

COMENEGO (Corpus Multilingüe de Economía y Negocios): corpus estable vs. metodologías *ad hoc* (*web as/for corpus*) aplicadas a la práctica de la traducción económica, comercial y financiera

DANIEL GALLEGO HERNÁNDEZ (*Universidad de Alicante*)

RAMESH KRISHNAMURTHY (*Aston University*)

Daniel Gallego Hernández (daniel.gallego@ua.es)

Universidad de Alicante

Dpto. de Traducción e Interpretación

Facultad de Filosofía y Letras

Universidad de Alicante

Apdo. Correos, 99 - 03080 Alicante – Spain

*La práctica de la traducción económica requiere desarrollar la competencia instrumental, que implica el uso de textos paralelos. Las actuales posibilidades tecnológicas permiten explotar dichos textos empleando metodologías de corpus. En este sentido, los corpus, concebidos como un conjunto de textos paralelos, pueden ayudar a satisfacer las necesidades informativas del traductor. Estos recursos pueden estar disponibles en Internet. En cambio, si el traductor se enfrenta a un texto cuyo campo no se encuentra entre los textos de dichos corpus ya compilados, es él quien puede buscar y compilar sus propios textos. En el caso de la traducción económica en francés y español, pocos son los corpus virtuales que pueden servir de recurso documental. COMENEGO puede ayudar a que el traductor reduzca el tiempo invertido con las metodologías *ad hoc* y explote directamente los textos. En este trabajo trataremos los temas relacionados con el diseño y la creación de este corpus, así como algunas de sus diferencias respecto de dichas metodologías.*

traducción económica, COMENEGO, textos paralelos, corpus multilingüe

*The practice of business translation requires the development of instrumental competence which involves using parallel texts. Current technological possibilities allow translators to exploit these texts by using corpus methodologies. In this sense, corpora conceived as sets of parallel texts may help the translator to satisfy his information needs. These corpora can be distributed via the Internet. However, if the translator is facing a text whose area of specialisation is not represented in the texts of the corpora currently available on the internet, he may compile his own *ad hoc* corpus. In the case of French-Spanish and Spanish-French business translation, there are very few corpora that can assist translation practitioners. COMENEGO may help the translator to save time when using *ad hoc* methodologies, by offering a ready-made variety of texts. The aim of this work is to discuss the issues relating to the design and creation of this corpus, as well as its differences from *ad hoc* methodologies.*

business translation, COMENEGO, parallel texts, corpus, multilingual

1. INTRODUCCIÓN

El traductor de textos económicos que trabaja en francés y español puede acceder, a través de internet, a una importante cantidad de información económico-financiera para satisfacer las

necesidades informativas que, en forma de problemas y dificultades de traducción, pueden presentarse durante su trabajo.

A día de hoy, las posibilidades tecnológicas permiten diversas estrategias para aprovechar al máximo estas fuentes de información. El traductor puede navegar por la web y consultar los textos paralelos (textos comparables respecto de la función, tema o situación comunicativa de los textos originales objeto de traducción) que se encuentra, utilizar la web como si fuera un corpus, compilar en su ordenador dichos textos y consultarlos con aplicaciones de análisis de corpus y, por supuesto, consultar los corpus ya compilados disponibles en línea.

Sobre utilizar la web como si fuera un corpus (*web as corpus*), son pocos los trabajos aplicados a la práctica de la traducción que se han llevado a cabo (Gallego Hernández, 2010a). Esta metodología supone emplear Google u otros buscadores como si fueran extractores de concordancias. Requiere el establecimiento de una serie de parámetros referidos tanto al texto original como a los textos paralelos, así como el uso estratégico de las funcionalidades y operadores del buscador utilizado. Los textos no se descargan en el ordenador, sino que se consultan a partir de los descriptores y resultados que devuelven los buscadores al interrogarlos con ecuaciones de búsqueda.

En cuanto a utilizar la web para compilar corpus (*web for corpus*), son diversos los trabajos que, en su aplicación a la práctica de la traducción, han debatido el tema (Bernardini & Zanettin, 2000; Corpas Pastor, 2002; Zanettin *et al.*, 2003; Sánchez Gijón, 2004; Beeby *et al.*, 2009; Rodríguez Inés, 2009; Gallego Hernández, 2010b). El diseño y la compilación de este recurso documental suponen, en esencia, reunir en un tiempo reducido una serie de textos paralelos cuya información término-fraseológica y conceptual permita al traductor resolver aquellos problemas o dificultades de traducción que puedan surgirle durante un encargo de traducción específico. En líneas generales, su compilación implica localizar los textos, descargarlos al ordenador, convertirlos a un formato reconocible por los programas de concordancias y, si cabe, limpiarlos.

Respecto de utilizar corpus ya compilados y disponibles en línea, los corpus de corte económico que pueden ser de aplicación a la traducción francés-español y español-francés escasean. En este sentido, el corpus técnico del IULA, aunque es de libre acceso, solo contiene un subcorpus español de economía de alrededor de un millón de palabras (Cabré & Martorell, 2004: 174). Por su parte, CLUVI permite consultar un subcorpus español de economía, EGAL (0,4 millones de palabras), y otro de consumo, CONSUMER (1,8 millones de palabras en español). Sin embargo, no tiene textos en francés. El MLCC Multilingual and

Parallel Corpora contiene un subcorpus genérico de artículos financieros de periódicos en francés y español, pero es de pago y los textos son de 1990-1994. Vicente (2007) posee un corpus representativo del lenguaje comercial en francés y español de la prensa generalista y especializada, pero es privado.

2. COMENEGO: CORPUS MULTILINGÜE DE ECONOMÍA Y NEGOCIOS

Ante este panorama, como formadores de traductores para el ámbito de la economía y los negocios, nos vemos obligados actualmente a implementar metodologías de consulta *ad hoc* de textos paralelos que implican destinar parte del tiempo disponible en el aula a su explicación y desarrollo. Por ello, nos parece interesante compilar COMENEGO, un corpus con el que, entre otras cosas, pretendemos reducir el espacio de tiempo dedicado al desarrollo de subcompetencias tecnológicas y aumentar el tiempo dedicado a la traducción de la especialidad.

El acrónimo se corresponde con *Corpus Multilingüe de Economía y Negocios*, donde *corpus* es, en términos de Sinclair (1996), «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», *multilingüe* equivale a francés y español, al tiempo que deja espacio a otras lenguas, y *economía y negocios* puede entenderse en el sentido de Mateo Martínez (2007), referido al lenguaje de la economía teórica, es decir, un lenguaje académico propio de expertos investigadores que utilizan un discurso poco transparente de gran especialización terminológica, y al de los negocios, un lenguaje profesional propio de usuarios no solo teóricos, sino también especialistas del mundo del comercio y las finanzas, periodistas e incluso personas sin ninguna especialización.

2.1. Finalidad de COMENEGO

Como acabamos de comentar, COMENEGO surge de la necesidad que, como formadores de traductores, tenemos de reducir el tiempo dedicado a la compilación de corpus *ad hoc* tanto en el aula como en casa. Ahora bien, el corpus no solo está concebido como recurso de documentación o fuente lingüística especializada. También puede servir como herramienta docente (detección de terminología y fraseología de interés, elaboración de

ejercicios de terminología, traducción o revisión, etc.), así como objeto de análisis con fines de investigación.

2.2. Fases en la compilación de COMENEGO

En líneas generales, la compilación de un corpus implica diferentes fases que, en conjunto, suponen un proceso cíclico: 1) documentación (identificación, por ejemplo, en una hoja de cálculo, de las características esenciales de los textos: idioma, variedad, campo, modo, etc.); 2) selección de textos en función de la finalidad y disponibilidad del corpus, de criterios externos relativos a características textuales como el idioma y sus variedades, el modo (discursos orales, escritos, textos electrónicos, etc.), el tipo de texto (libros, artículos, etc., en el caso de textos escritos; seminarios, presentaciones, etc., en el caso de discursos orales), la especialidad, el año y otras características; o de criterios internos fundamentados en las características propias de los textos; 3) obtención de permisos de usos (si se va a distribuir el corpus); 4) adquisición de textos (descarga de archivos, inclusión de metadatos, etc.) y 5) conversión a texto plano y limpieza de archivos (eliminación de ruido, caracteres no reconocidos, etc.).

En nuestro caso, los recursos textuales vienen acompañados de una hoja de cálculo que contiene diversos campos (URL, fecha de descarga, tipo de texto, etc.). La selección de textos, como hemos comentado, responde fundamentalmente a fines de formación y práctica de la traducción. La tabla 1 del apéndice contiene ejemplos de los tipos de textos recuperados, así como las categorías en las que nos hemos basado para agruparlos en diferentes subcorpus.

Algunas de estas categorías se corresponden con los tipos funcionales de comunicación en las organizaciones que, según unos criterios pragmático-discursivos (función, interlocutores, estructura y estilo), identifica Cassany (2004: 53-55). Se trata, en concreto, de las categorías *técnico (TEC)* y *científico (SCI)*, que tienen una función principalmente referencial, intentan transmitir información de manera objetiva y se dan en áreas técnicas de la organización (proyectos, auditoría, investigación). Nuestra clasificación también tiene en cuenta su concepción del discurso *organizativo (ORG)*, que puede tener una función conativa, referencial o metalingüística, trata de ordenar y regular la actividad de la organización, y suele tener cabida en áreas concretas (dirección, personal, administración, evaluación, calidad). Aunque el autor incluye dentro de esta misma categoría lo que denomina *lenguajes*

administrativo y jurídicos, hemos creado una categoría aparte, *legal (LEG)*, muy relacionada con la organizativa, pero referida, en esencia, a textos que tienen como objetivo no regular la actividad de una organización concreta, sino del conjunto de organizaciones de uno o varios países. De su clasificación también hemos recuperado el discurso *comercial (COM)*, cuya función ronda entre la conativa y la referencial, pretende influir sobre la opinión y la conducta del destinatario, y suele ser propio de áreas específicas (*marketing*, publicidad, comunicación, ventas).

A estas categorías hemos añadido otras dos, *didáctica (DID)* y *prensa (PRS)*, concebidas principalmente según criterios pragmáticos. La categoría *didáctica* se refiere a textos que surgen del círculo formativo y que, por tanto, tienen unos fines de formación no solo en el ámbito académico, como pueden ser cursos o apuntes, sino también en el ámbito profesional, como pueden ser textos que explican al inversor qué es la bolsa o guías para consumidores. La categoría *prensa*, por su parte, se corresponde con textos de carácter informativo de corte económico, comercial o financiero, publicados por la prensa general y especializada, así como con los comunicados o notas de prensa de las organizaciones.

La tabla 2 del apéndice muestra la distribución de estas categorías en ambos idiomas según el número de archivos (*files*), palabras (*tokens*), media de palabras por archivo (*ave*), tipos de palabra (*types*) y ratio *token-type (ratio)*. Estas cifras se refieren a los archivos TXT convertidos y limpiados (ruido y caracteres no reconocidos). Respecto de la obtención de permisos de usos, hemos llevado a cabo algunas peticiones que han recibido una respuesta afirmativa. En cualquier caso, se trata de un proceso todavía por completar, que probablemente hará variar estas cifras.

2.3. Análisis de COMENEGO

La lingüística de corpus dispone de diversas herramientas de análisis, como WordSmith Tools, AntConc (utilizado en el presente análisis), etc., que permiten extraer diversos datos en forma de 1) listados de palabras frecuentes: «the rank of a word-form in a corpus frequency list has some relationship to the importance of that word-form in the linguistic system» (Krishnamurthy, 2001), 2) concordancias (listados de palabras clave contextualizadas, que pueden ordenarse o expandirse según las necesidades del análisis), 3) colocaciones (otras palabras que aparecen inesperadamente con más frecuencia alrededor de palabras clave) y 4) n-gramas (listados de secuencias de palabras más frecuentes).

A continuación presentamos un breve análisis preliminar que puede servir para ilustrar el tipo de análisis que es posible llevar a cabo en investigaciones futuras, al tiempo que resaltamos en algunos puntos la necesidad de dar nuevos pasos en el estudio del corpus.

En la tabla 2 del apéndice podemos ver las similitudes de cada corpus respecto de su tamaño (9115352 *tokens* en español frente a 9086627 en francés), pero hemos de tener en cuenta que el cómputo de palabras puede verse afectado por las diferencias morfológicas de cada idioma.

El tamaño de los subcorpus o categorías es muy parecido: una media de 1,3 millones de palabras con variaciones de entre 1,19 y 1,37 millones. Estos datos pueden verse afectados por las posibles diferencias en la tipología textual de cada lengua.

Ahora bien, existen algunas diferencias significativas entre el corpus español y el francés: 1) el español tiene 10050 archivos, pero el francés solo 8880, lo que significa que los textos franceses son, de media, más largos que los españoles (1023 palabras frente a 907 palabras); 2) aun así, hay variaciones en las diferentes categorías: mientras que los subcorpus COM, DID, LEG y TEC tienen más textos en español que en francés (5255/3909, 1491/1121, 211/21 y 351/133, respectivamente), y, por tanto, los textos españoles son más cortos, los subcorpus ORG, PRS y SCI tienen, en cambio, más textos en francés que en español (634/429, 2859/2214 y 203/99, respectivamente), y, por tanto, sus textos en francés son, de media, más cortos que los españoles; por último 3) los textos más cortos en ambas lenguas pertenecen al subcorpus COM; por su parte, los más largos en español se encuentran en el subcorpus SCI, y los más largos en francés, en el subcorpus LEG.

2.3.1. Palabras frecuentes

En cualquier idioma las palabras más frecuentes suelen ser palabras gramaticales, como determinantes, preposiciones, etc. Sin embargo, precisamente por su elevada frecuencia, su polifuncionalidad y su complejidad de uso, conviene analizarlas mejor con más detenimiento. Por ello nos centramos ahora en las palabras no gramaticales más frecuentes, también denominadas *palabras de contenido* (*content words* o *vocabulary words*). La tabla 3 del apéndice contiene las 15 palabras más frecuentes en cada idioma (nótese que hemos optado por no distinguir entre mayúsculas y minúsculas). Este tipo de listados puede ayudar a identificar, por una parte, las palabras que los traductores deberían conocer dentro de un amplio abanico de contextos, significados y usos, y, por otra parte, un elevado número de palabras que, por su frecuencia mínima dentro del corpus, pueden, en principio, no resultarles de tanto interés.

Observemos, en primer lugar, los ítems similares en cada lengua. Podemos ver que 7 de las 15 palabras más frecuentes de cada corpus son cognados y, por tanto, los primeros equivalentes que proponen la mayoría de los diccionarios bilingües (ahora bien, hemos de tener en cuenta que pueden esconder diferentes significados y usos en ambas lenguas). También apreciamos algunas variaciones menores: en español, la forma singular *cuenta* coincide con las formas *compte* en singular y *comptes* en plural. Por supuesto, debemos consultar el listado completo de formas en español, pues *cuentas* aparece con una frecuencia algo menor. Ocurre algo parecido con el plural *empresas*, representado en francés con su forma en singular (*entreprise*). En análisis posteriores, será importante comparar un mayor número de ítems y analizar de manera más precisa las diferencias en la posición que ocupan.

Observemos ahora los ítems que aparecen en un solo idioma: *información, mercado, valores, millones, general, valor, riesgo*, en el caso del español, y *assurance, actions, conseil, france, conditions, ans*, en el caso del francés. De nuevo, deberemos analizar los listados de palabras completos para comprobar si se trata de diferencias significativas o de simples variaciones menores de frecuencia en cada corpus.

2.3.2. Concordancias

Como dijimos anteriormente, las concordancias permiten observar cualquier ítem aparecido en los listados de palabras frecuentes, tal como aparece en los textos, con los contextos que le rodean. Se trata del nivel de análisis más detallado, y permite investigar cada palabra en cada texto. Con ellas es posible identificar clases de palabras, significados, usos, colocaciones y fraseologismos, patrones gramaticales, usos pragmáticos y características específicas de los géneros textuales.

Incluso en el pequeño listado de 10 de las 11879 concordancias de *cuenta* (tabla 4 del apéndice), podemos apreciar algunos patrones: 4 líneas contienen el verbo *cuenta* seguido de *con*, 2 tienen la secuencia *de ser cliente de cuenta Nómina*. En futuros análisis, el estudio de la totalidad de sus contextos ayudará a determinar qué patrones referidos a *cuenta* pueden ser de utilidad para los traductores, extraer la terminología especializada de interés, etc.

También apreciamos algunos patrones en el caso de las 12 concordancias en francés (de 10342) aparecidas en la tabla 5 del apéndice: 4 vienen seguidas de la preposición *de*, 3 de las cuales vienen precedidas de *tenir/tenant*, y una, de *rendant* (también encontramos un ejemplo de *compre rendu*); 3 ejemplos de *votre compre*, dos de los cuales vienen seguidos de *sera* y uno, precedido de *sur*; un ejemplo con *compte-titres*; además *titres* aparece en otra posición dentro de otra línea; un ejemplo con *ouvrir un compte* e incluso uno con el verbo *qui compte*.

Nuevamente, una investigación más detallada revelará qué patrones son significativos y cuáles no. Por su parte, la elevada frecuencia de la forma en plural *comptes* (9149) también puede ayudar a identificar distintos usos de esta forma.

2.3.3. Colocaciones

Como acabamos de apreciar, las concordancias pueden ayudarnos a obtener diferentes informaciones sobre las características y comportamientos de una palabra. Ahora bien, también existen otras herramientas que pueden acelerar los procedimientos iniciales de detección de información. Se trata de las colocaciones, una herramienta de mayor complejidad que requiere del analista conocimientos lingüísticos avanzados. Esta función ofrece un análisis cuantitativo de las palabras que aparecen en las proximidades del núcleo o palabra clave. Algunos trabajos, como el editado por Krishnamurthy (2004), muestran que los colocativos más significativos suelen aparecer dentro de una distancia de 4 palabras alrededor del núcleo o palabra clave. AntConc nos permite seleccionar esta distancia.

Si observamos el listado de la tabla 6 del apéndice, observamos que parece haber menos correspondencias formales que en el listado anterior de palabras frecuentes (tabla 3): solo *corriente-courant* y *depósito-dépôt*. Tras un examen más detenido, apreciamos dos equivalentes semánticos potenciales: *ahorro-épargne* y *vivienda-logement*, así como la presencia del nombre de entidades bancarias: *Caixa-Bred*. El hecho de que no sea posible asociar fácilmente los colocativos de una y otra lengua sugiere que las palabras *cuenta* y *compte* se emplean en diferentes contextos y fraseologismos en ambos idiomas. Ello demuestra la necesidad de ser cautos en futuros análisis: las palabras con similitudes formales/etimológicas pueden ser falsos amigos.

Existen otros aspectos de la herramienta de colocaciones que pueden utilizarse para futuros análisis, como la posibilidad de determinar exactamente en qué posición aparece cada colocativo respecto de la palabra clave o núcleo.

2.3.4. 4-gramas

Otra herramienta que puede ser útil para identificar las características lingüísticas de los textos son los n-gramas. En lugar de crear listados de palabras frecuentes, esta herramienta crea listados de secuencias de dos palabras (2-gramas), tres palabras (3-gramas), etc.

Tal como ocurre con los listados de colocativos (tabla 6), los listados de 4-gramas aparecidos en la tabla 7 del apéndice muestran más diferencias que similitudes. Ahora solo queda un ítem paralelo del listado de palabras frecuentes (tabla 3): *artículo-article*. Algunas palabras continúan apareciendo en una de las listas, pero no en la otra: por ejemplo, *euros*

aparece en el listado español, pero no en el francés; *comptes* sigue estando en el listado francés, el verbo *compter* aparece en su forma infinitiva, pero *compte* desaparece, tal como hace *cuenta* en el listado español. Asimismo, en este nuevo listado aparecen ahora nuevos ítems, como *cadre*, que no aparecían en los listados de palabras frecuentes. Nótese que hemos tenido en cuenta la diferencia entre mayúsculas y minúsculas para mostrar que ahora aparecen algunos nombres propios de instituciones y órganos de gobierno. También es posible identificar algunos cognados: *dispuesto-dispositions*.

3. CONCLUSIONES

En este trabajo hemos presentado COMENEGO en contraposición con metodologías *ad hoc web as/for corpus* y hemos llevado a cabo un análisis preliminar con herramientas de explotación de corpus.

Respecto de sus diferencias con las metodologías *ad hoc web for corpus*, creemos que, con COMENEGO, tenemos posibilidades de omitir, entre otras cosas, la fase de compilación tanto para traductores como para formadores. En consecuencia, es posible disponer de más tiempo para dedicar a la práctica de la traducción propiamente dicha y a la planificación docente. Asimismo, el desarrollo de la competencia instrumental se ve reducido en su dimensión tecnológica, pues, una vez compilado y disponible el corpus, los usuarios no requerirán de conocimientos informáticos avanzados. Otro punto que puede jugar a favor es la fiabilidad de los textos, que puede ser mayor respecto de las metodologías *ad hoc*, con las que es posible recuperar textos no fiables o ruidosos. Desde el punto de vista del acceso a los recursos documentales, la centralización de los textos en un servidor puede suponer una ventaja respecto de las metodologías *ad hoc*, que requieren tener en el ordenador del usuario no solo los archivos del corpus, sino también el *software* de lectura. Además, la explotación de corpus puede verse mejorada, al menos respecto de las metodologías *web as corpus* que emplean buscadores comerciales y cuyas estrategias de búsqueda (truncación, palabras en contexto, visualización de resultados, etc.) suelen verse limitadas.

Por su parte, el análisis presentado tan solo es la punta del iceberg y muestra el tipo de estudio que puede llevarse a cabo sobre COMENEGO. No hemos querido ir más allá, pues todavía estamos ajustando y equilibrando el corpus y sus componentes. Ahora que disponemos de un corpus piloto, podemos embarcarnos en un análisis global más riguroso

aprovechando al máximo las herramientas de explotación de corpus. Las observaciones y datos mostrados simplemente dejan entrever la variedad de resultados que pueden obtenerse con el análisis, así como sus potenciales beneficios para los traductores.

En cualquier caso, COMENEGO es un corpus piloto todavía en construcción. El análisis de sus distintos subcorpus puede ayudar a definir con más precisión las categorías textuales establecidas. Es posible complementar la selección de textos incluyendo nuevas tipologías a partir, por ejemplo, de encuestas dirigidas tanto a los iniciadores de traducción como a los propios traductores profesionales con el propósito no solo de conocer sus necesidades de traducción, sino también de conocer qué tipos textuales, temas, campos, etc., son los que suelen traducirse en el ámbito profesional. Por otra parte, las funciones del corpus son, de momento, limitadas, pues la fase de solicitud de permisos de uso de los recursos textuales no está completada, lo que hace que se trate actualmente de un corpus privado con fines de investigación y de uso propio, y tampoco se dispone de una plataforma o sistema virtual que permita la consulta de los textos.

4. APÉNDICE

Tabla 1: categorías textuales en COMENEGO

TIPOS DE TEXTOS	CATEGORÍAS
promoción de productos bancarios, productos financieros y de seguros, páginas web corporativas, etc. (sitios web comerciales)	comercial
cursos en línea, guías para consumidores, inversores y clientes, etc. (sitios web comerciales e informativos: personales de profesores, universidades, instituciones)	didáctico
leyes, códigos, decretos, etc. (sitios web informativos: ministerios y agencias)	legal
estatutos, reglamentos, actas de juntas, etc. (sitios web corporativos e informativos)	organizativo
notas de prensa, noticias, boletines de noticias, etc. (sitios web corporativos y de periódicos)	prensa
artículos y trabajos, etc. (sitios web informativos: revistas especializadas)	científico
cuentas e informes anuales, resultados, informes de gestión, análisis técnicos, planes de marketing, informes sectoriales, etc. (sitios web corporativos e informativos)	técnico

Tabla 2: distribución de recursos en COMENEGO

CAT	ESPAÑOL					FRANCÉS				
	FILES	TOKENS	AVE	TYPES	RATIO	FILES	TOKENS	AVE	TYPES	RATIO
COM	5255	1329915	253	35321	37,65	3909	1325544	339	29316	45,21
DID	1491	1276089	856	40641	31,39	1121	1304585	1164	35937	36,30
LEG	211	1342698	6363	23077	58,18	21	1293704	61605	14772	87,57
ORG	429	1337822	3118	29417	45,47	634	1365468	2154	21885	62,39
PRS	2214	1329029	600	37314	35,61	2859	1308418	458	37928	34,49
SCI	99	1311731	13250	34483	38,03	203	1301102	6409	32710	39,77

TEC	351	1188068	3385	40777	29,13	133	1187806	8931	24646	48,19
TOTAL	10050	9115352	907	113100	80,59	8880	9086627	1023	89133	101,94

Tabla 3: listado de palabras frecuentes no gramaticales en COMENEGO

ESPAÑOL			FRANCÉS		
RANK	FRECUENCIA	PALABRA	RANK	FRECUENCIA	PALABRA
29	17540	información	28	26843	article
30	17531	mercado	41	15921	cas
31	17339	artículo	42	15294	société
35	15557	euros	50	11854	capital
36	15015	caso	52	11535	assurance
38	13921	sociedad	55	10898	actions
39	13582	valores	61	10379	entreprise
40	13378	millones	62	10342	compte
41	12421	capital	63	10335	conseil
42	12407	general	69	9556	france
43	11879	cuenta	70	9532	groupe
45	11777	valor	73	9149	comptes
47	11358	empresas	74	9027	conditions
49	11141	grupo	76	8821	ans
50	10969	riesgo	78	8607	euros

Tabla 4: listado de concordancias en español (*cuenta*)

...so de ser cliente de cuenta Nómina) o con Certifi...
 ...eneral de la Policía cuenta con 350 Oficinas de E...
 ...so de ser cliente de cuenta Nómina) o con el Cert...
 ...comendamos tengan en cuenta. Ante cualquier duda,...
 ...total o parcial, por cuenta de gobiernos o autori...
 ... capital liberada. A cuenta Complementario Total ...
 ...07/07/880,1800,144 A cuenta Complementario Total ...
 ...d, abertis logística cuenta con cerca de 933.000 ...
 ...dad El Grupo abertis cuenta con una plantilla med...
 ... de los trabajadores cuenta con contrato indefini...

Tabla 5: listado de concordancias en francés (*compte*)

...st ajusté pour tenir compte de la modification du...
 ...inatif pur Ouvrir un compte au nominatif pur Pour...
 ...'Administration, qui compte désormais 12 membres ...
 ... du dividende. Votre compte sera crédité dans les...
 ...irectement sur votre compte. Au nominatif adminis...
 ...e 17 mai 2010. Votre compte sera crédité dans les...
 ... suivants pour tenir compte des délais de traitem...
 ...ts et en lui rendant compte de son examen : organ...
 ... d'une action tenant compte des opérations ayant ...
 ...ier pour la tenue du compte-titres. Ils représent...
 ...s titres inscrits en compte nominatif pur. Droit...
 ...umenté au moyen d'un compte rendu écrit Avec le ...

Tabla 6: listado de colocativos (*cuenta y compte*)

COLOCATIVOS DE CUENTA	COLOCATIVOS DE COMPTE
nómina	unités
corriente	bancaire

servicio	épargne
naranja	titres
vivienda	bred
caixa	dépôt
condiciones	courant
ahorro	logement
producto	numéro
crédito	livret
vista	ouverture
tarjeta	gestion
depósito	titulaire
valores	relevés

Tabla 7: listado de 4-gramas

ESPAÑOL	FRANCÉS
de millones de euros	dans le cadre de
del Mercado de Valores	du Code de commerce
de la Ley de	dans les conditions prévues
la Ley de de	à compter de la
del Consejo de Administración	alinéa de l'article L
Nacional del Mercado de	Crédit Agricole S A
Comisión Nacional del Mercado	de commerce et d'industrie
la Comisión Nacional del	droit préférentiel de souscription
el Consejo de Administración	en application de l'article
en el caso de	valeurs mobilières donnant accès
artículo de la Ley	la mise en place
Consejo de Administración de	par lettre recommandée avec
el artículo de la	dans la limite de
los millones de euros	l'Autorité des marchés financiers
lo dispuesto en el	le commissaire aux comptes
en el artículo de	des commissaires aux comptes
En el caso de	le cadre de la
que se refiere el	dispositions de l'article L

5. BIBLIOGRAFÍA

Beeby, A.; Rodríguez Inés, P. & Sánchez Gijón, P. (Eds.). (2009). *Corpus Use and Translating*. Amsterdam/Philadelphia: John Benjamins.

Bernardini, S. & Zanettin F. (Eds.). (2000). *I corpora nella didattica della traduzione: Corpus use and learning to translate*. Bologna: CLUEB.

Cabré Castellví, M. & Bach Martorell, C. (2004). El corpus tècnic del IULA: corpus textual especializado plurilingüe. *Panacea*, 16, 173-176.

- Cassany, D. (2004). Explorando los discursos de las organizaciones. In van Hoofst Comajuncosas, A. (Ed.). *Textos y discursos de especialidad. El español de los negocios* (pp. 49-70). Amsterdam/New York: Rodopi.
- Corpas Pastor, G. (2002). Traducir con corpus: de la teoría a la práctica. In García Palacios, J. & M.^a T. Fuentes Morán (Eds.). *Texto, terminología y traducción* (pp. 189-226). Salamanca: Almar.
- Gallego Hernández, D. (2010a). *Traducción económica y textos paralelos en internet. Aproximación teórica y metodológica*. Tesis doctoral, Universidad de Alicante.
- Gallego Hernández, D. (2010b). Acquiring instrumental sub-competence by building do-it-yourself corpora for business translation. Paper presented at the *Using Corpora in Contrastive and Translation Studies*.
- Krishnamurthy, R. (2001). Size Matters: creating Dictionaries from the World's Largest Corpus, pp 169-180 in 8th Annual KOTESOL Conference Proceedings, Taegu: KOTESOL.
- Krishnamurthy, R. (Ed.) (2004). English collocation studies: The OSTI Report by John Sinclair, Susan Jones and Robert Daley. London: Continuum Books.
- Mateo Martínez, J. (2007). El lenguaje de las ciencias económicas. In Alcaraz Varó, E., *et al.* (Eds.). *Las lenguas profesionales y académicas* (pp. 191-203). Barcelona: Ariel.
- Sánchez Gijón, P. (2004). *L'ús de corpus en la traducció especialitzada: compilació de corpus ad hoc i extracció de recursos terminològics*. Barcelona: Universitat Pompeu Fabra.
- Sinclair, J. (1996). Preliminary recommendations on Corpus Typology. Retrieved from <http://www.ilc.cnr.it/EAGLES/corpusstyp/corpusstyp.html>.
- Vicente, C. (2007). Lingüística de corpus y traducción especializada: aplicaciones a la traducción francés-español de la economía. Paper presented at the *XXV Congrès international de linguistique et de philologie romanes*.