistung</b> erhält.
Español	El comprador debe atribuirse los derechos de garantía que ha recibido por parte del vendedor en referencia a la aplicación de la garantía.
Similitud: 73%	Recuperado de: J:\Elexxon\8509\trwork\ww3\723_Allg.-Geschaeftsbedingungen1
Alemán	Es gilt die gesetzliche <b>Gewährleistungsfrist</b> von 2 Jahren.
Español	Tiene vigencia el plazo de garantía legal de 2 años.

**Fig. 101: Muestra obtenida mediante la búsqueda por concordancias en un subcorpus paralelo concreto**

Como se puede ver, el sistema nos proporciona, por una parte, el término de búsqueda *Gewährleistung* contextualizado, un segmento resaltado donde también se encuentra la equivalencia de ese término y otro tipo de informaciones, como el documento de donde se ha obtenido y el porcentaje de coincidencia de las unidades de traducción equivalentes.

Resumiendo, los componentes del sistema son un gestor documental de diccionarios de gestión semiautomática, un gestor de proyectos automático y un gestor de corpus paralelos que explota las posibilidades que ofrece el almacén de memorias de traducción. Mediante la interacción de las tres herramientas de implementación secuencial y la flexibilidad de *Transit* para la explotación de concordancias en la integración en entornos ajenos, se multiplica y amplía la capacidad de gestión automatizada de los recursos

lingüísticos. Así, la generación de subcorpus *ad hoc* a petición de los usuarios, con los criterios de búsqueda deseados, reutiliza la indización existente en las memorias de traducción para asociar los pares bilingües o los bitextos y, por otra, reutiliza la marcación semántica que incorporamos a los diccionarios para su identificación documental.

Se trata de tres proyectos secuenciales cuyos logros y ventajas se pueden comprobar al final de cada fase. El control y explotación del conocimiento corporativo global acerca de los contenidos de los trabajos que llevamos a cabo puede gestionarse de forma óptima desarrollando los tres módulos de COGNImática.

### **4.3. Fase de pruebas**

En el momento de la ejecución de la fase que abordamos a continuación, tanto los aspectos funcionales como técnicos del sistema, ya estaban preparados para evitar la improvisación. El término *fase de pruebas de software* —*software testing*, en inglés— se define como *el proceso de cuestionar un producto para poder evaluarlo* (Wikipedia, 2005).

Así, tras la implementación del sistema, entramos en la fase de pruebas que consideramos como parte integrada del ciclo de vida del proyecto total COGNImática. En esta fase se incorporan nuevos agentes, algunos usuarios finales con funciones y perfiles diferenciados de los diseñadores y de los programadores.

La realización de las pruebas sigue una metodología para probar de forma homogénea las funcionalidades de la herramienta y, sobre todo, debe anticipar la detección de errores antes de dar a conocer a todo el equipo las funciones del sistema. De esta forma, se puede menguar el impacto que pueden producir los cambios de los sistemas de información y minimizar costes, ya que el coste de corregir los errores crece exponencialmente según

avanza el proyecto, siendo ese coste, en consecuencia, significativamente menor cuanto antes se detectan los fallos.

Esta fase se ha estructurado en dos partes: la selección de los tipos de evaluadores o betatesters y los objetivos de los betatesters.

### **4.3.1. Selección de betatesters**

La elección de los evaluadores ha ido en función de los tipos de usuario que deben utilizar la herramienta que, como hemos explicado en el apartado 4.1.1. Diseño de la información e implementación de la arquitectura de la información son de tres tipos. En este sentido, se ha tenido también en cuenta que no es posible hacer pruebas acerca de la usabilidad de la función de consulta, hasta no haber recogido datos, esto es, hasta no haber catalogado los diccionarios, cuyos datos vehiculan las funciones del gestor de proyectos y, éste, del gestor de corpus paralelos. Por eso, en una primera instancia de la fase de pruebas, se han seleccionado solo dos tipos de usuarios y no los tres perfiles de usuarios previstos.

Por una parte, se han designado dos gestores con experiencia no sólo como gestores de proyectos de traducción, sino como betatesters para proyectos de localización de software (véase el apartado 2.3.3. Fase de traducción de la gestión de documentación multilingüe). Éstos catalogaron los diccionarios de diez clientes concretos seleccionados en función del volumen de trabajo realizado para esos clientes.

Por otra parte, se ha determinado que la persona que hace las funciones de terminólogo de la empresa haría la función de administrador del sistema. La función del administrador es la de controlar (editar, modificar y eliminar) los datos introducidos por los gestores y cumplimentar el campo para el que solo él tiene acceso; esto es, el campo para introducir los descriptores del tesoro *Spines*.

En cuanto al nivel de sólo consulta, solo se llevan a cabo las pruebas sistemáticas con más agentes tras haber recogido la mitad de los diccionarios inventariados. Y esto es así, porque de otra manera no se contaría con datos suficientes para su explotación.

#### **4.3.2. Objetivos de los betatesters**

Si bien existe una norma estándar específica donde se clasifican los criterios que se deben tener en cuenta para evaluar cualquier sistema de software (ISO 9126: 2003), los usuarios deben llegar a la evaluación de la herramienta mediante una metodología de inducción. Esto es, mediante la realización de tareas concretas de catalogación con ejemplos de casos particulares. De esta manera los usuarios que hacen la función de betatesters pueden llegar incluso a conclusiones que repercuten en el diseño original del sistema propuesto.

La fase de pruebas se comienza con la catalogación de los diccionarios por parte de los usuarios del segundo nivel. Se comienza pues la catalogación de los diccionarios existentes. La primera fase de catalogación se había iniciado por los desarrolladores de la herramienta antes de iniciar la fase de pruebas con los betatesters elegidos. Esto significa que se ha ido catalogando entre la etapa de trazado de los criterios de clasificación y la de implementación de la primera propuesta informática. De otra manera, los betatesters habrían contado con demasiados errores que habría ralentizado la fase de pruebas propiamente dicha. Estos primeros elementos registrados sirven de referencia a los usuarios seleccionados para hacer las primeras pruebas.

Así, se dan las instrucciones de catalogación a los cinco gestores de proyectos, quienes continúan la labor de catalogación desde la perspectiva de usuario real.

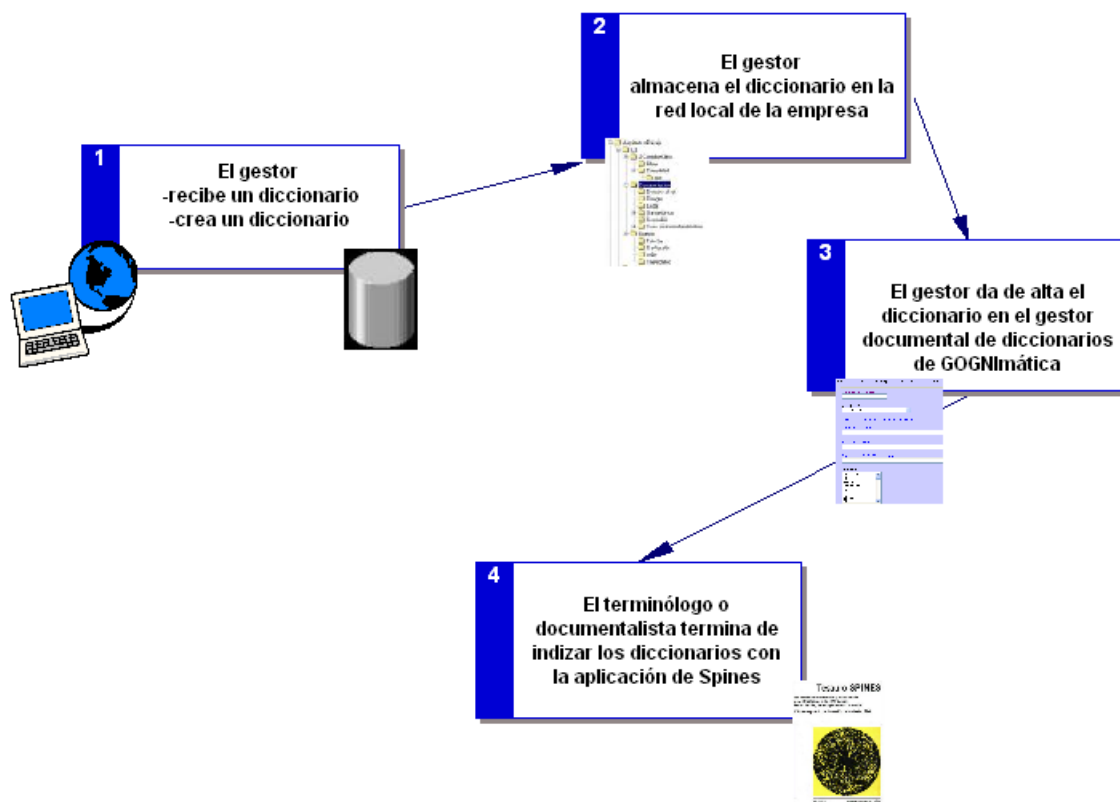
Una vez que los gestores de proyectos han llevado a cabo la recogida de los datos asociados a los diccionarios —combinación de lenguas, formato, ruta



de acceso, nomenclatura, datos de mantenimiento del diccionario, comentarios y palabras clave—, el administrador del sistema debe proceder a terminar la anotación semántica asociada a los diccionarios. Esta tarea, como ya hemos explicado, la asume el terminólogo de STAR SL, quien lleva a cabo la introducción de los descriptores del tesauro *Spines*. Y las palabras clave por parte de los gestores le deben servir de referente para la selección de los términos descriptores predefinidos del *Spines*.

La descripción del escenario de aplicación del sistema COGNImática es imprescindible para poder dar las instrucciones a los betatesters. Para ello, se han descrito los dos flujos generales: por una parte, las fases de registro en el gestor documental de COGNImática y, por otra, los tipos de explotación de los recursos informacionales y lingüísticos almacenados.

Durante la tarea de registro de los diccionarios en el gestor documental se establecen cuatro fases para la recogida de datos con la intervención de dos perfiles de usuarios, como se muestra en el siguiente diagrama de flujo:



**Fig. 102: Fases de registro en el gestor documental de COGNimática**

Como se desprende del diagrama, en este flujo intervienen los dos tipos de usuarios mencionados, cuyos perfiles profesionales están vinculados aunque sean diferentes. Estos usuarios son agentes indizadores de algunos de los recursos almacenados, concretamente de los recursos terminológicos. En las memorias de traducción estos agentes ya no participan, ya que el proceso de indización de estos recursos se hace de forma automática reutilizando la catalogación semántica hecha a los diccionarios.

Al perfil de terminólogo hemos añadido el de *documentalista*. En STAR SL no hay ninguna persona con perfil de documentalista, sin embargo, las tareas vinculadas al mantenimiento de COGNimática que se adjudican al terminólogo, podrían perfectamente estar desempeñadas por una persona cuya formación se haya centrado en la gestión de la información, como se lleva a cabo en las disciplinas de biblioteconomía y documentación y como abordaremos más adelante en el apartado 4.3.1. Perfiles y competencias de los usuarios.

Una vez que todos los betatesters han terminado sus tareas evalúan el gestor documental de diccionarios, pero no sólo los elementos de interacción existentes, como la interfaz de alimentación específicamente, sino también los no existentes. Y esto es así, porque ambos perfiles de usuarios mencionan en su evaluación que precisan además contenidos de fácil acceso que especifiquen los procedimientos para explotar las posibilidades de la herramienta. Es decir, el conocimiento de procedimientos debe convertirse en una guía de usuario en línea para poder ofrecer asistencia al usuario en procesos de interacción mediante la documentación de fácil acceso.

Ambos tipos de usuario enuncian la necesidad de mejorar el acceso al sistema en sí y a la documentación del sistema para optimizar los procesos. Asimismo, el terminólogo aborda la necesidad de tener herramientas de comunicación integradas en el propio sistema que contribuyan a la gestión de las preguntas frecuentes, los problemas y las sugerencias relacionadas con los procedimientos. Mediante tales herramientas se debe proporcionar al usuario la posibilidad de manifestar sus propuestas de mejora o preguntas de utilización del sistema y que de éstas se aproveche el resto de los usuarios.

El administrador del sistema, es decir el terminólogo, también alude a los problemas de indización directos e indirectos achacados también a un problema de formación previa. Es decir, los usuarios no introducen palabras clave que sean lo suficientemente representativas que puedan servir de referencia para que el terminólogo pueda seleccionar los descriptores *Spines*, a eso se refiere el administrador de la herramienta cuando habla de problemas indirectos de indización. El administrador del sistema debe resolver entonces la introducción de los descriptores de dos maneras. O bien se dirige por correo electrónico o de forma oral a los gestores para que le proporcionen más datos sobre el cliente; o bien recaba información acerca del cliente en Internet, lo que supone una ocupación mayor de tiempo en la catalogación semántica. En este sentido, considera necesario tener una herramienta de control de registros.

Asimismo, el administrador considera un problema directo de indización al resultante de su propia carencia de formación en análisis documental, área de conocimiento hasta ahora impartida más a los documentalistas que a los terminólogos.

Por último, el terminólogo aborda la carencia de herramientas de control que optimice una parte del flujo descrito en la *Fig. 101: Fases de registro en el gestor documental de COGNImática*. Esto es que de forma automática se pueda saber qué diccionarios se han registrado en el gestor documental de diccionarios suprimiendo el procedimiento de aviso mediante correo electrónico.

Además de la evaluación aportada por los betatesters derivadas del desempeño de las tareas asignadas, también se detecta el no desempeño de las tareas por parte de algunos agentes. El no cumplimiento de los objetivos por parte de dos usuarios de segundo nivel lo explican por *falta de tiempo y recursos*. De una entrevista más con los gestores que no han llegado a catalogar se deriva que han interpretado la tarea de catalogación de los diccionarios no como la gestión terminológica de las macroestructuras, sino como la edición y revisión de las microestructuras; esto es los registros terminológicos que componen cada diccionario o macroestructura. Ciertamente, el control de calidad de los términos de un diccionario requiere de pautas generales y específicas, así como también de una dedicación de tiempo mayor que el que hay que invertir para dar de alta un diccionario en el catálogo. Esta interpretación nos lleva a pensar que se debería invertir tiempo en una mínima formación avalando así la necesidad de documentación acerca de los procedimientos en línea y la formación en indización.

Asimismo, se detecta que la información apuntada en el botón:

A rectangular button with a light blue background and a thin black border. The text 'Introducir nuevos datos terminológicos' is centered within the button in a dark grey font.

es ciertamente equívoca y puede inducir a pensar más en la edición de entradas terminológicas de diccionarios

que en la catalogación de esos diccionarios. Un cambio de ese texto a *Introducir nuevos diccionarios en el catálogo* podría prevenir los equívocos.

Una vez que los usuarios de segundo y tercer nivel hayan realizado sus tareas, deben intervenir en último lugar los usuarios de primer nivel, es decir, aquéllos que sólo recuperan y extraen información del sistema sin alimentarlo. El perfil de las personas seleccionadas son los tres tipos de usuarios finales: gestores de proyectos, revisores y traductores.

A los gestores de proyectos se les ha dado instrucciones de cómo pueden utilizar el gestor documental de diccionarios. Si necesitan saber, por ejemplo, si cuentan con recursos lingüísticos para poder hacer un presupuesto con posibilidades de negociación o si cuentan con recursos lingüísticos susceptibles a ser definidos en la configuración de proyectos determinada.

Si necesitan informaciones asociadas a los proyectos, como por ejemplo, qué proyecto han realizado para un cliente, qué clientes lleva un gestor concreto, etc. pueden hacer una consulta en el gestor de proyectos.

A los traductores así como a los revisores se les ha explicado cómo pueden utilizar el gestor documental de diccionarios, por ejemplo, para buscar diccionarios adicionales, más allá de los definidos en un proyecto. Asimismo, a través del gestor de corpus paralelos pueden acceder a los corpus paralelos en busca de concordancias en diferentes idiomas.

Así como la recuperación de la información de los diccionarios y de los proyectos de traducción no ha presentado problemas al usuario del primer nivel, la extracción sí que ha resultado complicada para los usuarios. Y derivado de las evaluaciones inferimos que más allá de entender los pasos para activar una función específica; esto es *Abrir proyectos* para poder acceder a los contenidos textuales del corpus paralelo y extraer concordancias, deben entender previamente el concepto de corpus textual y, más aún, de corpus textual paralelo.

De las propuestas específicas propuestas por los diferentes usuarios que han probado el sistema se deriva además que el sistema se aplica o pretende ser aplicado como:

- un entorno web con una interfaz orientada a las tareas (gestión del conocimiento de procedimientos) y
- un sistema de información hipertextual orientado a la explotación de la información asociada a las bases terminológicas y traductológicas.

#### **4.4. Fase de rediseño y desarrollo de la versión 1.1.**

El objetivo de esta fase es mejorar las bondades del sistema teniendo en cuenta las necesidades apuntadas por los usuarios seleccionados. Por una parte, los usuarios que han servido de *conejiillos de indias* se inclinan a que se fomente la gestión del conocimiento; esto es, a conseguir condiciones que mejoren la gestión de la información y la comunicación. Y esto se afronta mediante tres estrategias que mencionamos a continuación:

- la formación de los agentes
- la mejora del acceso a la documentación para la explotación del sistema
- el diseño e implantación de instrumentos de colaboración

Por otra parte, y siguiendo las necesidades propuestas del usuario administrador, se diseñan instrumentos de control para poder optimizar el flujo tipificado en la *Fig. 101: Fases de registro en el gestor documental de COGNImática*.

Los siguientes apartados los dedicamos a explicar el desarrollo de las propuestas de afinamiento de la herramienta.

#### 4.4.1. Fomento de la gestión del conocimiento

La gestión del conocimiento es un concepto que engloba la generación, la representación, la adquisición —término también conocido como gestión del aprendizaje—, y la transferencia de saber, por tanto, es obvio inferir que la formación es una de las estrategias más relevantes para la gestión del conocimiento (Cámara, 2004). Para formar a las personas, el conocimiento debe estar previamente representado, con el fin de poderlo transferir y cumplir así el objetivo de la formación: adquirir y generar conocimiento en las personas receptoras del saber.

La propuesta repetida de los agentes de recibir mayor formación nos hace replantearnos los canales de transferencia de conocimiento existentes en el entorno de la herramienta. Este replanteamiento obliga a promover espacios, lenguajes y entornos de comunicación que ofrezcan soporte al aprendizaje colaborativo. En este sentido, se plantea cómo se debe llevar a cabo la transferencia del conocimiento y se establece una formación introductoria presencial. Esta formación presencial tiene como objetivo establecer un acercamiento a la tecnología desarrollada donde se expongan los objetivos y perspectivas del sistema que se va a implantar en el entorno de trabajo.

El principal objetivo de la formación presencial para los usuarios del primer nivel es conseguir la identificación de los corpus paralelos —las memorias de traducción— con las bases de conocimiento traductológico y terminológico como primer paso para que exploren y exploten el sistema.

Para los usuarios que tienen permisos de segundo nivel, además de lo expuesto en el párrafo anterior, el objetivo de la sesión presencial es la de abordar los beneficios de la indización documental y, de esta manera, alentar a los gestores en la introducción de los diccionarios. En estas sesiones presenciales se deben realizar prácticas para interiorizar los procedimientos de la indización.

Los usuarios de tercer nivel deben contar también con una buena formación en indización para controlar tanto la fase de análisis conceptual de los recursos lingüísticos, como la de traducción a identificadores semánticos de contenido al sistema conceptual decidido. Para ello, y siguiendo a Lancaster (1995. 286), el indizador debe comprender, por una parte, la materia del documento y, por otra, debe conocer muy bien las necesidades de los usuarios.

Adicionalmente los usuarios de tercer nivel deben contar con formación teórica más especializada en gestión de bases de datos documentales para entender también la forma de trabajo de los motores de búsqueda.

Esta base teórica debe estar abocada a posibilitar a los destinatarios a llevar a cabo un mejor control de calidad; además de proporcionar *ad hoc* la formación de los usuarios de segundo nivel.

Se establece también mejorar el conocimiento representado; esto es, mediante una documentación en línea pertinente que al mismo tiempo sea de fácil acceso para que repercuta en el fomento de la adquisición y transferencia del conocimiento de procedimientos.

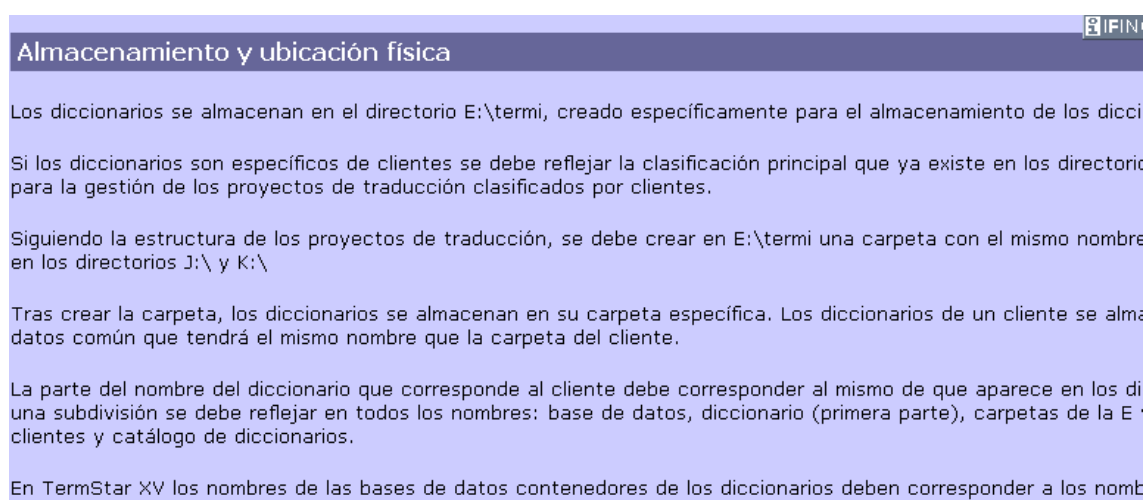
Por tanto, el fomento de la gestión del conocimiento implica una mejora de la gestión de la información (conocimiento representado) y de la comunicación (transferencia de conocimiento), temas que abordamos en los siguientes apartados.

#### **4.4.1.1. Acceso a la información: documentación en línea**

Ya en la primera propuesta informática de COGNIMática se había incluido un acceso directo a algunos criterios de clasificación, como se deriva de la *Fig. 83: Interfaz de búsqueda del gestor documental de diccionarios (1ª propuesta)*. También es cierto —y por ello están justificadas las sugerencias de los usuarios— que estos criterios de clasificación se remitían solo a los criterios según el estado del diccionario.

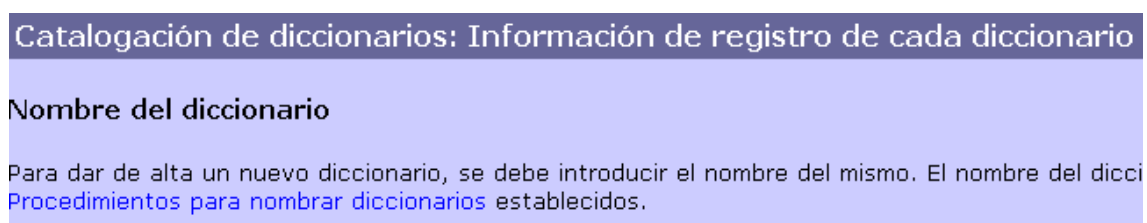


Esta nueva propuesta que tiene como objetivo mejorar el soporte procedimental a los usuarios del sistema amplía la documentación de la herramienta. Ésta se compone de un manual de usuario y utilización: guía de búsqueda y guía de almacenamiento. Esta información alude a procedimientos referentes a la organización física (ubicación o almacenamiento) de los diccionarios que se deben catalogar, como se muestra en la siguiente captura de pantalla:



**Fig. 103: Criterios de almacenamiento de los diccionarios en línea**

Además de procedimientos a los que se accede mediante la tecnología del hipertexto, se resaltan los hipervínculos, tal y como se muestra en la siguiente captura:



**Fig. 104: Hipervínculos de acceso a los contenidos de procedimiento**

La ayuda en línea contribuye al mantenimiento y actualización de los registros de los diccionarios accesibles en el entorno de la herramienta, como también se muestra en la siguiente captura de pantalla:

<p><b>Estado del diccionario</b></p> <p>Los diccionarios se deben marcar según su estado. Los tipos de diccionario según su estado (ACT, FRL, HIS, REF, VAL) de su gestión e intercambio entre STAR Servicios Lingüísticos SL, clientes y autónomos.</p> <p><b>Contenido temático del diccionario</b></p> <p>Los diccionarios se deben marcar según su contenido. El área temática en la que se inscriben los diccionarios se detern campos asociados a la información de cada diccionario. Los tres campos que dan información semántica acerca de los diccionarios son: Tema de traducción, palabras clave y descriptor Spines.</p> <p><b>Combinación de lenguas</b></p> <p>Los diccionarios catalogados deben contener la información relativa a las lenguas del diccionario. La información acerca seguir los códigos de lengua estandarizados también usados en TermStar XV y STARTeam</p> <p><b>Sugerencias de mantenimiento / Comentarios</b></p> <p>La información acerca de cómo se debe seguir gestionando el diccionario registrado es optativo, aunque se recomienda. En este apartado sí se deberían incluir, sin embargo, el número de entradas principalmente en los diccionarios ACT y V/</p>
---

**Fig. 105: Procedimientos de catalogación en línea**

Resumiendo, esta documentación en línea atiende a la gestión por procesos avalados por procedimientos de trabajo que repercute en una sistematización de los flujos de trabajo para su optimización.

#### **4.4.1.2. Fomento de la comunicación: herramientas colaborativas**

Para el fomento de la comunicación se ha previsto la incorporación de una herramienta colaborativa que dé soporte a la transferencia de los contenidos de procedimiento accesible en el entorno habitual de trabajo para que todos los usuarios pueden generar, cuestionar, sugerir y manipular.

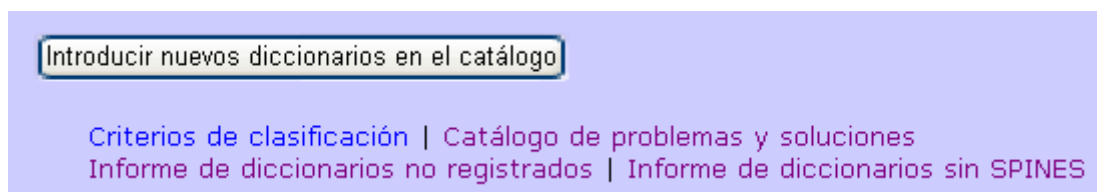
Basado en uno de los supuestos de este trabajo de que la mejora al acceso a cualquier tipo de información también repercute en la mejora de la comunicación y de la gestión del conocimiento, en la propuesta afinada de COGNImática, el sistema queda incorporado en el entorno de la intranet de STAR SL. Esto quiere decir que se accede a COGNImática a través del entorno de la intranet de la empresa que se ha definido también en función de las necesidades de los usuarios.

Asimismo, en la interfaz principal que corresponde con la interfaz de búsqueda del catálogo de los diccionarios, se ha incorporado, además del

hiperenlace que nos lleva a los *Criterios de clasificación*, un enlace que nos lleva a un *Catálogo de problemas y soluciones* y dos más que enlazan con las herramientas de control automático que generan informes de diccionarios no registrados e informes de diccionarios sin descriptores *Spines* —tema que abordamos en el siguiente apartado 4.2.4.2. Herramientas de control—.

Además de estos enlaces mencionados, se ha realizado el cambio de texto de un botón con la siguiente leyenda descriptiva: *Introducir nuevos diccionarios en el catálogo*.

De esta manera, como ya se había propuesto en el apartado 4.2.3.2. Objetivos de los betatesters, prevenimos posibles equívocos. En la siguiente captura de pantalla se pueden apreciar los cambios que acabamos de mencionar:



**Fig. 106: Acceso a procedimientos del gestor documental**

El catálogo de problemas y soluciones es un sencillo instrumento de comunicación de base colaborativa que contribuye a gestionar mejor el conocimiento. En la siguiente captura se muestra qué tipo de información se gestiona mediante esta herramienta colaborativa:

[->Japanese][Top Page][Edit][All Pages][Recent Changes] Search: <input type="text"/>		
Problema	Comentario	Solución
Caracteres especiales	por alguna razón algunos caracteres del alemán (ü) y otros se desconfiguran (&)	Evitar este tipo de caracteres.
El nombre que aparece registrado en el catálogo no coincide exactamente con el nombre de la base de datos	Si no coinciden las cadenas de caracteres, la funcionalidad de la herramienta se ve reducida	Cuando se cree el diccionario en TermStar y se escriba el nombre, se debe copiar y pegar el nombre en el catálogo al mismo tiempo. También se puede tomar el nombre del prj.
Los datos del diccionario se han dejado en TermStar received	Si no almacenamos el diccionario recibido, se corre el riesgo de perder la información relacionada con ese diccionario. La información asociada al diccionario previsiblemente sólo la tendrá la persona que recibe este diccionario en local, si no se renombra y cataloga	Se debe crear un diccionario interno, importarlo, nombrarlo y catalogarlo de acuerdo con la nomenclatura.

**Fig. 107: Catálogo de errores y soluciones de COGNimática**

La base técnica de este catálogo es un editor *wiki*. El término *wiki* que significa *rápido* en la lengua hawaiana se utiliza como prefijo en muchos sitios de Internet para nombrar una colección de páginas web hipertextuales donde al usuario se le permite crear y editar contenido de páginas con formato web, accediendo a ellas usando un navegador, como por ejemplo, la *Wikipedia*<sup>1</sup>. Un editor *wiki* permite la colaboración en grupos que comparten espacios de discusión para poder informarse a tiempo real de cualquiera de las actualizaciones realizadas o llevar a cabo trabajos en equipo. Este canal de intercambio y participación entre los usuarios en red muestra como virtud esencial la descentralización de toda actuación y la posibilidad de compartir todo el proceso durante la toma de decisiones (Castells, 2001). Asimismo, el fomento del intercambio y la participación, es decir, la comunicación basada en el formato de red aumenta la eficacia en la gestión del conocimiento (Lozano, 2004).

<sup>1</sup> Para más información véase [http://en.wikipedia.org/wiki/Main\\_Page](http://en.wikipedia.org/wiki/Main_Page)

#### 4.4.2. Herramientas de control

El procedimiento que hasta ahora habían iniciado los usuarios de segundo nivel, es decir, los gestores de proyectos que llevan a cabo una parte de la catalogación semántica, era el de enviar al terminólogo una notificación de un nuevo registro de diccionario mediante un correo electrónico. El terminólogo entonces procedía al control de los datos del registro y a la introducción del campo de los descriptores. Sin embargo, esta forma de proceder no garantiza que todos los diccionarios creados se registren, así como tampoco que todos los diccionarios registrados sean notificados a la instancia correspondiente. Por eso, se han desarrollado dos pequeñas aplicaciones que simplifique a todos los implicados el trabajo. Por una parte, se ha dispuesto una herramienta de control para los diccionarios no catalogados; y por otra, una herramienta de control para capturar solo aquellos diccionarios que carecen de descriptor *Spines* y así generar un informe.

Los elementos no catalogados pueden capturarse comparando el campo del nombre de los diccionarios registrados en el módulo del gestor de diccionarios, con los elementos que aparecen en la etiqueta [*Dictionaries*] que aparecen en los archivos \*.prj. A la morfología y estructura de los archivos \*.prj nos hemos remitido ampliamente en el apartado 3.3.3.1.2. Gestión del corpus paralelo con *Transit* y el papel crucial que juegan estos archivos en el funcionamiento de COGNImática, también se ha visto en el apartado 3.4.3.2. Diseño del gestor de proyectos. La implementación de esta sencilla herramienta se realiza mediante un script. Y la herramienta se activa haciendo clic en el enlace generando así el informe de los diccionarios no registrados en COGNImática, a pesar de que existen. En la siguiente captura de pantalla mostramos un informe resultante tras activar la herramienta descrita:

Informe diccionarios   Informe clientes	
Thu Jul 28 10:53:09 CEST 2005	
Número de resultados: 12	
diccionario	
BMW motorrad basis	
BMW_Schulungsunterlagen_BA1	
BoschEW_Herramientas-electricas_FRL	
BoschKFZ_Fahrsicherheitssysteme_FRL	
BoschKFZ_Global_FRL	
BoschKFZ_Grundlagen-Physik_FRL	
BSH_Hornos_ACJ_FRL	
Buehler_Muellerei_VAL	
Cooper_Tools_ACT	
Cooper-tools_PT_VAL	
TD	
TZK_0403	

**Fig. 108: Informe de diccionarios no registrados**

Basándose en este informe, el terminólogo puede obtener diferentes informaciones relevantes para su trabajo, por ejemplo, qué diccionarios no deben registrarse aunque aparezcan en el informe —como el caso de los diccionarios tipo FRL, tal y como se explicaba en el apartado [3.4.1.2.1. Clasificación de diccionarios por estado](#)—; qué diccionarios no siguen la sintaxis de la nomenclatura prescrita, como es el caso de *TD* y *TZK\_0403*; y, por último, qué diccionarios requieren terminar su registro, como es el caso, por ejemplo de *Cooper\_Tools\_ACT* y *Cooper-tools\_PT\_VAL*.

La herramienta para capturar los diccionarios catalogados cuyo registro está incompleto, es decir, que carecen de descriptores *Spines*, se compone también de un sencillo script. Esta vez busca los campos vacíos, es decir, sin

texto, del campo descriptores *Spines*. En la siguiente captura de pantalla mostramos un informe resultante tras activar la herramienta descrita:

Informe: diccionarios sin SPINES	
Informe diccionarios   Informe clientes   Diccionarios sin SPINES	
Thu Jul 28 11:04:02 CEST 2005	
Número de resultados: 11	
nombre	ruta
BMW_DIS-USP_ACT	E:\termi\BMW\BMW.mdb
BSH-Mont_induccion-vitro_ACT	BSH-Mont.mdb
Gucci_ACT	Gucci.mdb
Illig_ACT	Illig.mdb
ISA_ACT	ISA.mdb
Mobile_ACT	Inverline.mdb
Nextpage_displays-DVDPlayer_VAL	E:\termi\Nextpage\Nextpage.mdb
Renault_REX_FRA_VAL	E:\termi\Renault\Renault.mdb
Renault_RIN_FRA_VAL	E:\termi\Renault\Renault.mdb
Toyota_OM-2005_ESM_VAL	E:\termi\toyota\CMC-Toyota_OM.mdb
Wiedenmann_ACT	Wiedenmann.mdb

**Fig. 109: Informe de diccionarios sin *Spines***

De esta manera, el terminólogo podrá organizar y priorizar mejor las tareas de catalogación pendientes sin temor a posibles olvidos o que se traspapele la información, como sucedía con el procedimiento anterior de notificación por correo electrónico, explicado anteriormente.

Resumiendo, de las primeras pruebas se manifiestan los primeros fallos y carencias detectados por dos tipos diferentes de usuarios que nos llevan a formular las propuestas de mejora. Una de estas propuestas es la de hacer más usable el sistema y la de mejorar la formalización por parte de los gestores de proyectos de los nuevos procedimientos de trabajo. No obstante y, a pesar de los problemas detectados y los cambios que se han llevado a

cabo, se consideran muy positivos los resultados derivados de la implementación de COGNImática por las siguientes razones:

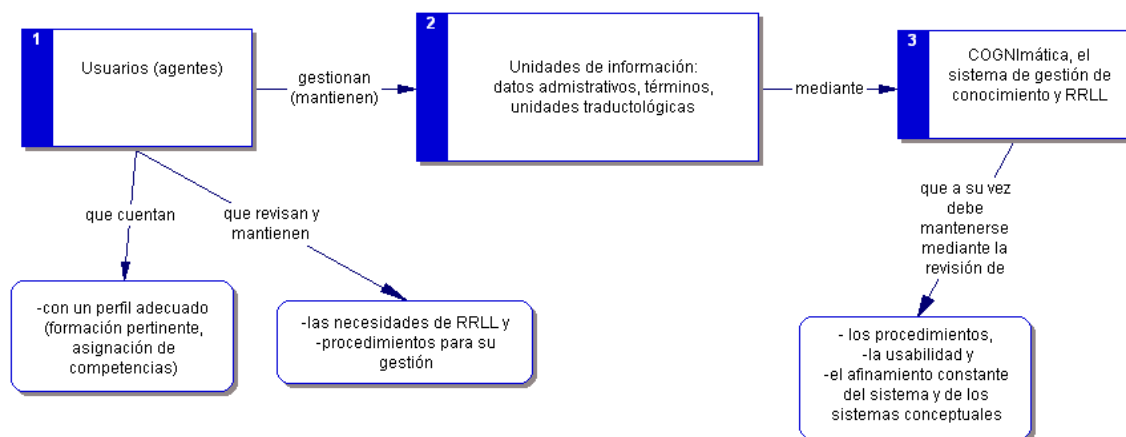
- la información llega en el momento oportuno al usuario adecuado: lo que fomenta el trabajo colaborativo;
- el enfoque orientado al usuario y a la mejora continua lleva a la excelencia en la gestión;
- la mejora en la automatización de las tareas y almacenamiento automatizado de los recursos lingüísticos y sus informaciones asociadas mejora los procesos de negocio;
- la buena organización de los recursos lingüísticos permite su posterior localización y reutilización;
- los recursos lingüísticos centralizados ofrecen una visión global del negocio y del conocimiento mejorando la calidad y la cantidad de la información disponible por contar con datos más precisos y actuales;
- los recursos lingüísticos centralizados garantizan además el control de privilegios de acceso por tipos de usuario;
- la eficiencia en la toma de decisiones operativas y en las relaciones con clientes y traductores autónomos mejora;
- la gestión de estos activos favorece también la comunicación, tanto la interna como la externa con clientes, proveedores y otros colaboradores;
- el diseño es sencillo, fácil de utilizar en un entorno amigable.

#### ***4.5. Seguimiento y mantenimiento de los recursos lingüísticos***

En el presente apartado abordamos la última fase de la auditoría de la información; asimismo abordamos la última fase del ciclo de vida de



COGNmática. Como explicábamos en el apartado 3.2.1.6. Seguimiento y mantenimiento, esta última fase se inicia tras la implementación de los protocolos de trabajo, de las propuestas técnicas y tras haber realizado las primeras pruebas de las herramientas implementadas por los usuarios previstos. Los objetivos de la fase de seguimiento y mantenimiento es, como su propio término indica, una fase de continuación de un proyecto iniciado. El proyecto de auditoría que comienza con la decisión de aplicar una política de la información, como explicábamos en el apartado 3.2.1.1. Implicación y decisión por parte de la dirección de la entidad en implantar una política de la información, culmina con la revisión de esa política ampliando el concepto de política de la información a *política de formación e información*, con el fin de gestionar el conocimiento interno de la entidad analizada. Para afrontar este cambio, se debe abordar el seguimiento de toda la política de la (in)formación sobre las tres patas sobre las que se apoya: los usuarios implicados, las unidades que se gestionan y los sistemas implementados para automatizar los flujos de trabajo. El desarrollo de la interacción de estos tres pilares y sus puntos clave que todavía no se han desarrollado a lo largo del trabajo, ocupará la última parte de nuestra auditoría, tal y como proponemos en el siguiente mapa conceptual:



**Fig. 110: Seguimiento y mantenimiento de una política de gestión de los RRLL**

Por último, describimos las propuestas de afinamiento del sistema técnico y conceptual para sofisticar sus funcionalidades y con ello el rendimiento.

### 4.5.1. Perfiles y competencias de los usuarios

Para hacer un seguimiento adecuado de cómo se mantienen los recursos lingüísticos, se debe hacer un seguimiento de los agentes implicados en su gestión y de las tareas implicadas en los flujos, tema ya abordado en el apartado 4.2.3.2. Objetivos de los betatesters y sintetizado en parte en el diagrama *Fig. 101: Fases de registro en el gestor documental de COGNImática*.

También, tal y como se ha visto en el apartado 4.2.3.2. Objetivos de los betatesters y después en el apartado 4.2.4.1. Fomento de la gestión del conocimiento, se ha insistido en la necesidad de ofrecer formación a los usuarios para mejorar la alimentación y explotación de las unidades de información registradas en COGNImática. Los traductores, gestores de proyectos y el terminólogo son los usuarios implicados a los que se les ha tenido que formar para poder desempeñar con satisfacción sus tareas. Partiendo de una situación ideal, consideramos imprescindible la reformulación y formación de los perfiles profesionales implicados, mucho antes de asumir las tareas ya especificadas para la gestión de los recursos lingüísticos y de todos los recursos informacionales que se gestionan en el ecosistema de una organización que se dedica a la gestión de proyectos de traducción. Una situación ideal relegaría a todos los gestores de información *improvisados* a profesionales de la información con especialidades adicionales.

Nuestra propuesta ideal se basa en tres perfiles: el de gestor del conocimiento o gestor de la (in)formación, el de terminólogo y documentalista y, por último, un técnico para asumir las competencias que explicamos a continuación:

El gestor de conocimiento o gestor de (in)formación debería contar con conocimientos teóricos o aplicados de lingüística, traducción, documentación e informática aplicada. Su actividad principal debe centrarse en diseñar un

modelo conceptual y de procedimientos para implantar una política de la información. Para ello debe ser capaz de realizar las siguientes tareas:

- ❑ analizar los flujos de trabajo;
- ❑ detectar los cuellos de botella en los flujos de trabajo;
- ❑ recoger los problemas desde diferentes ópticas de los usuarios;
- ❑ intercambiar información con los diferentes responsables de la dirección de la empresa;
- ❑ canalizar sugerencias de los usuarios para la optimización de los sistemas de información;
- ❑ trazar propuestas de mejora;
- ❑ fomentar y sistematizar la gestión del conocimiento mediante la oferta continuada de formación;
- ❑ identificar el conocimiento tácito y representarlo para hacerlo explícito mediante su registro en los sistemas de información;
- ❑ diseñar sistemas de información de acuerdo con las necesidades del entorno y en consonancia con la dirección estratégica de la organización;
- ❑ definir las entidades básicas de los sistemas de información;
- ❑ diseñar modelos conceptuales para la gestión de bases de datos;
- ❑ estructurar la documentación de los sistemas de información y documentar las especificaciones de las herramientas (entidades, atributos e interacción entre ellos);
- ❑ gestionar la formación de los miembros de la plantilla y asignar recursos humanos para impartir los cursos de formación;
- ❑ tener un profundo conocimiento del contenido de los recursos de información, incluyendo la capacidad de evaluarlos y filtrarlos críticamente;

- mejorar continuamente los servicios de información en respuesta a las necesidades cambiantes;
- contribuir con su asesoramiento en temas de información y optimización de la producción.

El terminólogo y/o documentalista, más allá de que todos los traductores desempeñen tareas de gestión terminológica cuando traducen o de gestión de la información en el caso de los gestores de proyectos cuando gestionan diccionarios o alimentan los sistemas de información, debe contar con una formación específica en terminología y documentación en pro a la unificación y optimización de procesos. La persona con este perfil debe llevar a cabo las siguientes tareas:

- intervenir y evaluar el flujo de trabajo relacionado con la gestión de terminología, la gestión de documentos y el mantenimiento de las bases de datos;
- evaluar los resultados del uso de la información e investigar para solucionar los problemas de gestión de la información mediante, por ejemplo, encuestas a los usuarios;
- monitorizar las extracciones terminológicas de las revistas especializadas a las que está suscrita la empresa;
- preparar diccionarios para su publicación en línea;
- servir de soporte a los proyectos de traducción para la manipulación de bases terminológicas o extractos de memorias de traducción;
- analizar la oferta editorial de diccionarios generales y especializados (en formato papel y electrónico) y hacer propuestas de adquisición a la dirección;
- analizar la oferta en línea (WWW) de diccionarios generales y especializados y divulgación de esta información entre los miembros de la empresa;

- analizar la oferta editorial de libros de texto y obras didácticas relacionados con los temas que se traducen en la empresa y propuestas de adquisición;
- analizar la oferta en línea (WWW) de páginas de contenido didáctico o de interés y divulgarla entre los otros miembros de la empresa;
- mantener la biblioteca de la empresa;
- gestionar diferentes bases de datos con información crítica para la empresa como: la base de datos con información de gestión de proyectos, base de datos con información acerca del perfil de los traductores internos y autónomos, base de datos de recursos terminológicos propios y de los clientes asociada a las memorias de traducción;
- realizar propuestas de mejora de las bases de datos: entidades, atributos, enlaces, acceso, lenguaje de interrogación;
- proponer el diseño de bases de datos específicas, en el caso de no existir;
- impartir cursos de formación en gestión de terminología y documentación.

El técnico debe contar con conocimientos aplicados de lingüística y conocimiento de informática aplicada para el desarrollo de tecnología web, motores de búsqueda y desarrollo de interfaces y bases de datos. La actividad del técnico con este perfil es la de proporcionar soporte técnico para implementar técnicamente los modelos y procedimientos diseñados en el marco de implantación y mantenimiento de una política de la información. En este sentido, el técnico debe llevar a cabo las siguientes tareas:

- implementar las herramientas diseñadas propuestas para la gestión de la información y conocimiento para el uso interno o externo de la organización o para usuarios individuales;

- actualizar en la intranet los cambios propuestos;
- producir documentación técnica para la tecnología desarrollada;
- desarrollar productos de información especializados para el uso interno o externo de la organización o para usuarios individuales (filtros, scripts, bases de datos, etc.);
- evaluar aplicaciones de software para la mejor implantación de las herramientas diseñadas;
- aplicar la tecnología apropiada para adquirir, organizar y difundir la información;
- impartir formación.

La incorporación de estos tres perfiles definidos no sólo dan soporte al control continuado de la información en general, sino proporciona la base para el desarrollo conceptual y técnico adicional de otros instrumentos para mejorar todos los flujos de trabajo implicados en la gestión de la información general de la empresa.

#### **4.5.2. Unidades de información: los recursos lingüísticos y su información asociada**

En la *Fig. 56: Información asociada a los recursos lingüísticos* del apartado 3.4. Diseño e implementación de las propuestas metodológicas, conceptuales y tecnológicas se sintetizan los recursos informacionales asociados a los recursos lingüísticos implicados en los flujos de trabajo de nuestro análisis.

Estos dos tipos de unidades informacionales tienen necesidades específicas que, aunque asociadas, son diferentes. Por una parte, como ya se ha visto, las memorias de traducción y las bases terminológicas requieren, por una parte, de profesionales que las creen y exploten y, por otra, de

procedimientos concretos para poder unificar la forma de trabajar de los usuarios implicados.

La fase de inventariar los recursos lingüísticos existentes en el nuevo sistema ya ha concluido y es, sin lugar a dudas, la fase más laboriosa y larga, en contraste, con el seguimiento y mantenimiento de los nuevos recursos lingüísticos que no requiere mucha dedicación, pero sí la intervención de diversos agentes.

Durante el proceso de inventariado también se han documentado los procedimientos y las decisiones acerca del modelado de la información, así como, con ayuda de los usuarios que han probado el sistema, se ha identificado el tipo de contenido relevante para los procedimientos de mantenimiento que deben tener en cuenta los agentes implicados en la gestión del sistema implementado. Mediante unos procedimientos claros de trabajo se evitan los posibles errores de registro en el catálogo. Asimismo, aunque la tarea de ponerse al día con el registro de todos los recursos lingüísticos haya concluido, así como la implementación de los procedimientos para el mantenimiento de la herramienta, sigue siendo necesario contar con una persona encargada de supervisar y coordinar el mantenimiento del catálogo, es decir, el terminólogo.

Es fundamental gestionar con rigor los datos introducidos para que la explotación del gestor de proyectos y del corpus paralelo sea satisfactoria. Los instrumentos de trabajo propuestos por el terminólogo e implementados para afinar COGNImática, como se ha visto en el apartado 4.2.4.2. Herramientas de control, contribuye a mejorar las condiciones de trabajo del responsable del mantenimiento de los recursos lingüísticos que se registran en COGNImática.

#### **4.5.2.1. Perspectivas de mantenimiento y seguimiento a medio y largo plazo**

Una de las perspectivas de mantenimiento y seguimiento a medio plazo que supondría una mejora en la explotación de los recursos lingüísticos es la anotación sistemática de las memorias de traducción que componen los corpus paralelos con información adicional de la tipología textual. Si las prestaciones de los corpus en general se multiplican si están debidamente anotados, se prevé que la identificación de los módulos textuales en el entorno de un proyecto de traducción proporcione mejores filtros para la creación de subcorpus. Partimos además de la idea de que esta propuesta es un procedimiento de trabajo asumible en cuanto a la carga de trabajo de los agentes implicados en la gestión de los proyectos de traducción. Asimismo, se considera pertinente dar formación a los agentes para que puedan llevar a cabo la identificación de los tipos de texto a partir de unos modelos como, por ejemplo, los sintetizados en la *Fig. 16: Tipología de los módulos textuales posibles en la documentación de un producto*, que se encuentra en el apartado 2.3.5. Los tipos / módulos textuales de la documentación técnica. Una vez instaurado el procedimiento de anotación textual, se debería también adaptar la interfaz de búsqueda de los corpus para poder filtrar las memorias pertinentes.

Se prevé también la optimización de las búsquedas mediante su expansión automatizada. La expansión de las consultas tiene como fin asistir de forma automatizada a la generación de términos de consulta asociados a la ecuación de búsqueda del usuario. Tal como exponíamos en un trabajo anterior (Cámara, 2002b: 90-91), la expansión de las consultas se asemejaría a la generación de asociaciones que, previsiblemente, realiza el usuario en el cerebro —redes semánticas—, si dispusiera del conocimiento activo en su memoria. Por tanto, estas expansiones deben entenderse como propuestas de simulación de la actividad del cerebro para asistirnos en el proceso de análisis y captura.



La expansión de las búsquedas se podrían derivar mediante:

- la utilización de los equivalentes de todos los idiomas registrados en el tesoro *Spines*; como habíamos mencionado en el apartado 4.2.3. Desarrollo de la primera propuesta informática, los usuarios pueden introducir términos en el idioma de la traducción, siempre y cuando se introduzca también la equivalencia en español. No obstante, hasta ahora no se ha previsto la introducción manual de los descriptores en otra lengua que no sea la española. Ahora bien, el hecho de utilizar los términos de indización extraídos del tesoro *Spines* en línea que, como se ha explicado, está disponible en una versión trilingüe (español, inglés y francés), abre perspectivas de afinamiento de COGNImática. En este sentido, previsiblemente en el futuro, también se podrán extraer del tesoro las equivalencias en los otros idiomas de los descriptores asignados. Mediante un script adicional se marcarían automáticamente los documentos, sin tener que introducir en los tres idiomas diferentes la marcación semántica de los documentos. De esta manera, el sistema nos permitiría recuperar diccionarios y memorias de traducción utilizando indistintamente tres idiomas, situación que ahora se limita principalmente al español;
- la extensión derivativa a partir de flexiones morfológicas a las que puede ir asociado un lema (lematización), lo que requiere de una anotación previa a nivel morfosintáctico de las memorias que constituyen el corpus;
- la extensión conceptual a partir de las asociaciones semánticas que se pueden inferir a partir de un término que no anota semánticamente ningún diccionario —y por tanto, ninguna memoria de traducción—, pero contenido en el tesoro, haciendo una expansión de las búsquedas no capturables mediante un razonamiento morfológico, pero sí conceptual. De este último razonamiento se pueden derivar las asociaciones conceptuales, para luego generar redes semánticas;

- la generación de nexos entre las palabras clave y los descriptores.

El problema para llevar a cabo las expansiones estriba en identificar los patrones de razonamiento requeridos y traducirlos computacionalmente. Asimismo, para ello previsiblemente se utilizarían algunos de los patrones usados en inteligencia artificial, como el razonamiento probabilístico, la lógica difusa, el razonamiento inductivo y deductivo.

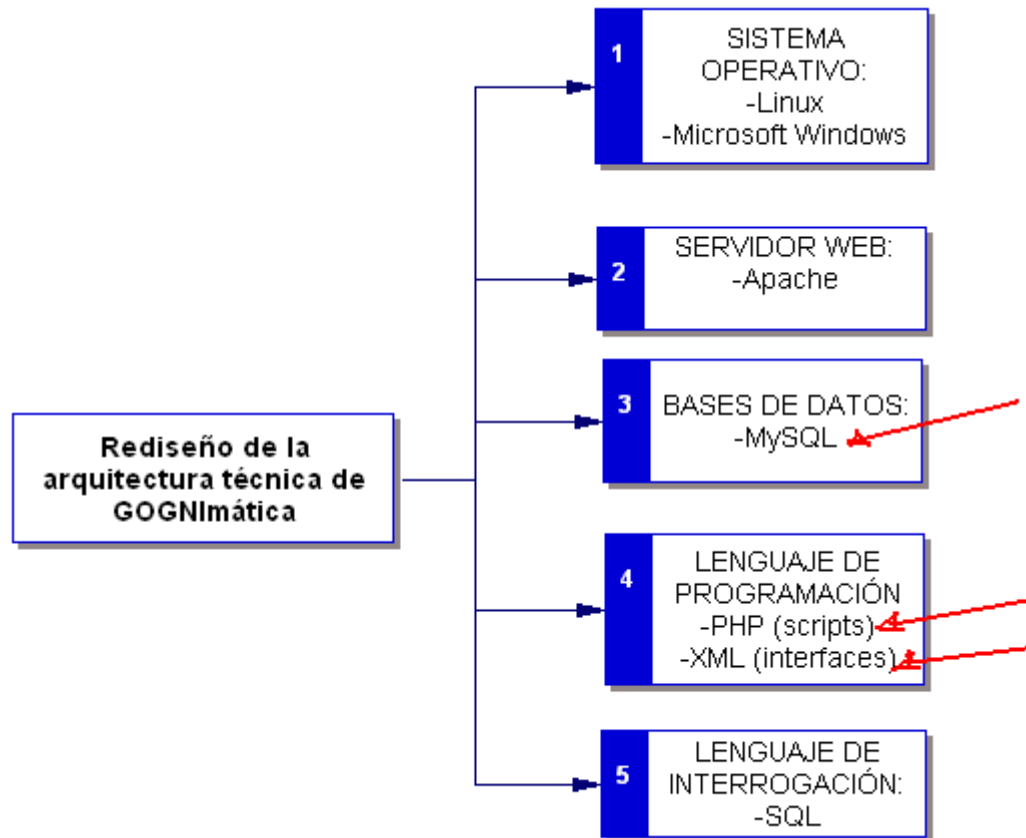
Como todos estos recursos de información se gestionan con COGNImática, en el planteamiento de seguimiento y mantenimiento también se deben atender a las posibles mejoras del sistema. Muchas de las mejoras son dependientes de un afinamiento de la herramienta, ya que existe una sinergia entre las unidades de información y el sistema, como veremos en el siguiente apartado.

### **4.5.3. Seguimiento de COGNImática**

Con referencia a las previsiones a corto, medio y largo plazo para mejorar la herramienta se establecen dos tipos de propuestas; por una parte, propuestas para refinar el sistema COGNImática en sí y, por otra, propuestas vinculadas a la mejora de la gestión empresarial, más allá del sistema diseñado e implementado.

#### ***4.5.3.1. Perspectivas técnicas de afinamiento a corto plazo***

En la *Fig. 78: Tecnologías aplicadas en COGNImática* del apartado 4.1.1. Diseño tecnológico: arquitectura técnica mostrábamos la arquitectura técnica en la que se basa nuestro sistema. En el siguiente gráfico, muy similar al mencionado, mostramos marcados con una flecha la tecnología web que ha sido propuesta en sustitución a la anterior con el fin de su rediseño e implementación técnica inmediata:



**Fig. 111: Futuras tecnologías aplicadas en COGNimática**

Como se ve en el diagrama, la siguiente relación es la propuesta para el cambio de tecnología:

- la base de datos *PostgreSQL* por *MySQL*
- el lenguaje de programación *Ruby on Rails* por *PHP*
- el lenguaje de programación para interfaces *HTML* por *XML*

Las razones para estas sustituciones se han derivado de las propuestas esbozadas por los responsables del departamento informático una vez evaluado el funcionamiento de COGNimática. En este sentido, su evaluación se resume en que la pareja de desarrollo *PHP* y *MySQL* tiene un uso más extendido, además de ser más estable en un entorno web que la pareja utilizada en la primera implementación —*Ruby on Rails* y *PostgreSQL*— que no ha sido lo suficientemente probada en este entorno. El uso más extendido de *PHP* y *MySQL* para implementar aplicaciones web repercute en la

búsqueda más eficiente de recursos humanos con estos conocimientos y experiencia en programación. El rendimiento y velocidad que ofrece *MySQL* al recuperar información e interacción con la arquitectura cliente/servidor es superior que mediante *PostgreSQL*. Estas pruebas se han llevado a cabo con otra aplicación desarrollada en el entorno de STAR SL que cuenta con la interacción de alrededor de 8000 registros y de 60 campos.

#### **4.5.3.2. Perspectivas técnicas de afinamiento a medio plazo: explotación de archivos TMX**

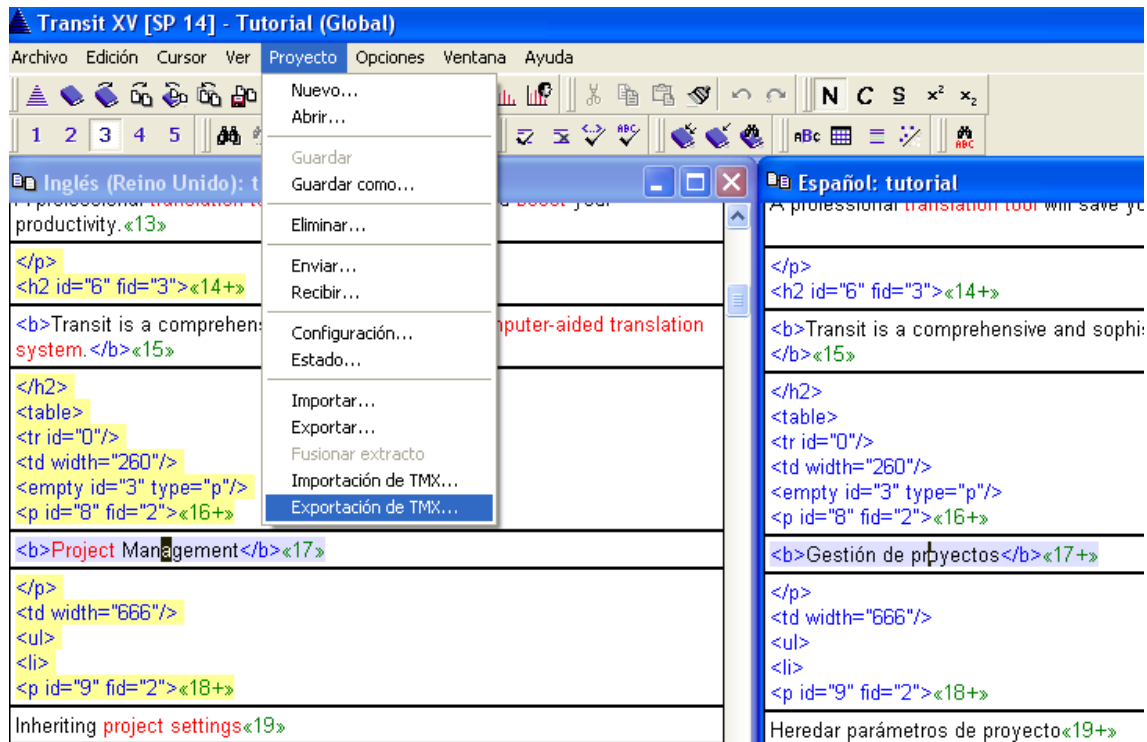
A excepción del programa *Transit*, que es una aplicación que utiliza como soporte el sistema operativo *Windows*, el resto de las funcionalidades de COGNlímica se basan en tecnología web. Los datos se recuperan mediante un navegador web y se extraen en cadenas de concordancias gracias al programa *Transit*, como gestor de los corpus paralelos. Esto es, el acceso a los datos de información documental del corpus paralelo se realiza a través de un navegador, sin embargo abrimos los subcorpus y accedemos a su contenido total mediante la activación de *Transit* instalado en una red local. Esto implica que su explotación solo puede hacerse en el entorno de la empresa donde se aplica. Esta solución local de explotación de corpus tiene algunas desventajas, ya que los traductores o gestores de proyectos empleados que trabajan de forma puntual fuera de las oficinas, aquéllos que trabajan en régimen de teletrabajo<sup>1</sup> y, por último, los traductores autónomos no pueden explotar los corpus de forma remota a través de un navegador de Internet.

---

<sup>1</sup> El teletrabajo es una forma flexible para la organización del trabajo que consiste en el desempeño de la actividad profesional sin la presencia física del trabajador en la empresa durante una parte importante de su horario laboral. La actividad profesional en este régimen laboral implica el uso frecuente de métodos de procesamiento electrónico de información y el uso permanente de algún medio de telecomunicación para el contacto entre el teletrabajador y la empresa.

La solución técnica que proponemos es la de convertir los archivos *Transit* en archivos TMX, estándar ampliamente descrito en el apartado [2.5.3. TMX](#). El almacenamiento del corpus utilizando este estándar de intercambio de las memorias de traducción basado en XML hace posible la compatibilidad de formato a través de Internet.

Para agilizar la conversión y publicación proponemos automatizar el proceso mediante un script que convierta los proyectos *Transit* al formato de intercambio TMX, ejecutando la función específica del programa *Transit* para la exportación de archivos *Transit* a TMX, que mostramos en la siguiente captura de pantalla:



**Fig. 112: Función de exportación a archivos TMX en *Transit***

Tras la conversión, el script puede llevar el almacenamiento de los archivos TMX a una base de datos XML preparada para estos efectos. Una vez convertido y almacenado, el usuario podrá extraer y explotar las cadenas de caracteres requeridas en concordancia y con sus equivalentes correspondientes en otros idiomas. Para las búsquedas simples o

compuestas se podría utilizar el programa *Search & Replace*, programa de búsquedas de gran rendimiento y fácil uso al que ya nos hemos remitido y documentado mediante la *Fig. 52: Resultado de las capturas de concordancias mediante Search & Replace* en el apartado 3.3.3.2.3. Gestión de los diccionarios (recursos terminológicos).

Recapitulando, como resultado de la aplicación del método auditor descrito en el capítulo anterior, en este capítulo se ha implementado COGNImática, un sistema modular orientado a los diferentes tipos de usuario.

Los módulos del sistema son un gestor documental de diccionarios de gestión semiautomática, un gestor de proyectos automático y un gestor de corpus paralelos que explota las posibilidades que ofrece el almacén de memorias de traducción. Mediante la interacción de las tres herramientas de implementación secuencial y la flexibilidad de *Transit* en la integración en entornos ajenos, se multiplica y amplía la capacidad de gestión automatizada de los recursos lingüísticos.

Asimismo, tras una fase de pruebas, COGNImática se ha visto sustancialmente mejorado mediante la implementación de muchas de las propuestas de los betatesters, quienes abordan el manejo del sistema desde la usabilidad de su entorno de trabajo concreto.

Las propuestas de seguimiento, mantenimiento y mejora en cuanto a la gestión general de los recursos lingüísticos y en cuanto a la mejora del rendimiento del sistema implementado, se resumen en los siguientes puntos:

- ❑ mejora de la formación de los agentes implicados;
- ❑ transparencia y colaboración en los procesos;
- ❑ mejora de la accesibilidad de los recursos lingüísticos almacenados mediante tecnología web y formatos estándares.

La gestión de los recursos terminológicos y memorias de traducción en la empresa analizada comprende una serie de actividades organizadas

(procesos) para satisfacer las necesidades terminológicas y de pretraducción de los proyectos de traducción. Todos los traductores tienen el perfil de terminólogos, en el sentido de que crean, evalúan, modifican y editan diccionarios terminológicos para su explotación en el entorno de producción, Los traductores además tienen el perfil de gestores documentales en el sentido de tener que almacenar, identificar y encontrar toda la información almacenada. Para hacerlo de una forma homogénea se precisan de los perfiles descritos para el establecimiento de la política de información y el fomento de la gestión del conocimiento, además de procedimientos transparentes y de colaboración avalados mediante tecnología web y formatos estándares.

## 5. Conclusiones y reflexiones finales

En el presente capítulo final sintetizamos las conclusiones que se derivan del trabajo que hemos llevado a cabo; asimismo reflexionamos acerca de las posibles líneas de investigación y aplicación que ofrecen los resultados obtenidos.

Tal y como se ha expuesto en el apartado correspondiente —1.2.Objetivos— mediante este trabajo de investigación nos proponíamos:

- ❑ desarrollar procedimientos para mejorar y automatizar la gestión de los recursos lingüísticos para los proyectos de traducción y localización en una empresa de traducción, y, —como consecuencia del desarrollo de los procedimientos—
- ❑ diseñar e implementar un sistema para optimizar la gestión de los recursos lingüísticos y, al mismo tiempo, dar soporte a la gestión del conocimiento con una orientación a los usuarios.



Creemos que en el trabajo, especialmente en el cuarto capítulo, se ha demostrado haber logrado con éxito nuestro propósito.

Para ello, en primer lugar, mediante una auditoría de la información en una empresa, se han identificado los cuellos de botella de los flujos de gestión de los recursos lingüísticos; se ha detectado la implicación de las herramientas de traducción asistida como las principales herramientas que los gestionan, así como los límites de estos sistemas para dar respuesta a las necesidades expresadas por los usuarios e identificadas por el analista.

En segundo lugar, y a partir de los resultados del análisis de los flujos de trabajo, se han formulado propuestas de mejora de la gestión que hemos traducido en forma de:

- una serie de protocolos de trabajo que afectan principalmente a la forma de almacenar, nombrar, ponderar y catalogar los recursos lingüísticos;
- diseño conceptual de un lenguaje de indización abocado a la recuperación de la información;
- diseño de COGNImática, sistema dotado de conocimiento experto que da respuesta a algunos de los límites de las herramientas de traducción asistida.

El conocimiento experto del que hemos dotado a COGNImática se basa en las necesidades de los diferentes tipos de usuarios y en un sistema conceptual para anotar semánticamente los recursos gestionados y facilitar así el acceso más intuitivo a la información.

El sistema conceptual para la anotación semántica se basa en un tesoro con terminología estandarizada y términos de selección libre. Es decir, se ha sistematizado la gestión de la terminología como recurso para organizar puntos de acceso de tipo semántico y mejorar el acceso a los recursos lingüísticos almacenados.

El trabajo realizado en cada una de las partes anteriormente especificadas nos ha permitido llegar a las siguientes conclusiones:

En primer lugar, la aplicación del método auditor ha puesto de manifiesto la relevancia de las memorias de traducción, diccionarios terminológicos e informaciones asociadas desde dos puntos de vista:

- como activos lingüísticos desde la perspectiva de una empresa como la analizada,
- como bases de conocimiento desde la perspectiva de los traductores y gestores para el desempeño de su trabajo.

En segundo lugar, se ha constatado que la sistematización de los flujos de trabajo, la anotación semántica, como soporte a la gestión humana, y la oferta de acceso intuitivo a los recursos almacenados deben constituir el punto central de la propuestas de gestión. En este sentido, concluimos que cuanto mayor sea la sistematización, es decir, la descripción semántica y estructuración lógica de los recursos lingüísticos almacenados, mayor es la precisión obtenida en la recuperación de información.

En tercer lugar, mediante COGNImática ofrecemos una solución a la infrutilización de los recursos lingüísticos basada en las técnicas básicas de gestión documental. Esta solución la integramos en el sistema de traducción asistida empleado en nuestro caso de estudio, resolviendo una de las funciones de las que carecen los sistemas de traducción asistida. La aplicación de técnicas de gestión documental tiene repercusiones en los profesionales implicados.

En este sentido y en cuarto lugar, llegamos también a la conclusión de que los nuevos perfiles profesionales dentro del entorno de la traducción especializada requieren de la incorporación de recursos humanos formados en técnicas de documentación, pero no sólo como consumidores de información, sino como organizadores de ésta. Esta doble competencia

contribuye a la adquisición, transferencia y comunicación de los contenidos de los documentos mejorando la calidad.

En quinto lugar, se ha mostrado que la política de la (in)formación aplicada en la empresa estudiada, más allá de los manuales de procedimientos y el despliegue tecnológico, ha fomentado un cambio abocado a compartir, consultar, sumar, unificar, proponer y construir; esto es, un cambio de proceder dentro de la cultura informativa en la empresa.

Teniendo en cuenta la orientación al usuario, existe una mejora en la implicación de los agentes en cuanto a la concienciación del problema desde la perspectiva comunicativa y formativa en la solución. Esto significa que en última instancia todo el equipo ha sido y sigue siendo portador y beneficiario de propuestas y soluciones. El desarrollo de COGNImática es, por tanto, una implementación de una buena gestión del conocimiento.

En este sentido y en sexto lugar, COGNImática no sólo es un producto de una buena gestión del conocimiento, sino que además es una herramienta que da soporte a la gestión del conocimiento mediante la incorporación de herramientas colaborativas en red. Estas herramientas fomentan el trabajo en grupo a través de una red, contribuyendo a la mejora de los procesos formativos en tecnología, además de fomentar el intercambio y la colaboración entre los miembros del equipo.

Y en séptimo lugar, la contextualización y descripción del sector industrial en el que se encuadra nuestro estudio de caso ha sido básico para poder entender el problema planteado, a la vez que ha servido para comprenderlo en toda su dimensión. Aunque no era uno de los objetivos que nos habíamos marcado, podemos concluir que la introducción contextual ha logrado sacar del anonimato el sector industrial dedicado a la gestión de contenidos multilingües, además de haber aportado el desarrollo de un modelo de tipología textual de base modular en respuesta a la forma de gestión de

contenidos que se lleva a cabo en el entorno de producción de documentación técnica.

Como resultado del proceso vivido a través de todo el estudio, de las experiencias y conocimientos adquiridos en el mismo y de las interacciones con los usuarios después de usar el sistema, finalizamos con algunas reflexiones.

En primer lugar, para ampliar el estudio y avanzar en la línea de solución propuesta, consideramos que la relevancia que hemos otorgado a la optimización del acceso a la información mediante las técnicas de recuperación y extracción debe seguir desarrollándose. En este sentido, consideramos que la exhaustividad de las respuestas que ofrece COGNImática se vería mejorada mediante la implementación de bases deductivas que posibilitarían a los usuarios, así como al motor de búsqueda, inferir nuevo conocimiento no explicitado previamente, recuperando un mayor número de recursos relevantes para una consulta dada.

En segundo lugar, desde el punto de vista técnico, para la implementación de COGNImática se ha empleado software de código abierto con formatos abiertos y públicamente accesibles como plataforma de desarrollo para nuestra aplicación de soporte a la comunicación y procesamiento de información. Si se tiene en cuenta que la tendencia se inclina al desarrollo de este tipo de software a causa de su escalabilidad, fiabilidad y elevado rendimiento, creemos que las implementaciones posteriores que se hagan a COGNImática deben seguir la misma línea de desarrollo empleadas hasta ahora.

En tercer lugar, cierto es que COGNImática garantiza una optimización en el flujo de trabajo de los proyectos de traducción porque agiliza una de sus tareas implicadas —la gestión de recursos lingüísticos mediante la implementación de funciones a modo de biblioteca virtual creada a la medida de las necesidades de los usuarios—. Sin embargo, la automatización del

resto de las labores implicadas en los flujos de trabajo en toda empresa de traducción es apremiante. Esta urgencia se debe al precio que está dispuesto a pagar el cliente por una traducción. Las empresas de traducción ven mermada su rentabilidad debido a que los costes humanos de producción aumentan cada año, mientras que los precios de los trabajos facturados no se ven incrementados. Ante estas condiciones y para poder seguir garantizando la calidad de la producción se hace acuciante el intercambio de propuestas como la presentada entre el sector profesional y académico implicado.

Finalmente, esta tesis en cuanto aplicada evidencia las diferencias existentes entre la ciencia y la tecnología que en este trabajo, más que dicotómicas, se presentan como complementarias. Así, si la ciencia parte y se orienta a la generación de conocimiento con tendencia generalista, la tecnología se orienta a necesidades concretas aprovechándose de las enseñanzas de lo que produce la ciencia. La tecnología puede llegar a ser general, aunque, como parte de la utilidad, se adapta *ad hoc* a las necesidades del entorno de aplicación. La ciencia da respuestas a abstracciones de problemas. La tecnología, basándose en esas abstracciones, pero sin perder de vista el entorno real de aplicación soluciona problemas prácticos. Así, del nuevo conocimiento generado a partir del análisis científico, se deriva un producto tecnológico resultante de la síntesis científica y el pragmatismo del entorno de aplicación, como se ha pretendido llevar a cabo con este trabajo de tesis de carácter aplicado e interdisciplinar.

Asimismo, el carácter interdisciplinar de este proyecto de investigación es inmanente al problema planteado cuya realidad es compleja y cuya comprensión y resolución ha requerido de la cooperación de diferentes disciplinas. El desafío que ha planteado el estudio interdisciplinar ha consistido en tomar como punto de partida las disciplinas abordadas, teniendo en cuenta la especificidad de conceptos, métodos y lógicas de cada uno de ellas, creando entre ellos un puente común de comunicación.

El carácter aplicado e interdisciplinar de la tesis no excluye que, a partir de este trabajo, se profundice en las diferentes líneas de conocimiento tanto de forma teórica, descriptiva como aplicada.

*Y aquí paz y después gloria...*



## 6. Bibliografía

Abaitua, J. 1995. "Prólogo a la edición española", en *Introducción a la Traducción automática*, Hutchins, W. J. y Sommers, H. L. Madrid: Visor.

Abaitua, J. 2002. "Tratamiento de corpora bilingües", en *Tratamiento del lenguaje natural*, Martí M. A. y Llisterri, J. (eds). Barcelona: Edicions Universitat de Barcelona: 61-90.

Ahmad, K., Davies, A. E., Fulford, H., Holmes-Higgin, P. R., Rogers, M. y Höge, M. 1992. "User Participation in Software Development", en *Multilingual Documentation and Communication*. Final Report of ESPRIT Project No. 2315 TWB, Kugler, M., Ahmad, K., Heyer, G., Rogers, M. y Thurmair, G. (eds). Bruselas-Luxemburgo: 7-15.



- Ahmad, K., Ogonowski, A., Dauphin, E., Sta, J.-D. y Arppe, A. 1996. "Engineering Terminology - A Case for a Linguistically Informed Database", en *Proceedings of the 4th Congress on Terminology and Knowledge Engineering*. Galisky, Ch. y Schmitz, K. D. (eds). Viena: TKE: 166-178.
- Allied Business Intelligence. 1998. "World Market Overview, Current Developments and Competitive Assessment", *Language Translation*.
- Arevalillo, J. J. 2003. "La gestión de proyectos en el ámbito de la traducción empresarial", en *Entornos informatizados de la traducción profesional*, Corpas, G. y Varela, M. J. (eds). Granada: Atrio: 5-46.
- Arevalillo, J. J. 2004. "Gestión de la documentación en proyectos de programas informáticos", en *Manual de documentación y terminología para la traducción especializada*, Gonzalo García, C. y García Yebra, V. (eds). Madrid: Arco libros: 445-469.
- Arrarte, G. 1995 "Tendencias actuales de la ingeniería lingüística en Europa", *II Jornadas de Informática e Investigación Operativa*, Montevideo: Universidad de la República Montevideo. [En línea] <[http://iling.torreingenieria.unam.mx/curso2002\\_2/lecturas/arrarte.pdf](http://iling.torreingenieria.unam.mx/curso2002_2/lecturas/arrarte.pdf)> [Consulta: 01.10.05].
- Arrouart, C. y Bédard, C. 2001. "Éloge du bitexte", *Circuit*, 73: 30-37.
- Austermühl, F. 2001. *Electronic Tools Form Translation*. Manchester: Jerome Publishing.
- Baeza-Yates, R. y Ribeiro-Neto, B. 1999. *Modern Information Retrieval*. Maryland: Addison-Wesley-Longman.
- Baeza-Yates, R., Loaiza, C. y Velasco, J. 2004. "Arquitectura de la información y usabilidad de la web", *El Profesional de la Información*, Vol. 13, N° 3: 168-178.

Baker, M. 1995. "Corpora in Translation Studies: An Overview and Some Suggestions for Future Research", *Target*, 7(2): 223-243.

Baldwin, C. y Clark, K. B. 2000. *Design Rules: The Power of Modularity*, Vol. 1. Massachussets: MIT Press.

Barite, M. 1982. *Glosario sobre organización y representación del conocimiento*. Montevideo: CDU.

Belkin, N., Oddy, R. y Brooks, H. 1982. "ASK for Information Retrieval: Background and Theory", *Journal of Documentation*, N° 38.

Berrocal, J. L., Figuerola, C. y Zazo, A. 2000. "SGML/XML: Desarrollo en entornos documentales", *Coloquio de la Asociación Internacional de Bibliología*. [En línea] <<http://www.fundacionyuste.org/acciones/aeb/XV/AIB2000/BERROCAL.pdf>> [Consulta: 01.10.05].

Blair, D. C. 1990. *Language and Representation in Information Retrieval*. Amsterdam: Elsevier.

Bolognini, M. y Pedrini, E. 1993. *Manuale del catalogatore: una guide per le biblioteche pubbliche*. Milán: Bibliografica.

Boon J. H. 1990. "Information management; an education perspective", *South African Journal of Library and Information Science*, N° 58(4): 319–327.

Bowker, L. 1997. "You say 'flatbed colour scanner', I say 'colour flatbed scanner': A Descriptive Study of the Influence of Multidimensionality on Term Formation and Use with Special Reference to the Subject Field of Optical Scanning Technology", *Terminology*, Vol. 4(2): 275-302.

Brace, C. 1992. "The Esprit Translator's Workbench Project", *Language Industry Monitor*. [En línea] <<http://www.lim.nl/monitor/esprit-twb.html>> [Consulta: 01.10.05].

Buchanan, S. y Gibb, F. 1998. "The information audit: an integrated strategic approach", *International Journal of Information Management*, Nº 18(1): 29-47.

Budin, G. y Wright, S. E. 1997. *Handbook of Terminology Management*, Vol. 1: *Basic Aspects of Terminology Management*. Amsterdam: John Benjamins B.V.

Budin, G. y Wright, S. E. 2001. *Handbook of Terminology Management*, Vol. 2: *Applications-oriented Terminology Management*. Amsterdam: John Benjamins B.V.

Budin, G. 2004. Global Content Management. Universidad de Viena. [En línea] <<http://www.ifi.unizh.ch/cl/yuste/postworkshop/repository/gbudin.pdf>> [Consulta: 01.10.05].

Bukowitz, W. R. y Williams, R. L. 1999. *The Knowledge Management Fieldbook*. Londres: Financial Times/Prentice Hall.

Burk, C.F. y Horton, B. 1988. *InfoMap: A Complete Guide to Discovering Corporate Information Resources*, New Jersey: Prentice Hall.

Burnard, L. 2003. "Text Encoding for Interchange: A new Consortium", *Ariadne*, Nº 23. [En línea] <<http://www.ariadne.ac.uk/issue24/tei>> [Consulta: 01.10.05]

Bustelo, B. y García-Morales. 2001. "Tendencias en la gestión de la información, la documentación y el conocimiento en las organizaciones", *El Profesional de la Información*, Vol. 10, Nº 12: 4-7.

Cabré, M. T. 1992. *Terminologia: la teoria, els mètodes i aplicacions*, Barcelona: Editorial Empúries.

Cabré, M. T. 1997. "Metodología de la Terminología" en *l Escuela de Verano de Estudios de terminología*, Barcelona: Universidad Pompeu Fabra.

- Cabré, M. T. 1999. *La terminología. Representación y comunicación. Una teoría de base comunicativa y otros artículos*. Barcelona: IULA-Universidad Pompeu Fabra.
- Cabré, M. T. 2002. “Análisis textual y terminología, factores de activación de la competencia cognitiva en la traducción”, en *La traducción científico-técnica y la terminología en la sociedad de la información*. Alcina, A. y Gamero, S. (eds). Castellón de la Plana: Publicacions Universitat Jaume I: 87-106.
- Cabré, M. T. 2004. “La terminología en la traducción especializada”, en *Manual de documentación y terminología para la traducción especializada*, Gonzalo García, C. y García Yebra, V. (eds). Madrid: Arco libros: 89-125.
- Cámara, L. 2001. “El papel de las herramientas TAO en la documentación técnica multilingüe”, *Tradumática*, Nº 0. [En línea]  
<<http://www.fti.uab.es/tradumatica/revista/num0/articles/lcamara/art.htm>>  
[Consulta: 01.10.05].
- Cámara, L. 2002a. “Tecnología para la gestión terminológica multilingüe: TermStar y Webterm”, en *La traducción científico-técnica y la terminología en la sociedad de la información*, Alcina, A. y Gamero, S. (eds). Castellón de la Plana: Publicacions Universitat Jaume I: 271-276.
- Cámara, L. 2002b. *Herramientas para la gestión del conocimiento*. [Proyecto para la suficiencia investigadora] Barcelona: IULA-Universidad Pompeu Fabra.
- Cámara, L. 2003a. “Solucions globals per la comunicació técnica multilingüe”, en *I Jornada: Estratègies de multilingüisme a l'empresa: “Mètodes, tecnologia i recursos en la internacionalització d'aplicacions informàtiques”*, Barcelona: ESADE.
- Cámara, L. y Chrupala, G. 2003b. “STAR Transit XV”, en *Entornos informatizados de la traducción profesional*. Corpas, G. y Varela, M. J. (eds). Granada: Atrio: 119-138.

- Cámara, L. 2004. "La representación lingüística del conocimiento y su relevancia en la ingeniería lingüística", *Hipertext.net*, Nº 2, [En línea] <<http://www.hipertext.net>> [Consulta: 01.10.05]
- Cámara, L. 2005a. "Necesidad de políticas de información y de sus profesionales para la automatización de la producción de documentación técnica en el entorno de la industria GILT", *El Profesional de la Información*, Vol. 14, Nº 2: 128-138.
- Cámara, L. 2005b. "Estrategias semánticas para la organización de las memorias de traducción" *Symposium On Technology And Translation*, 28-29 November 2003, Pym, E. (ed). Tarragona: Universitat Rovira i Virgili. [En proceso de publicación].
- Cámara, L. 2005c. "Sistema de gestión de recursos lingüísticos" en *Manual de traducción y de documentación*. Salas, D. (ed). Madrid: Ariel. [En proceso de publicación].
- Carbajo, S. 2004. "Instrumentos informáticos para la traducción especializada, I: Trados", en *Manual de documentación y terminología para la traducción especializada*, Gonzalo García, C. y García Yebra, V. (eds). Madrid: Arco libros: 473-489.
- Castells, M. Cerf, V., D'Elia, C., Branco, M. *et al.* 2004. *Declaración de Barcelona para el avance del software libre*. [En línea] <<http://www.dsd0.com/modulos/areasnegocio/index.php?area= analisis&categ =declaracion%20de%20barcelona>> [Consulta: 01.10.05].
- CEN/BTTF 138. 2002 – 2005 [En proceso de elaboración]. *Translation Services*. Comité Européen de Normalisation.
- Ciolek, T. M. 1997. "Today's WWW, tomorrow's MMM: the specter of multi-media mediocrity", *Educom review*, Vol. 32 (3). [En línea] <<http://www.educom.edu/web/pubs/review/reviewArticles/32323.html>> [Consulta: 01.10.05]

Codd, E. 1970. "A Relational Model of Data for Large Shared Data Banks", *Communications of the Association for Computing Machinery*, Vol. 13, Nº 6: 377-387.

Codina, L. 1997. *Proyecto docente: documentación en los medios de comunicación*. Barcelona: Universitat Pompeu Fabra. [Xerocopia].

Codina 2000a. "Evaluación de recursos digitales en línea: conceptos, indicadores y métodos", *Revista Española de Documentación Científica*, 1. Vol. 23: 9-44.

Codina, L. 2000b. "Parámetros e indicadores de calidad para la evaluación de recursos digitales", en *La gestión del conocimiento: retos y soluciones de los profesionales de la información, VII Jornadas Españolas de Documentación*. Bilbao: Servicio Editorial de la Universidad del País Vasco: 135-144.

Computerworld, 05/2003. "Hay que olvidarse de los cálculos ROI. Su estimación no conduce a ninguna parte".

Cornellá, A. 1994. *Los Recursos de Información: Ventaja Competitiva de las Empresas*. Madrid: McGraw-Hill.

Cornellá, A. 2000a. "Ecosistemas informacionales", *Extranet. El periódico de Internet Caracas-Venezuela*. [En línea]  
<<http://www.lared.com.ve/archivo/extra47.html>> [Consulta: 01.10.05].

Cornellá, A. 2000b. "El rol del infonomista", *El Profesional de la Información*, Vol. 9, Nº 1-2: 32-40.

Cornellá, A. 2000c. "La información no es necesariamente conocimiento", *Infonomía.com: La empresa es información*.

Corpas, G. 2001. "Compilación de un corpus ad hoc para la enseñanza de la traducción inversa especializada", *Trans.- Revista de traductología*, 5. Málaga: Universidad de Málaga: 155-184.

*CRATER: Corpus paralelo sobre telecomunicaciones en inglés, español y francés.* [En línea]

<<http://www.comp.lancs.ac.uk/linguistics/crater/spanish000-toc.html>>

[Consulta: 01.10.05].

Daille, B., Habert, B., Jacquemin, C. y J. Royauté, J. 1996. "Empirical Observation of Term Variations and Principles for their Description", en *Terminology*, 3(2): 197-258.

Davenport, T. 1997. *Information Ecology: Mastering the Information and Knowledge Environment*. New York : Oxford University Press.

Delgado, E. 2002. *La investigación en biblioteconomía y documentación*. Gijón: Trea.

Der Spiegel on line. 2004. "Microsofts Sprachprobleme: Windows für Huren", Hamburg. [En línea] <<http://www.spiegel.de>> [Consulta: 24.08.04].

Deutscher Fachverband für Technische Kommunikation, Dokumentation und Informationsentwicklung (Tekom). 1978. " [En línea] <<http://www.technikredaktor.ch/index.php?id=87>> [Consulta: 01.10.05]

Díez, P. L y Scholand, M. 2003. "Componentes de los sistemas integrados de herramientas de asistencia a la traducción (SIHAT): Transit", en *Traducción y nuevas tecnologías, herramientas auxiliares del traductor encuentros en torno a la traducción*, Valero Garcés, C. y de la Cruz Cabanillas, I. (eds). Madrid: Universidad de Alcalá: 215-232.

Díez, P. L. 2004. "Instrumentos informáticos para la traducción especializada, II: Transit", en *Manual de documentación y terminología para la traducción especializada*, Gonzalo García, C. y García Yebra, V. (eds). Madrid: Arco libros: 491-505.

DIN 2345. 1998. *Verwaltung für Übersetzungs- und Lokalisierungsprojekte*. Berlin: Das Deutsche Institut für Normung.

Doc6. 2004. "Habilidades y cualidades del profesional de la información en el siglo XXI". [En línea] <[http://www.doc6.es/enlaces/habilidades\\_conte2.html](http://www.doc6.es/enlaces/habilidades_conte2.html)> [Consulta: 01.10.05].

DOCUTES, 2000. *Tesaurus de la ciencias de la documentación de la Universidad de León*. [En línea] <<http://www3.unileon.es/dp/abd/tesauro/pagina/tesdocumentacion/docutes.htm>> [Consulta: 01.10.05].

Dörner, D. 1979. *Problemlösen als Informationsverarbeitung*. Stuttgart: Kohlhammer.

EAGLES (Expert Advisory Group on Language Engineering Standards), Text Representation subgroup, 2000. XML. *Corpus Encoding Standard - Document CES 1*. [En línea] <<http://www.cs.vassar.edu/CES/>> [Consulta: 01.10.05]

EAGLES (Expert Advisory Group on Language Engineering Standards), Text Representation subgroup, 2002. XML *Corpus Encoding Standard Document XCES 0.2*. [En línea] <<http://www.cs.vassar.edu/XCES/>> [Consulta: 01.10.05].

Eíto, R. 2003. "El contexto de la información. Herramientas y útiles para el proceso de auditoría", *El Profesional de la Información*, Vol. 12, Nº 4, 302-312.

ELDA (Evaluations and Language resources Distribution Agency), 2003. "Catalogue of language resources". [En línea] <<http://www.elda.fr/rubrique6.html>> [Consulta: 01.10.05].

Elosua, M. y Plágaro, J. 2000. *Diccionario de tecnologías de la información y comunicación*. Madrid: LID.

ELRA (European Language Resources Association). 2001-2003. "Language Resources Definition". [En línea] <<http://www.elra.info/>> [Consulta: 01.10.05].



Esselink, B. 2000. *A Practical Guide to Localization*. Amsterdam: John Benjamins Publishing Co.

Estopà, R., Vivaldi, J. y Cabré, M. T. 1998. *Sistemes d'extracció automàtica de (candidats a) termes: Estat de la qüestió.* ), Barcelona: IULA-Universidad Pompeu Fabra.

Estopà, R. 2001. "Elementos lingüísticos de las unidades terminológicas para su extracción automática", en *La terminología científico-técnica*, Cabré, M. T. (ed) (2001), Barcelona: IULA-Universidad Pompeu Fabra: 67-80.

Fernández Cepedal, J. M. 1976. "Categoría", en *Diccionario de filosofía contemporánea*, Salamanca: Ediciones Sígueme.

Fernández Muerza, 2000. "La necesidad de especialistas en información y documentación en Internet es acuciante", en *Divulcat, portal de ciencia y tecnología*. [En línea]  
<[http://www.100cia.com/divulgacion/\\_la\\_necesidad\\_de\\_especialistas\\_en\\_informacion\\_y\\_documentacion\\_en\\_internet\\_es\\_acuciante\\_\\_200.html](http://www.100cia.com/divulgacion/_la_necesidad_de_especialistas_en_informacion_y_documentacion_en_internet_es_acuciante__200.html)> [Consulta: 01.10.05].

Firth, J. 1957. "A Synopsis of Linguistic Theory", *Studies in Linguistic Analysis*, Volumen monográfico.

Free software foundation, 2002-2005. "Definition". [En línea]  
<http://www.gnu.org/philosophy/free-sw.html> [Consulta: 01.10.05].

Freigang, K. H. 2001. "Automation of Translation: Past, Presence, and Future", *Tradumática*, Revista N° 0. [En línea]  
<<http://www.fti.uab.es/tradumatica/revista/num0/articles/khfreigang/art.htm>> [Consulta: 01.10.05].

GALA (Globalization and Localization Association). 2002. "About us. Who we are". [En línea] <<http://www.gala-global.org/>> [Consulta: 01.10.05].

Galinsky, C. y Budin, G. 1998. "Introducción", en *Introducción a la teoría general de la terminología y a la lexicografía terminológica*, Wüster, E. / Cabré, M. T. (ed). Barcelona: IULA-Universidad Pompeu Fabra: 15-17.

Gamero, S. 2001. *Traducción de textos técnicos*. Barcelona: Ariel.

Ganzmann, J. 1990. "Criteria For The Evaluation Of Thesaurus Software" *Classification*, N° 3/4, International Society for Knowledge Organization (ISKO):148-157.

García, A. 1984. *Lingüística documental*. Barcelona: Mitre.

Garrett, J. J. 2000. "Los elementos de la experiencia del usuario". [En línea] <[http://www.jjg.net/ia/visvocab/spanish.html/elements\\_es.pdf](http://www.jjg.net/ia/visvocab/spanish.html/elements_es.pdf)> [Consulta: 01.10.05].

Garrido, M. R. 1996. "Fundamentos del análisis documental", en *Manual de información y documentación*. López, J. (ed). Madrid: Pirámide: 229-250.

Gesellschaft für Terminologie und Wissenstransfer. 1996. *Guidelines for the Design and Implementation of Terminology Data Banks*. Bozen: Europäische Akademie.

Ghenghea, V. A. 1998. "The impact of effective diagram design on LPS-Text understanding", en *Proceedings Of The 11<sup>th</sup> European Symposium on Lsp: Identity and Interface Research, Knowledge and Society*, Copenhagen: Copenhagen Business School: 783-793.

Gil, B. 1996. "Lenguajes documentales", en *Manual de información y documentación*, López Yepes (ed). Madrid: Pirámide: 354-374.

Gómez, J. 2001. "Una guía al TMX", *Tradumática*, N° 0. [En línea] <<http://www.fti.uab.es/tradumatica/revista/num0/articles/jgomez/art.htm>> [Consulta: 01.10.05].

- González, J., Seoane, J. y Robles, P. 2003. *Introducción al software libre*. Barcelona: Universitat Oberta de Catalunya.
- Gonzalo, C. 2004. "Fuentes de información en línea para la traducción", en *Manual de documentación y terminología para la traducción especializada*, Gonzalo García, C. y García Yebra, V. (eds). Madrid: Arco libros: 275-307.
- Google Scholar, 2004. "About Google Scholar". [En línea]  
<<http://scholar.google.com/scholar/about.html#about>> [Consulta: 01.10.05]
- Green, W.B. 1993. *Introduction to Electronic Document Management Systems*. Boston: Academic Press.
- Grossmann, R. 1992. *The Existence of the World: An Introduction to Ontology*. Londres: Routledge. [En línea]  
<<http://disputatio.tripod.com/articles/book71.html>> [Consulta: 01.10.05].
- Guzmán, R. 2003. "Tools' Review: LCT Organiser 4.6", *Localisation Focus*, Localisation Research Centre's (LRC), June 2003, Vol. 2, Nº 2.
- Harnessing the Power of Language*. Luxemburgo: DG XIII/E 1998. *Ingeniería lingüística. Cómo aprovechar la fuerza del lenguaje*. [En línea]  
<<http://www2.echo.lu/langeng/en/broch/harness.html>> [Consulta: 01.10.05].
- Harris, B. 1988a. "Are you bi-textual?", *Language Technology*, Nº 7.
- Harris, B. 1988b. "Bi-texts: A new concept in translation theory", *Language Monthly*, Nº 54.
- Harris, D. 1996. "Creating a Knowledge Centric Information Technology Environment", en Harris Training & Consulting Services Inc. (eds). Seattle, WA.
- Heartlight (2004). *La Biblia paralela*. [En línea]  
<<http://spanish.searchgodsword.org/par/>> [Consulta: 01.10.05].

Henczel, S. 2000. *The Information Audit : A Practical Guide*. Munich: K. G. Saur.

Höge, M. y Kroupa, E. 1991a. "Towards the Design of a Translator's Workstation - Organisational Background and User Implications", en *Human Aspects in Computing: Design and Use of Interactive Systems and Information Management, 18B. Proceedings of the Fourth International Conference of Human-Computer Interaction*, H.-J. Bullinger (ed). Stuttgart, Germany, Elsevier, Amsterdam:1036-1040.

Höge, M. 1991b. *Towards a Framework for the Evaluation of a Translators Aids' Systems*. Helsinki: Helsinki University Press.

Hofmann, C. y Mehnert, T. 2000. "Multilingual Information Management at SchneiderAutomation", en *Translation into success*, Sprung, R. (ed). Amsterdam: John Benjamins B.V.: 59-79.

Hubbard, S. 1987. "Information Skills for an Information Society: A Review of Research", en *ERIC Clearinghouse on Information Resources*, Kuhlthau, C. C. (ed). Syracuse: ERIC: 216-327.

Hurtado, A. 2001. *Traducción y Traductología. Introducción a la Traductología*. Madrid: Ediciones Cátedra.

Hutchenson, H. 1988. "La terminologie et Internet: profession en évolution", *Terminologies Nouvelle. Terminologie et diversité culturelle*, Vol. 18: 11-14.

Hutchins, J. y Somers, H. 1995. *Introducción a la traducción automática*. Madrid: Visor.

IBM, 2004. *The IBM Internationalization Glossary*. [En línea] <<http://www-106.ibm.com/developerworks/library/glossaries/unicode.html>> [Consulta: 01.10.05].

- Irazazábal, A. 2000. "Elaboración automática de tesauros multilingües", en *Documentación, terminología y traducción*, Gonzalo García, C. y García Yebra, V. (eds). Madrid: Síntesis: 123-132.
- ISO 2384. 1977. *Documentation – Presentation of translations*. Ginebra: International Organization for Standardization.
- ISO 639. 1988. *Code for the representation of names languages –Part 1: Alpha-2code Part 2: Alpha-3code*. Ginebra: International Organization for Standardization.
- ISO 9241-11. 1998. *Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability*. Ginebra: International Organization for Standardization.
- ISO 12200. 1999. *Computer applications in terminology - Machine-readable terminology interchange format (MARTIF)*. Ginebra: International Organization for Standardization.
- ISO 12620. 1999. *Computer applications in terminology – Data categories*. Ginebra: International Organization for Standardization.
- ISO 9000. 2000-2002. *ISO Standards Compendium: Quality management*. Ginebra: International Organization for Standardization.
- ISO 12616. 2002. *Translation Orienteted Terminography*. Ginebra: International Organization for Standardization.
- ISO 9126. 2003. *Software engineering -- Product quality*. Ginebra: International Organization for Standardization.
- Iverson, S. 2003. "Evaluating Electronic Project Management Tools", *ALC Link Newsletters*, Vol. 1, Nº 3.
- Jackendoff, R. 1998. *La conciencia y la mente computacional*. Madrid: Visor.

Jakobs, E. M. 2000. "Technische Dokumentation", en *Linguistische Berufe – Ein Ratgeber zu aktuellen linguistischen Berufsfeldern*. Becker-Mrotzek M., Brünner, G. y Cölfen. H. (eds). Frankfurt/Main: Lang: 71-78.

Kashyap, V. y Shet, A. 2000. *Information Brokering Across Heterogeneous Digital Data: A Metadata-Based Approach*. Londres: Kluwer Academic Publishers.

Kay, M. y Röscheisen, M. 1988. "Text-Translation Aligment -Technical Report". Palo Alto, CA: Palo Alto Xerox Research Center.

Kiraly, D. 2000. *A Social Constructivist Approach To Translator Education. Empowerment from theory to practice*. Manchester: St Jerome.

Knopp, S. 2000. "Aufbau, Gestaltung und Struktur bei Online-Hilfesystemen", en *Im Kontext der Mensch-Computer-Interaktion*, Hennig, J. y Tjarks-Sobhani, M. (eds). Lübeck: Schmidt-Römhild.

Lancaster, F. W. 1995. *El control de vocabulario en la recuperación de información*. Valencia: Universitat de Valencia.

Lauriston, A. 1995. "Criteria for Measuring Term Recognition", en *Proceedings of the seventh conference on European chapter of the Association for Computational Linguistics*. Dublín: Morgan Kaufmann Publishers. [En línea] <<http://acl.ldc.upenn.edu/E/E95/E95-1003.pdf>> [Consulta: 01.10.05].

Lewis, C. y Rieman, J. 1993. *Task-Centered User Interface Design - A Practical Introduction*. [En línea] <[http://www.eat.lth.se/Kurs/Material/Diverse/Lewis\\_Rieman/Publ\\_1993.htm](http://www.eat.lth.se/Kurs/Material/Diverse/Lewis_Rieman/Publ_1993.htm)> [Consulta: 01.10.05].

LISA (Localisation Industry Standards Association). 2002. "TBX". [En línea] <<http://www.lisa.org/tbx/>> [Consulta: 01.10.05].

- LISA (Localisation Industry Standards Association). 2004. [En línea] <<http://www.lisa.org/>> [Consulta: 01.10.05].
- LISA (Localisation Industry Standards Association). 2004. "TMX 1.4b Specification". [En línea] <<http://www.lisa.org/tmx/tmx.htm>> [Consulta: 01.10.05].
- Lörscher, W. 1991. *Translation Performance, Translation Process, and Translation Strategies. A Psycholinguistic Investigation*. Tübingen: Narr.
- Lörscher, W. 1996. "A Psycholinguistic of Translation Process", *Meta* 41(1): 36-42.
- López Yepes, J. 1996. "Hacia una teoría de la documentación", en *Manual de información y documentación*. López, J. (ed). Madrid: Pirámide: 63-93.
- López Yepes, J. 2002. "Presentación", en *La Investigación en biblioteconomía y documentación*, Delgado, E. Gijón: Trea: 11-12.
- Loughman, L. 1997. "The Expanding World of Globalization", *Language International*, 9.2.
- Lozano, A. 2004. "Comunidades de aprendizaje en red: diseño de un proyecto de entorno colaborativo", en el monográfico "Aprendizaje y construcción del conocimiento en la red", *Teoría de la educación*, Nº 5.
- Mallorquín, C. 1999. "¿Metodología o Ciencia Social?", *Revista Electrónica de Epistemología de Ciencias Sociales*, 6.
- Marcos, M. C. 2004. *Interacción en interfaces de recuperación de información*. Gijón: Trea.

Martín, P. 1998. "De la auditoría de Información a Intranet: Claves para la implantación de sistemas de gestión de información en las empresas", *VI Jornadas Españolas de Documentación. Los Sistemas de Información al servicio de la sociedad*, FESABID. [En línea] <[http://fesabid98.florida-uni.es/Comunicaciones/p\\_martin.htm](http://fesabid98.florida-uni.es/Comunicaciones/p_martin.htm)> [Consulta: 01.10.05].

Martínez de Sousa, J. 1993. *Diccionario de bibliotecología y ciencias afines*. Madrid: Fundación Germán Sánchez Ruipérez.

Martínez de Sousa, J. 1999. *Manual de edición y de autoedición*. Madrid: Pirámide.

Mayer, F. 1996. *Criteria for the Evaluation of Terminology Management Systems*. Gesellschaft für Terminologie und Wissenstransfer (GTW). Bozen: Europäische Akademie.

Mayoral, R. 1997, "La traducción especializada como operación de documentación", *Sendebarr*, Nº 8-9:137-153.

Melby, A. K. 1982. "A bilingual concordance System and its use in linguistics studies", en *The Eight Lacus Forum 1982*, Gutwinsky, W. y Jolly, G. (eds). Columbia: Hornbeam Press: 541-549.

Melby, A. 2001a. "Standards-based Access service to multilingual Lexicons and Terminologies", *Translation, Theory, and Training*. [En línea] <<http://www.ttt.org/salt/description.html>> [Consulta: 01.10.05].

Melby, A. 2001b. "Interchange using TBX", *Presentations of the Open Forum on Metadata Registries*. [En línea] <[http://www.lisa.org/term/tbx\\_intro/tbx\\_files/v3\\_document.htm](http://www.lisa.org/term/tbx_intro/tbx_files/v3_document.htm)> [Consulta: 01.10.05].



Melby, A., Schmitz, K. D. y Wright, S. E. 2001. "Terminology interchange", en *Handbook of Terminology Management*, Vol. 2: *Applications-oriented Terminology Management*, Budin, G. y Wright, S. E. Amsterdam: John Benjamins B.V.: 613-642.

Melby, A. K. 1992. "The Translator Workstation", en *Computers in Translation*, Newton, J. (ed). Londres: Routledge Press.

Méndez, E. 2002. *Metadatos y recuperación de información*. Gijón: Trea.

Merlo, J. A. 2003. "La evaluación de la calidad de la información web: aportaciones teóricas y experiencias prácticas", en *Recursos informativos: creación, descripción y evaluación*. Mérida: Junta de Extremadura: 101-110.

Microsoft. 2004. *The Microsoft glossary*. [En línea] <<http://www.microsoft.com/globaldev/reference/glosshome.msp>> [Consulta: 01.10.05].

Milstead, J. 1991. "Specifications for thesaurus software", *Information Processing and Management*, Vol. 27, Nº 2-3: 165–175.

Ministerio de Educación y Cultura. 2004. Banco de imágenes. [En línea] <<http://recursos.cnice.mec.es/bancoimagenes/>> [Consulta: 01.10.05].

Monterde, A. M. 2004a. "Evolución de modelos de formas de representación del conocimiento a nivel terminológico: propuesta de un modelo actual", *LSP & Professional Communication*. Vol. 4, Nº 1: 49-68.

Monterde, A. M. 2004b. "Importancia de la ilustración para la traducción técnica", en *Manual de documentación y terminología para la traducción especializada*, Gonzalo García, C. y García Yebra, V. (eds). Madrid: Arco libros: 259-274.

Moreno, A., Ares, F. y Marcos, F. 1999. *Recuperación de información: consulta a recursos lingüísticos*. Madrid: CICYT- Universidad Autónoma de Madrid.

Moreno Ortiz, A. 2000. *Diseño e implementación de un lexicón computacional para lexicografía y traducción automática*. Madrid: Estudios de Lingüística Española.

Moreno-Torres, I. 1996. "Textos dentro del ordenador", *Trans*, 1: 97-109.

Moscoso, P. 1996. "Bases de datos documentales: Estructura y organización", en *Manual de información y documentación*, en López Yepes (ed). Madrid: Pirámide: 393-403.

Moya, G. y Gil, I. 2001. "Evaluación de software de gestión de tesauros" *Ciencias de la Información*. [En línea]  
<<http://ttt.upv.es/isgil/Articulo%20software%20gestion%20de%20tesauros.pdf>  
> [Consulta: 01.10.05].

Murray Jr. H. 1996 "Methods for Satisfying the Needs of the Scientist and the Engineer for Scientific and Technical Communication". Washington, D.C.

Nkwenti-Azeh, B. 1994. "Positional and combinational characteristics of terms Consequences for corpus-based terminography", *Terminology* N° 1,1: 61-96.

Nord, C. 1991. *Text analysis in translation. Theory, methodology, and didactic application of a model for translation-oriented text analysis*. Amsterdam: Atlanta.

Olarán, M. y Martínez, M. 1999. *Manual de catalogación*. Madrid: Arco libros.

ÖNORM D 1200.2000. *Anforderungen an die Dienstleistung und an die Bereitstellung der Dienstleistung*. Viena: Österreichische Dienstleistungszentrum für Normen.

ÖNORM D 1201.2000. *Übersetzungsleistungen – Übersetzungsverträge*. Viena: Österreichische Dienstleistungszentrum für Normen.

Orna, E. 1999. *Practical information policies*. Londres: Gower.

Pérez, Ch. 2002. *Explotación de los corpórea textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento*. Madrid: Estudios de Lingüística Española.

Pinto, M. 2000a. "Documentación para la traducción en la sociedad de la información", en *DoIS-Documents in Information Science*. [En línea] <<http://dois.mimas.ac.uk/DoIS/data/Papers/julmnibkd2899.html>> [Consulta: 01.10.05].

Pinto, M. 2000b. "Premisas para la implantación de sistemas de calidad en los servicios de traducción" en *Documentación, terminología y traducción*, Gonzalo García, C. y García Yebra, V. (eds). Madrid: Síntesis: 61-72.

Recoder, M. J. 2003. "La documentació aplicada a la traducció especialitzada i a la traducció literaria", *Tradumática*, Nº 2. [En línea] <<http://www.fti.uab.es/tradumatica/revista/num2/articles/02/02.pdf>> [Consulta: 01.10.05].

Rinsche, A. 1997. "Translation Technology: The State of the Art" en *Localization Resources Centre. Yearbook 1997*. Dublin: Localization Resources Centre

Robertson, G. 1994. "Information audit: a broader perspective." *Managing Information*, 1(5): 34-36.

Roets Y. y Boon Y. 1992. "Information Manager, an exploratory study." *South African Journal of Library and Information Science*, 60 (3): 137-145.

Ros, M. 2004. "La ¿crisis? de la profesión del Documentalista / Bibliotecario en España", *El blog el documentalista enredado*. [En línea] <<http://www.documentalistaenredado.org/enredado/index.php?p=39>> [Consulta: 01.10.05].

Sánchez-Gijón, P. 2001. “Catàleg de sistemes de memòries de traducció”, *Tradumática*, N° 0. [En línea]

<<http://www.bib.uab.es/pub/tradumatica/15787559n0a6sanchez-gijon.pdf>>

[Consulta: 01.10.05].

Sánchez-Gijón, P. 2003. *Els documents digitals especialitzats: utilització de la lingüística de corpus com a font de recursos per a la traducció especialitzada*.

Barcelona: IULA-Universidad Pompeu Fabra.

Sandra, K., Roe, A. 2004. “Introducción”, en *The Thesaurus : Review, Renaissance, and Revision*. Sandra, K., Roe, A. (eds). Londres: Haworth: 4-13.

Sanz, E. 1996. “Necesidades informativas y estudios de los usuarios”, en *Manual de información y documentación*, López Yepes (ed). Madrid:

Pirámide: 183-199.

Schäler, R. “Language Resources and Localisation”, *Language Resources for Translation Work, Research and Training*. Ginebra: LR4Trans-2. [En línea]

<[http://www.ifi.unizh.ch/cl/yuste/lr4trans-2/WKS\\_PAPERS/3.pdf](http://www.ifi.unizh.ch/cl/yuste/lr4trans-2/WKS_PAPERS/3.pdf)> [Consulta:

01.10.05].

Scholand, M. 1996. *Computerspiele als Form der interaktiven Unterhaltung*.

*Globalisierung, Lokalisierung, Übersetzung*. Germersheim: Johannes

Gutenberg-Universität Mainz.

Scholand, M. 2002. Localización de videojuegos. *Tradumática*, N° 1.

Traducción del alemán de Lidia Cámara. [En línea]

<<http://www.fti.uab.es/tradumatica/revista/articles/mscholand/mscholand.pdf>>

[Consulta: 01.10.05].

Scholand, M. 2004. “La internacionalización de software y de sitios Web”,

*Contextos*. Madrid: Linguaserve, N° 16.

- Schwartz, H. y Toon, A. 2004. "Global Content Management", en *White Papers*, Trados (ed).
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Somers, H. y Fernández Díaz, G. 2003. "Diferencias e interconexiones entre los sistemas de memorias de traducción y la EBMT", en *Entornos informatizados de la traducción profesional*, Corpas, G. y Varela, M. J. (eds). Granada: Atrio: 167-192.
- Society for technical communication (STC). 2004. "About us. Mission". [En línea] <<http://www.stc.org/about.asp>> [Consulta: 01.10.05].
- Soy, C. 2003a. "La auditoría de la información, componente clave de la gestión estratégica de la información", *El profesional de la información*. Vol. 12, Nº 4: 261-268.
- Soy, C. 2003b. *Auditoría de la información. Análisis de la información generada en la empresa*. Barcelona: Editorial UOC.
- STAR Servicios Lingüísticos SL. [En línea] <<http://www.star-spain.com>> [Consulta: 01.10.05].
- STAR, 2004. WebTerm. *Technical White Paper*. Suttgart.
- Strandvik, I. 2001. "Las memorias de traducción en la Comisión Europea", *Tradumática*, Nº 0. [En línea] <<http://www.fti.uab.es/tradumatica/revista/num0/articles/istrandvik/art.htm>> [Consulta: 01.10.05].
- Surmacz, J. 2003. "Business intangibles", en el portal CIO. [En línea] <<http://www.cio.com>> [Consulta: 01.10.05].
- Tamis, 2003. Corporate terminology with XST. [En línea] <[http://www.temis-group.com/temis/attachments/factsheets/XTS\\_en.pdf](http://www.temis-group.com/temis/attachments/factsheets/XTS_en.pdf)> [Consulta: 01.10.05].

Termcat. 1990. *Metodología del treball terminològic*. Barcelona: Departament de Cultura. Generalitat de Catalunya.

TRADOS. 2004. TRADOS Mutiterm iX. *Technical White Paper*.

Tradumatica, 2001. "Editorial". [En línea]  
<<http://www.fti.uab.es/tradumatica/revista/num1/Editorial.htm>> [Consulta: 01.10.05].

Tradumatica, 2002. "Editorial". [En línea]  
<<http://www.fti.uab.es/tradumatica/revista/num2/Editorial.htm>> [Consulta: 01.10.05].

Tramullas, J. 2001. *Introducción a la documática*. Zaragoza: Kronos. [En línea] <<http://tramullas.com/documatica/1.html>> [Consulta: 01.10.05].

Tramullas, J. 2000. "Planteamiento y componentes de la disciplina *Information Design*". [En línea] <<http://www.tramullas.com/ai/concepto.pdf>> [Consulta: 01.10.05].

Tramullas, J. 1995. "Una introducción a la Informática Documental", *Apuntes del Centro de Cálculo de la Universidad de Zaragoza*, Nº 8: 6-10.

Unicode. 1991-2004. "The Unicode Consortium". [En línea]  
<<http://www.unicode.org/consortium/consort.html>> [Consulta: 01.10.05].

UNI 10574. 1995. *Definizione dei servizi e delle attività delle imprese di traduzione e di interpretariato*. Roma: Organismo di Normazione e di Certificazione di Sistemi di qualità italiano.

Valdés, M. C. y Urra, P. 2000. "El localizador de información como herramienta para gerenciar los recursos de información", *ACIMED*, Vol. 8(2): 116-27.

Vargas Sierra, Ch. 2002. "Utilización de los programas de concordancias en la traducción especializada", *El español, lengua de traducción. I Congreso Internacional*. Bruselas: Servicio de Traducción de la Comisión Europea: 468-483.

Vasconcellos, M. 1988. "Introduction", en *Technology as translation strategy*, Vasconcellos, M. (ed). Binghamton, NY.: State University of New York Press, (American Translators Association Scholarly Monograph Series, Vol. 2).

Vellucci, S. L. 1998. "Metadata", *Annual Review of Information Science and Technology*, N° 33: 187-222.

Villanueva, E. 1996. "Bases de datos y bibliotecología: cómo deshacer la innecesaria incomunicación", *Investigación bibliotecológica*, 10 (20): 27-32.

W3C, *Extensible Markup Language (XML) 1.0* (W3C recommendation 10-February-1998). [En línea] <<http://www.w3.org/TR/1998/REC-xml-19980210>>[Consulta: 01.10.05].

Wahle, K. 2000. "Wie wird Software lokalisiert?", en *Software Lokalisierung*, Schmitz, K.D. y Wahle, K. Tübingen: Stafuffenburg: 49-68.

Ward, T. 2004. "Measuring the ROIs of intranets: mission possible?" [En línea] <<http://www.cio.com>>. *Resumen del White Paper "Intranet ROI"*. [En línea] <<http://www.IntranetInsight.com>> [Consulta: 01.10.05].

Wassmer, T. 2003. *LTC Organiser*, en *Multilingual Computing and Technology magazine*. Vol. 15 (1), N° 61.

Webopedia: Online Computer Dictionary for Computer and Internet. 2005. [En línea] <<http://www.webopedia.com/>> [Consulta: 01.10.05].

Wikipedia, the Free Enciclopedia. 2005. [En línea] <[http://en.wikipedia.org/wiki/Software\\_testing](http://en.wikipedia.org/wiki/Software_testing)> [Consulta: 01.10.05].

Willpower Information, 1998. "Thesaurus principles and practice", en *Thesauri for museum documentation. Proceedings of a Workshop held at the Science Museum*. Cambridge: The Museum Documentation Association, Terminology Working Group, MDA. [En línea] <<http://www.w3.org/TR/1998/REC-xml-19980210>> [Consulta: 01.10.05].

Wilss, W. 1992. *Übersetzungsfertigkeit, Annäherungen an einem komplexen übersetzungspraktischenbegriff*. Tübingen: Narr.

### **6.1. Herramientas utilizadas para el diseño e implementación de COGNmática**

Search and Replace, Funduc. 2004. [En línea] <[http://www.funduc.com/search\\_replace.htm](http://www.funduc.com/search_replace.htm)> [Consulta: 01.10.05].

Transit XV, Star Group, Suiza. 1999-2004. [En línea] <<http://www.star-spain.com/es/tecnologia/transit.html>> [Consulta: 01.10.05].

Tesauro *Spines*, Vocabulario controlado y estructurado para el tratamiento de información sobre ciencia y tecnología para el desarrollo. UNESCO-CINDOC 1984-2004 [En línea] <<http://pci204.cindoc.csic.es/tesauros/SpinTes/Spines.html>> [Consulta: 01.10.05].