

LA PECULIARIDADES TEXTUALES DE LAS PÁGINAS WEB COPORATIVAS TRADUCIDAS AL ESPAÑOL.

Miguel A. Jimenez Crespo

University of Granada, Spain

University of North Carolina at Wilmington

Abstract: a translated webpage is textually different from the one originally developed in the target language. The paper studies the textual, cultural and pragmatic differences in a small parallel corpus of corporate web pages developed in Spanish and translated into Spanish. The use of parallel corpora in translation is discussed along with strategies to better localize web pages into several languages.

Introducción

El proceso de traducción de páginas Web está mediatizado por el acceso al texto original. Dependiendo del contenido, al traductor se le puede proporcionar un texto que se puede considerar como completo, tal como la traducción del cuerpo de una página; segmentos textuales independientes como los que aparecen en los *scripts* que pueden aparecer en cadenas sueltas, o simplemente segmentos de una sola palabra como ocurre en la traducción de menús de navegación, menús desplegables, etc. (Esselink, 2001:57-58). Estas peculiaridades afectan al proceso cognitivo de la traducción de manera especial, puesto que a menudo conllevan un mayor aislamiento de las cadenas y una falta de contexto que, según nuestra hipótesis, se refleja en el producto, el sitio web traducido.

El centro del presente estudio es la recopilación y análisis de un pequeño corpus comparable (Baker, 1995) de páginas web en español y su posterior análisis.

En primer lugar, necesitamos aclarar el concepto de corpus. Una de las definiciones más aceptadas de corpus es la propuesta por Sinclair (1991:171), “[...] a corpus is a collection of naturally-occurring language text, chosen to characterize a state or variety of a language”. Leech (1992: 106) introduce el concepto de corpus informatizado de la siguiente manera: “On the face of it, a computer corpus is an unexciting phenomenon; a helluva lot of text, stored on a computer”. Nuestro corpus lo definimos como un corpus comparable según la definición de

Baker (1995), que es “ a structured electronic collection of texts originally written in a particular language, alongside texts translated into that same language”¹.

El primer núcleo del corpus está compuesto por páginas web traducidas al español de empresas multinacionales con intereses en España. El segundo núcleo comparable de nuestro corpus lo componen páginas web en español de empresas españolas redactadas originalmente en español. Cabe destacar las limitaciones de nuestro corpus, puesto que éste ha sido recopilado por un solo investigador, el autor de este trabajo, y se discutirá con detenimiento en la sección de diseño de corpus las implicaciones teóricas y prácticas de las mismas. Con ayuda del paquete de software Wordsmith tools, desarrollado por Mike Scott (1996), se estudiaron las peculiaridades léxicas y comunicativas de los textos traducidos comparándolos con los textos originales producidos en español. Los estudios de corpus que se centran en las características de los textos traducidos comparados con los textos originalmente producidos en el idioma meta cuenta con varias investigaciones, destacando los estudios de Baker (1995, 2001) sobre los “universales de la traducción” en textos traducidos al inglés, y continuado por otros investigadores que han estudiado las peculiaridades textuales de las traducciones en otros idiomas (Toury, 1980; Puurtinen, 1995; Laviosa, 1996, 1997, 1998). En estas investigaciones con corpus comparables se ha demostrado que en los textos traducidos suelen aparecer estructuras y frecuencias que no son típicas, si los comparamos con textos que se han producido originalmente en la misma lengua (Gellerstain, 1996; Laviosa-Baithwate, 1996).

El primer paso en cualquier investigación basada en corpus es especificar exactamente qué es lo que se desea estudiar (Kenny, 1998); nuestra hipótesis de trabajo se basa en que en las páginas traducidas aparecerá estadísticamente unas frecuencias diferentes de unidades léxicas y colocaciones de las páginas españolas, debiéndose a una influencia clara de la lengua de origen, el inglés en nuestro caso. Esto se debe a que la traducción es un “communicative event which is shaped by its own goals, pressures and context of production” (Baker, 1996:175), por lo que los textos que se producen en de esta manera deben mostrar rasgos lingüísticos distintivos que se pueden observar con técnicas de corpus (Zanettin, 2000. Puesto que al traductor se le presentan

¹ Esta idea de Baker se ha cristalizado en la creación del English Comparable Corpus (ECC) que consiste en dos colecciones de textos, una es el Translational English Corpus (TEC) de textos traducidos al inglés desde diferentes idiomas, y otro corpus denominado Non-Translational English Corpus (NON-TEC) compuesto por textos originales en inglés de similares características producidos durante el mismo periodo de tiempo.

los segmentos más aislados y a menudo con contexto limitado², le es más difícil construir su plan global mental de la traducción que se refleja al procesar cada segmento. Al ser éstos más aislados, existe una mayor tendencia a realizar una traducción asociativa directa, más horizontal, y no incluir en el procesamiento cognitivo controlado del segmento más esquemas mentales de la lengua y cultura meta. Así, las traducciones que se producirán se perciben como correctas gramaticalmente y semánticamente tanto por el traductor en su actividad metacognitiva de planeamiento, revisión y evaluación (Ericsson y Simmon, 1984), como por los receptores de la traducción, pero mostrarán una mayor influencia del texto origen.

Objetivos del estudio

Los objetivos de estudio son los términos y las colocaciones que aparecen en todas las páginas web y que cumplen las funciones pragmáticas básicas de navegación. Por colocación entendemos: “ Collocation is the occurrence of two or more words within a short space o each other in a text” (Sinclair, 1991)³. Los términos y colocaciones que aparecen en las páginas en inglés suelen estar muy fijados, y los mismos aparecen casi siempre en los menús de navegación de las páginas producidas en un contexto anglosajón. En nuestro estudio nos centramos en la función de contactar con la empresa, “contact us”, en la opción conocer mejor la empresa, “about us”, en conocer las directrices de confidencialidad o de privacidad de la misma, “privacy policy”, en indicar las condiciones de uso del sitio web, “terms of use”, y en la vuelta a la página principal⁴, “home”. En contraste con la gran fijación que existe en estos términos y colocaciones en inglés, la variación en español es mayor, existiendo un claro contraste entre las páginas traducidas y las páginas originadas en España. Según nuestra hipótesis, el proceso de traducción tendrá una influencia clara en la elección de las unidades léxicas (Baker, 1995), y estos textos traducidos presentarán frecuencias diferentes a las páginas originadas en España, así como unidades léxicas que no aparecieran nunca en textos originados en español. Debemos destacar en este punto, que al ser la cultura anglosajona dominante en el campo de Internet, aparecerá

² En el caso de sitios web corporativos, los programas de actualización con memorias de traducción extraen los segmentos que no tienen traducción y presentan al traductor solamente estos segmentos independientes.

³ Lopez Rodriguez (2001: 104) define colocación de la siguiente manera: “ Colocación es la ocurrencia de palabras cercanas en un texto o la relación que establece un lexema con otros lexemas que aparecen con una probabilidad significativa en su co-texto inmediato. Normalmente, en lingüística de corpus (Sinclair, 1991), este cotexto abarca cuatro palabras a ambos lados del lexema en cuestión.”

⁴ Tomamos la terminología de los glosarios Microsoft que consideramos es la más extendida.

asimismo una influencia en las páginas producidas en España por cuanto a la terminología empleada.

Bases Teóricas Para La Creación Del Corpus Comparable

Nuestro corpus lo consideramos un corpus comparable⁵ monolingüe (Baker, 1995). En nuestro caso, contamos con un corpus de páginas Web traducidas al español, todas ellas pertenecientes a sitios web de empresas multinacionales con intereses en España, junto con otro corpus de páginas web que se han desarrollado y escrito originalmente en España usando el español⁶, este último componente de textos se ha modelado según la composición del primero (Laviosa, 1997:293).. Los corpórea comparables no han estado libres de críticas, puesto que al constar de textos traducidos, se alega que los mismos no presentan el lenguaje natural en su uso real sino textos influenciados por el texto origen (Hartmann, 1994). Nuestro objeto de estudio es precisamente el estudio de esta influencia del texto origen, con lo que queda más justificada si cabe la recopilación de nuestro corpórea comparable.

En nuestro caso no contamos con las páginas originales en inglés, lo que hubiera convertido a nuestro corpus en un corpus paralelo⁷ (McEnery, 1996:58), puesto que nuestro objetivo de estudio es el producto final (páginas traducidas) y las peculiaridades del texto traducido al español.

Como se indica anteriormente, el corpus está compuesto por dos componentes, el corpus de páginas traducidas al español que cuenta con 71 páginas Web de inicio y el corpus comparable de páginas originales en español con 84 páginas Web de inicio. Ambos se encuentran en formato HTML. El corpus ha sido recopilado por un solo investigador, el autor del presente estudio. Somos conscientes de las limitaciones del presente corpus puesto que ,como indica Sinclair (1991), el estudio de colocaciones y otras estructuras requiere normalmente usar corpórea de gran

⁵ En la lingüística de corpus nos encontramos con otra definición diferente de corpus comparable (Peters et al. 1996: 69), siendo éste un conjunto de textos producidos en diferentes lenguas que coinciden en el la motivación, el tema, la función comunicativa y que se usan para comparar dos o más lenguas o para la extracción de terminología.

⁶ Debido a la internacionalización (i18n) y globalización (g10n), son varias las empresas españolas que han desarrollado originalmente sus sitios web en inglés, adaptándolos más tarde al español.

⁷ Un corpus paralelo es aquel en el que se presenta el mismo texto en más de una lengua, siendo una de ellas la lengua de origen y las demás las lenguas meta.

tamaño, más aún cuando el objeto de nuestro análisis, las cadenas y los menús de navegación que aparecen en todas las páginas web es un campo de estudio muy limitado.

Así nos embarcamos en la recopilación de este corpus, que se ha guiado por diversos criterios. Los criterios de representatividad, estandarización y tipología de los corpórea han sido tres conceptos que más se han debatido en el seno de la comunidad científica (Pérez Hernández, 2002). El criterio más importante en nuestro caso es la representatividad, concepto esencial en la recopilación. En palabras de Biber (1993), “a corpus is put together in a principled way so as to be representative of a larger textual population, in order to make it possible to generalize findings concerning that population”, y así, para ser representativo este debe incluir “the full range of variability in a population” (Biber 1993: 243). Nuestro corpus se compone de páginas web de multinacionales con intereses en España, y para conseguir la representatividad, hemos incluido empresas de diferentes ámbitos económicos como la automoción, la alimentación, la informática, la información, la banca, la electrónica etc. Las páginas web se han encontrado en directorios de empresas españolas y extranjeras de www.google.com y en www.Yahoo.com y en listados de empresas en prensa. En el apéndice al final del presente trabajo se encuentra el listado de páginas web ordenado por campos de actividad económica. Estas páginas web representan una misma situación comunicativa, que se basa en la relación comercial entre una empresa multinacional y los clientes de la misma. El esquema general de la situación comunicativa sería:

EMPRESA MULTINACIONAL QUE PROPORCIONAN PRODUCTOS O SERVICIOS



La tipología textual está definida por la función comunicativa y pragmática de estas páginas web. Éstas están muy influenciadas por el marketing y la imagen, puesto que la misma representa la imagen corporativa. Se concentran en la venta, en la ayuda al cliente y en proporcionar información adicional sobre los productos, servicios⁸ de diverso tipo o sobre la propia empresa en cuestión. Puesto que hemos realizado una selección aleatoria de páginas para incluirlas en el corpus, debemos asegurar que son representativas de la población textual total (Zanettin 1998). Con este fin, necesitamos establecer unos criterios claros de inclusión en el mismo. Uno de los criterios es la recepción, es decir, a quién va dirigida la página. Nuestras páginas se dirigen como

⁸ Consideramos que empresas como los motores de búsqueda Google.com proporcionan servicios de información.

hemos indicado anteriormente a un público general que tiene un interés claro en conocer más de la empresa, búsqueda de servicios adicionales pre y post venta, o contactar con la misma. La condición lingüística del texto es otro criterio importante. Hemos mencionado que el componente traducido se compone de páginas web traducidas al español. El primer criterio que incluimos es que las traducciones deben haber sido producidas por “traductores profesionales”, y entendemos este concepto en contraposición con “estudiantes” o “traductores no-profesionales”⁹. Asumimos, puesto que las empresas tienen un interés económico manifiesto, que la traducción ha sido encargada y realizada por traductores profesionales. Además, la direccionalidad de las traducciones debe haber sido hacia la primera lengua¹⁰ del traductor (Baker, 1992, 1995), con lo que se excluyó cualquier página Web que presentara errores graves propios de una traducción inversa. De esta manera, avanzamos el siguiente criterio, el que las traducciones sean textos que ocurran de manera natural, y esto se asegura puesto que según la teoría del scopos, la traducción de la página se ha negociado con el cliente que ha encargado la acción (Vermeer, 1989:174). Puesto que estas páginas Web están muy influenciadas por el marketing, necesitamos tener unos criterios claros a la hora de determinar si la página Web es traducida o no. Por cuanto a nuestro corpus, nos hemos guiado por los siguientes criterios para categorizar a una página Web como traducida:

1. La empresa multinacional tiene su sede principal en un país en el que el español no es idioma oficial y usa el inglés como “Lingua franca”
2. La posibilidad de elegir idioma o sitio global en España desde la página principal en inglés. Tanto en menú desplegable como en menú gráfico con banderas¹¹.
3. La existencia de una página web original en inglés, junto con una página en español que tenga exactamente la misma composición. Normalmente la página en inglés tiene el dominio .com y la traducción al español .es
4. La presencia de comentarios¹² y variables en inglés en el archivo fuente formato .txt en la página en español. Los comentarios de los desarrolladores y las variables de los scripts

⁹ En traductología se ha usado el concepto de “novice,” (Kiryal 1995; Shreve 1997) así como el de “no profesional” o “estudiantes de traducción” para contraponerlo al experto.

¹⁰ Siguiendo a Baker (1992, 1995), tomamos el concepto de “primera lengua” y no hablante nativo.

¹¹ La W3C desaconseja el uso de banderas para la elección del lenguaje debido a que, como el español, se hablan la misma lengua en diferentes países.

¹² Ejemplo de línea de comentarios en la página traducida al español de Samsung [n <http://www.samsung.es/cgi-bin/sesa/main/samsunges.jsp>]

son de gran utilidad para saber si las páginas Web que se suponen traducidas han sido originalmente desarrolladas en España. Por ejemplo, la página española de la multinacional Adeco¹³, incluye en sus javascript variables en español, con lo que asumimos que la página se desarrolló en español.

```
cerrado=true;
function despe(){
  if(cerrado){
    bmen.frenadoTo(null,67,4,5,'cerrado=false');
  }
  else bmen.frenadoTo(null,48,4,5,'cerrado=true;')
```

5. La presencia de partes en inglés de la página web traducida al español. A menudo uno o varios marcos [frames] se quedan sin traducir.
6. La no inclusión de caracteres exclusivos del español tales como ï, ð, ñ.
7. Los receptores de la traducción son un público general y no especializado.
8. Y como último, el que no exista un lenguaje publicitario claramente destinado a la cultura meta, España. Como ejemplo podemos poner a Nissan España¹⁴, cuya página web en español, aún siendo muy parecida a la página principal en inglés Nissan.com, contiene una tipología textual que la nos indica que es una creación original en español.

Una de las condiciones indispensables para la inclusión en nuestro corpus es la existencia de un menú de navegación en la página Web. Cabe destacar que el diseño de estas páginas Web se basa en las recomendaciones del consorcio W3C para la creación de las mismas, y todas incluyen los elementos básicos de navegación y organización indispensables y comunes a todas las páginas web accesibles¹⁵.

</TD></TR></TBODY></TABLE></DIV><!-- end Main Promo --><!-- end Main Promo -->

¹³ <http://www.adecco.es/>

¹⁴ <http://www.nissan.es/servlet/nissan.convergence.servlets.extended.PromotionServlet?successURL=/homepage.jsp&failureURL=/homepage.jsp&action=1>

¹⁵ Una de las recomendaciones más repetidas entre los desarrolladores de páginas web es la inclusión del texto exclusivamente necesario. Los usuarios necesitan encontrar visualmente lo que buscan rápidamente. Si no, se cumple la máxima de que el botón más usado en la barra de navegación es el de “atrás” an el navegador.

Herramientas de Análisis del Corpus

La herramienta básica para el análisis del corpus es WordSmith Tools, desarrollada por Mike Scott (1996). Para la presente investigación sólo disponemos de la versión demo del programa, con lo que las funcionalidades del mismo son limitadas.

No obstante, debemos recordar que gran cantidad de términos y colocaciones pueden aparecer en formato gráfico o en formato Flash con lo que Wordsmith Tools no los recuperaría. En un proyecto futuro se ampliará el corpus y se incluirá el texto de las presentaciones en Flash, usando herramientas especialmente diseñadas con tal fin.

Resultados del Análisis Contrastivo

Para nuestro análisis hemos contabilizado la aparición por página web de las unidades léxicas, es decir, que no hemos tomado en consideración la frecuencia total de la aparición de una unidad léxica en el corpus, sino la frecuencia relativa por página. Esto se debe a que muchos de los términos se repiten gran cantidad de veces en una misma página en nombres de archivos a los que se enlaza, nombres de variables en javascripts etc. En la siguiente ilustración de los resultados con “inicio” en las páginas traducidas, se observa la repetición del mismo término en la misma página en las líneas de concordancia de la 13 a la 25. Para nuestro análisis contaríamos sólo una aparición de inicio en la página www.lonelyplanet.com/geoplaneta

El análisis de los resultados muestran una clara diferencia en la frecuencia de las unidades léxicas más usadas para cumplir las cinco funciones básicas de navegación que centran nuestro estudio. Existen diferencias por tanto por la frecuencia de uso de las unidades léxicas como por la variabilidad de los mismos.

Si comenzamos con el análisis de los términos y colocaciones que cumplen la función de contactar con la empresa, observamos que los dos términos más usados en el corpus español son “contacto” y “contactar”, mientras que en el corpus traducido, la unidad léxica más usada contiene el pronombre personal “nosotros”, reflejo de la colocación anglosajona “contact us”. Así las unidades que contienen las diversas formas del pronombre personal “nosotro” aparecen con

mayor frecuencia en las páginas traducidas, “contáctenos”, “contacta con nosotros” y “contáctanos”. Observamos aquí un claro reflejo de la tendencia en la cultura anglosajona a un trato más distendido e informal con el cliente, cuyo reflejo se plasma en el texto traducido de estas páginas web. Este registro más informal aparece en con mucha menor frecuencia en las páginas originadas en España, en donde aparece con mayor frecuencia la forma más impersonal “contacto”. Y aún cuando encontramos “contacta con...” en las páginas españolas, aparece en nombre de la empresa (“contacta con BBK”, “contacta con Fénix”, etc.), dando una mayor formalidad al texto.

Tabla de frecuencias. CONTACT US

	Minicorpus traducido (71 páginas)
Contáctenos	16.9%
Contacto	11.26%
Contactar	11.26%
Contacta con nosotros	9.8%
Contáctanos	5.6%
Contactos	4.22%
Contacte	1.4% 'Contacte con Toshiba'
Contacta	0%

	Minicorpus español (84 páginas)
Contacto	<29.76% <25+
Contacta	4.76%
Contactar	4.76%
Contáctanos	4.76%
Contáctenos	4.76%
Contacta con nosotros	3.57%
Contacte	1.19%
Contactos	0%

Es interesante reseñar la aparición en tres casos de la forma de la variación del lexema “contactos”, ya que en España se usa en páginas web para indicar contactos personales, “personals” en inglés.

Por cuanto a las directrices de privacidad o confidencialidad de las empresas, existe una gran variabilidad en las páginas traducidas. En general, esta colocación y sus variaciones aparecen con gran frecuencia en las páginas traducidas, sobre todo el calco “política de privacidad”, que aparece muy poco en las páginas españolas. En general, en las páginas españolas no nos encontramos con el término “privacidad” con tanta frecuencia puesto que en la cultura española la privacidad no llega a tener tanta importancia como en la cultura anglosajona, lo que se refleja en una mayor variabilidad en las traducciones. Esta variabilidad en las traducciones de “privacy policy”, se muestra en que se han encontrado en español un total de seis posibles variables. La variación aparece tanto en el primer elemento de la colocación, “ política” (términos, condiciones, cláusula, compromiso), como en el segundo, “privacidad” (privacidad, confidencialidad). Esta falta de un realidad clara para la colocación inglesa en la cultura española se muestra también en la frecuencia de aparición de la misma en las páginas, puesto que “política de privacidad”, la colocación más usada, aparece en un 19.71% de las páginas traducidas y solo en 5.9% de las páginas originadas en España.

Tabla de frecuencias. PRIVACY POLICY

	Minicorpus traducido (71 páginas)
Política de privacidad	19.71%
Politica de privacidad (2 sin acento)	
Privacidad	14.08%

	Minicorpus español (84 páginas)
Política de privacidad	5.95%
Privacidad	3.57%

Declaración de privacidad	4.2%
Cláusula de privacidad	1.4%
Compromiso de privacidad	1.4%
Política de confidencialidad	0%

Política de confidencialidad	1.19%
Declaración de privacidad	0%
Cláusula de privacidad	0%
Compromiso de privacidad	0%

Pasamos ahora a la función para obtener más información sobre la empresa. Aquí existe una menor diferencia entre las páginas de los dos corpus, puesto que aparece “¿Quiénes somos?” aparece con más frecuencia en las páginas traducidas y como segunda opción más frecuente en las páginas españolas, pero ambas muestras la misma frecuencia. En las páginas traducidas aparece el calco “sobre nosotros”, y en el corpus español aparece “nuestra compañía” o “nuestra organización”. De nuevo, una diferencia de registro al preferir en español un registro más formal. Otro grupo que aparece en el corpus español es “conoce...” para obtener información sobre diferentes aspectos de la empresa.

Tabla de frecuencias. ABOUT US

	Minicorpus traducido (71 páginas)
¿QUIÉNES SOMOS?	9.8%
Nuestra compañía Nuestra organización	14.08%

	Minicorpus español (84 páginas)
Conoce ...	10.71%
¿QUIÉNES SOMOS?	9.52%

Sobre nosotros	4.2%
Conoce ...	0%
Nuestra compañía Nuestra organización	0%

Nuestra compañía Nuestra organización	2.38%
conócenos	0%
Sobre nosotros	0%

Por cuanto a las condiciones generales de uso de las páginas web, “Terms of use”, suele aparecer la colocación “términos de uso” y “términos y condiciones” que no aparecen en nuestro corpus de páginas producidas en España. En este corpus español suele indicarse con la colocación “Condiciones de uso” pudiendo aparecer “Condiciones generales de uso”. Vemos aquí con “términos de uso” un claro ejemplo de una colocación que nunca aparece en las páginas originadas en España. Al igual que en el caso de “privacy policy”, el proceso de traducción produce una gran variación en la forma, mientras que las páginas españolas suelen concentrarse en dos formas. Cabe destacar también que al ser “terms of use” un concepto más anglosajón, la frecuencia de aparición total en las páginas traducidas es mucho mayor que en las páginas originadas en España.

Tabla de frecuencias. TERMS OF USE

	Minicorpus traducido (71 páginas)
Condiciones de uso	8.45%
Términos y condiciones	5.6%

	Minicorpus español (84 páginas)
Condiciones de uso	3.57%
Condiciones generales de uso	3.57%

Términos de uso	5.6%
Condiciones de utilización	1.4%
Condiciones generales de uso	0%

Términos y condiciones	0%
Condiciones de utilización	0%
Términos de uso	0%

El enlace para llegar a la página principal, “home” o “ homepage” en inglés, existe una mayor frecuencia de la aparición del término “inicio” en las páginas españolas. Este término también se usa en las páginas traducidas pero con mucha menor frecuencia. No obstante aparece “home” directamente en inglés en las páginas traducidas.

Tabla de Frecuencias. HOME

	Minicorpus traducido (71 páginas)
Inicio	11.26%
Home	9.85%
Página principal	5.6%
Página de inicio	4.2%

	Minicorpus español (84 páginas)
Inicio	26.19%
Página de inicio	8.33%
Página principal	3.57%
Home	0%

Discusión de los Resultados

El presente estudio se inició con la finalidad de observar y contrastar preliminarmente las peculiaridades de las páginas Web traducidas al español. Aún cuando somos conscientes de las limitaciones de este estudio preliminar, podemos constatar las diferencias en frecuencia, variabilidad y elección de términos entre el lenguaje traducido y el natural en las páginas Web destinadas a un público español. Las frecuencias en los cinco casos estudiados son diferentes en ambos corpus, lo que reafirma los descubrimientos de investigaciones previas por cuanto a las diferencias de aparición de estructuras y frecuencias en textos originales y traducidos (Gellerstain, 1996; Laviosa-Baithwate, 1996). Por cuanto a la variabilidad, el caso de “privacy policy” es un buen ejemplo de las particularidades de los textos traducidos frente a los originados en otra lengua. La variabilidad se observa tanto en el primer componente de la colocación, “privacy”, donde aparece “confidencialidad”, y “privacidad”, como en el segundo componente, “policy”, donde nos encontramos “cláusula”, “compromiso”, “política”, y “declaración”. Esta variación es inexistente en las páginas españolas con lo que nos encontraríamos con una variabilidad debida a la no fijación de la unidad léxica “privacy policy” en el dominio de las páginas de internet en España.

Las diferencias se observan también en el registro del lenguaje empleado, apareciendo en las páginas traducidas unas construcciones más informales y directas, reflejo de la cultura anglosajona (“contacta con nosotros, contáctanos”), siendo las páginas españolas más formales en el trato (“contacta con Fénix, contacto”). Encontramos además términos y colocaciones que son exclusivos de las páginas traducidas, fiel reflejo del texto origen.

Este estudio preliminar apoya nuestra hipótesis de trabajo más general, el impacto de la reducción y aislamiento de los segmentos textuales a traducir y la falta de contexto de los mismos en el proceso de traducción. Con nuestro estudio preliminar de corpus hemos observado una influencia del proceso de traducción en localización en la traducción de segmentos que aparecen en las páginas Web. Éstos muestran unas peculiaridades propias del lenguaje traducido, que atribuimos a las características propias de este proceso cognitivo de traducción, caracterizado por una mayor falta de contexto y un aislamiento de los segmentos que dificultan la construcción de un plan global más centrado en la cultura y lengua meta durante el proceso.

Es nuestra intención futura la ampliación de ambas partes del corpus y una extracción del texto en formato Flash, para dar un mayor peso en términos de frecuencia a nuestro estudio, y poder así

construir un corpus que sirva como referencia tanto a los futuros investigadores como a estudiantes y traductores de este tipo de sitios Web.

Bibliografía

Atkins, S. Clear, J. Y N. Olster. 1992. "Corpus Design Criteria". *Literary and Linguistic Computing*, 7 (1), 1-16.

Baker, M. y Maeve Olohan. 2001. "El uso del that en textos traducidos en inglés: La explicitación como proceso cognitivo en traducción". En Carmen Valero Garcés e Isabel de la Cruz Cabanillas (eds.) *Traducción y Nuevas Tecnologías. Herramientas Auxiliares del Traductor*, Alcalá: Universidad de Alcalá, P. 77-101

Baker, Mona 1995. "Corpora in Translation Studies: An Overview and some Suggestions for Future research". En *Target* 7 (2): Amsterdam - Philadelphia: John Benjamins, p. 223-243

Baker, Mona. 1992. *In Other Words*. London: Routledge.

Baker, Mona. 1996. "Corpus-based Translation Studies: The Challenges that Lie Ahead". En Harold Somers (ed) *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*. Amsterdam - Philadelphia: John Benjamins.

Biber, D. 1993. "Representativeness in Corpus design". En *Literary and Linguistic Computing*, 8 (4). 243-257.

Bowker, L. 1998. "Using Specialised Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study". *Meta* 43 (4)

Brown, P. et al. 1993. "Machine-readable dictionaries and research in computational linguistics". *Computational Linguistics* 19(2).

Budin, G. y S.E. Wright. 1997. *Handbook of Terminology Management*. Volumen 1 y 2. Amsterdam/Philadelphia: John Benjamins.

Cabré Castellví, M.T. 1999. La terminología: representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Ericsson, K. A., y Simon, H. A. 1984. Protocol analysis: Verbal reports as data. Cambridge, MA: Bradford Books/MIT Press.

Esselink, B. 2001. A Practical Guide to Localization. Amsterdam - Philadelphia: John Benjamins.

Firth, J. R. 1957. Papers in Linguistics. London: Oxford University Press, p. 196.

Gellerstam, M. (1986). "Translationese in Swedish novels translated from English". En Wollin, L. Y H. Lindquist (eds.) Translation Studies in Scandinavia, Proceedings from the Scandinavian Symposium on Translation Theory (SSOTT). Lund: CKW Gleerup. 88-95.

Hartmann, R.R.K. 1994. "The use of parallel text corpora in the generation of translation equivalents for bilingual lexicography". EURALEX 1994 Proceedings. Amsterdam: 291-297.

Johanson, Stig. 1991. "Times Change, and so do Corpora". En Aijmer, K. Y Altengerg, B. (eds.) English Corpus Linguistics: Studies in Honour of Jan Svartvik. London y Nueva York: Longman. P. 305-314.

Kenny, Dorothy. 1998. "Creatures of Habit? What translators usually do with words". Meta 43 (4), p. 515-523.

Kenny, Dorothy. 2000. "Lexical Hide-and-Seek: Looking for creativity in a parallel corpus". En Maeve Olohan (ed.) Intercultural Faultlines. Research Models in Translation Studies I: Textual and Cognitive Aspects, Manchester: St. Jerome. P. 93-104

Kiraly, Donald. 1995. Pathways to Translation. Pedagogy and Process. Kent, Londres: Kent State University Press

Laviosa, Sara. 1996. The English Comparable Corpus (ECC): A resource and a Methodology for the Empirical Study of Translation, Tesis Doctoral. Manchester: Centre for Translation Studies, UMIST.

Laviosa, Sara. 1997. "How Comparable Can 'Comparable' Corpora Be?". *Target* Vol. 9 (2). P. 289-319.

Laviosa, Sara. 1998. The English Comparable Corpus: a Resource and a Methodology. *Unity in Diversity: Current Trends in Translation Studies*. En Bowker, L., Cronin, M., Kenny, D. Y Pearson, J. (eds.) Manchester: St. Jerome Publishing. P. 101-112.

Leech, G. 1992. "Corpora and Theories of Linguistic Performance". En Svartvik (ed.), *Directions in Corpus Linguistics*. La Haya: Mouton de Gruyter. P. 105-122.

Lopez Rodriguez, Clara Ines. 2001. *Tipologías textuales y conexión en la traducción biomédica inglés-español: un estudio de corpus*. Granada: Editorial Universidad de Granada.

Malmkjaer, K. 1998. "Love thy neighbour: will parallel Corpora Endear Linguists to Translators". En *Meta*, 43 (4)

McEnery, A. y A. Wilson. 1996. *Corpus Linguistics*. Edimburgo: Edinburgh University Press.

Pérez Hernández, C. (1994). *Corpus-based Bilingual Lexicography: The Use of Computerized Corpora for the Identification of Translation Equivalents between English and Spanish*. MA Dissertation. University of Exeter.

Perez Hernandez, M. Chantal. 2002. *Explotación de los corpórea textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento*. Madrid: CSIC/Elides, volumen 18 [disponible en : <http://elies.rediris.es/elies18/>]

Puurtinen, T. 1995. *Linguistic Acceptability in Translated Children's Literature* (University of Joensuu Publications in the Humanities N. 15), University of Joensuu. Atk

Rosch, E. 1978. Principles of categorization. En Rosch, E., L.B. Lloyd (eds.) Cognition and categorization. Erlbaum, Hillsdale.

Scott, Mike, 1996. WordSmith Tools (Program). Oxford: University Press Baker.

Sinclair, John. 1991. Corpus, Concordance, Collocation. Oxford: Oxford University Press.

Stewart, Dominique. 2000. "Conventionality, Creativity and Translated Text: The Implications of Electronic Corpora in Translation". En Maeve Olohan (ed.) Intercultural Faultlines. Research Models in Translation Studies I: Textual and Cognitive Aspects, Manchester: St. Jerome. P. 73-91

Stubbs, M. 1996. Text and Corpus Analysis: Computer-assisted studies of language and culture. Londres: Blackwell Publishers.

Tognini-Bonelli, E. (1996a). "Towards Translation Equivalence from a Corpus Linguistics Perspective". International Journal of Lexicography. Vol. 9 (3). P. 197-217.

Toury, G. 1980. In Search of a Theory of Translation. Tel Aviv: The Porter Institute for Poetics and Semiotics.

Toury, G. 1985. "A rationale for Descriptive Translation Studies". En Hermans (ed.) The manipulation of Literature: Studies in Literary Translation. London y Sydney: Croom Helm. 16-41.

Vermeer, H. 1989. "Scopos and Commission in Translational Action". En Chesterman (ed.), Readings in Translation Theory. Finland: Oy Finn Lectura Ab. 173-187.

W3C Consortium. Disponible en [<http://www.w3c.es/>]

Zanettin, Federico. 1998. "Bilingual Comparable Corpora and the Training of Translators". Meta 43 (4)

Zanettin, Federico. 2000. "Parallel Corpora in Translation Studies: Issues in Corpus Design and Analysis". En Maeve Olohan (ed.) *Intercultural Faultlines. Research Models in Translation Studies I: Textual and Cognitive Aspects*. Manchester: St. Jerome, 105-118.