

La extracción de conocimiento y terminología a partir de corpus *ad hoc*: el uso de documentos digitales de la web pública

Pilar Sánchez-Gijón

Universitat Autònoma de Barcelona

Electronic communication accelerates the exchange of knowledge and information in areas of specialized knowledge. This state of affairs forces anyone involved in such communication (e.g. technical writers, technical translators) to remain up to date with new developments. Not only do professionals belonging to this group of people have to master the standard terminology of each specialized domain, they must also assimilate and understand the subject matter within which they are working. This article proposes a method for assembling and using specific corpora with a view to extracting from them systematic and bilingual knowledge relating to terminology, the conceptual relations between terms, and the knowledge that they represent. Special attention is devoted to the strategies that will enable professionals to use such corpora in English and in Spanish.

1. El uso de corpus en traducción

Los córpora en traducción pueden tener muchas y muy diversas aplicaciones (como por ejemplo las recogidas por Corpas 2001). En este artículo nos centraremos en la utilización de los córpora *ad hoc* como recurso para la obtención de información de carácter especializado. En concreto nos van a interesar tanto la información lingüística como la información factual, es decir, el conocimiento sobre un ámbito especializado necesario para que un traductor entienda un texto original y sea capaz de elaborar un texto final. En este sentido debemos diferenciar previamente entre *información y conocimiento*; para ello recurriremos a Neumüller (2001):

- Data is facts. A datum, a unit of data, is one or more symbols that are used to represent something.
- Information is interpreted data. Information is data placed within a meaningful context. [...]
- Information is necessarily subjective. Information must always be set in the context of its recipient. The same data may be interpreted differently by different people depending on their existing knowledge.
- Knowledge is derived from information by integrating information with existing knowledge.

Según esta gradación, al documentarse para una traducción técnica o científica el traductor generalmente busca información que le permita adquirir conocimientos factuales sobre el ámbito en el que se enmarca la

traducción. En otras palabras, el traductor busca información que le permita entender el mensaje original y que, por lo tanto, le aporte conocimientos relativos a la materia, es decir, un aporte cognitivo, mientras que simultáneamente adquiere conocimientos lingüísticos. Para poder caracterizar la utilidad de los *córpore ad hoc* en la traducción especializada deberíamos comenzar por establecer qué entendemos por traducción especializada. Hurtado Albir define la traducción especializada como “la traducción de textos dirigidos a especialistas y pertenecientes a los llamados lenguajes de especialidad” (Hurtado Albir 2001: 59); de hecho prefiere denominarla *traducción de textos especializados*. No se espera que el traductor tenga el mismo dominio sobre un ámbito temático que un experto, sino que el traductor debe limitarse a mediar entre interlocutores en una situación especializada (Mayoral 1997/1998: 137).

Así pues, por un lado el traductor no tiene que convertirse en un experto en el tema sobre el que traduce; pero por otro lado, para que el traductor realice su cometido con éxito debe entender la materia sobre la que traduce. De ahí surge la necesidad del traductor de documentarse, es decir, de seguir un proceso que le permita obtener todos los conocimientos necesarios para comprender el texto original y elaborar el texto final.

Según Gamero (1998: 80), los tres factores que obligan al traductor especializado a documentarse antes de acometer su traducción son: los conocimientos temáticos, los conocimientos terminológicos, los conocimientos de las convenciones de género o tipos de textos. El traductor debe saber reconocer cuándo necesita obtener conocimientos en cualquiera de estos tres aspectos, así como contar con las habilidades necesarias para poder cubrir esta necesidad. Estas habilidades documentales forman parte de la *competencia traductora*, que PACTE define como “the underlying system of knowledge and skills needed to be able to translate” (2000: 100). En cuanto a la competencia relativa a los conocimientos temáticos, Hurtado Albir indica que:

[...] se trata de una competencia sobre todo de comprensión, ya que, a diferencia del especialista, no es necesario que sea capaz de producir por sí sólo textos especializados. En caso de carecer de esos conocimientos, debe saber suplirlos mediante su capacidad de documentación, que le permitirá adquirir los conocimientos necesarios. (Hurtado Albir 2001: 61)

El segundo aspecto de los textos especializados que puede suponer un problema para el traductor es la terminología. Los términos son las unidades de conocimiento propias del lenguaje de especialidad, por lo que su ámbito natural son los textos especializados. De este modo, en tanto que unidades de conocimiento enmarcadas en un área temática, no forman parte ni de la competencia lingüística general del traductor (ya sea en la lengua de origen o en la lengua meta), ni de sus competencias cultural y temática. De hecho, Faber (2002) vincula los dos aspectos anteriores, el tema y los términos, haciendo alusión a la *competencia terminológica*, que encuadra como módulo de la competencia traductora

basado en la aplicación de los postulados de la representación del conocimiento a la traducción. La propuesta de Faber incluye:

[...] procesos tales como la adquisición y asimilación en estructuras cognitivas previas de conocimiento especializado, mediante su extensión a niveles más específicos. Igualmente, supondría la capacidad de relacionar este conocimiento con representaciones lingüísticas en una o varias lenguas. (Faber 2002: 4)

La adquisición de términos en una o varias lenguas y la de conocimiento sobre un ámbito temático van de la mano. Según Faber, la competencia terminológica consiste en la asimilación de conocimiento especializado en la memoria, los automatismos propios de la adquisición de terminología, la creatividad en la formación de términos y la capacidad del traductor para resolver problemas de adquisición de conocimiento durante el proceso de traducción. A pesar de que la teoría determina con qué conocimientos debe contar el traductor para llevar a cabo su cometido, todavía no ha sido posible fijar un proceso documental estándar, debido a que este proceso depende en cualquier caso de los conocimientos factuales y lingüísticos que el traductor ya posee. Si el traductor está familiarizado con el área temática y las características lingüísticas del texto de especialidad no necesitará documentarse. No obstante, si no cuenta con estos conocimientos, el traductor debe llevar a cabo un proceso de documentación a partir de un mínimo cognitivo, sus propios conocimientos factuales y lingüísticos, que podemos denominar de cultura general:

Al igual que podemos distinguir entre conocimientos factuales ‘de cultura general’, que previsiblemente forman parte del acervo de conocimientos de un tipo determinado de ciudadanos (por ejemplo el traductor) y que forman parte tanto de la experiencia común a todos los miembros de una misma comunidad, de los conocimientos especializados adquiridos en una formación académica característica y de los conocimientos factuales especializados, propios del objeto de la comunicación especializada en la que participa el traductor; también podremos distinguir entre los elementos lingüísticos generales y los elementos lingüísticos especializados relacionados con dichos conceptos. (Mayoral 1997/1998: 140)

Podemos afirmar que, por un lado, el traductor cuenta con conocimientos que comparte con el resto de miembros de las dos comunidades, la de la lengua origen y la de la lengua meta – aunque generalmente los conocimientos sobre su cultura materna serán más profundos – por lo que no deberá documentarse sobre estos aspectos. Por el contrario, los conocimientos especializados y los lingüísticos vinculados a estos “son demasiado numerosos y dispersos para formar parte del acervo del traductor, aunque cuanto mayor sea su familiaridad con ellos más eficaz será su trabajo profesional” (Mayoral *ibid.*: 140).

Por tanto, la clave del proceso documental radica en cubrir el vacío existente entre conocimiento general y conocimiento especiali-

zado de manera objetiva, o entre conocimiento adquirido y conocimiento por adquirir de manera subjetiva, es decir, en el caso de cada traductor.

Otro aspecto difícil de establecer es en qué momento podemos considerar que el traductor ya ha adquirido todos los conocimientos que necesita. En este sentido, tanto Mayoral (ibid.) como Hurtado Albir (2001) y Gamero (1998) coinciden en considerar que el traductor debe tener un nivel de *conocimiento suficiente* para comprender el texto original y elaborar un texto final adecuado, por lo que este conocimiento está más próximo al nivel de comprensión pasivo de la materia que al activo. El nivel de conocimiento suficiente, tal como Mayoral lo entiende, es un nivel prácticamente imposible de concretar con exactitud; no obstante, este conocimiento es el resultado de la confluencia del conocimiento del traductor, que incluye desde su formación educativa y cultural hasta sus experiencias vitales, y el ámbito temático de la traducción que va a realizar.

Sin embargo, este conocimiento factual pasivo debe materializarse en una competencia lingüística que le permita hacer un uso preciso de las denominaciones especializadas y demás elementos lingüísticos en relación al encargo de traducción:

El traductor no está interesado ni en las denominaciones posibles ni en las denominaciones idóneas o normalizadas; el traductor está interesado en la solución idónea para una situación de comunicación concreta, y estas soluciones idóneas para una situación concreta pueden tener poco que ver con las soluciones lexicográficas y con las soluciones terminológicas. (Mayoral 1997/1998:143)

En nuestra opinión, el uso de *corpóra ad hoc* se postula como uno de los recursos que mejor pueden ayudar a adquirir los conocimientos propios de la competencia terminológica tal como Faber la describe, sobre todo teniendo en cuenta que las necesidades informativas de cada traductor son diferentes, tanto en lo que al encargo de traducción se refiere como al punto cognitivo de partida del profesional de la traducción.

2. La compilación de un corpus *ad hoc*

La compilación de un corpus *ad hoc* que va a utilizarse como fuente de información terminológica, tanto lingüística como conceptual, debe cumplir varias premisas. De forma resumida, un corpus *ad hoc* de estas características que deba proporcionar información a un traductor o a un redactor técnico debe contener el mayor número de textos posible y se debe poder compilar en el menor lapso de tiempo posible, ya que no se debe olvidar que la compilación de un corpus *ad hoc* no es un fin en sí mismo, sino un instrumento para obtener otro objetivo, ya sea una traducción o la redacción de un texto técnico o científico. Es por esto que,

actualmente, el modo más rápido de compilar un corpus de este tipo consiste en recurrir a las páginas públicas de la WWW.

Si bien es el modo más rápido de obtener textos en formato digital que traten de un tema determinado, el uso indiscriminado de este tipo de recursos textuales no está libre de ciertos riesgos. En primer lugar, para asegurarse de que la información que proporciona el corpus no es únicamente correcta sino que también se adecua a la situación comunicativa del encargo de traducción (o en la que se consumirá el texto técnico), los textos que lo configuran deben cumplir en la medida de lo posible las mismas condiciones pragmáticas de un texto paralelo (Nord 1991). Para ello, es necesario conocer en profundidad las características pragmáticas y comunicativas de los documentos que se publican en la web pública (Sánchez-Gijón 2004).

Sin embargo, una de las principales carencias de la mayoría de textos que se hacen públicos en la WWW es su falta de fiabilidad. La mayor parte de ellos no se revisan ni pasan por ningún proceso editorial de publicación, por lo que no existe ninguna garantía de calidad, ni en cuanto a aspectos formales o lingüísticos ni en cuanto a aspectos factuales o conceptuales. Es por ello que en muy pocas ocasiones se pueden utilizar estos textos como textos paralelos de manera individual.

A continuación se abordan los principales aspectos que caracterizan los textos digitales publicados en Internet, sobre todo en contraposición a los de los textos en formato papel, y posteriormente se propone una metodología de obtención de textos digitales para la compilación de un corpus.

2.1 Los recursos textuales de la web pública

Larsen (2001) presenta el concepto de *alfabetización informativa* consistente en la capacidad de saber buscar la información que se necesita, encontrarla, evaluarla, procesarla y utilizarla para tomar las decisiones adecuadas, y esto es precisamente lo que debe hacer un traductor durante el proceso de documentación. En ese mismo artículo, el autor también pone de relieve que la WWW es, en realidad, un depósito en formato digital de libros, conjuntos de datos, enciclopedias, bibliotecas, así como “any disparate piece of text, graphic, or sound byte that someone chose to put on-line” (Larsen, *ibid.*). Este es el motivo por el que muchos internautas piensan que cualquier información está al alcance de sus dedos, cosa que Codina ha bautizado como *ciberingenuidad* (1995), dado que este es un razonamiento que no parece justificarse en la práctica, ya que no siempre se encuentra la información que se necesita.

Incluso en el caso de encontrar la información que se busca, tampoco hay garantías de hallarla en la lengua que el usuario necesita. Por todos es conocido que el entorno web está dominado por la lengua ingle-

sa, aunque algunos estudios revelan que la preponderancia de esta lengua está comenzando a ceder (por ejemplo Funredes 1998, 2002).

Otro aspecto que el traductor debe tener en cuenta a la hora de acceder a la WWW para documentarse sobre un tema determinado es la volatilidad de las páginas web (Caslon Analytics 2002). La información no permanece de manera perenne en Internet; de hecho, la vida media de una página web es inferior a dos años, y aunque mantengan su dirección URL, el contenido puede variar en cualquier momento. La volatilidad es una de los causantes de la redundancia en la WWW, cosa que afecta a la compilación de un corpus *ad hoc* puesto que un mismo texto puede recogerse a partir de diferentes URLs y como páginas web distintas. Estas repeticiones pueden deberse tanto a la duplicación de sitios web completos para evitar la pérdida de información en el caso de que un servidor deje de estar en funcionamiento, como al hecho de que a menudo se copia parcial o totalmente la información de una página y se presenta en otra sin indicar su origen (este hecho, que es tan habitual como ilegal –al menos en la mayoría de los países occidentales– se conoce como *transclusión*). Por todo ello, se calcula que aproximadamente el 30% de las páginas web son duplicados o copias parciales de otras páginas (Baeza-Yates y Ribeiro-Net 1999: 368).

Con la aparición de la WWW los sistemas de comunicación se democratizan, ya que la línea divisoria entre los productores y los consumidores de información se diluye: cualquier usuario que consuma información publicada también puede producir información que, a su vez, pueda ser consumida por otro usuario. Sin embargo, este hecho es la causa de ciertos inconvenientes que ya hemos puesto de manifiesto anteriormente. El principal inconveniente, sobre todo si hablamos de textos de temática técnica o científica, es la ausencia de un proceso editorial de publicación, que incluye desde una simple corrección ortográfica hasta la revisión per peers. El proceso editorial de publicación de textos en formato papel garantiza su calidad, cosa que también sucede con muy pocos textos publicados en la WWW pública. Por este motivo, muchos centros de documentación han realizado propuestas metodológicas para la evaluación de los recursos que se encuentran en Internet (por ejemplo Alexander y Tate 2001).

Por último, si se observan los recursos textuales digitales en tanto que documentos, el espectro de publicaciones en formato digital puede contemplarse como una especie de continuo que abarca desde la publicación más trivial hasta la más especializada, igual que los textos en formato papel. La diferencia entre los dos formatos radica en que los textos en formato digital pueden incluir mucha más variación textual en un mismo documento, por lo cual resulta mucho más complicado identificar los diferentes tipos de recursos en la red que entre los otros medios de comunicación tradicionales. Por otro lado, los diferentes recursos textuales tradicionales suelen diferenciarse de manera pragmática por la situación comunicativa en la que intervendrán, aspecto que resulta difícil

analizar en un entorno virtual en el que cualquier usuario puede acceder a cualquier información en cualquier momento y situación.

A modo de resumen, las características más relevantes de la WWW y de los textos que contiene que deben tenerse en cuenta al recurrir a páginas web como fuente de información:

- La WWW no es un entorno ordenado de información, y no contiene absolutamente todo el conocimiento, por lo que no siempre se encuentra la información que se desea.
- Los recursos textuales son volátiles, y en este entorno suele ser habitual la redundancia de información y la transclusión.
- Los recursos textuales no suelen pasar por ningún proceso de corrección o revisión por lo que se debe acabar por recurrir a criterios de calidad para evaluarlos.
- Se da la dificultad de diferenciar entre géneros o tipos textuales, que en gran medida no se corresponden plenamente con los géneros textuales tradicionales.

Dado que la información que pensamos extraer del corpus *ad hoc* es de carácter terminológico, el mayor inconveniente de los cuatro señalados es la falta de fiabilidad. Sin embargo, pensamos que, si bien los recursos textuales que formaran parte del corpus no presentan los indicadores necesarios para garantizar su calidad y, por lo tanto, no son suficientemente fiables de manera individual para el traductor, el conjunto de textos sí que lo es. Es decir, mantenemos que el criterio de cantidad inherente a la metodología de obtención de información que propondremos en el apartado 3 nos garantiza la calidad de la información del conjunto, a pesar de que los textos por separado no lo sean.

2.2 La compilación de un corpus *ad hoc*

La manera más inmediata de compilar un corpus a partir de la WWW consiste en realizar búsquedas basadas en palabras clave y descargar las páginas identificadas. La búsqueda de documentación digital se realiza en función de un propósito concreto; no se busca información de manera abstracta o descontextualizada, sino con un objetivo determinado. Por este motivo, antes de iniciar una búsqueda es necesario conocer las necesidades informativas de cada uno. En este sentido, Chowdhury llega a las siguientes conclusiones:

1. information need is a relative concept. It depends on several factors and does not remain constant;
2. information needs change over a period of time;
3. information needs vary from person to person, from job to job, subject to subject, organization to organization, and so on;
4. people's information needs are largely dependent on the environment. For example, information needs of those in an academic environment are different from those in an industrial, business or government/administrative

environment; 5. measuring (quantifying) information need is difficult; 6. information need often remains unexpressed or poorly expressed; 7. information need often changes upon receipt of some information. (Chowdhury 1999: 181)

Durante el proceso de documentación del traductor especializado el concepto de necesidad de información también es relativo, no solamente porque los traductores tienen necesidades informativas diferentes a las de otras profesiones, sino sobre todo porque estas necesidades varían en función del encargo de traducción. Si bien por información de calidad se entiende aquella que es objetivamente completa y actual, las características de la traducción pueden obligar al traductor a buscar información que no coincida con estas premisas; a modo de ejemplo, si el original es un texto científico anterior a un descubrimiento relevante que modifica los fundamentos teóricos de esa materia, el traductor deberá documentarse a partir de textos coetáneos o que traten el tema desde la misma perspectiva. Por todo ello, el recurso documental de calidad en cada caso debe cubrir una necesidad informativa adecuándose a la situación comunicativa del encargo de traducción, o en otras palabras, debe ser pertinente en relación a las necesidades informativas del traductor.

En el momento de compilar un corpus *ad hoc* resulta complicado atender a todos estos condicionantes. De nuevo, el valor cuantitativo del corpus permitirá realizar consultas que faciliten la recuperación de información de recursos pertinentes y pasen por alto la de recursos no pertinentes a la consulta realizada.

Con el objeto de llevar a cabo este estudio, hemos compilado dos corpora *ad hoc*, uno en inglés y otro en castellano, sobre el fenómeno astronómico conocido como Leónidas. La obtención de los recursos textuales se realizó mediante búsquedas a partir de las palabras clave “Leónidas” y “astronomía” en castellano y en inglés. Las búsquedas se ejecutaron a través de un agente inteligente de búsqueda, en concreto Copernic (Copernic Technologies, Inc.), que permite la realización de búsquedas simultáneas sobre un gran número de motores de búsqueda. Después de realizar esta operación, descargamos los documentos web a los que pertenecían las páginas identificadas y obtuvimos los siguientes resultados:

Lenguas	Volumen de información descargada
Inglés	6.522 mb
Castellano	409 mb
Total	6.931 mb

Tabla 1: volumen total de información descargada de Internet

Cabe señalar que no se descargaron únicamente las páginas que se identificaron con las búsquedas a partir de palabras clave, sino que se descargaron los documentos web a los que pertenecían, incluyendo cualquier

unidad de información que se encontrase en forma no textual (gráficos, vídeos, etc.). Después de obtener toda esta información, se realizó una operación de filtrado que permitió identificar los recursos textuales que finalmente formarían parte de cada uno de los dos *córpora ad hoc*. Los criterios de filtrado seguidos fueron:

- que incluyesen la palabra “Leónidas”, en la lengua correspondiente, en el cuerpo del recurso, y no como mero enlace hipertextual.
- que no tuviesen información en más de un idioma.
- que no fuesen mensajes electrónicos.
- en el caso del castellano, que “Leónidas” no se utilizase como nombre propio, sino en referencia al fenómeno astronómico que nos ocupa.

Después de esta operación, el resultado del filtrado y la composición final de los *córpora* fue la siguiente:

Lenguas	Recursos Textuales
Inglés	922
Castellano	116
Total	1038

Tabla 2: recursos textuales que conforman los dos *córpora monolingües*

Tal como era de esperar, el corpus en lengua inglesa es cuantitativamente superior al corpus en castellano. Esto se debe a que la información que en la WWW se encuentra mayoritariamente información en inglés. Antes de profundizar en el análisis de estos dos *córpora*, es necesario tener en cuenta que las características de las páginas web provocan ciertos inconvenientes a la hora de explotarlas como corpus. Los principales inconvenientes son:

- Las etiquetas propias del lenguaje HTML, y sobre todo las metaetiquetas: el programa de análisis de corpus puede interpretar estas etiquetas como parte del texto, cosa que provocaría una gran cantidad de ruido a la hora de realizar el análisis.
- Los caracteres especiales: caracteres como las vocales acentuadas o la “ñ” del castellano no siempre se representan en el lenguaje HTML de acuerdo con las reglas ortográficas. En el caso de las vocales acentuadas, por ejemplo, a menudo se sustituye el acento por “acute;” para que el navegador, al mostrar el documento, interprete que debe mostrar esa vocal acentuada (p. ej.: “oacute;” es “ó”). Este hecho provoca la alteración de la grafía de muchos vocablos (p. ej.: “Leónidas” sería “Leoacute;nidas”), cosa que hay que tener en cuenta al realizar, por ejemplo, una extracción de concordancias. En nuestro caso, sustituimos todos los caracteres especiales en formato HTML por su forma ortográficamente correcta.
- Los enlaces hipertextuales: desde un punto de vista lingüístico, los enlaces hipertextuales son segmentos de texto que en la mayor parte de los casos aparecen de manera independiente y descontextualizada,

como si de un listado de títulos se tratase. Además, a menudo las páginas de un mismo documento recogen los mismos enlaces, por lo que su frecuencia de aparición es muy alta, cosa que puede alterar los resultados del análisis desde el punto de vista cuantitativo.

3. La extracción de información terminológica y factual

Tal como hemos señalado anteriormente, el traductor especializado o el redactor técnico se documentan ya sea porque desconocen tanto el tema como la terminología o porque aun conociendo el tema desconocen la terminología (en la lengua origen o en la lengua meta). Tanto en un caso como en el otro, la documentación a partir de *córpore ad hoc* requiere que el usuario adquiera ciertas estrategias de consulta que le permitan identificar la información que necesita. Es por ello que nos centraremos en identificar qué expresiones van a facilitar la obtención de información de manera sistemática; estas expresiones, que implícitamente representan relaciones conceptuales, permitirán afianzar los conocimientos del traductor sobre el ámbito en el que se enmarca la traducción, a la vez que adquiere la información terminológica (y fraseológica) que necesita, tanto para comprender el original como para elaborar el texto meta.

El proceso que hemos seguido para identificar esas expresiones se basa en el análisis de los contextos de “Leónidas” en ambos *córpore ad hoc*, el inglés y el castellano. Al analizar estos contextos hemos obtenido un catálogo de expresiones que nos permiten identificar definiciones en contexto, rasgos caracterizadores, relaciones conceptuales jerárquicas (de género – especie, metonímicas y de material – objeto) y relaciones conceptuales no jerárquicas (infralógicas, genéticas y causales). Una vez identificadas expresiones de este tipo podremos realizar estudiar el término “Leónidas” o cualquier otro término y obtener de manera estructurada toda la información posible, de manera que el traductor, a partir de sus propios conocimientos, pueda asimilar el conocimiento especializado pasivo que necesita para llevar a cabo su cometido. Debe tenerse en cuenta que ambos *córpore ad hoc* son conjuntos de textos no marcados ni etiquetados, por lo que no se pueden ejecutar búsquedas a partir de unidades semánticas o de elementos morfosintácticos. (La tarea de etiquetado de un corpus de estas características ralentizaría la obtención de información de tal modo que el corpus dejaría de ser un recurso rentable para el traductor.)

3.1. Definiciones en contexto

El análisis de las concordancias de “Leonids” y “Leónidas” nos ha permitido identificar las siguientes expresiones que introducen una definición en contexto.

INGLÉS	CASTELLANO
[DENOMINACIÓN] are a [DEFINICIÓN]	[DEFINICIÓN] conocid(*) como [DENOMINACIÓN]
[DEFINICIÓN] are called [DENOMINACIÓN]	[DENOMINACIÓN] es (ART. INDEF.) [DEFINICIÓN]
[DEFINICIÓN] is named (the) [DENOMINACIÓN]	[DENOMINACIÓN] es (ART. DEF.) [DEFINICIÓN]
[DEFINICIÓN] is known as (the) [DENOMINACIÓN]	[DENOMINACIÓN], (APOSICIÓN – DEFINICIÓN)
[DEFINICIÓN] which is called [DENOMINACIÓN]	[DENOMINACIÓN] son un(*) [DEFINICIÓN]
The [DENOMINACIÓN] is/are [HIPERÓNIMO],	[DEFINICIÓN] llamad(*) [DEFINICIÓN]
which [CARACTERÍSTICA]	
[DENOMINACIÓN] is/are [HIPERÓNIMO],	
that [CARACTERÍSTICA]	
The [DENOMINACIÓN], [APOSICIÓN]	

Tabla 3: expresiones que presentan definiciones en contexto

En ambas lenguas aparece de manera recurrente el esquema típico de la definición por genus *et differentiae*. Este esquema permite definir un término a partir de la clase a la que pertenece, añadiendo al menos una característica que lo diferencie de los términos conceptualmente más cercanos, ya sea su hiperónimo como sus posibles cohipónimos. La búsqueda de coapariciones a partir de “Leonids” y “are a” en el corpus inglés arroja como respuesta el siguiente listado.

Most years	<u>the Leonids are a</u>	minor event — boasting no more than 10 to 15 shooting stars per hour. That's what happens when Earth passes through the rarefied zones between Tempel-Tuttle's debris trails.
This means that	<u>the Leonids are a</u>	stream to which the outburst predictions can be applied.
	<u>the Leonids are a</u>	shower of short duration, lasting only one week centered on November 18. The Leonids are particles from Comet Temple-Tuttle, which will reach perihelion in 1998. A great concentration of particles exists near the parent comet. In years when t
	<u>the Leonids are a</u>	meteor shower. They are called the Leonids because they appear to radiate out of the constellation Leo. A Meteor, also known as a "shooting star," is a particle from space. Its typical size ranges from that of a grain of sand to that of a pea.
	<u>the Leonids are a</u>	meteor shower that can be most impressive every 33 years or so. While experts are predicting that best show will be seen from the Europe, meteor showers can be somewhat unpredictable. So you may want to keep a careful watch.
	<u>the Leonids are a</u>	swarm of meteoroids that intercept Earth's orbit to some degree every year, usually with little more fanfare than a spectacular night of watching shooting stars. This year, however, the storm will reach levels not seen in over thirty years as Earth
	<u>the Leonids are a</u>	meteor shower that appear to radiate out of the constellation Leo. A Meteor, also known as a "shooting star. All meteors appear as brief streaks of light moving a short distance across the sky. Some meteors move slow and some move fast.
In most years	<u>the Leonids are a</u>	rather insignificant annual meteor stream. Rates peak at 13 per hour on November 17. The main activity is between November 13 and 20, but Leonids occur annually at a rate larger than 1 per hour in the peri
A:	<u>the Leonids are a</u>	meteor stream. If you go out during the nights 16/17 and 17/18 November after midnight, you will see many more meteors, or shooting stars, than normal
;	<u>the Leonids are a</u>	good example - the parent comet is comet Temple-Tuttle which has a 33 year period, the dust is not evenly spread out around the orbit, but is thought to be bunched in denser ribbon-like strands the densest of which follow the comet

Figura 1: listado de coapariciones de "Leonids" y "are a"

Como ejemplo en castellano hemos elegido una de las expresiones menos productivas en el corpus compilado, pero que puede resultar tan o más productiva en el análisis de cualquier otro corpus *ad hoc*.

Se trata de una gran tormenta de meteoritos	llamada	las Leónidas
Debido a que esta lluvia de meteoritos aparentemente viene de la constelación de Leo, se le ha	llamado	Leónidas.

Figura 2: coapariciones de "llamad(*)" y "Leónidas"

En la tabla 3 se recogen todas las expresiones que introducen una definición en contexto que se han encontrado en los contextos de “Leónidas” y “Leonids”. Estas expresiones no son exclusivas del discurso especializado, sino que son formulaciones propias de la lengua y que el traductor puede deducir a partir de su bagaje lingüístico. Por ello, en la tabla 3 no se recoge un listado cerrado de expresiones, ya que se podría ampliar con cualquier otra expresión que cumpla la misma función en el discurso.

3.2. Relaciones conceptuales

En muchas ocasiones, las unidades terminológicas no son los únicos elementos que representan el conocimiento especializado en un texto. Tan importantes como estas unidades son las relaciones que se establecen entre ellas, ya que ponen en relación los diferentes términos y, por lo tanto, las diferentes unidades cognitivas, de manera que permiten la representación del conocimiento. Al aparecer de manera explícita en los textos especializados, las relaciones conceptuales facilitan la comprensión de la materia por parte del traductor. Sin embargo, estas relaciones no siempre aparecen de forma transparente en el texto, sino que quedan diluidas por las características del propio discurso.

Una vez se ha introducido un término en contexto, probablemente acompañado por una definición en contexto, el hilo argumental del texto puede facilitar el proceso de identificación de nuevos rasgos conceptuales que caractericen el término. Las formulaciones identificadas en el entorno de “Leónidas” y “Leonids” en cada uno de los corpóra son las siguientes:

	INGLÉS		CASTELLANO
[DENOMINACIÓN] is characterized by [CARACTERÍSTICA]	[CARACTERÍSTICA] is typical of [DENOMINACIÓN]	The reason that [DENOMINACIÓN] is/are X is [CARACTERÍSTICA]	[DENOMINACIÓN] es [CARACTERÍSTICA]
[HIPERÓNIMO] is called the [DENOMINACIÓN] because [CARACTERÍSTICA]	(THE) [ADJ.] [DENOMINACIÓN]	There (be) [CARACTERÍSTICA] [DENOMINACIÓN]	[ARGUMENTO GENÉRICO]. En el caso de (ART.) [DENOMINACIÓN] [ARGUMENTO ESPECIFICATIVO]
[CARACTERÍSTICA EN CONTEXTO] is the [DENOMINACIÓN]	(The) [DENOMINACIÓN], which is/are the [CARACTERÍSTICA]	(Most) [DENOMINACIÓN] is/are (not) that [ADJ.]	[DENOMINACIÓN], (APOSICIÓN – DEFINICIÓN / CARACTERÍSTICA)
That is why [DENOMINACIÓN] are [CARACTERÍSTICA – REF. CATAFÓRICA]	[DENOMINACIÓN], which [VERBO]	The main [CARACTERÍSTICA] with the/of the [DENOMINACIÓN] is that [EXPLICACIÓN CARACTERÍSTICA]	
MOST [DENOMINACIÓN] ARE (CARACTERÍSTICA HABITUAL)	[DENOMINACIÓN] is/are [HIPERÓNIMO] that [CARACTERÍSTICA]	[ARGUMENTO]. In the case of [DENOMINACIÓN], [ESPECIFICACIÓN DEL ARGUMENTO]	
HOW (CARACTERÍSTICA) ARE/IS [DENOMINACIÓN]	A (ADJ. MÁS HABITUALES) [DENOMINACIÓN]	THE [CARACTERÍSTICA] OF THE [DENOMINACIÓN]	

Tabla 4: expresiones que permiten identificar características conceptuales

Si bien todas estas formulaciones permiten identificar rasgos conceptuales concretos, algunos pueden resultar de carácter más prototípico que otros. Este es el caso de las coapariciones de “most Leonids”, expresión a partir de la cual se recogen características comunes de los referentes del concepto:

Despite last year's fireball storm,	most Leonids	<u>are not that bright</u> . The more light pollution, the fewer meteors you will see! So
on a mid-November night.	Most Leonids	<u>are seen</u> when the radiant is high, during the early-morning hours. In 2000, ther
as a result of which	most Leonid	<u>meteoroid impacts</u> will be on the back side of the Moon.
Fortunately, we believe that	most Leonid	<u>meteoroids will be smaller</u> than this. Figure 1 - Infrared image of comet Tempel/2
Though	most Leonid	<u>meteoroids are smaller</u> than a grain of sand, they zoom across space at a menacing 155,000 miles per hours.
Arne Danielsen Explanation:	Most Leonid	<u>meteoroids</u> , the bits of comet debris which produce the annual Leonid meteor shower, <u>range from</u> a mere millimeter to a centimeter in diameter.
meteors were very faint and	most Leonid	meteors were three or less magnitude. No fireball appeared during the observation. On Nov. 18/19, many Leonid meteors appeared.
	Most Leonids	will burn up in the upper atmosphere, posing no danger to humans
While	most Leonid	meteors are no larger than grains of sand, some may be as large as a meter in diameter.
Bright Arcturus low in ENE.	Most Leonids	should be seen in a "belt" around the Moon - at 30-60° distance - as the radiant (red cross) is just to the east of the Moon.
	Most Leonids	we saw were +1 magnitude or fainter.
the previous night when	most Leonids	were relatively faint. A bright -8 Leonid fireball was recorded at 7:46:59 UT in south-western direction
large as a meter in diameter.	Most Leonids	will burn up in the upper atmosphere, posing no danger to humans on Earth.

Figura 3: coaparición de "most" y "Leonids"

Dado su tamaño reducido, el corpus castellano no fue capaz de arrojar más de un ejemplo por expresión, cosa que resulta claramente insuficiente.

3.3. Relaciones conceptuales jerárquicas

Las relaciones conceptuales jerárquicas son las más fáciles de asimilar para un no experto en el tema, ya que facilitan la creación de estructuras

simples de representación del conocimiento. Para el traductor especializado son esenciales en la fase de comprensión del original, y también facilitan la delimitación conceptual de cada unidad terminológica.

3.3.1. Relaciones conceptuales de género – especie

Esta relación conceptual es el eje en el que se basan las definiciones de *genus et differentiae* y, por lo tanto, gran parte de las definiciones en contexto, tal como hemos señalado anteriormente. Las expresiones identificadas en este caso son las siguientes:

INGLÉS	CASTELLANO
[HIPERÓNIMO] is the [DENOMINACIÓN]	[DENOMINACIÓN] es un(*) [HIPERÓNIMO]
[DENOMINACIÓN] is/are a [HIPERÓNIMO] that [CARACTERÍSTICA]	[DENOMINACIÓN], (ARTÍCULO INDEF.) [HIPERÓNIMO]
[DENOMINACIÓN], a [HIPERÓNIMO] that [CARACTERÍSTICA]	[DENOMINACIÓN], (ARTÍCULO DEF.) [HIPERÓNIMO]
Like many [HIPERÓNIMO] the [DENOMINACIÓN] [CARACTERÍSTICA]	[DENOMINACIÓN] and other [HIPERÓNIMO]

Tabla 5: expresiones que permiten identificar hipónimos – hiperónimos

Una de las expresiones más comunes del lenguaje es la de relacionar un término determinado con sus cohipónimos mediante el hiperónimo, de manera que sin citar cada uno de los cohipónimos se hace referencia a un conjunto cerrado de nociones. Eso es lo que sucede en el siguiente listado de coapariciones:

s 50 years ago. Enough is known about the	<u>Leonids and other</u> i	mportant showers to get it partly right much of
gan looking for previous appearances of the	<u>Leonids and other</u>	meteor showers among ancient and medieval do
e's Leonid Page Information on viewing the	<u>Leonids and other</u>	meteor showers. Introduction Congressi
s with video systems, to determine of	<u>Leonids and other</u>	meteors, and to look at the orbits dependency of th
y orbit, going into re-entry, dealing with the	<u>Leonids and other</u>	showers .Their deliberations will feed through to
ded questions from the audience about the	<u>Leonids and other</u>	showers, such as the Giacobinids whose 1946 s

Figura 4: coapariciones de “Leonids” y “other”

Otro de los mecanismos previstos por la lengua es la aposición, cuando se establece una equivalencia entre un término y un sintagma o frase subordinada que lo describe. Esta es una formulación difícil de identificar en un corpus no etiquetado como el que hemos analizado, aunque sí hemos podido identificar algunos ejemplos:

gran máximo de las Leónidas,	<u>una de las lluvias</u>	de estrellas fugaces más espectacular de la historia.
Las Leónidas,	<u>la lluvia</u>	de estrellas de la noche de ayer, resultó ser un fen
Las leónidas,	<u>la lluvia</u>	de estrellas de la noche del 17 al 18 de noviembre,

Figura 5: coapariciones de “Leónidas” seguida de coma y aposición.

3.3.2. Relaciones meronímicas

Después de observar las relaciones entre hipónimos e hiperónimos, otra de las relaciones jerárquicas que permite elaborar un esquema representativo sencillo son las que se establece entre un término unidad y los términos que forman parte de dicha unidad, en otras palabras, las relaciones parte – todo.

INGLÉS	CASTELLANO
[DENOMINACIÓN], which contain [PARTES]	(ART.) [PARTE] que form(*) (ART.) [DENOMINACIÓN]
[DENOMINACIÓN] contain [PARTES]	[DENOMINACIÓN], contiene [PARTE]

Tabla 6: expresiones que permiten identificar relaciones meronímicas

En el caso de las relaciones meronímicas, las expresiones identificadas, tanto en inglés como en castellano, son escasas. Esto no quiere decir que los corpóra compilados no contengan otras expresiones que representen esta relación conceptual, sino únicamente que esas otras expresiones no han aparecido en los contextos de “Leónidas”.

lar ZHRs by our model).To reiterate:	<u>contain</u>	many young dust trails The geometry of t The Leonids
on and Planets that some Leonid meteoroids do	<u>contain</u>	unusually tough pieces of matter. A particu
,000 mph. Besides being fast, the Leonids usually	<u>contain</u>	a large number of very bright meteors. The
red days of passing Earth’s orbit, Leonid showers	<u>contain a</u>	n above-average proportion of fireballs and bolides (bright meteors that flare or explode).

Figura 6: coapariciones de “Leonid” y “contain”

Además de contar con pocos ejemplos, en este caso existe una dificultad añadida: no resulta sencillo diferenciar entre el uso de estos verbos cuando representan una relación metonímica de cuando no lo hacen. De todos modos, las siguientes coapariciones en castellano también ponen de manifiesto una relación de meronimia entre “material” o “partícula” y “Leónidas”:

1966. Aunque la geometría del material que <u>forman</u> las Leónidas no será especialmente favorable,
aciones de la dinámica de las partículas que <u>forman</u> las Leónidas. Los resultados de sus investig

Figura 7: coapariciones de “formar” y “Leónidas”

3.3.3. Relaciones material – objeto

La última de las relaciones conceptuales jerárquicas que hemos podido analizar a partir de los contextos de “Leónidas” en cada corpus es la que se establece entre un objeto y su material.

INGLÉS	CASTELLANO
[DENOMINACIÓN] (is/are) made of [MATERIAL]	[MATERIAL] forma(*) [DENOMINACIÓN]
[DENOMINACIÓN] material [CARACTERÍSTICA DEL MATERIAL]	
The material in the [DENOMINACIÓN] [CARACTERÍSTICA]	

Tabla 7: expresiones que permiten identificar la relación material – objeto

En el caso de estas expresiones, la proximidad entre el término sobre el cual nos estamos documentando y la expresión que denota la relación, “material”, no coaparecen con una proximidad regular, como en la mayoría de contextos vistos hasta el momento.

ite, not interplanetary dust or cometary	<u>material</u>	.Wednesday, November 18 Leonids Afte
rs University in Piscataway. “The Leonids’	<u>material</u>	is much more clumped. The stream of debris is
quite. The Leonids are more like a spiral of	<u>material</u>	left behind by comet Tempel-Tuttle, and Earth
al-time analyses of the composition of the	<u>material</u>	coming off the comet. Like the Leonids balloon
Leonids are caused by cometary particles. This	<u>material</u>	has a strong tendency to fragment and disinteg
ronomers to discover that the Leonids were	<u>material</u>	left behind Comet Tempel-Tuttle. In 1899

Figura 8: coapariciones de “material” y “Leonids”

De nuevo, igual que sucedía en el caso de los rasgos conceptuales, el corpus en castellano no arroja suficiente información como para ilustrar esta relación conceptual.

3.4. Relaciones conceptuales no jerárquicas

Las relaciones conceptuales no jerárquicas ahondan en las circunstancias que rodean al término y lo caracterizan. Profundizar en estas características facilita al traductor la comprensión tanto del término como también del ámbito temático sobre el que se documenta, ya que de este modo puede prever algunos de los argumentos que probablemente aparecerán en el texto que debe traducir o redactar.

3.4.1. Relaciones infralógicas

Entre las relaciones infralógicas nos interesaremos por las de carácter cronológico, las genéticas y las causales. Las expresiones que permiten obtener información cronológica sobre “Leónidas” que hemos identificado son las siguientes:

INGLÉS	CASTELLANO
During [EXPRESIÓN TEMPORAL] the [DENOMINACIÓN]	[DENOMINACIÓN] en [EXPRESIÓN TEMPORAL]
[DENOMINACIÓN] is in [EXPRESIÓN TEMPORAL]	[DENOMINACIÓN] de (ART.) [EXPRESIÓN TEMPORAL]
The [DENOMINACIÓN] (may/might) happen [EXPRESIÓN TEMPORAL]	[DENOMINACIÓN] durante [EXPRESIÓN TEMPORAL]
[DENOMINACIÓN] lasts [EXPRESIÓN TEMPORAL – PERIODO]	
[DENOMINACIÓN] (AUX.) occur in [EXPRESIÓN TEMPORAL - MOMENTO EXACTO]	
The [DENOMINACIÓN] occurs between/every [EXPRESIÓN TEMPORAL - PERIODO]	
[DENOMINACIÓN] take place [EXPRESIÓN TEMPORAL – MOMENTO EXACTO]	

Tabla 8: expresiones temporales en los contextos de “Leónidas”

En el caso de este término, el aspecto temporal parece tener mucha importancia, motivo por el que abundan las expresiones de este tipo, principalmente en inglés.

wers, the Perseids, Leonids and Geminids will all	<u>happen</u>	without any interfering moonlight during 19
mous showers like the Leonids and the Perseids	<u>happen</u>	once a year at about the same time.
basics of what the Leonids are and what might	<u>happen</u>	on November 17. Join our growin
occurrence. High activity from the Leonids can	<u>happen</u>	over several years every 33 years or so, but
nd so have lower impact velocities. The Leonids	<u>happen</u>	every year, but peak every 33 years after the
r Brown indicates a peak in bright Leonids may	<u>happen</u>	around 11h UT on November 18, with a main
rs predicted that Leonid meteor storm should be	<u>happen</u>	on 17/18 November 1998 and/or 1999
was wonderful. Leonid meteor storms do not	<u>happen</u>	often. Scientists do not know when the n
gain. However, Leonid meteor storms do not	<u>happen</u>	very often. In fact, scientists can not exa
Introduction The long-awaited storm did	<u>happen</u>	after all on 1999 November 17-18, with Leonid Euro
g, so there is a general belief that Leonid storms	<u>happen</u>	at 33-year intervals, though the chain is not un
entists can not exactly say when Leonids will	<u>happen</u>	for sure. Every 33 years, the earth passes the thro
s ahead of schedule. Why didn't a Leonid storm	<u>happen</u>	in 1998? The seven Leonid storms (displays fe

Figura 9: coapariciones de “Leonid” y “happen”

También en castellano aparecen diferentes expresiones que sitúan temporalmente los referentes de “Leónidas”, como la que se recoge a continuación:

de duración, comparable al máximo de las	<u>Leónidas en</u>	1966” dijo. Diez millones de toneladas de roca
spués del fenomenal retorno de leónidas	<u>Leónidas en</u>	1966, con una THZ (Tasa Horaria las Zenital: es
r más privilegiado para la observación de las	<u>Leónidas en</u>	1998 va a ser la parte oriental asiática. El máx
tormenta de (meteoros o estrellas fugaces)	<u>leónidas en</u>	1998 y/o en 1999. Por supuesto, no hay ning
del 17 nov 1998. El comportamiento de las	<u>Leónidas en</u>	1998 es muy similar al que se produjo en 1966
os obtenidos de la lluvia de meteoros las	<u>Leónidas en</u>	1998. El primer análisis de las observaciones r
/leon99.htm * Tormenta de meteoros: Las	<u>Leónidas en</u>	1999, Observatorio ARVAL (Venezuela). [+] ht

Figura 10: coaparición de “Leónidas” y “en” seguido de un año.

La información sobre el origen del término, el productor que lo genera o el producto que elabora o la consecuencia que provoca también proporciona información adicional sobre la materia sobre la que el traductor se documenta. Las expresiones que explicitan relaciones genéticas o causales identificadas en los corpóra inglés y castellano son las siguientes:

INGLÉS	CASTELLANO
[PRODUCTOR] is the source of [PRODUCTO]	[PRODUCTOR] producir [PRODUCTO]
[PRODUCTOR] is attributed to [PRODUCTO]	
[PRODUCTOR] is responsible for the [PRODUCTO]	
The [PRODUCTOR] responsible for the [PRODUCTO]	
[PRODUCTOR] produce the [PRODUCTO]	

Tabla 9: expresiones que permiten identificar relaciones genéticas y/o causales

Algunos de los contextos siguientes, por ejemplo, recoge información que se puede interpretar como consecuencia o producto de “Leonids”:

verage snowstorm”, while in the 1966 the <u>Leonids</u>	produced “a <u>hail of meteors</u> too numerous to count
re capable of creating such a deluge. The <u>Leonids</u>	produced a magnificent <u>storm</u> on the nights of Nov.
unate. Early indications are that the 1999 <u>Leonids</u>	produced a more intense display than had been pre
Met passed 2 years ago, and last year the <u>Leonids</u>	produced an astonishing number of <u>bright meteors</u> .
e peak location considerably. In 1998 the <u>Leonids</u>	produced an astonishing number of <u>bright meteors</u> .
ce for satellites. Thirty-two years ago, the <u>Leonids</u>	produced an estimated 150,000 <u>meteoroids per hour</u>
ching. In November the much-anticipated <u>Leonids</u>	produced an outburst of nearly 600 <u>meteors per hou</u>

Figura 11: coapariciones de “Leonid” y “produced a*”

Y lo mismo sucede en castellano:

Eoros se atribuye al espectacular despliegue <u>producido</u>	por las Leónidas de 1833. Se descubrió que las “Lluvias de Meteoros” a
Las Leónidas están <u>producidas</u>	por partículas de polvo que se mueven a 71 km/s en el espacio.
Aún si no se presenta, las Leónidas <u>producirá</u>	meteoritos a una tasa muy superior a cualquier otra lluvia periódica

Las Leónidas	<u>producen</u>	una gran “Tormenta de Meteoros” cada 33 años aproximadamente, co
n una lluvia de meteoros que en ocasiones	<u>producía</u>	tormentas de meteoros (miles de meteoros por hora)
atraviesa la región más densa de partículas	<u>produciendo</u>	el máximo de las Leónidas. Durante cualquier año esta lluvia presenta
la AAGC. Recuerda que este año se podría	<u>producir</u>	una tormenta en las Leónidas, con varios cientos de estrellas
Las Leónidas 2000 podrían	<u>producir</u>	abundantes meteoros. También que la abundancia de meteoros será ma
66. “Cada encuentro con una nube de polvo	<u>producirá</u>	un estallido de Leónidas en algún lugar de nuestro planeta”
nde el ángulo orbital de las Leónidas puede	<u>producir</u>	impactos (zona punteada). Es recomendable
enas expectativas para 1999, cuando puede	<u>producirse</u>	la verdadera tormenta de Leónidas. En cas
meteoros son en realidad fenómenos que se	<u>producen</u>	dentro de la atmósfera terrestre (por eso llevan ese nombre,
con tanta exactitud el momento en que se	<u>produciría</u>	el máximo de la tormenta. Las Leónidas de este año han sido menos bril
La madrugada del 17 al 18 de noviembre se	<u>producirá</u>	la importante tormenta de meteoros, conocida como las Leónidas.

Figura 12: coapariciones de “producir” y “Leónidas”.

Por otro lado, también hemos identificado expresiones puramente causales en los contextos de “Leónidas”. Son las siguientes:

INGLÉS	CASTELLANO
[EFECTO]. This is because [CAUSA]	[EFECTO] es (ART.) responsable de [CAUSA]
[CAUSA] is the cause of [EFECTO]	[CAUSA] causar [EFECTO]
[EFECTO] occur as [CAUSA]	[EFECTO] ocurrir [CAUSA]
[EFECTO] occur when [CAUSA]	

Tabla 10: expresiones que permiten identificar relaciones causales.

De nuevo, nos encontramos con formulaciones propias del lenguaje, no exclusivas del discurso especializado. Sin embargo, el conocimiento que vehiculan en este contexto no es de conocimiento general, tal como muestra el siguiente ejemplo del corpus inglés:

The Leonids	occur when	the Earth encounters debris left from comet Tempel-Tuttle, and are the cosmic equivalent of bugs hitting the windshield of an automobile.
The Leonids meteor shower	occurs when	the Earth crosses the wake of Comet Tempel- Tuttle. The comet throws out thousands of dust grains that slam into the Earth’s atmosphere at over 70 kilometres per second.

The Leonids storm will	occur when	the earth enters the dense debris behind comet 55P/Tempel-Tuttle over a roughly two-hour span on 17 November and again on 18 November 1999.
This maximum observed rate	occurs when	the Leonids peak at astronomical twilight with the radiant at 71 deg. elevation, vertically up from the Sun (which is at 18 deg. below the horizon).

Figura 13: coapariciones de “occur when” y “Leonids”

Del mismo modo que sucede en el ejemplo inglés, en el corpus castellano también hemos encontrado expresiones muy habituales en el uso de la lengua diaria que, no obstante, en este contexto proporcionan información especializada:

nidas . Dado que el cometa Tempel-Tuttle,	<u>responsable</u>	de la lluvia de Leónidas, recién cruzó la órbita
pel-Tuttle el 4 de marzo de 1997, cometa	<u>responsable</u>	de las leónidas, ha incrementado aún más las
teoros. El cometa 55P/Tempel-Tuttle es el	<u>responsable</u>	de Las Leónidas. El camino de polvo que deja
erseidas, etc). En el caso de las leónidas, el	<u>responsable</u>	es el cometa 55/P Tempel-Tuttle, el que al vi

Figura 14: coapariciones de “responsable” y “Leónidas”.

4. Conclusiones

En este artículo hemos descrito el uso que se puede hacer de los corpus *ad hoc* como recurso documental para la traducción especializada y también para la redacción técnica. La utilidad de este recurso depende de las estrategias de consulta que se sigan. Saber cómo extraer la información de un corpus de estas características es tan importante como saber compilarlo. Las expresiones recogidas en el apartado anterior no son más que un ejemplo de las estrategias que se pueden utilizar para extraer información de un corpus *ad hoc*. Un estudio similar sobre el entorno contextual de cualquier otro término, en los corpóra con los que hemos trabajado o en cualquier otro corpus, pondría de manifiesto otras expresiones que pueden dar lugar a nuevas estrategias de consulta tan productivas como las que hemos presentado.

Para el usuario de este tipo de recurso es tan importante la extracción de información concreta como el saber identificar y utilizar las estrategias de consulta apropiadas. De hecho, consideramos que la habilidad de compilar un corpus *ad hoc* y de utilizar las estrategias correctas está dentro de la competencia traductora en tanto que mecanismo de documentación. En cualquier caso, y siendo aun muy reducido el catálogo de expresiones que hemos presentado, consideramos que puede convertirse en un instrumento útil tanto en el aula de traducción especializada como en la formación general del traductor.

Bibliografía

- Alexander, J. and M. A. Tate (2001). *Evaluating web resources*. Wolfram Memorial Library. <http://www2.widener.edu/Wolfram-Memorial-Library/webevaluation/webeval.htm> (consulted: 15.8.2004).
- Baeza-Yates, R. And B. Ribeiro-Net (1999). *Modern Information Retrieval*. Addison Wesley: Reading.
- Caslon Analytics (2002). *Net metrics & statistics guide*. <http://www.caslon.com.au/metricsguide1.htm> (consulted: 15.8.2004).
- Codina, L. (1995). "La ciberingenuidad". *Boletín Euromonitor* 19, 11-12.
- Corpas, G. (2001). "Compilación de un corpus *ad hoc* para la enseñanza de la traducción inversa especializada". *Trans. Revista de traductología* 5, 155-184.
- Corpas, G. (2002). "Utilización de corpus multilingües en traducción: introducción al tipologizador textual automático para textos jurídicos". S. Gamero & A. Alcina (eds) *La traducción científico-técnica y la terminología en la sociedad de la información*. Castellón: Servicio de Publicaciones de la Universidad Jaume I, 155-162.
- Faber, P. (2002). "Investigar en Terminología". P. Faber and C. Jiménez Hurtado (eds) *Investigar en Terminología*. Interlingua. Granada: Editorial Comares, 3-23.
- Funredes (1998). *El segundo estudio de la cultura y la Internet*. Observatorio de las lenguas y las culturas. <http://funredes.org/LC/espanol/C2.html> (consulted: 15.8.2004).
- Funredes (2002). *El tercer estudio de la cultura y la Internet*. Observatorio de las lenguas y las culturas. <http://funredes.org/LC/espanol/C3/> (consulted: 15.8.2004).
- Gamero, S. (1998). *La traducción de textos técnicos (alemán-español)*. Géneros y subgéneros. Tesis doctoral. Universitat Autònoma de Barcelona.
- Hurtado Albir, A. (2001). *Traducción y traductología: introducción a la traductología*. Madrid: Cátedra.
- Larsen, L. (2001). *Information Literacy: The Web is not an Encyclopedia* [en línea]. University of Maryland. <http://www.inform.umd.edu/LigInfor/literacy/default.htm> (consulted: 5.12. 2001).
- Mayoral, R. (1997/1998). "La traducción especializada como operación de documentación". *Sendeban. Boletín de la Facultad de Traducción e Interpretación* 8/9, 137-153.
- Neumüller, M. (2001). *Hypertext Semiotics in the Commercialized Internet*. Tesis doctoral, Doktorat der Sozial- und Wirtschaftswissenschaften, Wirtschafts-universität Wien. 11 d'agost de 2001 [en línea]. <http://sammelpunkt.philo.at:8080/archive/00000023/> (consulted: 15.8.2004)
- Nord, C. (1991). *Text analysis in translation. Theory, methodology, and didactic application of a model for translation-oriented text analysis*. Amsterdam/Atlanta: Rodopi.
- PACTE (2000). "Acquiring translation competence. Hypotheses and methodological problems of a research project". A. Beeby *et al.* (eds) *Investigating translation. Selected papers from the 4th International Congress on Translation, Barcelona, 1998*. Amsterdam/Philadelphia: John Benjamins Publishing, 99-106.
- Sánchez-Gijón, P. (2004). *L'ús de corpus en la traducció especialitzada*. IULA, Grup Tradumàtica, Departament de Traducció i d'Interpretació.