

Corpus electrónico, género textual y traducción: metodología, concepto y ámbito de la Enciclopedia electrónica para traductores GENTT

Isabel García Izquierdo

Volume 50, numéro 4, décembre 2005

Pour une traductologie proactive — Actes
For a Proactive Translatology — Proceedings
Por una traductología proactiva — Actas

URI : <https://id.erudit.org/iderudit/019918ar>
DOI : <https://doi.org/10.7202/019918ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)
1492-1421 (numérique)

[Découvrir la revue](#)

Citer cet article

García Izquierdo, I. (2005). Corpus electrónico, género textual y traducción: metodología, concepto y ámbito de la Enciclopedia electrónica para traductores GENTT. *Meta*, 50(4). <https://doi.org/10.7202/019918ar>

Résumé de l'article

L'incorporation des nouvelles technologies au travail du traducteur et l'utilité qu'on peut en obtenir sont des faits constatés. Parmi les différents outils à disposition du traducteur, on peut remarquer les corpus textuels électroniques. Dans ce travail nous réfléchissons, en général, aux multiples possibilités que les corpus donnent au traducteur, en essayant d'établir une relation entre corpus, spécialité (médicale, technique et juridique) et genre textuel, comme méthodologie, domaine et point de départ conceptuel pour l'analyse. Et, plus concrètement, on applique ces réflexions au corpus textuel électronique GENTT.

Corpus electrónico, género textual y traducción: metodología, concepto y ámbito de la Enciclopedia electrónica para traductores GENTT¹

ISABEL GARCÍA IZQUIERDO
Universitat Jaume I, Castelló, Spain
igarcia@trad.uji.es

RÉSUMÉ

L'incorporation des nouvelles technologies au travail du traducteur et l'utilité qu'on peut en obtenir sont des faits constatés. Parmi les différents outils à disposition du traducteur, on peut remarquer les corpus textuels électroniques. Dans ce travail nous réfléchissons, en général, aux multiples possibilités que les corpus donnent au traducteur, en essayant d'établir une relation entre corpus, spécialité (médicale, technique et juridique) et genre textuel, comme méthodologie, domaine et point de départ conceptuel pour l'analyse. Et, plus concrètement, on applique ces réflexions au corpus textuel électronique GENTT.

ABSTRACT

The incorporation of new technologies for the translator's daily task is a fact. And the usefulness they may offer to that work is unquestionable. Among the tools that the translator has at his/her disposal one can enhance, in the last years, the electronic textual corpuses. This paper reflects, in general terms, on the possibilities that corpuses provide to the translator, establishing a relationship between the fields of specialty (medical, technical and legal) and the concept of textual genre, as methodology, context and point of conceptual departure for the respective analysis. In particular, these reflections are applied to the case of the GENTT electronic textual corpus.

MOTS-CLÉS/KEYWORDS

linguistics applied to translation studies, textual genre, textual corpuses, technologies of information and communication (TIC)

1. Introducción²:

La traducción, como el resto de actividades humanas, está en constante evolución. Así, si nos planteamos cómo están evolucionando los estudios sobre la traducción en los últimos años, la característica que de manera más significativa llama la atención es la incorporación de la tecnología en la tarea diaria del traductor y de los docentes y aprendices de la traducción. Más allá de la polémica surgida en los años 70-80 en torno a la traducción automática (véase, p.e. Bowker 2002), y que aún hoy en algunos círculos pervive, la opinión más extendida es la que aboga por la integración en la tarea del traductor de todos aquellos medios que puedan facilitarle el trabajo. Y uno de ellos son, de manera indiscutible, las TIC. En palabras de Samuelsson-Brown:

In fact, the translator is becoming more and more dependent on information technology and, if the translator does not adapt to change, he or she may become uncompetitive (1996: 280).

Esta afirmación, con la que estoy de acuerdo, resulta ocho años después de actualidad absoluta, puesto que la realidad tanto de la docencia como del mercado profesional nos confirma que la actividad traductora se está vinculando de manera indisoluble con las nuevas tecnologías. Como afirman Bernardini et al. (2003: 3): “[...] *translators and other language service providers have access to a wide range of computational resources*”. Pero, como en otros ámbitos, también aquí se ha producido el *Translator's Turn* y la relación entre las TIC y la traducción no es unidireccional:

si las nuevas tecnologías (o tecnologías de la información) están aportando avances significativos a la actividad traductora, ésta también está proporcionando nuevos campos de análisis para la tecnología, de manera que se produce una retroalimentación entre ambas.

Ahora bien, las llamadas tecnologías de la información y la comunicación abarcan un ámbito de análisis tan amplio que no sería adecuado realizar generalizaciones sobre su utilidad para la traducción. Por ello, en el próximo apartado intentaré delimitar los aspectos de las nuevas tecnologías que ocuparán mi reflexión.

2. Los corpus textuales en traducción

Como acabamos de mencionar, son muchas las herramientas que podemos incluir bajo la denominación de *nuevas tecnologías*. Como afirma Bowker (2002: 4), entre los diferentes tipos de tecnología utilizados por el traductor, hay que distinguir entre la llamada **traducción automática** (TA, *Machine Translation*), en la que de algún modo se sugiere el reemplazar la tarea humana por las máquinas; la **traducción (humana) asistida por ordenador**, conocida como CAT (TAO, *Computer-Aided Translation*), en la que el traductor es el responsable de realizar la traducción pero utiliza una serie de herramientas computerizadas que le ayudan a completar su tarea y aumentan su nivel de productividad. Y la **Traducción Humana** (TH), referida a las herramientas que forman parte del conocimiento general del traductor (como de otros profesionales), como el uso de procesadores de texto, la www, el correo electrónico, etc.

Ahora bien, incluso si nos situamos en el nivel de la tecnología CAT, deberemos realizar distinciones entre diferentes metodologías y herramientas. Entre los diferentes tipos de herramientas que podemos incluir en las CAT destacan: herramientas de captura de datos, memorias de traducción, herramientas de traducción de páginas web y localización, sistemas de administración de terminología o herramientas de análisis de corpus. Todas ellas de mucha utilidad para el traductor, pero con finalidades y usos diferentes.

Me centraré, por tanto, únicamente, en el análisis de corpus que, como ya he apuntado en líneas superiores, se está imponiendo como herramienta fundamental para el análisis de los textos originales y traducidos (véanse los trabajos de Baker 2003 ; Kenny 2001 ; Zanettin et al. 2003, o Olohan (2004), Alcina (e.p), entre otros). Siguiendo la definición propuesta por EAGLES (*Expert Advisory Group on Language Engineering Standards*) 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*. (1996: 4).

L. Bowker lo plantea en términos muy similares:

[...] a corpus is not simply a random collection of texts. Rather, the texts are selected according to explicit criteria in order to be used as a representative sample of a particular language or subset of that language (2002: 45).

Y R. Luna, se atreve con una definición específica del corpus para el ámbito de la traducción y la interpretación. Así, propone la siguiente definición de *corpus traductor*:

Conjunto de enunciados producidos en un determinado código con sus versiones o variantes de traducción en otro(s) código(s) que serán objeto de análisis, desde una determinada perspectiva teórica, con el fin de extraer constantes o variantes respecto del comportamiento interlingüístico (2002: 59).

En mi opinión, no obstante, la definición de Luna alude claramente a los corpus paralelos. 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 hacer listas lematizadas de palabras, listas de términos que queremos ignorar, analizar los usos contextuales de las palabras, a través de los programas de concordancias (monolingües o bilingües); las funciones, las colocaciones, los préstamos en un género textual concreto; las redundancias en los textos; la variación sintáctica o la fraseología específica de un género (**anotación**); la complejidad de las oraciones en diferentes tipos textuales, la macroestructura característica del género (a través de la

marcación del corpus), etc. 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 variada³. En I. García Izquierdo y E. Monzó (2003b), repasamos la propuesta de tipologías de G. Corpas (2001: 3 ss.) quien, a su vez, utiliza la de otros autores como Baker (1995). Según estas propuestas, podemos distinguir los siguientes tipos de corpus:

- 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 avanzar que el corpus GENTT que estamos recopilando, del que enseguida me ocuparé, se definiría como un corpus **comparable** (puesto que en él se incluyen textos originales, equivalentes, paralelos y traducciones en las cuatro lenguas de trabajo en cuatro lenguas: español, catalán, inglés y alemán), **especializado** (porque incluye diferentes ámbitos de especialidad), **textual** (con textos completos), **anotado** (etiquetado lingüística o metatextualmente) y **documentado** (cada documento textual lleva asociado una cabecera o *Header*, descriptiva de su procedencia). No obstante, como veremos, posee algunas características que lo hacen diferente de otros corpora comparables. A ello podríamos añadir que se trata de un corpus **dinámico** (Olohan 2004: 45) puesto que recoge una gran cantidad de textos que se podrá ir actualizando constantemente.

Las ventajas que la utilización del corpus pueden tener para la docencia y la investigación han sido puestas de manifiesto en numerosas ocasiones.

Applied corpus-based translations studies are finally becoming established as a framework for research and pedagogy (Bernardini et al., 2003: 2).

Prueba de ello son los volúmenes que últimamente se han dedicado precisamente a analizar la importancia y utilidad de la utilización de corpora y, en general, de las CAT (*Computer Aided Translation*) en la traducción (Bowker (2002), Zanettin et al. (2003), Olohan (2004), entre otros). Ahora bien, como afirma K. Malmkjaer (2003), el uso irreflexivo de los corpora puede resultar muy peligroso y, en ocasiones, puede ser preferible romper una norma antes que utilizarla. Por ello, es fundamental que los corpora sean herramientas de apoyo que puedan utilizarse de manera flexible, en función de las necesidades del usuario. Y esto sólo será posible si los criterios de construcción del corpus han sido minuciosamente valorados y el resultado final es una herramienta que permite al traductor explotar diferentes posibilidades y gozar de mayor autonomía y flexibilidad. En opinión de autores como Aston (1999), si bien es cierto que el corpus facilita la tarea del traductor, metodológicamente puede tener sus límites como herramienta y, en ocasiones, llevará al traductor/traductólogo a problematizar, más que a simplificar su tarea, actuando como elemento de pensamiento-provocación frente a determinadas cuestiones.

Por último, como afirma M. Olohan (2000: 2) –citando a A. Chesterman (2000)–, existen diferentes *modelos de traducción* (modelo comparativo, modelo procesual y modelo causal) y la elección de cada uno de estos modelos puede tener implicaciones para el modelo de investigación escogido, la metodología utilizada para investigarlo y la subsiguiente construcción de las teorías de traducción. Así, el caso de los corpora, se incluiría en los llamados *modelos comparativos* (véase también Olohan 2004: 9), cuya principal finalidad es describir los procesos y los productos de

traducción o, lo que es lo mismo, descubrir las correlaciones entre las características de los sistemas lingüísticos, los textos o conjunto de textos.

En general, en mi opinión, las ventajas de la utilización de corpora son, hoy por hoy, mucho mayores que sus posibles inconvenientes, si bien creo que hay que insistir en la importancia fundamental de una buena construcción de los mismos (tanto desde el punto de vista teórico o conceptual como práctico).

3. El concepto de género y su relación con la traducción

A partir de los 90, diferentes aproximaciones empíricas ponen el énfasis en la detección y análisis de las características sistemáticas de las traducciones. Y la lingüística de corpus les proporciona la herramienta adecuada, como metodología que focaliza en la identificación de patrones recurrentes de comportamiento lingüístico (véase, p.e. Baker 1993).

Partiendo de estas consideraciones, el equipo de investigación GENTT amplía esta perspectiva y se interesa no sólo por las características sistemáticas de las traducciones, sino también de los originales y su interrelación. Metodológicamente, por las razones ya explicadas, teníamos muy claro que la utilización del corpus era la opción más adecuada. Pero ¿de dónde debíamos partir en nuestra investigación para dar cabida tanto a textos originales como a textos paralelos y posibles traducciones en diversas lenguas e ir más allá de la delimitación de patrones de comportamiento lingüístico? No queríamos caer en el error de compilar una gran cantidad de textos que no tuviesen otro valor que el cuantitativo porque, como afirma M. Olohan (2004: 22), siguiendo a autores como I. Mason (2001) o T. Hermans (1999): *The need of contextualize and co-textualize translation and our study of it means that purely quantitative studies of corpus data are regarded as limited in their usefulness.*

No obstante, reconoce el interés que, para algunos propósitos, los estudios con corpus pueden tener y cita a L. Malmkjaer, para quien:

The use in translation studies of methodologies inspired by corpus linguistics has proved to be one of the most important gate-openers to progress in the discipline since Toury's re-thinking of the concept of equivalence (2003: 119).

Esa necesidad de contextualización hizo que *a priori*, como se afirma en I. García Izquierdo y Monzó (2003b: 32), el *texto* fuese la unidad que parecía más adecuada para comenzar el análisis, como componente fundamental de la situación profesional de la traducción. Sin embargo, enseguida advertimos que un texto es sólo una ocurrencia aislada, que nos puede ayudar a delimitar cuestiones fundamentalmente lingüísticas, pero no a prevenir dificultades futuras y sus posibles soluciones, mientras que el género se define en GENTT como una interfaz entre el texto de partida y el texto de llegada y representa el recipiente en el que convergen todos los aspectos lingüísticos y culturales de las comunidades en contacto (I. García Izquierdo y V. Montalt 2002; I. García Izquierdo, e.p.; V. Montalt 2003). Por ello, nuestro interés se centró, en primer lugar, en demostrar la utilidad del concepto de *género textual* como punto de partida conceptual de la investigación.

El trabajo, pues, se fundamenta en el concepto de **género** en los ámbitos de comunicación especializada y pretende desarrollar las aplicaciones de la noción al ámbito de la traducción con la finalidad de crear 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 se plantea una herramienta de consulta textual, conceptual, terminológica y lingüística, como veremos.

Como ya se ha puesto de manifiesto en numerosas publicaciones de los miembros del equipo (véanse, entre otros, García Izquierdo y V. Montalt (2002), García Izquierdo y E. Monzó (2003 a, b y c), García Izquierdo (e.p.b) o V. Montalt (2003)), del concepto de género se ha ocupado ampliamente la llamada *teoría del género* en ámbitos como la teoría de la literatura y la lingüística. A partir del giro sociológico que nos aporta Bakhtin (véase, por ejemplo, 1978/1986), el estudio del género incorpora un elemento sociológico muy importante del que rinden cuenta en la lingüística estudios como los de Halliday y Hasan (1985), Swales (1990) o Bhatia (1993), pero

también en retórica (Miller 1984; Dunmire 2000), estilística (Crystal y Davy, 1969/1979) o comunicación audiovisual (Kress 1985)⁴. En estos estudios, los textos se entienden como productos semióticos (con lo que la convergencia con los intereses de la Traductología resulta evidente, desde el momento en que la traducción se entiende como operación intertextual e intercultural). Pero, además de productos semióticos, los géneros de los ámbitos de especialidad son representaciones de las unidades de comunicación de cada comunidad discursiva. Constituyen una síntesis de situaciones y manifestaciones discursivas prototípicas que constituyen una herencia aprendida por los agentes y un medio de aceptación por parte de la comunidad. En otras palabras, los géneros, como ha demostrado E. Monzó en algunos de sus trabajos, son medios de *socialización* del traductor, ya que a través de su conocimiento el traductor logrará que su acción social en el ámbito de especialidad de que se trate sea efectiva.

El género, pues, es un concepto semiótico, porque representa un signo para cada cultura. Pero no es sólo un concepto semiótico, porque la semiótica incide en el producto acabado y el género no es estático. Es una categoría dinámica e híbrida, y de ahí la dificultad de establecer los límites entre géneros, en especial en los ámbitos de especialidad (véase I. García Izquierdo, 2002). Ahora bien, las recientes investigaciones en el seno del proyecto GENTT nos han llevado a confirmar la posibilidad de establecer taxonomías de los géneros, más o menos permanentes, que ayudan a describir determinados aspectos relacionados con las macroestructura y microestructura textuales y que permiten un análisis sistematizado de la información contenida en los mismos, con aplicaciones útiles, sobre todo desde el punto de vista pedagógico.

La importancia de la detección de las convenciones en las diferentes manifestaciones textuales de las lenguas ha sido puesta de manifiesto en numerosas ocasiones (Reiß y Vermeer (1989/1991), Hatim y Mason (1990), Baker (1992), Hurtado Albir (1995) o Trosborg (1997) o García Izquierdo (1999, 2000, 2002)). Y, desde nuestro punto de vista, el género es la categoría que puede ayudar de un modo más efectivo a sistematizar dichas convenciones.

Así pues, comprobada la dificultad del traductor especializado a la hora de hacer explícitas las convenciones de los géneros de especialidad en diferentes lenguas, nos adentramos en el proyecto convencidos de la necesidad de un estudio sistemático de los diferentes géneros como paso previo a la traducción.

4. El proyecto GENTT

Textual Genres for Translation (GENTT). Annotation of textual microstructure of a multilingual comparable corpus in specialized fields

El equipo GENTT pretende desarrollar una completa herramienta auxiliar que permita obtener, en poco tiempo y con un esfuerzo reducido, la información que un traductor profesional consigue interiorizar tras largas horas de consulta, documentación, estudio y traducción. El equipo pretende reunir en un único lugar, de manera sistematizada, una información que, en la actualidad, en la mayoría de casos, se encuentra dispersa. Con estas premisas, los objetivos iniciales del proyecto fueron:

1. Proporcionar a los traductores profesionales de tres ámbitos de especialidad (traductores de textos jurídicos, técnicos y médicos) una herramienta que les permita reducir considerablemente el tiempo dedicado a la documentación.
2. Producir una Enciclopedia de géneros en la que puedan encontrarse instancias reales de los documentos que deben traducirse, junto con datos sobre sus contextos de cultura y situación, su desarrollo semántico y textual, etc. y la relación del género en cuestión con otros disponibles en el corpus, ya sean originales o también traducciones.
3. Conseguir mejoras para la traducción en el ámbito formativo (utilización de los textos del corpus *per se*), pero también en la práctica profesional de la traducción especializada, para la que plantearemos una herramienta de consulta textual, conceptual, terminológica y lingüística.

Establecidos, pues, cuáles eran los objetivos generales del proyecto, el proyecto se ha desarrollado en las tres fases claramente diferenciadas que suelen sucederse en el análisis con corpus (véase R. Luna 2002):

1. **Observación** (establecimiento de los criterios de selección) Se procedió a establecer los criterios para la compilación de ese corpus multilingüe de géneros de especialidad (jurídico, técnico y médico), con textos originales, paralelos y, en un futuro, traducidos, de evidente repercusión práctica tanto para la docencia (a modo de memoria de traducción), como para la investigación aplicada; y la comprobación de la pertinencia del enfoque teórico –géneros textuales–, en colaboración con los profesionales de los distintos campos estudiados.
2. **Constitución** (utilización de las técnicas de eliminación o extracción de información no pertinente y utilización de criterios de selección de textos basados en la exhaustividad, la homogeneidad y la representatividad). En esta segunda fase, el equipo advirtió la necesidad de contar con un programa de etiquetado para el *Header* o cabecera de los textos, que ayudase a la clasificación inicial. Así, se diseñó un programa de etiquetado en SGML, que se ha podido aplicar mediante el diseño de una herramienta informática específica, diseñada con *Visual Basic* (programa Gentt. Véase Anexo 1. fig 1).
3. **Análisis:** en la tercera fase, en la que nos encontramos en estos momentos (con un total aproximado de 450 textos revisados), el cometido es el desarrollo del corpus textual, mediante la ampliación del número de textos por especialidad y la definición y diseño de etiquetas, contempladas en el sistema TEILite, que permitan etiquetar ya no el *Header*, como se ha hecho hasta el momento, sino el cuerpo mismo del texto (*body markup*) (macroestructura y microestructura: aspectos de cohesión y coherencia, fraseología específica, terminología, expresiones referenciales, etc.). Para ello contamos con las herramientas de tratamiento de corpus ya adquiridas (programas de gestión, programas de concordancias, memorias -como el Wordsmith Tools 3.0, o el Wordbench–TRADOS–, herramientas de obtención de datos, como el *Adobe Acrobat writer*, o de diseño de aplicaciones, como el *Visual Basic*), si bien nos estamos planteando la posibilidad de desarrollar software propio integrado en las herramientas del programa GENTT.

Llegados a este punto, en mi opinión la labor realizada ha sido muy fructífera y, a pesar de que todavía falta trabajo por hacer, creo que el material almacenado y las reflexiones que han surgido en el seno del grupo hasta el momento ya son un buen resultado y pueden proporcionar múltiples **ventajas**. De hecho, en estos momentos el equipo está confeccionando lo que será una primera versión del Diccionario Enciclopédico, en la que se incluirá la definición de cada uno de los géneros de los diferentes ámbitos (por orden alfabético), atendiendo fundamentalmente a parámetros relacionados con la situación comunicativa, el contexto socio-profesional de producción; aspectos de macroestructura, contrastividad y correferencias entre los diferentes géneros. En efecto, esta primera versión ofrecerá la posibilidad al lector de acceder al corpus electrónico a través de una entrada codificada, y mediante la cumplimentación de una ficha de acceso, en la que se comprometa a utilizar los datos únicamente con fines investigadores y de consulta.

Ahora bien, no sería honesto no aludir a las **dificultades** con las que nos hemos encontrado en la investigación. Señalaré, a modo de ejemplo, algunas de las más significativas:

1. Como he señalado más arriba, a pesar de las dificultades que entraña una clasificación de los géneros de especialidad, las investigaciones del equipo GENTT nos llevaron al convencimiento de que es posible establecer taxonomías, más o menos permanentes, que ayudan a describir determinados aspectos relacionados con las macroestructura y microestructura textuales y que permiten un análisis organizado de la información contenida en los mismos. De hecho, como enseguida veremos, se ha establecido un listado de géneros y subgéneros para cada uno de los tres ámbitos de trabajo, que ya ha demostrado su utilidad para la programación de asignaturas de traducción, y no sólo de nuestro equipo (Véase Anexo 1. fig. 2) En varias ocasiones nos hemos planteado el posible componente *apriorístico* de nuestra investigación, puesto que establecemos una taxonomía al mismo

tiempo que recolectamos el corpus. Como afirma M. Olohan (2004: 47): *The subjectivity of decisions about what material is included or excluded from a corpus is no less problematic in corpus-based studies of translation*. Y esto tiene que ver tanto con la clasificación misma de los géneros como con la elección de los textos que van a conformar el corpus y van, por tanto, a validar su *representatividad*. Pero pensamos que, como participantes expertos de la comunidad lingüística y cultural en la que estos géneros se producen, y con la colaboración constante de los expertos en el campo de especialidad, estamos en condiciones de realizar esa primera clasificación que, obviamente, únicamente podremos confirmar tras el análisis pormenorizado del corpus.

2. El proceso de recolección de textos y el diseño de la ficha que constituye el etiquetaje del *Header* (véase Anexo 1. fig 3), se explican con detalle En I. García Izquierdo y E. Monzó (2003b). A pesar de la aparente simplicidad de la ficha de etiquetado, que recoge datos como autor, fuente, año de publicación, si se trata de una traducción o no, tipología textual, campo o género, entre otros, el equipo se ha encontrado con muchas dificultades y largas horas de discusión para delimitar qué elementos eran pertinentes para el etiquetado y, sobre todo, cómo plasmar la información referente a cada uno de esos elementos (especialmente interesante ha sido el problema de la información sobre fuente y autor y el caso de las tipologías textuales). En el momento de comenzar nuestra investigación, no conocíamos la existencia de corpus que pretendiesen ofrecer tanta información metadocumental en el *Header*, y este enfoque novedoso ha conllevado algunas complicaciones. Nos hemos encontrado, por tanto, ante lo que R. Luna llama *serendipia*, o:

flexibilidad cognoscitiva que debe adoptar el investigador a lo largo de su trabajo. Recuérdese que, en su condición de sujeto observador, el analista debe prestar atención a eventos accidentales que pudieran presentarse y que lo harán reflexionar sobre la validez del problema o hipótesis propuestos. En suma, el investigador debe tener el coraje suficiente para desechar el problema e hipótesis que consideraba relevantes, si durante la recolección de datos de la fuente surgieran aspectos de análisis más fructíferos y relevantes (2002: 67).

Y, como consecuencia, la ficha ha ido sufriendo diversas modificaciones, si bien en estos momentos estamos frente a una versión bastante pulida y, por tanto, cuasi definitiva de la misma, como se podrá comprobar en el acceso al corpus electrónico a partir del Diccionario Enciclopédico.

3. Respecto al corpus, si bien siguiendo las clasificaciones más extendidas hemos visto que podríamos clasificarlo como **comparable** (puesto que en él se incluyen textos originales, equivalentes, paralelos y traducciones en las cuatro lenguas de trabajo: español, catalán, inglés y alemán), **especializado** (porque incluye diferentes ámbitos de especialidad), **textual** (con textos completos), **anotado** (etiquetado lingüística o metatextualmente) y **documentado** (cada documento textual lleva asociado una cabecera o *Header*, descriptiva de su procedencia), la realidad es que posee características específicas que podrían llevarnos a hablar de un corpus combinado, entre el corpus comparable y el paralelo (véase a este respecto Olohan 2004: 43), puesto que, a pesar de no incluir de momento traducciones como tales, sí incluye ejemplares de género que se publican simultáneamente en varias lenguas (p.e. catálogos del ámbito técnico o textos jurídicos y administrativos de territorios con cooficialidad lingüística). Y si bien se les otorga a todos ellos estatus de original (puesto que así funcionan en su contexto sociocultural), en muchas ocasiones hay evidencias que corroboran que se trata de traducciones.
4. Por último, un problema todavía sin resolver con el que nos hemos encontrado ha sido el del *copyright*. Como se sabe, los textos están protegidos por la legislación sobre la propiedad intelectual y únicamente es posible utilizar extractos de los mismos, a no ser que la utilización sea únicamente con fines de investigación o docentes y no persiga, por tanto, un fin lucrativo. El hecho de estar ante un corpus que pretende mostrar géneros, y no ejemplares textuales aislados, ha supuesto que, en la práctica, nuestro corpus no esté accesible. En estos momentos estamos diseñando lo que será la primera versión del

Diccionario enciclopédico, como hemos visto, y hemos previsto la posibilidad de dar acceso al corpus electrónico a través de una web, con entrada restringida a investigadores que se comprometan a utilizarlo únicamente con los fines apuntados. No obstante, y para poder cumplir con algunos de los objetivos del proyecto, la finalidad última sería poder publicarlo para que pudiesen también utilizarlo los traductores profesionales y los profesionales de los ámbitos jurídico, técnico y médico.

Con todo, me gustaría terminar con un mensaje optimista respecto a las posibilidades del corpus GENTT que estamos construyendo y les emplazo a colaborar con nosotros a través de la página de sugerencias que encontrarán en la dirección <http://www.gentt.uji.es/>

NOTAS

1. El proyecto “Géneros Textuales para la traducción (GENTT). Etiquetado de la microestructura textual de un corpus multilingüe comparable de los ámbitos de especialidad.” (BFF202-01932), dirigido desde el año 2000 por la profesora Isabel García Izquierdo, de la Universitat Jaume I, está financiado por el Ministerio de Educación y Ciencia.
2. Este artículo es una versión revisada y ampliada de la conferencia “Estudios recientes en traductología basados en corpus: el corpus electrónico GENTT”, presentada al VII Congresso de tradução científico técnica em Língua portuguesa, organizado por la *Unión Latina* y celebrado en Lisboa en noviembre de 2004.
3. Véanse a este respecto los trabajos de Bowker (2002: 45 ss.), R. Luna (2002: 58 ss.) o Bernardini et al. (2003: 5 ss.).
4. Para una excelente revisión de estos aspectos, véanse los trabajos de E. Monzó, en especial E. Monzó (2001).

BIBLIOGRAFÍA BÁSICA

- ALCINA CAUDET, A. (e.p.): “La implementación del concepto de género textual en los corpus electrónicos para traductores”, en I. GARCÍA IZQUIERDO (ed.) (e.p.).
- ÁLVAREZ LUGRIS, A. (2001): “TECTRA: Teoría y práctica de las investigaciones con corpus en el marco de los Estudios de Traducción”, *Trans* 5, pp. 185-204.
- ANDERMAN, G. y M. ROGERS (1999): *Word, Text, Translation: Liber Amicorum for Peter Newmark*, Clevedon, Multilingual Matters.
- ASTON, G. (1999): “Corpus Use and Learning to Translate”, en S. BASSNETT, R. M. BOLLETTIERI y M. ULRICH (eds.): *Translation Studies Revisited, Textus 12-2*, Genova, Tilgher, pp. 289-314.
- AUSTERMÜHL, F. (2001): *Electronic Tools for Translators*, Manchester, St. Jerome.
- BAKER, M. (1993): “Corpus Linguistics and Translation Studies: Implications and Applications”, en M. BAKER, G. FRANCIS and E. TOGNINI-BONELLI (ed.): *Text and Technology: in Honour of John Sinclair*, Amsterdam-Filadelfia, John Benjamins, pp. 233-250.
- (1995): “Corpora in Translation Studies: An Overview and Some Suggestions for Future Research”, *Target* 7-2, pp. 223-243.
- (1996): “Corpus-based translation studies: the challenges that lie ahead” in H. SOMERS (ed.): *Terminology, LSP and Translation. Studies in Language Engineering in Honour of Juan Sager*, Amsterdam/Philadelphia, John Benjamins.
- BERNARDINI, S. et al. (2003): «Corpora in Translator Education», en ZANETTIN et al. (eds.): *Corpora in Translator Education*, Manchester, St. Jerome, pp. 1-13.
- BIBER, D. (1994): “Corpus-based Approach to Issues in Applied Linguistics”, *Applied Linguistics* 15-2, pp. 169-189.
- BATHIA, V. K. (1993): *Analysing Genre. Language Use in Professional Settings*, Essex, Longman.
- BORJA ALBI, A. (2000): *El texto jurídico inglés y su traducción al español*, Barcelona, Ariel.
- BOWKER, L. (1999): “The Design and Development of a Corpus-based Aid for assessing Translations”, *Teanga* 18.
- (2000): “A Corpus-Based Approach to Evaluating Student Translations”, *The Translator* 6-2, pp. 183-210.
- (2002): *Computer-Aided Translation Technology. A Practical Introduction*, Didactic of Translation Series, Ottawa, University of Ottawa Press.

- C. VALERO EL DE LA CRUZ CABANILLAS (eds.) (2001): *Traducción y Nuevas Tecnologías. Herramientas auxiliares del traductor*, Universidad de Alcalá, Servicio de Publicaciones.
- CABRÉ CASTELLVÍ, M. T. (1999): "Fuentes de información terminológica para el traductor", en M. PINTO Y J. A. CORDÓN (eds.), pp. 19-40.
- CORPAS PASTOR, G. (2001): "Compilación de un corpus ad hoc para la enseñanza de la traducción inversa especializada", *Trans* 5, pp. 155-184.
- DUNMIRE, P. L. (2000): "Genre as Temporally Situated Social Action. A Study of Temporality and Genre Activity", *Written Communication* 17-1, pp. 93-138.
- EAGLES (1996): "Preliminary Recommendations on Corpus Typology", Documento Eagles (*Expert Advisory Group on Language Engineering*)
- EBELING, J. (1998): "Contrastive Linguistics, Translation and Parallel Corpora", *Meta* 43-4, p. 602-615.
- 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*, Lynx 23.
- GARCÍA IZQUIERDO, I. (2000a): *Análisis textual aplicado a la traducción*, Valencia, Tirant lo Blanch.
- GARCÍA IZQUIERDO, I. (2000b): "The Concept of Texttype and its Relevance to Translator Training", *Target* 12-2, pp. 283-295.
- GARCÍA IZQUIERDO, I. (2002): "El género: plataforma de confluencia de nociones fundamentales en didáctica de la traducción", *Revista Discursos. Série Estudos de Tradução* 2, Universidade Aberta, Lisboa.
- GARCÍA IZQUIERDO, I. (2005): "Traducción", en A. López García (ed.): *Lenguaje y vida*, Valencia, Servei de Publicacions de la Universitat de València.
- GARCÍA IZQUIERDO, I. (ed.) (e.p.): *El género textual y la traducción. Reflexiones teóricas y aplicaciones pedagógicas*, Berna, Ed. Peter Lang.
- GARCÍA IZQUIERDO, I. (e.p.a): "El género y la lengua propia: el español de especialidad" en GARCÍA IZQUIERDO (ed.): *El género textual y la traducción. Reflexiones teóricas y aplicaciones pedagógicas*, Berna, Ed. Peter Lang.
- GARCÍA IZQUIERDO, I. et al. (e.p.b): "Enciclopèdia electrònica de gèneres. Anàlisi i descripció dels gèneres especialitzats i aplicacions a la traducció", *II Jornada Xarxa temàtica llenguatges d'especialitat*, setembre 2001, València, Universitat de València.
- GARCÍA IZQUIERDO, I. y E. MONZÓ (2003a): "Enciclopedia electrónica de géneros para la traducción", A. ALCINA y S. GAMERO (eds.): *La traducción científico-técnica y la terminología en la sociedad de la información*, Castellón, Servei de Publicacions de la Universitat Jaume I.
- GARCÍA IZQUIERDO, I. y E. MONZÓ (2003b): "Corpus de géneros GENTT. Una enciclopedia para traductores", E. SÁNCHEZ TRIGO y O. DÍAZ FOUQUES (eds.): *Traducción & Comunicación*, vol. 4, Servicio de Publicacións Universidade de Vigo, pp. 31-55.
- GARCÍA IZQUIERDO, I. y E. MONZÓ (2003c): "Una enciclopedia para traductores. Los géneros de especialidad como herramienta privilegiada del traductor profesional", R. MUÑOZ MARTÍN (ed.): *Actas del I Congreso Internacional de AIETI*, Granada, Universidad de Granada, pp. 83-97.
- GARCÍA IZQUIERDO, I. y E. MONZÓ (2004): "Traducir con corpus de géneros", *Revista de la Facultad de Lenguas Modernas* 7, Universidad Ricardo de Palma, Lima, Perú, pp. 45 ss.
- GARCÍA IZQUIERDO, I. y V. MONTALT (2002a): "Translating into Textual Genres", *Lingüística Antverpiensia* 1, Bruselas.
- GARCÍA IZQUIERDO, I. y V. MONTALT (2002b): "Multilingual, Corpus-based Research of Medical Genres for Translation Purposes: The Medical Corpus of the GENTT Project", *Actes del Congrés "traduir ciència"*, Universitat Pompeu Fabra.
- GARCÍA PALACIOS, J. (1996): "En los límites de la especialidad: los textos de divulgación científica", en BARGALLÓ et al. *Las lenguas de especialidad y su didáctica*, Tarragona, Universitat Rovira i Virgili, pp. 157-168.
- HALLIDAY, M. A. K. and R. HASAN (1985): *Language Context and Text. Aspects of Language in a Social-semiotic Perspective*, Victoria, Deakin University Press.
- HATIM, B. y I. MASON (1990/1995): *Teoría de la traducción. Una aproximación al discurso*, Barcelona, Ariel. Traducción de Salvador Peña.
- HURTADO ALBIR, A. (2001): *Traducción y traductología. Introducción a la traductología*, Madrid, Cátedra.
- HYLAND, K. (2000): *Disciplinary Discourses: Social Interactions in Academic Writing*, Harlow, U.K, Pearson Education.
- JOHANSSON, S. and S. OKSEFIELL (eds.) (1998): *Corpora and Cross-linguistic Research: Theory, Method and Case Studies*, Amsterdam/Atlanta, Rodopi.
- KENNY, D. (2001): *Lexis and Creativity in Translation. A corpus-based Study*, Manchester, St. Jerome.
- KRESS, G. (1985): *Linguistic Processes in Sociocultural Practice*, Victoria, Deakin University Press.

- LAVIOSA, S. (ed.) (1998): *L'approche basée sur le corpus / The corpus-based approach*, *Meta* 43-4.
- LEE, D. Y. (2001): "Genres, Registers, text Types, Domains and Styles: Clarifying the Concepts and navigating a Path Trough the BNC Jungle", *Language Learning and Technology* 5, pp. 37-72.
- LINDQUIST, H. (1999): "Electronic Corpora as Tools for Translation" in G. ANDERMAN and M. ROGERS (eds.), pp. 179-189.
- LÓPEZ CHOLLET, M. J. (2001): "Utilización de corpus en la formación del traductor", en C. VALERO E I. DE LA CRUZ CABANILLAS (eds.), pp. 317-325.
- LUNA, R. (2002): *Temas de traducción*, UNIFE, Lima (Perú).
- MALMJAER, K. (2003): "On a Pseudo-subversive Use of Corpora in Translator Training", in ZANETTIN *et al.* (eds.): *Corpora in Translator Education*, Manchester, St. Jerome, pp. 119-134.
- MONTALT, V. (2003): "El gènere textual com a interfície pedagògica en la docència de la traducció científicotècnica", en R. SAMSON (ed.): *Actes de les VII Jornades de Traducció a Vic*, Vic, Eumo Editorial.
- MONZÓ I NEBOT, E. (2001): *El concepte de gènere discursiu. Una aplicació a la traducció jurídica*, Trabajo de investigación, Castellón, Universitat Jaume I.
- MONZÓ I NEBOT, E. (2003): "Las socializaciones del traductor especializado: el papel de los géneros". *Revista de la Facultad de Lenguas Modernas, Universidad Ricardo Palma (Lima-Perú)* 6, pp. 15-29.
- NORD, C. (1997): *Translation as a Purposeful Activity*, Manchester, St. Jerome.
- OLOHAN, M. (ed.) (2000): *Intercultural Faultlines, Research Models in Translation Studies 1. Textual and Cognitive Aspects*, Manchester, St. Jerome.
- OLOHAN, M. (2000): "Shifts, But Not as We Know Them? Research Models and Methods in Translation Studies", in OLOHAN, M. (ed.) (2000), pp. 1-15.
- OLOHAN, M. (2004): *Introducing Corpora in Translation Studies*, Londres, Routledge.
- ORLINOVSKI, W. y J. YATES (1998): "Genre Systems: Structuring interaction through Communicative Norms", <http://ccs.mit.edu/papers/CCSWP205>.
- PALTRIDGE, B. (1997): *Genre, Frames and Writing in Research Settings*, Amsterdam/Philadelphia, John Benjamins.
- PINTO, M. y J.A. CORDÓN (eds.) (1999): *Técnicas documentales aplicadas a la traducción*, Madrid, Síntesis.
- SWALES, J. (1990): *Genre Analysis: English in Academic and Research Settings*, Cambridge, Cambridge University Press.
- SWALES, J. y C. FEAK (2000): *English in Today's Research World: A Writing Guide*, Ann Arbor, MI, The University of Michigan Press.
- TEMMERMAN, R. (2000): *Towards New Ways of Terminology Description. The Socio-cognitive Approach*, Erasmus Hogeschool Brussel.
- TEUBERT, W. (ed.) (2001): *International Journal of Corpus Linguistics* 6, Institute für Deutsche Sprache, Mannheim.
- TODOROV, T. (1990): *Genres in Discourse*, Cambridge, Cambridge University Press.
- TORRUELLA, J. y J. LLISTERRI (1999): "Diseño de corpus textuales y orales", en BLECUA, J. M., CLAVERÍA, G., SÁNCHEZ, C. y J. TORRUELLA (eds.), pp. 45-77.
- TROSBORG, A. (1997): "Text Typology: Register, Genre and Text Type" en TROSBORG (ed.) *Text Typology and Translation*, Amsterdam, John Benjamins.
- TROSBORG, A. (ed.) (1997): *Text Typology and Translation*, Amsterdam, John Benjamins.
- TYMOCZKO, M. (1998): "Computerized Corpora and the Future Translation Studies", *Meta* 43-4, pp. 652-659.
- WOLLMAN-BONILLA, J. (2000): "Teaching Science Writing in First Grade: Genre Learning and Recontextualization", *Research in the Teaching of English* 35, pp. 35-66.
- ZANETTIN, F., BERNARDINI, S. y D. STEWART (eds.) (2003): *Corpora in Translator Education*, Manchester, St. Jerome.

ANEXO 1

Figura 1. Programa GENTT

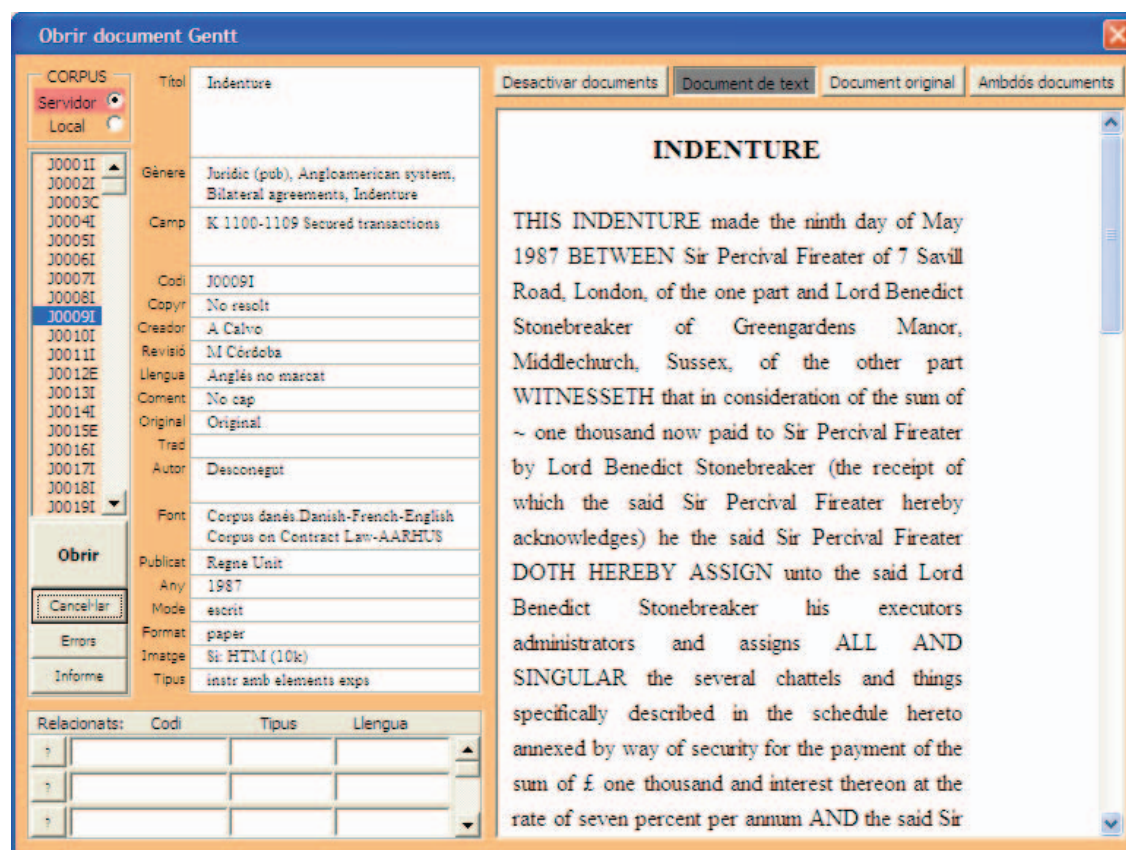


Figura 2. Listado de géneros

Ej. Médico

> Clínicas

- >> Característiques de producte per a professionals
- >> Carta de resultats
- >> Citació
- >> Dieta
- >> Fitxa de seguretat de medicament
- >> Full de vigilància de malalties
- >> Guia clínica
- >> Història clínica
- >>> Consentiment informat
- >>> Electrocardiograma
- >>> Estudi post mortem
- >>> Evolució d'infermeria
- >>> Full clínicoadministratiu d'hospitalització
- >>> Full clínicoadministratiu de consulta externa
- >>> Full d'administració de medicaments
- >>> Full d'anamnesi i exploració
- >>> Full d'anestèsia

- >>> Full d'autorització d'assaig clínic
- >>> Full d'autorització d'autòpsia
- >>> Full d'evolució
- >>> Full d'examen radiològic
- >>> Full d'infecció hospitalària
- >>> Full d'informe citològic
- >>> Full d'informe d'alta
- >>> Full d'informe quirúrgic
- >>> Full d'interconsulta
- >>> Full d'informe anatomopatològic
- >>> Full d'observacions i cures
- >>> Full d'ordres mèdiques
- >>> Full d'urgències
- >>> Full de circulat
- >>> Full de codificació
- >>> Full de constants
- >>> Full de control d'exàmens complementaris
- >>> Full de demanda quirúrgica programada
- >>> Full de determinacions analítiques
- >>> Full de donació d'òrgans
- >>> Full de llista de problemes
- >>> Full de petició d'alta voluntària
- >>> Full de postanestèsia
- >>> Full de postanestèsia
- >>> Full de preanestèsia
- >>> Full de preanestèsia
- >>> Full de seguiment intensiu
- >>> Full de sol·licitud d'estudi anatomopatològic
- >>> Full de sol·licitud d'estudi citològic
- >>> Full de sol·licitud d'ingrés
- >>> Full de tractament convencional
- >>> Full de transfusió
- >>> Full de treball social
- >>> Full de valoració d'infermeria
- >>> Fórmula lucocitària
- >>> Gràfica anestèsica
- >>> Hemograma
- >>> Informe clínic
- >>> Informe d'exploracions especials
- >>> Informe preoperatori
- >>> Part d'assistència per lesions
- >> Informe forense
- >> Manual d'instruccions
- >> Manual mèdic
- >> Part
- >>> Informe de consulta i hospitalització
- >>> Part d'alta
- >>> Part de baixa
- >>> Part de confirmació
- >> Programa informàtic de diagnòstic
- >> Prospecte de medicament
- >> Protocol
- >> Qüestionari sanitari
- >> Reconeixement mèdic periòdic

- >> Reconeixement neonatal
- >> Vademècum
- > **Divulgatius**
- >> Article d'opinió
- >> Article temàtic
- >> Atles visual
- >> Característiques de producte per a pacients
- >> Carta
- >> Comunicat de premsa
- >> Conferència
- >> Conte mèdic infantil
- >> Enciclopèdia divulgativa
- >> Fullet informatiu
- >> Guia divulgativa
- >> Informe anual
- >> Llibre divulgatiu
- >> Notícia
- >> Part epidemiològic
- >> Preguntes més freqüents (FAQ)
- >> Resum per a pacients
- >> Tutorial
- > **Metagèneres**
- >> Base de dades mèdica
- >> Diccionari mèdic
- >> Manual d'estil
- >> Normes per a autors
- >> Tesaure
- > **Pedagògics**
- >> Altes anatòmic
- >> Llibre de text
- >> Tractat
- >> Tutorial
- > **Publicitaris**
- >> Anunci per a pacients
- >> Anunci per a professionals
- >> Article publicitari
- >> Catàleg de productes sanitaris
- >> Fullet publicitari
- >> Publireportatge
- > **Recerca**
- >> Article de revisió
- >> Article en acta de congrés
- >> Article especial
- >> Article original
- >> Carta al director
- >> Carta científica
- >> Cas clínic
- >> Conferència clínica
- >> Document de consens
- >> Editorial científic
- >> Expedient de registre farmacèutic
- >> Informe
- >> Informe farmacològic
- >> Nota clínica

- >> Notícia
- >> Original breu
- >> Patent
- >> Quadern de recollida de dades
- >> Ressenya bibliogràfica
- >> Tesi doctoral
- >> Treball d'investigació

Figura 3. Ficha etiquetado *Header*

Afegir o modificar dades

Codi d'aquest arxiu: J0010I [Veure usats >](#)

Creat per: A Calvo

Revisat per: M. Córdoba, A. Borja

Copyright: No resolts

1) En quina llengua està escrit aquest document?

Català Espanyol Anglès Alemany Una altra Multilingüe

☐ Britànic ☐ Americà ☒ No marcat

2) Aquest document és l'original?

☐ Sí ☐ No ☐ És una traducció amb estatus d'original ☐ No se sap

Si, aquest document és l'original

3) Aquest document està relacionat amb algun altre del corpus? [Documents >](#)

☐ Sí ☐ No

4) Quin n'és el títol? Mortgage of insurance policy

5) Qui n'és l'autor? PJM Fidler, H.I. Freeman

6) Quina n'és la font? Corpus danés.Danish-French-English C

7) Quin n'és el mode? escrit

8) Quin n'és el format original? electrònic

9) On es publicà? Estats Units

10) En quin any? Entre 1980 & 2000

11) Quin n'és el tipus textual? [Tipus >](#)

Text instructiu

12) Seleccionen-ne el gènere: [Gèneres >](#)

Jurídic, Angloamerican system, Bilateral agreements, Mortgage, Mortgage of insurance policy

13) Seleccionen-ne el camp temàtic: [Camps >](#)

K 1100-1109 Secured transactions

14) Afegiu-hi qualsevol comentari o observació:

català espanyol anglès

Cancel·lar Guardar Servidor ☒ Local ☒