

Una enciclopedia *para* traductores. Los géneros de especialidad como herramienta privilegiada del traductor profesional

Isabel GARCÍA IZQUIERDO Y Esther MONZÓ NEBOT
Universitat Jaume I

Como citar este artículo:

GARCÍA IZQUIERDO, Isabel y MONZÓ NEBOT, Esther (2003) «Una enciclopedia *para* traductores. Los géneros de especialidad como herramienta privilegiada del traductor profesional», en MUÑOZ MARTÍN, Ricardo [ed.] *I AIETI. Actas del I Congreso Internacional de la Asociación Ibérica de Estudios de Traducción e Interpretación. Granada 12-14 de Febrero de 2003*. Granada: AIETI. Vol. n.º 1, pp. 83-97. ISBN 84-933360-0-9. Versión electrónica disponible en la web de la AIETI:
<http://www.aieti.eu/pubs/actas/I/AIETI_1_IGI_EMN_Enciclopedia.pdf>.

Una enciclopedia para traductores. Los géneros de especialidad como herramienta privilegiada del traductor profesional

Isabel GARCÍA IZQUIERDO y Esther MONZÓ NEBOT
Universitat Jaume I
igarcia@trad.uji.es

Resumen

El equipo de investigación GENTT del Departamento de Traducción y Comunicación de la Universitat Jaume I lleva a cabo una investigación que tiene por objeto proporcionar a los traductores profesionales de tres ámbitos de especialidad (traductores de textos jurídicos, médicos y técnicos) una herramienta que les permita reducir considerablemente el tiempo dedicado a la documentación. Nuestro empeño se ha centrado en demostrar la utilidad del concepto de género en su aplicación a la tarea de los traductores profesionales y, desde los primeros resultados, hemos visto la necesidad de dotar al colectivo de una herramienta en la que no sólo puedan consultarse textos de especialidad en los idiomas de trabajo sino que pueda accederse también a la información más relevante para el profesional, convenientemente tratada y presentada. En el presente trabajo exponemos las bases teóricas y metodológicas de nuestro proyecto, damos cuenta de la situación de nuestro proyecto en la actualidad y planteamos los nuevos retos que el equipo se ha propuesto resolver de cara a producir una enciclopedia de géneros en la que puedan encontrarse instancias reales de los documentos que deban traducirse junto con datos sobre sus contextos de cultura y situación, su desarrollo semántico y textual, la utilización real que le dan los usuarios habituales y la relación del género en cuestión con otros disponibles en el corpus, ya sean originales o también traducciones. Proponemos también otras vías de investigación a las que la nuestra podría contribuir. En suma, nos planteamos favorecer los avances que puedan mejorar la vida del profesional, el investigador y el docente de traducción.

El encargo llega, una hipoteca británica. En la carrera decían que era un encargo común, pero todavía no me habían pedido ninguna. La verdad es que lo mío son estatutos, algún contrato y como mucho balances. Pero traducir hipotecas... Si lo acepto, y les gusta, tengo cliente nuevo, y lo cierto es que un banco cunde. Pero, ¿de dónde saco yo toda la documentación en tan poco tiempo? Yo sé de una que esta noche no duerme.

1. Introducción

Para una profesional, un encargo nuevo de traducción supone familiarizarse con una gran cantidad de información desconocida, pero quizás peor que esto es el tiempo que se requiere para encontrarla, por no hablar del trabajo que implica asegurarse de que se conserva la información para no tener que volver a buscarla de nuevo. GENTT no pretende ser un remedio contra el insomnio, aunque no nos importaría ayudar a conciliar el sueño a los profesionales. Lo que persigue nuestro equipo de investigación es desarrollar una herramienta auxiliar que permita obtener en muy poco tiempo y con un esfuerzo reducido la información que un traductor profesional consigue interiorizar tras largas horas de consultas, documentación, estudio y traducción. Si ahora mismo hay diferentes traductores buscando la misma información, si mañana otros tantos la volverán a buscar y, sobre todo, si esta documentación existe, aunque dispersa, y es sistematizable, ¿por qué no podemos reunirlos en un solo lugar para no tener que acudir a una infinidad de fuentes distintas?, ¿por qué no podemos hacer que ese lugar esté accesible a todos los traductores que lo necesiten?

Nuestro equipo se planteó estas preguntas y los 12 investigadores integrantes nos pusimos manos a la obra. Nuestro objetivo principal es proporcionar a los traductores profesionales una herramienta sencilla, a modo de diccionario o de enciclopedia, que les permita reducir considerablemente el tiempo dedicado a la documentación. Aunque nos gustaría llegar a todos los traductores, hemos acotado nuestro campo de trabajo a tres ámbitos de especialidad (traductores de textos jurídicos, médicos y técnicos).

Una cuestión previa que tuvimos que resolver era de dónde debía partir nuestra investigación, es decir, en qué debían consistir las entradas de nuestra enciclopedia. Para dilucidar esta cuestión nos fijamos en los componentes de la situación profesional, entre los cuales consideramos de mayor importancia el *texto*. Sin embargo, un texto es una ocurrencia aislada que no puede resultarnos útil para prevenir dificultades futuras y conocer los ingredientes que nos permitirían paliarlas. Por ello, necesitábamos acudir a un concepto que nos ayudara a *prever*, que recogiese los factores reiterativos de estos textos que, en la situación de traducción, dibujan los encargos del profesional. Necesitábamos, por tanto, acudir al *género textual*.

Partiendo de este razonamiento básico, nuestro empeño se ha centrado hasta el momento en demostrar la utilidad del concepto de *género* de los ámbitos de comunicación especializada en su aplicación a la tarea de los traductores profesionales y, desde los primeros resultados, hemos visto la necesidad de incluir en la herramienta resultante no sólo textos de especialidad en los idiomas de trabajo sino también la información *genológica* más relevante para el profesional, convenientemente tratada y presentada. El producto que pretendemos ofrecer con nuestra investigación es una arquitectura informática de explotación que permitirá conseguir mejoras para la traducción en el ámbito formativo, pero también en la práctica profesional de la traducción especializada, para la que plantearemos una herramienta de consulta textual, conceptual, terminológica y lingüística.

En el presente trabajo queremos dar cuenta de la situación de nuestro proyecto en la actualidad y plantear los nuevos retos que el equipo se ha planteado resolver de cara a producir una enciclopedia de géneros en la que puedan encontrarse instancias reales de los documentos que deban traducirse junto con datos sobre sus contextos de cultura y situación, su desarrollo semántico y textual, explicaciones sobre la utilización real que le dan los usuarios habituales (en los ámbitos especializados) y la relación del género en cuestión con otros disponibles en el corpus, ya sean originales o también traducciones. El modo de consultar esta información en la versión consumible del producto se pretende que sea lo suficientemente sencillo como para asegurar su correcta difusión entre los profesionales y los investigadores y lo bastante sofisticado para permitir búsquedas complejas, modos alternativos de visualización y explotación y la presentación clara y completa de los datos relevantes.

2. Bases teóricas

Como ya hemos remarcado, nuestro trabajo se plantea intereses tanto teóricos como prácticos. Para sustanciar las bases de este trabajo empezaremos, en primer lugar, por ofrecer una panorámica de índole teórica que nos permita trabajar en las aplicaciones de la noción de género al ámbito de la traducción. En nuestro proyecto, recogemos la tradición genológica, con presencia en diversas disciplinas y ámbitos de estudio, y hemos extraído de ella los aspectos que fundamentan nuestra hipótesis teórica, es decir, que los géneros nos permitirán obtener información de naturaleza varia que sea de utilidad para el traductor (para una panorámica sucinta de esta tradición y de sus aplicaciones a la traducción, véase Monzó 2001).

Este concepto constituye de por sí un ámbito de estudio que tiene una presencia prolongada y destacada en disciplinas como la teoría de la literatura (Bajtín 1986) o la lingüística (Halliday y Hasan 1985), pero que también ha llamado la atención de estudiosos en campos como la retórica (Miller 1984, Dunmire 2000), la estilística lingüística (Crystal y Davy 1969/1979) o los estudios de comunicación audiovisual (Kress 1985). Desde la introducción de conceptos culturales en estas disciplinas, el género, tratado en ocasiones como mera forma repetitiva, adquiere una dimensión sociológica en la que se cimienta la comunicación humana. Por otro lado, desde que en traductología se entiende la traducción como una operación intertextual e intercultural (Bassnett 1991) los puntos de convergencia entre ambos ámbitos hacen surgir el interés del estudio de esta noción en nuestro campo (Reiß y Vermeer 1989, Hatim y Mason 1990, Baker 1992, Hurtado 1995, Trosborg 1997).

En un trabajo anterior (Monzó 2002: caps. 2 y 3), se realizó un estudio multidisciplinar de esta tradición genológica y caracterizamos la noción del modo que a continuación exponemos. Un género es, ante todo, un acto de comunicación. Como cualquier concepto, lo que el género como noción añade a este acto es la abstracción, que se consigue a través de la tipificación de sus rasgos. Sin embargo, esta tipificación no surge de intereses investigadores, sino que se

desarrolla y fija en una comunidad determinada a través del uso que sus miembros hacen de estos *moldes* de comunicación. Los moldes se construyen progresivamente a partir de la recurrencia en un sistema sociocultural determinado de una situación particular en la que los participantes se adhieren a elementos formales que discursivizan sus propósitos comunicativos. La reiteración de estos rasgos por parte de los interlocutores no es gratuita, sino que se debe a su experiencia previa en situaciones y entornos socioculturales análogos, ya sea como intervinientes o como testimonios, que les permiten establecer relaciones eficientes entre sus enunciaciones y las reacciones de su interlocutor. Así, la recurrencia del acto discursivo constituye un punto de convergencia entre los individuos de la misma cultura que facilita la consecución de las metas comunicativas y reduce la incertidumbre en el proceso de comunicación. Por otro lado, en situaciones nuevas para el receptor, el género adopta el papel de transmisor de conocimientos, actitudes y emociones que el sistema sociocultural en que enraíza el género ha tipificado por un uso reiterado; evidentemente, estos factores y rasgos pueden variarse de acuerdo con el conjunto de propósitos comunicativos que el individuo quiere actualizar en una situación comunicativa que se produce en un ámbito de comunicación humana.

En resumen, pues, los géneros son las unidades de comunicación de una comunidad. Constituyen una síntesis de situaciones y manifestaciones discursivas prototípicas que constituyen una herencia aprendida y un medio de aceptación en las diferentes comunidades (son, al cabo, parte de nuestro capital cultural). Son, por supuesto, un instrumento de transmisión de contenidos a través de medios discursivos en el que se aprecia la situación de los interlocutores en la comunidad en la que participan (cumplen, por tanto, las funciones del lenguaje que identifican los funcionalistas). Y, por último, pertenecen a comunidades, que pueden ser definidas por su campo de actuación, cuyos miembros son clientes potenciales de los traductores en cualquier situación intercultural. Si el traductor quiere triunfar en su tarea de comunicación, debe maniobrar con todos estos factores, por lo que deberá, cuanto menos, conocerlos.

Desde esta perspectiva multicultural, algunos estudios realizados desde nuestro proyecto (García Izquierdo 1999) muestran con estudios empíricos cómo las diferentes culturas construyen los textos de acuerdo con una serie de parámetros preestablecidos, no sólo ateniendo a la estructura de la lengua, cosa que resulta evidente, sino también a cuestiones relacionadas con la ideología, el registro, etc.; en definitiva, con cuestiones determinadas culturalmente. Estas divergencias, de por sí, permiten que los géneros presenten un interés destacado para la traductología y, muy especialmente, para el traductor. Si éste debe conocer las características discursivas y culturales de sus clientes también deberá conocer las de aquella cultura por la que éste se interesa y en la que se inscribe el texto originario.

Desde el convencimiento empíricamente respaldado de que los rasgos de estos textos culturales corresponden a patrones discursivos y son, por tanto, sistematizables, otros miembros del equipo de investigación se han centrado en la clasificación de géneros de especialidad (Borja 2000, Gamero 2001). El intento de una taxonomía general, acotada a los campos que ha tomado nuestro

proyecto, es uno de nuestros objetivos procedimentales, que nos permitirá descubrir cuáles son las entradas relevantes de nuestra enciclopedia. Por el momento, hemos apreciado que no sólo la estructura interna de los géneros presenta divergencias interculturales, sino que el propio sistema genológico es distinto en cada sistema sociocultural.

2.1. Fijación terminológica de los conceptos relevantes

Hasta el momento, hemos manejado una serie de conceptos de extensa difusión en varias disciplinas, pero que quizá no sean excesivamente comunes en el propio seno de la traductología. Por ello, hemos considerado adecuado dedicar un pequeño apartado a definir someramente los conceptos básicos que cimientan nuestra investigación. Esta pequeña explicación de algunos términos se ha extraído en parte de trabajos anteriores (García Izquierdo 2000, Monzó 2002) y acoge perspectivas diversas cuya utilidad hemos comprobado con el desarrollo de la investigación.

En primer lugar, en cuanto al propio término *género*, ya hemos desarrollado su significado en el apartado anterior, pero lo que nos interesaría aquí es especificar su idiosincrasia con respecto a otro término, *tipo textual* o variantes, con el que se confunde a menudo. Una definición de género que nos permitiría dilucidar claramente las divergencias es la que ofrecen Hatim y Mason (1990) y que extraen de Kress (1985):

forma convencionalizada de texto que posee una función específica en la cultura en la que se inscribe y refleja un propósito del emisor previsible por parte del receptor.

Si bien *tipo textual* también se refiere a formas convencionalizadas de texto, en este caso tales formas se relacionan con la estructura lingüística de las mismas y no con su proyección sociocultural (véase, a este respecto, Castellà 1995). Es cierto que esta definición del género lo convierte en una categoría de difícil descripción para el traductor, puesto que las convenciones de cada género están determinadas culturalmente y ello comportará no pocas dificultades. Pero también es cierto que, así definida, la noción de género se revela como una noción central en cuya descripción convergen, casi sin quererlo, otras nociones contextuales relevantes para el análisis textual y, en consecuencia, para el análisis de la traducción en tanto que texto (García Izquierdo, en prensa).

En relación con este concepto central, aparecen en nuestro trabajo toda una serie de derivados que hemos ido creando por necesidad o adoptando por conveniencia. Entre ellos, *genología* (van Tieghem 1938), para referirnos a la corriente de estudio de la noción; *sistema genológico* (Opacki 1963/2000), para hacer referencia al conjunto ordenado y jerarquizado de géneros con los que cuenta una comunidad o una cultura; *macrogénero* (Martin 1992); *supragénero* o *subgénero*. A este último respecto, debemos destacar el conflicto entre macrogénero y supragénero. Este último grupo de términos nos ha planteado un problema ante el cual adoptamos una solución que nos hace divergir en cuanto a opciones terminológicas de una parte de estudiosos preocupados por la genología.

En concreto, *supragénero* se refiere a uno de los niveles de nuestra taxonomía genológica, que hemos dividido en tres niveles. Así, un género podría dividirse en varios subgéneros (como dentro del género *contrato* podríamos encontrar contratos de agencia o contratos de arrendamiento) y podría quedar agrupado con otros géneros en un mismo supragénero (en nuestra taxonomía, *contratos* y *poderes*, por ejemplo, se incluirían en el supragénero *acuerdos de voluntades*). Los criterios de la consideración de una determinada abstracción como género, subgénero o supragénero son por el momento difusos, pero esperamos poder fijarlos con más acierto a medida que avance nuestra investigación. Asimismo, también resulta complejo establecer claramente los límites entre distintos géneros en los ámbitos profesionales (García Izquierdo y Montalt, en prensa). En la bibliografía consultada, no obstante, esta categoría taxonómica (aplicada alternativamente de acuerdo con diferencias comunicativas o estructurales) se designaba con el término *macrogénero*. Nosotros, sin embargo, decidimos separarnos de esta tendencia al necesitar el concepto que los funcionalistas designaban *macrogénero*. Esta noción, que no utilizábamos en un principio, surgió de la necesidad de clasificar de algún modo, sin perder información, textos en los que se combinaban diversos géneros pero cuya entidad final no era la suma de ellos. Así, un catálogo publicitario, por ejemplo, incluye una carta de presentación que precede a la información técnica, en la que se combinan artículos divulgativos con publicidad. La información técnica no constituye ningún catálogo y la designación de artículos *divulgativos+publicidad+carta* no resulta transparente a la hora de determinar el propósito comunicativo o la estructura de tal amalgama de textos.

Finalmente, existen nociones relacionadas con la metodología utilizada que también nos llevaron a tomar decisiones terminológicas, pero que preferimos tratar en el apartado correspondiente a la metodología. Asimismo, hemos tenido que enfrentarnos a dilemas terminológicos en otros aspectos de nuestro trabajo. En concreto, en cuanto a la metodología, resultaron problemáticos el propio concepto de *corpus*, toma matices distintos en nuestro trabajo que en el de otros estudios, o el de texto paralelo. Por ser éstas cuestiones metodológicas las trataremos en el epígrafe siguiente.

3. Bases metodológicas

La metodología utilizada sigue las tendencias actuales de la investigación en traductología, donde la utilización de los *corpora* se está imponiendo como herramienta fundamental para el análisis de los textos originales y traducidos. Siguiendo la definición propuesta por EAGLES (*Expert Advisory Group on Language Engineering Standards*) (EAGLES), un corpus es:

A collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language

Las posibilidades que ofrece este tipo de estudios son múltiples. Podemos estudiar la frecuencia de determinadas palabras y contrastarla con estudios de otros corpus; podemos analizar los usos contextuales de las palabras; las funciones, las colocaciones, los préstamos en un género textual concreto, la complejidad de las oraciones en diferentes tipos textuales, la variación sintáctica o la fraseología específica de un género. Todo ello, teniendo en cuenta que los corpora actuales están digitalizados y adquieren por ello dimensiones inimaginables hace unos años.

La tipología de corpus es también muy variada. Corpus (2001:3ss.) propone una clasificación partiendo de la propuesta de Torruella y Llisterri (1999) y combinándola con las de Baker (1995) y Johansson (1998). Así, en su opinión, podemos distinguir:

- a) Según el porcentaje y la distribución de los diferentes tipos de texto: Corpus grande o extenso, Corpus equilibrado, Corpus piramidal, Corpus monitor, Corpus paralelo y Corpus comparable
- b) Según la especificidad de los documentos contenidos en el corpus: corpus general, corpus especializado, corpus canónico, corpus periódico o cronológico y corpus diacrónico.
- c) Según la cantidad de texto que se reúne en cada uno de los documentos que componen el corpus: corpus textual, corpus de referencia y corpus léxico.
- d) Según la codificación y anotación: corpus no anotado, corpus anotado
- e) Según la documentación que acompaña al corpus: Corpus documentado y corpus no documentado.

Siguiendo esta clasificación, podemos clasificar el corpus de investigación que estamos recopilando como un corpus comparable (incluye tipos similares de textos originales en diversas lenguas), especializado, textual (con textos completos), anotado (etiquetado lingüística o metatextualmente) y documentado (cada documento textual lleva asociado una cabecera o *Header*, descriptiva de su procedencia). Sin embargo, debemos distinguir este corpus de investigación del producto final que se ofrecerá en forma de enciclopedia, que constará también de textos completos, similares entre diversas lenguas, de campos especializados, con documentación auxiliar, pero no anotado. En lo sucesivo explicaremos la relevancia de estas opciones para los objetivos que nos planteamos en el corpus de investigación.

En primer lugar, se trata de un corpus *comparable*. En este sentido, destacamos la contribución a la traductología de los estudios que Baker (1995) ha desarrollado en el seno del proyecto TEC. En su opinión, a pesar de los inmejorables resultados que un corpus multilingüe puede reportar (sobre todo para la elaboración de diccionarios), desde el punto de vista de los estudios sobre la traducción es más útil dejar de lado el contraste entre lenguas originales y utilizar estos corpus comparables, conformados de textos paralelos y multilingües. Es decir, se trata de comparar los textos originales en una lengua x y esos

mismos textos cuando son producto del transvase. En ese sentido, el corpus que proponemos entraría dentro de la categoría de los que Baker denomina «comparable», puesto que en el corpus se incluyen textos originales, equivalentes, paralelos y traducciones en las cuatro lenguas de trabajo.

Por otro lado, hemos ceñido la selección de los textos a su naturaleza *especializada*. Así, en nuestro corpus incluimos los textos que tienen un uso real en los ámbitos profesionales de los campos jurídico, médico y técnico. El motivo para esta elección es trabajar con textos que puedan encontrarse los traductores en su actividad profesional, aunque para conservar nuestros objetivos dentro de unos límites controlables hemos acotado campos concretos y así, por ejemplo, dentro de la traducción técnica nos centramos en los ámbitos de la automoción y de la construcción.

En tercer lugar, se trata de un corpus *textual*. El objetivo que nos planteamos no es en realidad extraer patrones lingüísticos de una gran cantidad de datos (Sánchez 1995), sino crear una *enciclopedia* con un texto completo *prototípico* por género que dé cuenta de los rasgos genológicos. El traductor puede así conocer rápidamente el género a través del texto y, junto con el material complementario, socializarse genológicamente hablando y, al fin, mejorar su rendimiento en la traducción de los textos de especialidad. Por ello el volumen de nuestra enciclopedia se mide no por palabras sino por textos.

Se trata también de un corpus *anotado*. Como se explicará en el apartado 5, pretendemos analizar los textos disponibles con vistas a obtener datos sobre la macroestructura textual y la microestructura lingüística. Al definir el lenguaje del etiquetado nuestras prioridades fueron, por un lado, que nos ofreciera amplias posibilidades de análisis y, por otro, que fuera compatible con otros sistemas y trabajos de corpus. Por ello, utilizaremos como punto de partida el estándar de anotación diseñado por el consorcio TEI.

Finalmente, se trata de un corpus documentado porque cada documento textual lleva asociado una cabecera o *Header*, en la que se incluyen todos los datos relevantes para la identificación de los diferentes textos, como veremos más adelante.

4. El estado actual de la enciclopedia de géneros

En el siguiente diagrama de flujo podemos observar los objetivos generales que se han planteado para este proyecto. En estos momentos nos encontramos en el punto de etiquetado del cuerpo de los textos (*body markup*). A continuación exponemos brevemente el trabajo realizado en las fases previas.

En un primer momento, el grupo reunió una gran cantidad de textos susceptibles de ser traducidos, pertenecientes a los ámbitos acotados en el planteamiento del proyecto. Los textos se conseguían o bien en papel o ya en soporte digital, pero en cualquier caso, nuestro *modus operandi* daba como resultado tres versiones formalmente distintas de cada texto. La primera de ellas era el soporte papel, para archivar todos los textos conjuntamente con la documentación del equipo. En segundo lugar, se escaneaba el texto impreso y

se conseguían dos documentos: un documento pdf en el que se conservaba una fiel *imagen* del texto original y un documento html del que se extraía cualquier información gráfica. El texto html es que el se etiquetará a través del programa diseñado *ad hoc* para nuestros objetivos. El etiquetado que se introduce en el texto a través de este programa (véase García Izquierdo y Monzó Nebot, en prensa) se divide en dos partes, de las que hasta el momento hemos completado una, el *header markup*.

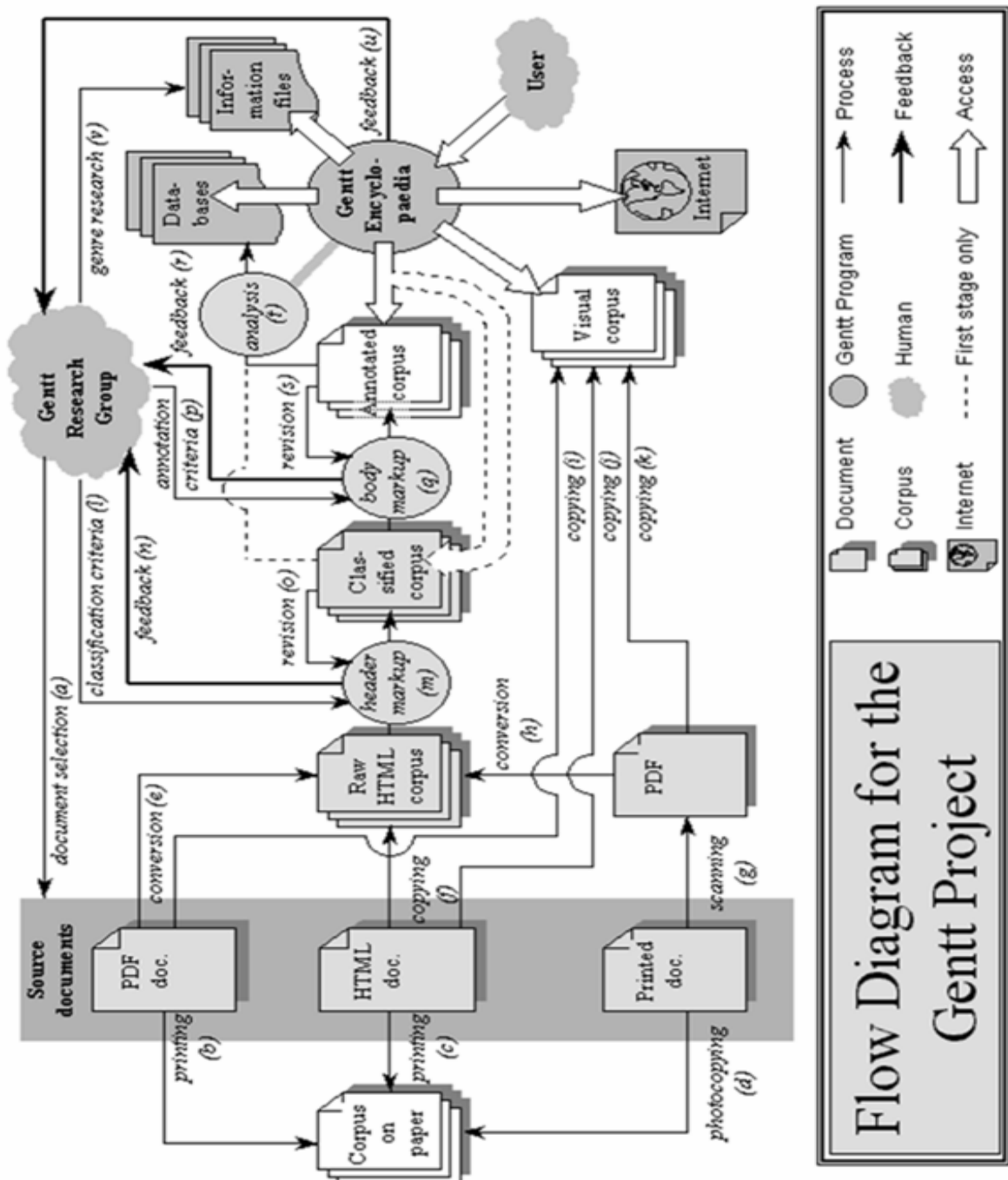


Figura 1. Diagrama de flujo del trabajo realizado por el grupo GENTT (Jennings, en proceso).

- a) El grupo GENTT selecciona documentos para el corpus.
- b) Los documentos originariamente en PDF se imprimen para archivarse en la copia impresa del corpus.
- c) Los documentos originariamente en HTML se imprimen para archivarse en la copia impresa del corpus.
- d) Los documentos originariamente en papel se fotocopian para archivarse en la copia impresa del corpus.
- e) Los archivos PDF originarios se convierten en archivos HTML para introducirse en el corpus de HTML simple.
- f) Los archivos HTML se copian en el corpus de HTML simple.
- g) Los documentos originariamente impresos se escanean (y posteditan) para crear archivos PDF.
- h) Los archivos PDF de documentos escaneados se convierten en archivos HTML para introducirse en el corpus de HTML simple.
- i) Los archivos originariamente en PDF se copian para incluirse en el corpus visual.
- j) Los archivos originariamente en HTML se copian para incluirse en el corpus visual.
- k) Los archivos PDF de documentos escaneados se copian para incluirse en el corpus visual.
- l) El grupo de investigación GENTT evalúa y modifica los criterios de clasificación del corpus documental. Estos criterios se incorporan al programa de etiquetaje del encabezado que se utiliza en el proceso *m*.
- m) Los archivos procedentes del corpus HTML simple se clasifican mediante el programa de etiquetaje del encabezado y se incorporan al corpus clasificado.
- n) El grupo GENTT se reúne para discutir acerca de diversas observaciones surgidas del procesamiento de los textos y, si procede, se repite el paso *l*.
- o) Se revisa el corpus clasificado.
- p) El grupo de investigación GENTT establece y modifica los criterios de notación para el corpus documental. Estos criterios toman forma en el programa de etiquetaje del cuerpo utilizado en el proceso *q*.
- q) Se etiquetan los archivos del corpus clasificado mediante el programa de etiquetaje del cuerpo y se incorporan al corpus anotado.
- r) El grupo GENTT se reúne para discutir acerca de diversas observaciones surgidas en el etiquetaje de los textos y, si procede, se repite el paso *q*.
- s) Se revisa el corpus anotado.
- t) Se analiza el corpus anotado mediante un módulo del programa Enciclopedia GENTT para producir una serie de bases de datos. (En estos momentos y puesto que todavía no se ha etiquetado el corpus, el módulo tan solo analiza el corpus clasificado.)
- u) El grupo GENTT se reúne para discutir acerca de diversas observaciones surgidas del uso del programa de la Enciclopedia GENTT.
- v) El grupo de investigación confecciona documentación adicional sobre los géneros trabajados y la incluye en el programa Enciclopedia GENTT.

Tabla 1. Breve explicación de los procesos indicados en el diagrama de flujo.

Paralelamente a la fase ya explicada, el grupo se reunió para establecer los criterios que deberían conformar esta descripción de los textos (*header markup*) que sería el primer paso a seguir tras su digitalización. La ficha resultante de

esta discusión, que fue modificándose en sucesivas sesiones de corrección, puede verse en la siguiente ilustración:

The screenshot shows a web-based form for cataloging documents. The form is titled "Afegeix o modifica dades" and contains the following sections:

- Metadata fields:**
 - Codi d'aquest arxiu: J0097C (with a "Veure usats >" button)
 - Creat per: A Calvo
 - Revisat per: M. Córdoba
 - Copyright: No resolts (dropdown menu)
- 1) En quina llengua està escrit aquest document?**
 - Buttons: Català, Espanyol, Anglès, Alemany, Una altra, Multilingüe
 - Radio buttons: Oriental, Occidental, No marcat
- 2) Aquest document és l'original?**
 - Buttons: Sí, No, És una traducció amb estatus d'original, No se sap
 - Text: Si, aquest document és l'original
- 3) Aquest document està relacionat amb algun altre del corpus?**
 - Buttons: Sí, No
 - Text: 1) Hi ha també un document amb codi J01651 i escrit en anglès (with dropdown menus for "traduït" and "anglès")
 - Button: Documents >
- 4) Quin n'és el títol?** Llei de marques
- 5) Qui n'és l'autor?** Parlament d'Andorra
- 6) Quina n'és la font?** http://www.spi-ad.com/
- 7) Quin n'és el mode?** escrit (dropdown menu)
- 8) Quin n'és el format original?** electrònic (dropdown menu)
- 9) On es publicà?** L'Estat espanyol (dropdown menu)
- 10) En quin any?** En 1995 (dropdown menu)
- 11) Quin n'és el tipus textual?** Text instructiu amb elements expositius (with a "Tipus >" button)
- 12) Seleccioneu-ne el gènere:** Jurídic, Normatiu, Lleis (with a "Gèneres >" button)
- 13) Seleccioneu-ne el camp temàtic:** K 7570-7582 Industrial property (with a "Camps >" button)
- 14) Afegeix-hi qualsevol comentari o observació:**
 - Buttons: català, espanyol, anglès
 - Text area for comments

At the bottom of the form are "Cancel·lar" and "Guardar" buttons.

Figura 2. Ficha utilizada para la catalogación de los textos mediante el etiquetado de su encabezado.

De los datos contenidos en esta ficha, explicaremos los que consideramos más relevantes o que han supuesto un grado mayor de dificultad de definición para el equipo. Estos campos son, por un lado, los relativos a la fuente y autor del documento y, por otro, la clasificación de géneros que se ha establecido.

En cuanto a la fuente y el autor, el debate que mantuvo el equipo se centró en determinar la diferencia de la persona/lugar de la que se obtenía el texto y quién lo había creado. Esto nos pareció especialmente trascendente para textos corporativos, especialmente en textos técnicos, puesto que si bien está muy claro quién proporciona el texto (Ford, por ejemplo), no siempre se hace constar quién lo ha creado. Por ello decidimos separar ambos campos y planteamos una serie de posibilidades para el caso del autor. Así, si es un autor conocido podemos explicitarlo o, por motivos de confidencialidad, consignar «Confidencial» junto con una mención a su profesión o estatus. Si, por el contrario, no conocemos al autor, utilizaremos la etiqueta «Desconocido». Respecto a la fuente, el debate fundamental se centró en los textos procedentes de la red. Finalmente, decidimos que la fuente sería la dirección *url* de dónde lo habíamos obtenido sólo cuando se tratase de textos específicamente creados para la red.

Por otro lado, se elaboró, conjuntamente con especialistas, una primera clasificación de géneros, en tanto que formas convencionalizadas de textos, en

función de su experiencia profesional. En la mayoría de los casos, nos dimos cuenta de que el profesional no era consciente de que el material con el que trabaja a diario puede constituir un conjunto coherentemente, y a veces jerárquicamente, relacionado. Por ello, fue necesario establecer sesiones conjuntas en las que el profesional explicaba su experiencia mientras nosotros tomábamos notas de los textos/géneros que iban apareciendo en su narración. En ocasiones fue necesario intervenir con preguntas para extraer más géneros. Finalmente, se estableció conjuntamente con los profesionales una primera clasificación que nos serviría de base para comenzar un trabajo de verificación empírica con métodos inductivos. Tras este trabajo, fue necesario modificar en algunos puntos las diferentes taxonomías. Como resultado de esta primera fase, contamos con algo más de 300 textos totalmente procesados a los que se han aplicado múltiples revisiones. No obstante, y a pesar de que hemos iniciado la segunda fase, este corpus continuará alimentándose regularmente con nuevos textos. Ya en estos momentos, estamos procesando más de 200 nuevos textos que, tras las revisiones pertinentes, se incorporarán al corpus en las próximas semanas.

5. Objetivos de la nueva etapa

Tras este trabajo y con el comienzo de un nuevo año, nos hemos planteado unos objetivos claros que hemos desglosado en las correspondientes actividades y tareas. Cada uno de los miembros del equipo, en razón a sus posibilidades, se encarga de una parte de las tareas de distinta naturaleza (de gestión, coordinación, investigación y producción) que, de acuerdo a nuestras expectativas, nos llevarán a conseguir en un plazo de tres años los objetivos que exponemos a continuación.

En primer lugar, pretendemos sentar y difundir las bases teóricas del proyecto mediante la publicación de los resultados en un volumen colectivo del equipo de investigación. Cada investigador, de acuerdo con el ámbito de trabajo realizado en el seno del grupo, se ha encargado de confeccionar una parte de un monográfico dedicado a explicar los avances y las posibilidades en cuanto a aspectos teóricos y metodológicos tratados en nuestro proyecto de la informática aplicada a la investigación con corpus, los géneros de la traducción técnica, de la traducción médica y de la traducción jurídica y administrativa.

Otro de nuestros objetivos relacionados con la difusión de nuestro trabajo es desarrollar una web en la que daremos a conocer nuestros propósitos y el trabajo realizado. Creemos también que esta plataforma nos servirá de ayuda para entrar en contacto con otros investigadores e iniciar un intercambio de información que pueda beneficiarnos a todos. Esta web está por el momento en construcción, pero puede consultarse ya en la siguiente dirección: www.gentt.uji.es.

En cuanto a objetivos de investigación, queremos seguir nutriendo nuestro corpus con nuevos textos procesados de los que podamos extraer las características genológicas con el fin de construir nuestro producto final, la enciclopedia de géneros para la traducción. Paralelamente, vamos a establecer cuáles serán las cuestiones que queremos analizar en nuestro corpus teniendo siempre pre-

sente que debemos dejar puertas abiertas para que los resultados nos puedan indicar nuevas oportunidades y objetivos de análisis. Por el momento, los factores a analizar incluyen la microestructura lingüística, con cuestiones como cohesión y coherencia (dentro de ellas, por ejemplo, terminología y fraseología), la macroestructura de los textos y, dentro de ésta, la progresión semántica, y los contextos de producción y recepción.

Tras conseguir resultados en relación a este propósito, el próximo paso será determinar el etiquetado necesario del cuerpo de los textos para poder analizarlos. Nuestra intención es limitar al máximo el número de etiquetas necesarias, puesto que gran parte de la información podrá extraerse, por ejemplo, con búsquedas simples de concordancias.

Al mismo tiempo, al trabajar con los textos iremos extrayendo datos que podrán dar lugar a la información que acompañará a los propios géneros en el producto final. Así, nos planteamos empezar a construir en este período algunas herramientas auxiliares como son glosarios, repertorios de fuentes de documentación o síntesis temáticas a modo de introducciones conceptuales a los temas tratados en los géneros susceptibles de traducción.

6. Perspectivas de futuro

Además de la herramienta que buscamos, estamos convencidos de que un corpus multilingüe de géneros especializados puede ser de gran utilidad como herramienta para otras investigaciones. En concreto, nos planteamos la introducción de traducciones en el corpus de investigación para determinar las similitudes y diferencias de los sistemas genológicos originarios y traducidos (Monzó 2002). A partir de aquí, podríamos construir redes de géneros en los que éstos se relacionen por su concurrencia en encargos de traducción o en los procesos profesionales, por su jerarquía formal o por otros criterios que puedan considerarse útiles. Otro trabajo posible sería observar en un sentido descriptivo las normas de traducción que imperan en nuestra sociedad o se podrían sancionar prácticas de calidad; asimismo, podrían construirse memorias de traducción de los distintos géneros trabajados y analizados en nuestra investigación que facilitarían en gran medida el trabajo de los profesionales; y, desde este trabajo, podrían confeccionarse diccionarios que ultrapasaran las concepciones tradicionales mediante el trabajo con concurrencias.

7. Conclusiones

En el presente trabajo hemos querido exponer cuáles son los objetivos del grupo GENTT y, especialmente, cuál será la utilidad que el producto final tendrá para el traductor profesional y el investigador. Para ello, hemos revisado someramente los fundamentos teóricos de nuestro proyecto y hemos querido hacer hincapié en algunas cuestiones terminológicas que nos permitiesen enunciar con exactitud el progreso de nuestro trabajo. Sin embargo, antes de entrar en

ese apartado, hemos querido dar a conocer los aspectos más relevantes de la metodología seguida y, a este respecto, hemos distinguido el corpus de investigación, construido con textos, del corpus/enciclopedia de géneros que constituirá el producto final. Posteriormente, hemos expuesto el trabajo desarrollado hasta el momento y, a continuación, el que prevemos llevar a cabo en el próximo período. Finalmente, hemos apuntado algunas posibilidades de investigación que se abren no solo para nuestro proyecto sino también para otros que pudieran estar relacionados.

Quisiéramos subrayar, para concluir, que nos sentimos muy esperanzados ante los beneficios este proyecto pueden proporcionar al profesional, el investigador y el docente de traducción. La primera etapa nos ha convencido de este hecho pero también nos ha hecho ver que todavía queda mucho por hacer. Esperamos que esta participación anime a quienes estén interesados en las aplicaciones de los corpus traductológicos a colaborar con nosotros a través de la web antes mencionada, www.gentt.uji.es.

Este trabajo se realiza en el seno del proyecto *Géneros textuales para la traducción*. Etiquetado de la microestructura textual de un corpus multilingüe comparable de los ámbitos de especialidad (*Ministerio de Ciencia y Tecnología, BFF2002-01932*).

Referencias

- BAJTÍN, M. M. 1986. The problem of speech genres. Trad. de V. W. MCGEE. En C. EMERSON y M. HOLQUIST, eds. *Speech Genres and Other Late Essays*. Austin: University of Texas Press, pp. 60-102.
- BAKER, M. 1992. *In Other Words: A Coursebook on Translation*. Londres: Routledge.
- . 1995. Corpus Linguistics and Translation Studies: Implications and applications. En M. BAKER *et al.*, eds. *Text and Technology: in Honour of John Sinclair*. Amsterdam: John Benjamins, pp. 233-250.
- BASSNETT, S. 1991. *Translation Studies*, Londres: Routledge.
- BORJA ALBI, A. 2000. *El texto jurídico inglés y su traducción al español*. Barcelona: Ariel.
- CASTELLÀ, J. M. 1995. Diversitat discursiva i gramàtica. Avantatges i mites de la tipologia textual. *Articles de Didàctica de la Llengua i de la Literatura*, 4, pp. 73-84.
- CORPAS PASTOR, G. 2001. Compilación de un corpus *ad hoc* para la enseñanza de la traducción inversa especializada. *TRANS*, 5, pp. 155-184.
- CRYSTAL, D., y D. DAVY. 1969/1979. *Investigating English Style*. Londres: Longman.
- DUNMIRE, P. L. 2000. Genre as temporally situated social action. *Written Communication*, 17 (1), pp. 93-138.
- EAGLES. 1996. *Preliminary Recommendation on Corpus Typology* [Documento Eagles (Expert Advisory Group on Language Engineering)].
- GAMERO PÉREZ, S. 2001. *La traducción de textos técnicos*. Barcelona: Ariel.
- GARCÍA IZQUIERDO, I. 1999. *Contraste lingüístico y traducción. La traducción de los géneros textuales*. [23], LYNX, Centro de Estudios de Comunicación Interlingüística e Intercultural. Valencia: Universidad de Valencia.
- . 2000. *Análisis textual aplicado a la traducción*. Valencia: Tirant lo Blanch.
- . (En prensa). El género: plataforma de confluencia de nociones fundamentales en didáctica de la traducción. *Discursos. Estudos de Tradução*, 2.

- GARCÍA IZQUIERDO, I. y V. MONTALT I RESSURECCIÓ. (En prensa). Translating into Textual Genres. *Linguistica Antverpiensia*, 1.
- GARCÍA IZQUIERDO, I. y E. MONZÓ NEBOT. (En prensa). Enciclopèdia electrònica de gèneres: Anàlisi i descripció dels gèneres especialitzats i aplicacions per a la traducció. ed. *Acta de la I Jornada de la xarxa Coneixement i Llenguatge*. València: Universitat de València.
- HALLIDAY, M. A. K. y R. HASAN. 1985. *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective*. Oxford: Oxford U. Press.
- HATIM, B. y I. MASON. 1990. *Discourse and the Translator*, 1ª ed. Londres: Longman.
- HURTADO ALBIR, A. 1995. La didáctica de la traducción. En E. LE BEL, ed. *Le masque et la plume. Traducir: reflexiones, experiencias y prácticas*. Sevilla: Servicio de Publicaciones de la Universidad de Sevilla, pp. 65-92.
- JENNINGS, S. (En preparación). *The Development of Software Tools for Corpus-Based Genre Research in Translation Studies: A Practical Application in the Context of the Gantt Project* [trabajo de investigación]. Castellón: Universidad Jaume I.
- JOHANSSON, S. y S. OKSEFIELL, eds. 1998. *Corpora and Cross-linguistic Research: Theory, Method and Case Studies*. Amsterdam: Rodopi.
- KRESS, G. 1985. *Linguistic Processes in Sociocultural Practice*. Victoria: Deakin U. Press.
- MARTIN, J. R. 1992. *Macro-Genres: The Ecology of the Page*, Department of Linguistics, Universidad de Sydney (Mimeo).
- MILLER, C. R. 1984. Genre as social action. *Quarterly Journal of Speech*, 70, pp. 151-167.
- MONZÓ NEBOT, E. 2001. Textos jurídics i traduccions: testimonis de coneixements i eines de formació per al traductor jurídic. *Revista de Llengua i Dret*. 36, p. 23-40.
- . 2002. *La professió del traductor jurídic i jurat. Descripció sociològica de la professió i anàlisi discursiva del transgènere* [tesi doctoral]. Universidad Jaume I.
- OPACKI, I. 1963/2000. Royal Genres. Trad. de D. MALCOLM]. En D. DUFF, ed. *Modern Genre Theory*. Londres: Longman, pp. 118-126.
- REIS, K. y H. J. VERMEER. 1989. *Grundlegung einer allgemeinen Translationstheorie*. 2ª ed. Tübinga: Niemeyer.
- SÁNCHEZ, A. 1995. *Cumbre, Corpus Lingüístico del Español Contemporáneo*. Madrid: SGEL.
- VAN TIEGHEM, P. 1938. La question des genres littéraires. *Hélicon*. 1, pp. 95-101.
- TORRUELLA, J. y J. LLISTERRI. 1999. Diseño de corpus textuales y orales. En J. M. BLECUA et al., eds. *Filología e informática. Nuevas tecnologías en los estudios filológicos*. Seminario de Filología e Informática, Departamento de Filología Española, Universidad Autónoma de Barcelona. Barcelona: Milenio, pp. 45-77.
- TROSBORG, A., ed. 1997. *Text Typology and Translation*. Amsterdam: John Benjamins.