

**El uso de corpus textuales en localización**

Miguel A. Jiménez-Crespo, Assistant Professor  
Rutgers University, The State University of New Jersey

**1. Introducció**

Durante las últimas dos décadas hemos sido testigos del surgimiento y consolidación de una nueva modalidad de traducción, la localización. Este nuevo proceso forma parte del ciclo global conocido como GILT<sup>1</sup> que incluye la internacionalización previa del producto digital, la localización entendida como proceso de adaptación global de textos digitales interactivos para su uso en una región lingüística meta, así como el propio proceso de traducción (Dunne, 2006). El espectacular auge de la localización de textos interactivos durante los años 90 supuso la consolidación del sector de la localización, una industria que representa hoy en día una parte considerable del volumen de traducción en el mundo (Schäler, 2008). No obstante, varios investigadores han apuntado que la estructuración de sus procesos globales se estableció sin una aplicación plena del cuerpo de conocimiento de los Estudios de Traducción (Pym, 2003; Dunne, 2006), de ahí que ciertos principios teóricos, metodológicos y prácticos ampliamente consolidados en esta disciplina no se hayan integrado plenamente en su práctica profesional. Éste es el caso del uso de corpus, a pesar de las considerables mejoras en los índices de calidad global a los que se asocia su uso (Zanettin, 1998; Bowker, 2001; Wilkinson, 2005).

Desde una óptica funcionalista, la práctica totalidad de los procesos de localización se pueden caracterizar como tipos de traducción instrumentales (Nord, 1997), es decir, que el receptor del texto meta interactúa con el texto localizado como si de una producción original en su lengua y cultura se tratara. Éste es precisamente el objetivo del sector de la localización: producir textos que se asemejen a los textos originalmente producidos en la región lingüística (*locale*) de destino (LISA, 2003). No obstante, a pesar de los esfuerzos de este sector, varias investigaciones recientes han demostrado que los textos presentan unas características propias tales como una mayor explicitación (Diéguez, 2008), la presencia de estructuras clonadas de los originales (Jiménez-Crespo, 2008a), una mayor variación terminológica o una menor presencia de convenciones de género (Jiménez-Crespo, 2009). Estas características específicas de los textos localizados entroncarían con la corriente que defiende la existencia de una lengua de traducción paralela (Baker, 1995; Baker y Olohan, 2000; Frawley, 1984; Duff, 1981), debida inevitablemente al proceso de traducción como un caso especial de producción textual subordinada (Baker, 1999). Así, y dado que el objetivo de toda localización de calidad ha de equilibrar el cumplimiento de las especificaciones del cliente-iniciador (ISO 9000, ASTM)<sup>2</sup> con las expectativas de la comunidad de discurso meta, los corpus constituyen una herramienta clave a la hora de asistir al traductor para conocer estas expectativas léxicas, sintácticas y estilísticas. Esta mayor aproximación que proporcionan los corpus redundará a su vez en una mejora en la apreciación de calidad por parte de la audiencia (Nobs, 2006).

Por otro lado, el estudio de los recursos utilizados durante el proceso de traducción ha cobrado especial interés en esta disciplina (Yuste Trigo, 2008). No obstante, el sector de la localización prima los equipos de localización o «localizator's workbench» con memorias de traducción, bases de datos terminológicas y acceso a Internet, sin una incorporación plena de los corpus textuales (Schäler, 2008; Esselink, 2001). En este sentido, y aun teniendo en cuenta restricciones debidas a la limitación de tiempo y esfuerzo de un proceso entendido como actividad industrial (Wright, 2006), el presente

<sup>1</sup> Globalización, Internacionalización, Localización y Traducción.

<sup>2</sup> En el sector de la localización, las definiciones operativas de calidad suelen ser las definidas por los estándares internacionales ISO 9000, «the totality of features and characteristics of a product or service that bears on its ability to satisfy stated or implied needs» (ISO 9000, apud Ørsted, 2001: 443), o los de la ASTM, que define la calidad como «the degree of conformance to an agreed upon set of specifications» (ASTM).

artículo presenta desde una óptica práctica la introducción de corpus textuales en el proceso de traducción como fuente indispensable de información conceptual, terminológica, fraseológica, ontológica y pragmática. La estructuración del presente artículo será la siguiente: en primer lugar se repasarán los escasos estudios sobre el uso de corpus en localización, para pasar a describir estos usos por medio de análisis en Corpus Web Comparable del Español (Jiménez-Crespo, 2008a), un corpus de 40.000 páginas web corporativas originales y localizadas.

## 2. Corpus y la práctica profesional de la localización

Los usos de corpus en la práctica profesional, didáctica de la traducción, o como metodología de investigación sobre el producto y el proceso la traducción, han sido extremadamente productivos durante las últimas dos décadas (Baker, 1995, 1996, 1999, 2004; Laviosa, 2002; Olohan, 2004; Beeby *et al.*, 2009). Su pertinencia en la práctica profesional se debe a que todo corpus proporciona datos cuantitativos y cualitativos objetivos sobre los que fundamentar el proceso de toma de decisiones (Shreve, 2006: 311; Baker, 2004: 184), ya sea durante la fase previa de documentación, durante el proceso de traducción o en su posterior evaluación. Sin embargo, su introducción en la práctica profesional no se ha producido en la misma medida que en la didáctica o en las investigaciones traductológicas (Bowker y Barlow, 2008; Bernardini y Castagnoli, 2008), una tendencia que se hace especialmente patente en la localización. De hecho, son escasas las publicaciones que promuevan su introducción en su ciclo global (Shreve, 2006; Jiménez-Crespo, 2007, 2008b, 2008c, 2009).

De entre las razones aducidas tras su escaso impacto en la localización, la falta de rentabilidad económica ocupa un lugar primordial (Biau y Pym, 2006: 18; Bowker y Barlow, 2008). Este hecho se justifica por el esfuerzo que supone construir un corpus sin que se produzca una mejora en la productividad inmediata, tal como ocurre con las memorias de traducción. Como recurso alternativo, Biau y Pym (2006) defienden que el uso de Internet como Corpus, o *Web as a Corpus* (Fletcher, 2007)<sup>3</sup>, puede sustituir a corpus representativos cuyos procesos de selección textual y compilación se realicen según principios teóricos sólidos (Ej. Sinclair, 1991; Biber, 1993; Pérez, 2002). Sin embargo, esta propuesta ignora la tendencia mayoritaria de traducciones del inglés al español en este ámbito tecnológico, de ahí que el uso de Internet como recurso documental entrañe el riesgo de justificar ciertas decisiones en un conjunto textual plagado de anglicismos léxicos, sintácticos y estilísticos, traducciones erróneas o inadecuadas (Jiménez y Tercedor, 2009), continuando así el ciclo encubierto de expansión de convenciones anglosajonas a través de la traducción (House, 2001).

Por otro lado, se supone que la mayor incidencia de las memorias de traducción o las bases terminológicas se deben a que esta tecnología la desarrollan empresas con amplios recursos económicos para su promoción, mientras que el software de análisis léxico de corpus es normalmente desarrollado en círculos académicos y, por tanto, con menos capacidad de promoción y un perfil menos tecnológicamente avanzado (Bowker y Barlow, 2008). En este sentido, la solución por la que abogan ciertos investigadores es la plena integración de las herramientas de análisis de corpus en las memorias de traducción (Bernardini y Castagnoli, 2008). Como último punto, la discreta presencia de los corpus en localización podría deberse a que el flujo de localización se solía producir mayoritariamente desde un contexto anglosajón hacia el resto de *locales* (Cronin, 2003), con una relativa inexistencia inicial de textos originalmente producidos en las regiones lingüísticas de destino. No obstante, la producción de textos digitales en cada región lingüística ha crecido exponencialmente

<sup>3</sup> En los recientes estudios sobre el uso de Internet como corpus textual (Kilgariff y Grefenstette, 2003), se distingue entre los usos de la Web para la creación de corpus, «Web for Corpus» (WfC), como fuente de textos digitales que se pueden recopilar para formar un corpus, o la Web como corpus, «Web as Corpus» (WaC), cuando la Web se consulta directamente en su conjunto como corpus (Fletcher, 2007).

durante la última década, de ahí que sea un momento óptimo para la creación y uso práctico de corpus textuales para su uso en tareas de localización. Dado que hoy en día se están consolidando gran cantidad de géneros digitales (Santini, 2007), esta metodología puede mostrar qué características lingüísticas poseen estos géneros digitales mediante las producciones naturales de hablantes de cada comunidad de discurso (Jiménez Crespo, 2008c).

Pasamos a continuación a describir los distintos tipos de corpus que se podrían recopilar a partir de géneros digitales y que serían de gran utilidad en el proceso de localización.

### 3. Tipología de corpus y la localización

De entre los posibles tipos de corpus utilizados en el campo de la traducción (Laviosa, 2002), los que tendrían una aplicación práctica durante el proceso de localización serían: (1) los corpus monolingües de control, (2) los paralelos y (3) los comparables. En primer lugar, un **corpus monolingüe** es aquel que contiene una sola lengua (Kenny, 2001: 58) y a su vez se puede subdividir entre los corpus monolingües traductológicos o naturales u originales (Laviosa, 2002: 35). Los corpus monolingües originales o no traducidos de textos especializados se han usado sobre todo en la enseñanza de la traducción (Beeby *et al.*, 2009; Rodríguez Inés, 2008; Corpas Pastor, 2001; Bowker, 1998; Pearson, 1998), como ayuda en la cuantificación de la calidad de las traducciones (Bowker, 1999; Bowker, 2001), para la extracción de terminología (Sánchez Gijón, 2004; Faber, López y Tercedor, 2001; Pearson, 1999), o como parte de las estaciones de trabajo para la creación asistida por ordenador de textos técnicos o especializados (Shreve, 2006). Debido a la necesidad de comparar los datos obtenidos (Kenny, 2001: 125), se usan además varios corpus monolingües de control como el *British National Corpus* o *Cobuild Bank of English* en inglés o el CREA en español.

En segundo lugar, un corpus paralelo lo conforman textos originales en una lengua con las traducciones de los mismos hacia una o varias lenguas (Laviosa, 2002: 37). Este tipo de corpus también se ha denominado «corpus de traducción» (Johansson, 1998) o «bitexts» (Bowker, 2002). Los corpus paralelos son de gran utilidad para comprobar cómo se ha traducido previamente un término o unidad fraseológica, una funcionalidad que comparte con las memorias de traducción. No obstante, los análisis por medio de programas de concordancias bilingües permiten búsquedas de una sola palabra o segmentos infraoracionales, además de permitir examinar párrafos completos o la traducción del texto completo en lugar de pares de oraciones independientes (Bowker y Barlow, 2008). A pesar de su amplio uso en la comunidad traductora, de entre las posibles desventajas de este tipo de corpus se encuentra que no presenta una variedad natural de la lengua (McEnery y Xiao, 2007; Sinclair, 1991) y que puede representar la idiosincrasia de un sólo traductor (Teubert, 1996: 247).

Finalmente, un corpus comparable se identifica con una colección estructurada de textos digitales en una lengua junto con traducciones hacia esta misma lengua. Según Baker, este tipo de corpus se define como «a structured electronic collection of texts originally written in a particular language, alongside texts translated into that same language» (1995: 234). Esta colección de textos se ha de recopilar de acuerdo a unos criterios que aseguren su comparabilidad, por lo que deben incluir un mismo campo, variedad de lengua, temporalidad y una extensión similar. En el caso de la localización, ha de asegurarse de que los géneros digitales seleccionados, tales como sitios web corporativos o ayudas en línea para programas de software, representan la población textual que será el objeto del proceso de localización posterior. Este tipo de corpus se ha utilizado mayoritariamente para estudiar las peculiaridades textuales de los textos traducidos *per se* (Baker, 1995; Baker y Olohan; Kenny, 2001), sobre todo con el objetivo de identificar los «universales de

traducción»<sup>4</sup>. No obstante, este tipo de corpus puede ser de utilidad durante la localización o evaluación puesto que se puede investigar en la sección original del corpus las distintas posibilidades para un término, unidad fraseológica o colocación, pudiendo al mismo tiempo comprobar en la sección traducida del corpus si una posible opción constituye un caso de «translationese» (Gellerstam, 1996) o caso recurrente de interferencia del texto fuente tal como veremos en la sección 4 (Toury, 1995). Un ejemplo de este tipo de corpus y sus posibles usos en localización lo constituye el Corpus Web comparable del español (Jiménez-Crespo 2008a). El siguiente apartado describe brevemente las características del mismo como paso previo a la descripción de sus posibles usos en esta modalidad de traducción.

### 3.1. El Corpus Comparable Web del español

El proceso de recopilación y selección textual del Corpus Web Comparable del español se guió por los parámetros teóricos que se han consolidado tras las aportaciones de un número significativo de investigadores (Sinclair, 1991; Biber, 1993; Atkins *et al.*, 1992; EAGLES, 1996; Olohan, 2004, etc.). El corpus web se compone de dos secciones paralelas: una primera sección que incluye una selección representativa del género «sitio web corporativo» de textos producidos originalmente en español de España, y una segunda sección en la que se incluyeron los sitios web corporativos estadounidenses localizados para una audiencia española. La primera parte del corpus se seleccionó tras un proceso sistemático de análisis en el directorio de negocios de Google España y en el mismo están representadas todas las áreas de negocio (Jiménez-Crespo 2008b, 2008c). Por su parte, los sitios web localizados representan el conjunto de sitios web de las mayores empresas estadounidenses según la lista Forbes en noviembre del 2006.

Puesto que el corpus ha sido ampliamente descrito en publicaciones previas (Jiménez-Crespo, 2008a, 2008b, 2008c; 2009), expondremos brevemente sus principales características. La sección original del corpus está compuesta por 172 sitios web, un total de 19.102 páginas web y una media de 111,50 páginas por sitio web. La sección localizada del corpus se compone de 95 sitios web, un total de 21.322 páginas web, con una media de 224,93 páginas por sitio. Así, y siguiendo estudios con corpus comparables previos (Olohan, 2004; Baker, 2004), el tamaño de la sección original y la localizada se asemejan en el número aproximado de palabras o «tokens», en lugar de ser similar en lo relativo al número de textos incluidos.

A su vez, y tras un estudio sobre la superestructura prototípica de los sitios web corporativos (Jiménez-Crespo 2008b), se procedió a la extracción de subcorpus representativos de cada sección comunicativa o «move-steps» (Swales, 1991; Askehave y Nielsen, 2004) dentro del sitio web, tal como se indica en la Figura 1. Esta subdivisión entre secciones comunicativas dentro de un mismo texto global o sitio web permite a su vez la extracción de subcorpus comparables que se pueden usar durante la traducción de una sección específica, tales como los avisos legales, los formularios de contacto o las descripciones de productos. En este caso se extrajeron corpus de las secciones de mayor representatividad, con lo que se obtuvo un corpus terminológico de todos los segmentos en los menús de navegación (Jiménez-Crespo, 2009), un subcorpus textual de páginas web de contacto y otro que incluye la totalidad de los textos legales en los sitios web

<sup>4</sup> Los universales de la traducción se definen como «features that typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems» (Baker, 1996: 176).



Figura 1. Superestructura prototípica del sitio web corporativo que permite la extracción de subcorpus de las distintas secciones o la búsqueda específica en una sección del sitio web global. Adaptado de Jiménez-Crespo (2008b).

Esta subdivisión por secciones comunicativas cobra especial relevancia en aquellos géneros híbridos en los que se mezclan tipologías textuales (Tercedor, 2005; Jiménez-Crespo, 2008c). En el caso de la localización de software, la descomposición del género textual «programa de software» de mayor nivel de concreción sobre la que sustentar estudios similares sería la presentada por el investigador Austermühl (2006: 80)<sup>5</sup>.

#### 4. El uso de corpus y la calidad en localización

A pesar de su escasa integración en la práctica profesional de la localización, el uso de corpus se asocia en las publicaciones del ramo a una mayor calidad en los textos traducidos a nivel léxico, morfosintáctico o estilístico (Bowker, 1998; Bowker y Barlow, 2002; Wilkinson, 2005). Este hecho se debe a que un corpus proporciona rápidamente ejemplos claros de uso natural de ciertas expresiones, su frecuencia o representatividad, así como su contexto inmediato. Por lo tanto, su aplicación puede ser de gran utilidad tanto en el proceso de traducción como en las tareas de evaluación y Análisis de Calidad (Bowker, 2001). La siguiente cita de Federico Zanettin (2002) resume brevemente las posibles aplicaciones prácticas de la metodología de corpus que se podrían dar en la localización:

If a corpus is appropriately designed, it can provide reliable evidence of authentic linguistic behavior and text-structuring conventions by highlighting recurrent patterns. Terminological and collocational information can be especially useful (Zanettin, 2002).

De esta manera, más allá de las memorias de traducción y los glosarios, los análisis en corpus representativos pueden ayudar en el proceso de toma de decisiones, mostrando las estructuras más frecuentes en su contexto y en textos originalmente producidos en lengua meta con el fin de producir localizaciones más ajustadas a las expectativas de la audiencia meta. En este sentido, las ventajas para la localización se centrarían en la consecución de su repetido desiderátum mediante (1) el ajuste con mayor precisión del texto a las convenciones de género esperadas por la audiencia meta, entre las que destacan las super y macroestructuras prototípicas junto con sus convenciones asociadas (Shreve, 2006; Jiménez-Crespo 2008b), (2) la plasmación en el texto localizado de los patrones de co-ocurrencia léxica y composición léxica prototípica en cada género digital en la cultura meta (Hoey, 2005), (3) la plasmación de la fraseología asociada a cada bloque comunicativo, y (4) el ajuste del texto a las convenciones de registro en cada cultura. Pasamos a describir cada uno de estos puntos.

##### 4.1. Terminología y convenciones de género

La investigación y documentación para la terminología es uno de los usos más extendidos de los corpus (Pearson, 1999; Bowker y Pearson, 2002; Faber *et al.*, 2001). En localización, el recurso de

<sup>5</sup> A pesar de que el investigador no utiliza el concepto de «género textual» al que se refiere, sino el de «text type», cuyo significado es diferente en los Estudios de Traducción (Trosborg, 1997).

mayor impacto son las bases de datos terminológicas proporcionadas por los clientes, e incluso ciertas empresas clave en la localización las han hecho públicas con el claro interés de estandarizar la terminología en este dominio (Ej. Microsoft, Sun). A pesar de que estas herramientas ayudan al traductor a proporcionar la coherencia terminológica necesaria (DIN 2345; Bass, 2006), sería imposible a priori identificar la totalidad de términos existentes en un texto digital, de ahí que el traductor se encuentre inmerso a menudo en procesos de toma de decisiones terminológicas. Para ilustrar este punto, proponemos como ejemplo la traducción de los frecuentes términos en sitios web en inglés, «legal disclaimer» y «sitemap». Comenzando con el primero, este término suele aparecer en menús de navegación en sitios web estadounidenses y asumiremos que el traductor ha propuesto preliminarmente como posible traducción el término «declaración legal». La sustentación de esta traducción se realizó tras la recuperación de 60.700 casos en páginas web según Google,<sup>6</sup> usando así como recurso documental la totalidad de la «Web como Corpus» (Fletcher, 2007).

No obstante, y tal como muestra el extracto de concordancias en la Figura 2, una búsqueda en el corpus original de control mediante la palabra «legal» puede proporcionar inmediatamente información más relevante sobre el término de mayor frecuencia en los sitios originales, «aviso legal» (51 casos), así como la gama posible de variación con menor frecuencia como «nota legal» (13 casos) o «información legal» (8 casos).

N	Concordance
4	s en el presente Aviso Legal, lesivos de los de
5	incluidas en este Aviso Legal. Los servicios y
6	rtal, así como su Aviso Legal. Por ello, EUROP
9	y cancelación. aviso legal © 2004 UNIO
10	ad   Aviso legal Aviso legal Denominació
41	s en el presente Aviso Legal y en el Portal.
42	uscholux.es AVISO LEGAL Y POLÍTICA DE 15.310
43	ágina Principal   Aviso Legal   El hecho de c
44	(en adelante, el "Aviso Legal") regula el uso
60	807 025. Información Legal -----
61	o legal 2. Información legal a la recogida de d
62	icado. - Información Legal   - Seguridad
63	Reservas Información Legal   contacto
68	ontacto   Ayuda   Nota Legal
78	* NOTA LEGAL* * Esta Nota
79	corresponder. *Nota Legal* *Gracias por vis
80	d Diseño *NOTA LEGAL* Los Logotipos
81	echos reservados. Nota Legal.

Figura 2. Extracto de concordancias para la palabra «legal» en el corpus original de textos legales en sitios web originales.

En este caso la traducción propuesta, «declaración legal», no se encuentra entre los términos que aparecen en los textos originalmente producidos en español y pertenecientes al mismo género textual y situación comunicativa. Un corpus comparable, con una sección de textos localizados, nos puede permitir además observar si esta propuesta constituye un calco del término de texto fuente debido a un caso de interferencia (Toury, 1995). En este caso, el análisis de concordancia con la palabra «legal» como núcleo nos muestra la presencia exclusiva en los textos localizados de esta combinación léxica, de ahí que se pueda claramente comprobar cómo «declaración legal» constituiría un caso de «unique item» (Tirkonnen-Condit, 2004), es decir, elementos que son correctos en el sistema de la lengua meta pero que, sin embargo, no aparecen en textos producidos directamente por miembros de la comunidad de discurso meta pertenecientes al mismo género textual. Este análisis que se recoge en la Tabla 1 puede identificar además otros «unique items», tales como «acuerdo legal» o «advertencia legal». La presencia de estos elementos en textos altamente convencionalizados, como los sitios web corporativos (Nielsen, 2000, 2001; Nielsen y Tahir, 2002), conllevarían su distanciamiento conjunto de combinaciones de elementos léxicos esperados por un receptor medio de este texto (Hoey, 2005), reduciendo subconscientemente la apreciación de calidad del mismo.

Corpus original		Corpus localizado	
Término	Frecuencia	Término	Frecuencia

<sup>6</sup> Según la búsqueda realizada el 10 de junio del 2009.

Aviso legal	51	Aviso legal	28
Nota legal	13	Declaración legal	13
Información legal	8	Información legal	11
		Acuerdo legal	5
		Advertencia legal	3
		Legal	2

Tabla 1. Términos que incluyen el nodo «legal» en una búsqueda de concordancias en el corpus original y localizado.

De hecho, volviendo a los sitios web ofrecidos por Google como justificación del calco originalmente propuesto, se puede observar que los mismos constituyen localizaciones al español de sitios web internacionalizados, es decir, que el uso de Google en la búsqueda documental se basa en multitud de ejemplos de textos traducidos, y por tanto que no representan la lengua de destino de forma natural y genuina (McEnergy y Xiao, 2008; Teubert 1996).

Otro caso de interés para demostrar la pertinencia de utilizar corpus con fines terminológicos en localización puede ilustrarse con la traducción de «sitemap», un término altamente convencionalizado en los sitios web estadounidenses (Nielsen y Tahir, 2002). Este término suele aparecer en los menús de navegación, unos segmentos textuales que presentan un alto nivel de convencionalización en su terminología tanto en español (Jiménez-Crespo, 2009), como en inglés (Nielsen y Tahir, 2002). En este caso, un análisis de concordancias con la palabra «mapa» como núcleo nos muestra cómo el término «mapa web» se encuentra convencionalizado en español, aunque otros seis términos se presentan como alternativas para expresar este concepto propio de todo sitio web. La Figura 3 muestra un análisis comparativo extraído a partir de estas concordancias, extrayendo una sola forma por sitio web para normalizar las frecuencias. Este análisis muestra cómo el calco «mapa del sitio» es el término más usado en los textos localizados, mientras que la convención identificada en los sitios españoles aparece sólo en un 5% de los sitios localizados. En este caso, y puesto que con el objetivo de mejorar la usabilidad y la calidad se recomienda ajustar los textos digitales interactivos a las convenciones establecidas en la totalidad de guías de estilo web (véase Jiménez-Crespo, 2008a: 210-225; Jeney, 2007) y manuales de usabilidad web (Nielsen, 2004; Nielsen.2000; Nielsen y Loranger, 2006; Brink *et al.*, 2002), sería más adecuado pragmáticamente el uso de la convención «mapa web» en lugar de otras opciones con menor frecuencia de uso, o incluso «unique items» tales como «guía web» o «plano del sitio».

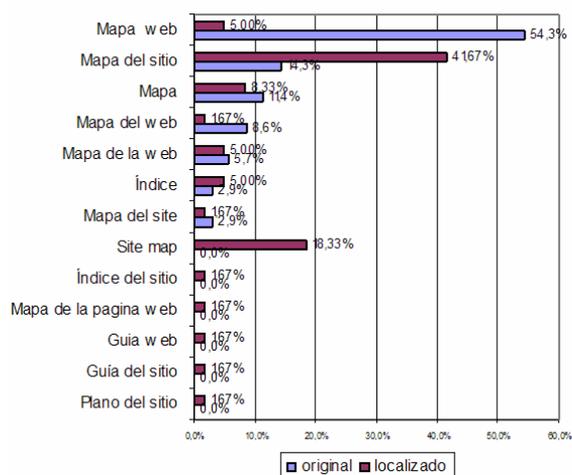


Figura 3. Análisis contrastivo de terminología en sección comunicativa «mapa web» en sitios web localizados (de Jiménez-Crespo 2008a).

## 4.2. Concordancias

El análisis léxico de corpus con programas de concordancias es de gran utilidad a la hora de producir textos de calidad (Zanettin, 2002; Bowker, 1998). En general, una concordancia se define como la aparición recurrente de una o más palabras en un entorno inmediato en un texto o «the occurrence of two or more words within a short space of each other in a text» (Sinclair, 1991:170), siendo el «nodo» la palabra que se está estudiando o analizando, y el «colocador», la palabra que co-ocurre en el entorno de un nodo determinado (ibid: 115). En este sentido, es posible ajustar los patrones de co-ocurrencia léxica en los textos traducidos para que éstos presenten lo más posible los mismos patrones de concurrencia que los textos originales (Hoey, 2005). Como ejemplo, presentamos el caso de las formas verbales que aparecen en el entorno de la palabra «formulario», un sustantivo que suele aparecer en los sitios web interactivos formando parte de expresiones que animan al receptor a rellenarlo con distintos fines. En este caso, la Figura 4 recoge un resumen del análisis de concordancias de las formas verbales que aparecen asociadas al sustantivo «formulario».



Figura 4. Verbos asociados al sustantivo «formulario» en sitios web originales y localizados.

El análisis muestra cómo en los textos originales la palabra formulario convencionalmente aparece asociada al verbo «rellenar» y en menor medida con «utilizar», mientras que en los textos localizados el verbo más frecuente es «completar», un caso claro de interferencia de la construcción inglesa «complete the form».

De igual modo, otro ejemplo sería la traducción de la expresión recurrente en textos web en inglés «free of viruses», ya que todo sitio web asegura en sus avisos legales la imposibilidad de garantizar que el mismo no presente virus o troyanos. Un análisis de concordancia con el lema «virus\*» en ambos corpus demuestra que el calco sintáctico «libre de virus» en azul en el extracto de concordancias (Figura 5), aunque de nuevo correcto en el sistema de la lengua, no aparece en textos originales. En estos últimos textos, la palabra «virus» aparece mayoritariamente asociada a los sustantivos «presencia» y «ausencia» y «existencia».

Corpus Localizado		Corpus Original	
N	Concordance	N	Concordance
17	vío constituye ni contiene virus informáticos, propaga	9	s por la introducción de virus. Caja Madrid no se
18	mpletos y están libres de virus u otros componente	10	ios. - La existencia de virus, programas malicios
19	(ES) ESTÉN LIBRES DE VIRUS U OTROS COMP	11	ios. • La existencia de virus, programas malicios
20	este Sitio estén libres de virus ni de ningún otro có	12	l Usuario la ausencia de virus, gusanos, caballos
21	OR ESTÉN LIBRES DE VIRUS INFORMÁTICOS.*	13	ventual existencia de virus u otros elementos n
22	té absolutamente libre de virus. Antes de bajar las	14	garantiza la ausencia de virus u otros elementos e
23	a estará siempre libre de virus, y no aceptará ning	15	NTO; (v) la presencia de virus u otros elementos
24	eb de GLASER libre de virus, GLASER no garanti	16	garantiza la ausencia de virus u otros elemento
25	oducir o difundir en la red virus informáticos o cuale	17	garantiza la ausencia de virus u otros elementos
26	ducir o difundir en la red virus informáticos u otros	18	r previo, la ausencia de virus o elementos en los
27	pueda introducirse algún virus en el Sitio Web me	19	r previo, la ausencia de virus o elementos en los
28	os por la introducción de virus y/o ataques infor	20	controlar la ausencia de virus o cualquier otro ele
29	ir inadvertidamente algún virus informático a través	21	berse a la presencia de virus o a la virus o a la pr

Figura 5. Análisis de concordancias en el corpus original y localizado del lema «virus\*».

En este caso, se puede identificar claramente cuáles son los sustantivos que aparecen ligados al sustantivo «virus», siendo posible ajustar la expresión de esta unidad de sentido recurrente a las convenciones propias del género digital meta.

#### 4.3. Fraseología

Como géneros textuales altamente convencionalizados (Kennedy y Shepherd, 2005; Santini, 2007), la mayoría de géneros digitales poseen además una fraseología con distinto grado de fijación y en constante evolución. Los géneros digitales presentan una gran mezcla de tipos textuales (Tercedor, 2005), como son descripciones técnicas, políticas de protección de datos, noticias o segmentos interactivos, por lo que el localizador puede no poseer una competencia activa a la hora de decidir en cuestiones fraseológicas específicas de cada sección comunicativa. Los análisis de concordancias se presentan como un mecanismo fundamental durante el proceso de traducción para la identificación de unidades fraseológicas (Zanettin, 2002; Tercedor, 1999), suponiendo una mejora en la calidad global del texto meta. Como paso previo a este análisis fraseológico, definiremos el concepto de unidad fraseológica como:

[A] lexicalized, reproducible bilinear or polylinear word group in common use, which as relative syntactic and semantic stability, may be idiomatized, may carry connotations, and may have an emphatic or intensifying function in a text (Gläser, 1998: 125).

Estas unidades fraseológicas pueden estar compuestas de dos o más lexemas, y normalmente poseen distinto grado de fijación. Entre las mismas se suelen distinguir entre las colocaciones, con un grado variable de fijación, y los modismos, tales como frases hechas, refranes, etc. Para ilustrar así las ventajas del uso de corpus para el análisis y la selección fraseológica en localización, presentamos un ejemplo utilizando la palabra «aceptación». El análisis de concordancias en el corpus de control muestra la existencia de la unidad fraseológica «implica su/la *aceptación* plena y sin reservas» en la sección de avisos legales de sitios españoles. Se observa además la gama de posible variación, tal como la posibilidad de sustituir el verbo «implicar» por «suponer», la posible exclusión del adjetivo «plena», y el mayoritario uso verbal en presente de indicativo.

Corpus Original	
N	Concordance
1	usuario expresa su <b>comprensión</b> y <b>aceptación</b> plena y sin reservas
2	delante, el "Usuario") e <b>implica</b> la <b>aceptación</b> plena y sin reservas
3	El mero uso del Portal <b>implica</b> la <b>aceptación</b> sin reservas por parte
4	El mero uso del Portal <b>implica</b> la <b>aceptación</b> sin reservas por parte
5	El acceso a dicho sitio <b>implica</b> su <b>aceptación</b> sin reservas. La utilizac
6	El acceso al mismo <b>implica</b> su <b>aceptación</b> sin reservas. * 3. S
7	et. El acceso al mismo <b>implica</b> su <b>aceptación</b> sin reservas. La utiliz
8	en el área de clientes <b>implicará</b> su <b>aceptación</b> expres y sin reserva
9	n de la Tarjeta A.G.C. <b>supondrá</b> a <b>aceptación</b> plena y sin reservas,
10	n de la Tarjeta A.G.C. <b>supondrá</b> a <b>aceptación</b> plena y sin reservas,
11	. La visita del sitio web <b>supone</b> la <b>aceptación</b> plena y sin reservas

Corpus Localizado	
N	Concordance
1	inados servicios <b>mplicará</b> <b>asimismo</b> la <b>aceptaci</b> ón, sin reserva alguna, de
2	(en adelante, el "Usuario") e <b>implica</b> la <b>aceptación</b> plena y sin reservas de
3	b, el visitante ("Usuario") <b>manifiesta</b> su <b>aceptación</b> sin reservas de las co
4	la página web el Usuario <b>manifiesta</b> su <b>aceptación</b> sin reservas de las pri
5	vegación por este sitio Web <b>supone</b> su <b>aceptaci</b> ón sin limitaciones de est
6	xpress.es" <b>implica</b> el conocimiento y a <b>aceptación</b> , plena y sin reservas,

Figura 6. Extracto de una KWIC que muestra los resultados de la búsqueda mediante el lema «accept\*» en sitios web originales y localizados.

Por su parte, el análisis del corpus localizado muestra posibles variaciones en los textos producto de un proceso de traducción, tales como la presencia de «sin reserva alguna» (línea 1) o «sin limitaciones» (línea 5), además de aparecer el verbo «manifiesta» inexistente en los textos originales. Dado el alto grado de fijación de las unidades fraseológicas en los textos legales en general (Borja Albí, 2000), sería por tanto pertinente el uso de la unidad fraseológica identificada en los sitios web españoles con el objetivo de mejorar estilísticamente la calidad global del texto traducido.

#### 4.4. Adecuación a la situación comunicativa

Los textos digitales se caracterizan por un alto grado de interactividad, en los que tanto el emisor del mensaje (la empresa-persona-colectivo tras el sitio web o programa) como el programa o el propio sitio web interactúan de forma paralela con el usuario final (Janoschka, 2003). Esta peculiaridad de la situación comunicativa que supone una constante interacción con el usuario hace que los textos digitales posean una función o foco contextual apelativo-exhortativo (Nord, 1997) que alterna entre secundario y primario (Gamero, 2001). Un corpus de control puede ayudar a ajustar las características del texto localizado a las convenciones establecidas en este tipo textual en español. Para ilustrar este particular, presentamos la traducción de un segmento altamente convencionalizado en las páginas de contacto, por el que el emisor insta al usuario a comunicarse con éste en caso de dudas, preguntas, etc. Este segmento suele aparecer en un párrafo previo a la presentación de la información de contacto en sí, y en inglés se expresa mediante la expresión «Please contact us...».

Corpus Original		Concordance	Corpus Localizado		Concordance
N	ulario.	"Para cualquier otro tema, se pueden poner en contacto con nosotros a	N	e expresa más arriba, me pondré	en contacto con
5	na impresora laser Hp o un plotter Hp puede ponerse en contacto con nosotros ll		6	tos autorizados deberían ponerse	en contacto con
6	7 lo desea puede rellenar este formulario para ponerse en contacto con nosotros.		7	plataforma, no dude en ponerse	en contacto con €
7	8 CR Class English Version "Si desea ponerse en contacto con nosotros o		8	Información sobre cómo ponerse	en contacto por te
8	9 "CONTACTAR" Rellene el formulario para ponerse en contacto con nosotros!		9	Información de Interés - Ponte	en contacto con r
9	10 *Clientes* Inicio > Contacto >	"Si desea ponerse en contacto con nosotros, p	10	Información de Interés - Ponte	en contacto con r
10	11 uiera de" "nuestras actividades, no dude en ponerse en contacto con nosotros ei		11	los negocios, por favor póngase	en contacto con l
11	12 para la agricultura. Información Si desea ponerse en contacto con nosotros pi		12		en contacto con
12	13 on alguna Delegación Comercial?... "Para ponerse en contacto con nosotros, p		13	Conecte con Avaya	Póngase en contacto con /
13	14 del site   Política de privacidad No dude en ponerse en contacto con nosotros pi		14	rt Services Directory	*Póngase en contacto con (
14	15 sa.com *» Cubra este formulario si desea ponerse en contacto con Reflexion /		15	y cuestiones técnicas.	Póngase en contacto con (
15	16 05 Jazztel - Jazz Telecom, S.A." Si desea ponerse en contacto conmigo o solic		16	istema que necesita?	Póngase en contacto con (
16	17 ORES _CONTACTO ***** Si quieres ponerte en contacto con nosotros ré		17	de compra y entregas.	Póngase en contacto con €
17	18 ORES _CONTACTO ***** Si quieres ponerte en contacto con nosotros ré		18	itio España - español »	Póngase en contacto con t
18	19 Apellidos:* Indique cómo quiere que nos pongamos en contacto con usted.* *P		19	ol »	Póngase en contacto con t
19	20 editaciones Investigación Referencias Póngase en contacto con nosotros pi		20	r Página imprimible	Póngase en contacto con t
20	21 Preguntas más frecuentes o en el sitio Web, ponte" en contacto con nosotros pi		21	ficinas de la empresa"	Póngase en contacto con t
21	22 io) Para cualquier tipo de consulta o duda póngase en contacto con nosotros a		22	ntacto con HP Inicio >	Póngase en contacto con r
22			23	ntacto con nosotros	Póngase en contacto con r

Figura 7. Extracto de una KWIC que muestra los resultados de la búsqueda mediante la expresión «en contacto» en sitios web originales y localizados.

Como se puede observar en la Figura 5, los sitios web españoles se suelen dirigir al usuario por medio de perífrasis verbales tales como «si desea ponerse en contacto» (líneas 8,10, 12, 15 y 16), «si quieres ponerte en contacto» (líneas 17 y 18) o «puede(n) ponerse en contacto» (líneas 5 y 6). Se observa además una escasa presencia de la forma verbal en modo imperativo. Por su parte, los textos localizados muestran una preferencia casi exclusiva en el uso de la forma verbal en modo imperativo a la hora de dirigirse al usuario en estos casos, tales como «póngase en contacto con nosotros» o «ponte en contacto con nosotros». El análisis de las líneas de concordancia nos ha permitido observar cómo gran cantidad de textos localizados publicados sujetos a un proceso de Análisis de Calidad presentan mayoritariamente una formulación de la apelación más directa y radicalmente diferente de como se expresa esta unidad de sentido recurrente en textos españoles. Así, la calidad de los textos localizados, a pesar de que el uso del imperativo sea correcto desde un punto de vista sintáctico, se vería afectada puesto que este uso no se corresponde con la convención establecida en los textos pertenecientes al mismo género y tipo textual en la región lingüística de destino (Nord 1997; Nobs, 2006; Gamero, 2001).

#### 5. Conclusiones

El objetivo de todo proceso de localización es producir textos que sean recibidos como si de producciones originales en la región lingüística de destino fuera (LISA, 2003), toda vez que el proceso de traducción supone inevitablemente la producción de textos con características diferenciadas al tratarse de procesos de producción textual de distinta naturaleza (Baker, 1999). Es por lo tanto lógico asumir que la calidad de los textos localizados está directamente relacionada con el mayor grado de similitud posible con la población textual originalmente producida en la *locale* meta. En este sentido, el uso de corpus y su análisis con programas de concordancia se presenta como una herramienta clave en la consecución de este objetivo, más allá de la erradicación de

errores de lengua o de transferencia típicos en la industria. Este artículo ha repasado las posibles ventajas tras la introducción práctica de corpus textuales representativos en el proceso de traducción y evaluación en localización, tales como su uso en la investigación terminológica, fraseológica, ontológica o sobre las peculiaridades pragmáticas a las que se han de ajustar los textos localizados.

A pesar de las reticencias del sector de la localización a su uso (Bowker y Barlow, 2008; Bernardini y Castagnoli 2008; Biau y Pym 2006), hemos ilustrado cómo un corpus de referencia o de control puede constituir una herramienta clave una vez integrada en el proceso de localización. Asimismo, mediante el uso de un corpus comparable con una sección de sitios localizados de las mayores empresas estadounidenses, se ha puesto de manifiesto que la integración de tal metodología supondría un gran salto cuantitativo en la consecución de unos textos localizados de mayor calidad.

## 6. Bibliografía

Askehave, I. y A. E. Nielsen. (2005). «Digital genres: a challenge to traditional genre theory», *Information Technology and People*, 18, pp. 120-141.

Atkins, J. S. Clear, y N. Olster. (1992). «Corpus Design Criteria», *Literary and Linguistic Computing*, 7, pp. 1-16.

Austermühl, F. (2006). «Training Translators to Localize», en Pym, A. Perekstenko, A. y Starink, B. (eds.) (2006), *Translation Technology and its Teaching*, Tarragona: Intercultural Studies Group, pp. 69-81.

Baker, M. (2004). «A corpus-based view of similarity and difference in translation», *International Journal of Corpus Linguistics* 9, pp. 167-193.

Baker, M. (1999). «The Role of Corpora in Investigating the Linguistic Behaviour of Professional Translators», *International Journal of Corpus Linguistics* 4 (2), pp. 281-298.

Baker, M. (1995). "Corpora in Translation Studies: An overview and some suggestions for future research", *Target* 7, pp. 223-243.

Baker, M. y Olohan, M. (2000). «Reporting that in Translated English: Evidence for Subconscious Processes of Explicitation?», *Across Languages and Cultures* 1, pp. 141-158.

Bass, S. (2006). «Quality in the Real World», en Dunne, K. (ed.) (2006), *Perspectives on Localization*, Ámsterdam-Filadelfia: John Benjamins, pp. 69-84.

Beeby, A., Rodríguez Inés, P., y Sánchez-Gijón, P. (2009). *Corpus Use and Translation*. Ámsterdam-Filadelfia: John Benjamins.

Bernardini, S. y Castagnoli, S. (2008). «Corpora for Translation Education and Translation Practice», en Yuste Trigo, E. (ed.), *Topics in Language Resources for Translation and Localization*, Ámsterdam-Filadelfia: John Benjamins, pp. 39-55.

Biau Gil, J. R. y A. Pym (2006). «Technology and translation: a pedagogical overview», en Pym, A., Perekstenko, A. y Starink, B. (eds.) (2006), *Translation Technology and its Teaching*, Tarragona: Intercultural Studies Group, pp. 5-20.

Biber, D. (1993). «Representativeness in corpus design». *Literary and Linguistic Computing* 8 (4), pp. 243-257.

- Borja Albí, A. (2000). *El texto jurídico inglés y su traducción al español*. Barcelona: Ariel.
- Bowker, L. (2002). *Computer Aided Translation Technology: A Practical Introduction* Ottawa: University of Ottawa Press.
- Bowker, L. (2001). «A Corpus-Based Approach to Translation Evaluation», *Meta* 46 (2), pp. 345-363.
- Bowker, L. (1998). «Using specialised native-language corpora as a translation resource: a pilot study», *Meta* 43. En línia < <http://www.erudit.org/meta/1998/v43/n4/index.html> >.
- Bowker, L. y Barlow, M. (2008). «A comparative evaluation of bilingual concordancers and translation memory systems», en Yuste Trigo, E. (Ed.), *Topics in Language Resources for Translation and Localization*, Àmsterdam-Filadelfia: John Benjamins, pp. 1-22.
- Bowker, L. & J. Pearson (2002). *Working with Specialized Language: a practical guide to using corpora*. London: Routledge.
- Brinck, T. D. Gergle and Wood S. D. (2002). *Usability for the Web*. San Francisco: Morgan Kaufman.
- Corpas Pastor, G. (2001). «Compilación de un corpus ad hoc para la enseñanza de la traducción inversa especializada», *Trans* 5, pp. 155-184. En línia < [http://www.trans.uma.es/Trans\\_5/t5\\_155-184\\_GCorpas.pdf](http://www.trans.uma.es/Trans_5/t5_155-184_GCorpas.pdf) >
- Cronin, M. (2003). *Translation and Globalization*. Londres: Routledge.
- Dieguez Morales, M.I: (2008). «Análisis terminológico de sitios web localizados del inglés al español: uso de técnicas de amplificación y elisión», *Tradumàtica* 6. En línia < <http://www.fti.uab.es/tradumatica/revista/num6/articles/09/9art.htm> >.
- DIN 2345 Edition: 1998. *Übersetzungsaufträge*.
- Duff, A. (1981). *The third language: recurrent problems of translation into English*. Oxford: Pergamon Press.
- Dunne, K. (2006). «A Copernican Revolution», en Dunne, K. (ed) (2006), *Perspectives on Localization*, Àmsterdam: John Benjamins, pp. 1-11.
- EAGLES (1996). «Text Corpora Working Group Reading Guide», *Documento Eagles* (Expert Advisory Group on Language Engineering) EAG-TCWG-FR-2.
- Esselink, B. (2001). *A Practical Guide to Localization*. Àmsterdam-Filadelfia: John Benjamins.
- Faber, P., López, C. I. y Tercedor Sánchez, M. (2001). «Utilización de técnicas de corpus en la representación del conocimiento médico». *Terminology* 7(2), pp.167-197.
- Fletcher, W.H. (2007). «Concordancing the Web: Promise and problems, tools and techniques», en Hundt, M. Nesselhauf, N. y Biewer, C. (eds.) (2007), *Corpus Linguistics and the Web*. Àmsterdam: Rodopi.
- Frawley, W. (1984). «Prolegomenon to a theory of translation», en Frawley W. (ed.), *Translation. Literary, linguistic and philosophical perspectives*. London-Toronto: Associated University Presses, pp. 159-175.

- Gamero Pérez, S. (2001). *La traducción de textos técnicos*. Barcelona: Ariel.
- Gellerstam, M. (1996). «Translation as a source for cross-linguistic studies» en Aijmer, K., Altenberg, B. y Johansson, M. (eds.), *Languages in Contrast*. Lund: Lund University Press, pp. 159-175.
- Gläser, R. (1998). »The stylistic potential of phraseological units in the light of genre analysis», en Cowie, A. P. (ed.) (1998), *Phraseology: Theory, analysis, and applications*. Oxford: Clarendon Press, pp. 125-144.
- Hoey, M. (2005). *Lexical Priming. A new theory of words and language*. Londres: Routledge.
- House, J. (2001). «Translation Quality Assessment: Linguistic Description versus Social Evaluation», *Meta* 46 (2), pp. 243-257.
- Izwaini, S. (2004). *Translation and The Language of Information Technology. A Corpus-based Study of the Vocabulary of Information Technology in English and its Translation into Arabic and Swedish*. Tesis Doctoral, UMIST, Manchester, Reino Unido.
- Janoschka, A. (2003). *Web Advertising*. Ámsterdam-Filadelfia: John Benjamins.
- Jeney, C. (2007). *Writing for the Web: a Practical Guide*. Columbus, OH: Pearson Prentice Hall.
- Jiménez-Crespo, M. A. (2008a). *El proceso de localización web: estudio contrastivo de un corpus comparable del género sitio web corporativo*. Tesis doctoral, Universidad de Granada, España. En línea < <http://hera.ugr.es/tesisugr/17515324.pdf> >.
- Jiménez-Crespo, M.A. (2008b). «Caracterización del género 'sitio web corporativo' español: análisis descriptivo con fines traductológicos», en Fernández Sánchez, M. y Muñoz Martín, R. (eds.) (2008), *Aproximaciones cognitivas al estudio de la traducción e interpretación*. Granada: Comares, pp. 259-300.
- Jiménez-Crespo, M. A. (2008c) «Web genres in Localization: a Spanish corpus study», *Localization Focus – The International Journal of Localization*. 6 (1), pp. 4-14.
- Jiménez-Crespo, M.A. (2009). «Conventions in localisation: a corpus study of original vs. translated web texts», *Jostrans: The Journal of Specialized Translation* 12, pp 79-102. En línea < [http://www.jostrans.org/issue12/art\\_jimenez.php](http://www.jostrans.org/issue12/art_jimenez.php) >.
- Johansson, S. y Oksefiell S. (eds.). (1998). *Corpora and Cross-linguistic Research: Theory, Method and Case Studies*. Ámsterdam-Atlanta: Rodopi.
- Kennedy, A. and M. Shepherd (2005). «Automatic Identification of Home Pages on the Web», Actas del XXXVIII Annual Hawaii International Conference on System Sciences, Maui, Hawaii. Los Alamitos, CA: IEEE-Computer Society.
- Kenny, D. (2001). *Lexis and Creativity in Translation. A corpus-based study*. Manchester: St. Jerome.
- Kilgarriff, A. y Grefensttete, G. (2003). «Introduction to the special issue on the web as corpus», *Computational Linguistics* 29 (3), pp. 333-347.
- Laviosa, S. (2002). *Corpus-based Translation Studies*. Ámsterdam: Rodopi.

LISA. (2003). *Localization Industry Primer*. D. Fry (ed.). Geneva: The Localization Industry Standards Association (LISA). En línia < <http://www.lisa.org/products/primer.html> >.

McEnery, A. y R. Z. Xiao. (2007). «Parallel and comparable corpora: what are they up to?», en *Incorporating Corpora: Translation and the Linguist. Translating Europe*. Clevedon: Multilingual Matters.

Nielsen, J. (2004). *The need for Web Design Standards*. En línia < <http://www.useit.com/alertbox/20040913.html> >.

Nielsen, J. (2000). *Designing Web Usability: the practice of simplicity*. Indianapolis: News Riders.

Nielsen, J. y Loranger, H. (2006). *Prioritizing Web Usability*. Indianapolis: News Riders.

Nielsen, J. y Tahir, M. (2002). *Homepage usability: 50 Websites deconstructed*. Indianapolis: News Riders.

Nobs, M. (2006). *La traducción de folletos turísticos: ¿Qué calidad demandan los turistas?*. Granada: Comares.

Nord, C. (1997). *Translating as a Purposeful Activity. Functionalist Approaches Explained*. Manchester: St. Jerome.

Olohan, M. (2004). *Introducing Corpora in Translation Studies*. Londres: Routledge.

Ørsted, J. (2001). «Quality and efficiency: Incompatible elements in Translation Practice?», *Meta* 46, pp. 438-447.

Pearson, J. (1998). *Terms in Context*. Ámsterdam: John Benjamins.

Pérez Hernández, C. (2002). *Explotación de los corpórea textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento*. Tesis doctoral, Universidad de Málaga. En línia < <http://elies.rediris.es/elies18/> >.

Pym, A. (2003). «What localization models can learn from Translation Theory», *The LISA Newsletter: Globalization Insider* 12.

Rodríguez-Inés, P. (2008). *Uso de corpus electrónicos en la formación de traductores (inglés-español-inglés)*. Tesis doctoral. Departament de Traducció i d'Interpretació. Universitat Autònoma de Barcelona.

Sánchez-Gijón, P. (2004). «La extracción de conocimiento y terminología a partir de corpus ad hoc: el uso de documentos digitales de la web pública», *Linguistica Antverpiensa* 3, p. 179-202.

Santini, M. (2007). *Automatic Identification of Genre in Webpages*. Tesis doctoral sin publicar. Universidad de Brighton.

Schäler, R. (2008). «Linguistic Resources and Localization», en Yuste Trigo, E. (ed.) (2008), *Topics in Language Resources for Translation and Localization*. Ámsterdam-Filadelfia: John Benjamins, pp. 195-225

Shreve, G. M. (2006). «Corpus Enhancement and localization», en Dunne, K. (ed.) (2006), *Perspectives on Localization*. Àmsterdam-Filadelfia: John Benjamins, pp. 309-331.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Swales, J. M. (1990). *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Toury, G. (1995). *Descriptive Translation Studies and beyond*. Àmsterdam-Filadelfia: John Benjamins.

Tercedor Sanchez, M. I. (2005). "Aspectos Culturales en la localización de productos multimedia", *Quaderns. Revista de Traducció* 12, pp. 151-160.

Teubert, W. (1996). «Comparable or Parallel Corpora?», *International Journal of Lexicography* 9 (3), pp. 238-265.

Tirkkonen-Condit, S. (2004). «Unique items – over – or under-represented in translated language?» In *Translation Universals, Do They Exist?*, en Mauranen, A. y Kujamäki, P. (eds.) (2004). Àmsterdam-Filadelfia: John Benjamins, pp. 177-184.

Trosborg, A. (2000). «Translating Hybrid Political Texts», en Trosborg A. (ed.) (2000), *Analysing Professional Genres*. Àmsterdam: John Benjamins, pp. 145-158.

Wilkinson, M. (2005). «Using a Specialized Corpus to Improve Translation Quality», *Accurapid* 9 (3).

Wright, S. E. (2006). «Language Industry Standards», en Dunne, K. (ed.) (2006), *Perspectives on Localization*. Àmsterdam-Filadelfia: John Benjamins, pp. 241-278.

Yuste Trigo, E. (ed.) (2008). *Topics in Language Resources for Translation and Localization*. Àmsterdam-Filadelfia: John Benjamins.

Zanettin, F. (2002). «DIY Corpora: The WWW and the Translator», en *Training the Language Service Provider for the New Millennium*, B. Maia, J. Haller y M. Urlrych, (eds). Porto: Faculdade de Letras, Universidade do Porto. En línia <http://www.dedericozanettin.net/DIYcorpora.html>.

Zanettin, F. (1998). «Bilingual Comparable Corpora and the Training of Translators», *Meta* 43 (4), pp. 616-630.