

# Testimonios de la evolución de los estudios traductológicos basados en corpus durante la última década (2003-2013)

## 1. Una realidad caleidoscópica

El cariz a primera vista caleidoscópico de esta obra no es sino buena muestra de la evolución de la disciplina de los estudios traductológicos basados en corpus durante su segunda década de existencia (2003-2013). Laviosa (2014) describe esta década como un periodo en el que la rama descriptiva de la disciplina se ha caracterizado por el mantenimiento del interés por los universales de la traducción y la caracterización de la lengua traducida –en este último caso con un interés particular por el estudio del comportamiento de los traductores, y especialmente los literarios–, así como por la consolidación de la relación entre los estudios de traducción e interpretación y la lingüística contrastiva por medio de los corpus. Dentro de la rama aplicada de la disciplina, Laviosa subraya la contribución de los corpus en esta década a «una cultura de la investigación en educación» (Laviosa 2014)<sup>3</sup>, la creación de nuevos y numerosos corpus, con la aparición de los primeros corpus de interpretación y la investigación correspondiente y, finalmente, el uso de los corpus como recursos para la evaluación de la calidad y el desarrollo de las herramientas de traducción asistida por ordenador.

Otro rasgo que ha caracterizado el cierre de esta segunda década delimitada por Laviosa (2014) es la aparición de voces críticas que apuntan a la necesidad de una mayor transparencia metodológica en la investigación con corpus en general y sobre los universales de traducción en particular. Un ejemplo de esta postura lo encontramos en Sutter et al. (2012), en la introducción al número 13 de *Across Languages and Cultures*. Marco (2014), por su parte, sintetiza y analiza magistralmente esas voces críticas e insinúa caminos para encauzar la futura investigación en estudios traductológicos basados en corpus teniendo en cuenta la valiosa reflexión que proporcionan esas críticas.

El propósito de esta introducción es mostrar cómo los trabajos que se recogen en esta obra encajan perfectamente en la representación de esta década de consolidación de los estudios traductológicos basados en corpus llevada a cabo por Laviosa (2014). En las siguientes líneas proporcionamos una breve clasificación de las aportaciones, que describiremos y comentaremos con mayor profundidad en los apartados 2 a 6 de esta introducción.

---

<sup>3</sup> La traducción es nuestra.



(i) Dentro de la rama descriptiva de los estudios traductológicos basados en corpus, este volumen recoge estudios sobre indicadores para la investigación de la lengua traducida (Ruiz y Motoki, Etxeberria), estudios sobre el estilo de traductores literarios (Novodvorski, Humblé) y trabajos que abordan el comportamiento traductor (Tonin y Castillo, Chretariu-Braescu). Por lo que respecta a la rama aplicada, destaca el hecho de que tres de los cuatro trabajos incluidos en este volumen con reflexiones y propuestas para la didáctica pertenezcan a los estudios de *interpretación* (Sánchez y Lázaro, Santamaría y Fernández, Araújo y Correia), siendo el estudio de Trujillo y Cabrera el único dedicado a la didáctica de la traducción (turístico-comercial, en este caso). Todavía dentro de la rama aplicada queremos llamar la atención sobre el trabajo de corpus aplicados a la (meta)lexicografía, enfocado a la evaluación de diccionarios generales bilingües (Ortego y Fernández).

(ii) En la práctica totalidad de los trabajos aquí recogidos se percibe un marcado interés por explicitar los detalles metodológicos. De hecho, el trabajo de Novodvorski está enteramente dedicado a la exposición paso a paso de la metodología empleada en su tesis doctoral. Destacan asimismo en este sentido los trabajos de Vilar, Ortego y Fernández, Tonin y Castillo – estos últimos con su metodología *corpus-driven* aplicada a la historia de la traducción– así como las aportaciones de Ruiz y Motoki, Sánchez y Lázaro y Humblé. Dentro de los aspectos metodológicos, el trabajo de Novodvorski es un ejemplo de cómo el lingüista-traductólogo desarrolla su propio sistema de etiquetado en función de su interés investigador. Asimismo, queremos destacar dentro de esta instantánea de una «realidad caleidoscópica» la convivencia de enfoques cualitativos y cuantitativos dentro de la investigación con corpus: La aportación de Humblé ejemplifica cómo ciertos análisis cuantitativos (concretamente aquí la relación *type/token*) pueden dar pistas para posteriores vías de investigación de corte cualitativo. Por su parte, Vilar hace uso de una metodología cualitativa y muestra cómo el análisis cualitativo con herramientas CAQDAS se constituye en el caso de corpus pequeños como alternativa al etiquetado de las concordancias (o *in-corpus annotation* en la terminología de Smith et al., 2008). Por último, en Tonin y Castillo vemos cómo el corpus (las ediciones sinópticas) no es el único objeto de análisis cualitativo, sino que junto a este se analizan también otros objetos diferentes, en este caso los paratextos de las ediciones, que permiten una explicación de las causas y efectos de ciertos rasgos de las traducciones. El trabajo de estas autoras demuestra que la llamada a la combinación de métodos cuantitativos y cualitativos formulada por Olohan (2002) ha dado su fruto a lo largo de esta década.

(iii) Por lo que respecta a la creación de recursos, y en correspondencia con el análisis de Laviosa (2014), en este volumen encontramos trabajos que informan sobre la creación de corpus, ya sea para describirlos o bien para informar sobre

el proceso de su compilación o los fines para los que fueran compilados. En ocasiones, la información técnica que aportan los autores de los trabajos sobre el proceso de compilación de sus respectivos corpus es muy inspiradora y ayuda a su replicabilidad en otros contextos de investigación (Araújo, Novodvorski). La gama de corpus considerados abarca desde los corpus *ad hoc* de referencia especializados (Arnaiz et al.) hasta grandes proyectos de acceso libre (Bilbao y Makazaga, Tonin y Castillo, Araújo y Correia) o aún en proceso de análisis (Medina y Ramírez) o incluso de compilación (Sánchez y Lázaro). De estos recursos ya creados o en proceso de creación, dos de ellos son corpus específicos de interpretación (Sánchez y Lázaro, Araújo y Correia).

(iv) Corpus de interpretación: De los tres trabajos que incluye el presente volumen sobre corpus de interpretación, uno de ellos está dedicado a la interpretación de conferencias (Araújo y Correia), y los otros dos a la interpretación en las instituciones públicas (Lázaro y Sánchez, Santamaría y Fernández). La incorporación de los corpus a los estudios de interpretación fue analizada por Shlesinger en su artículo de 1998 y está recibiendo un importante impulso en nuestros días, como demuestra el volumen de Straniero y Falbo, publicado en 2012. Los trabajos de Lázaro y Sánchez y Santamaría y Fernández suponen una novedad importante puesto que están dedicados a una modalidad de interpretación cuya investigación hasta ahora los corpus no habrían hecho entrada. Destaca en este contexto el interesante comentario que Santamaría y Fernández hacen de un recurso lingüístico-informático que, si bien no es un corpus en sentido estricto, sí está basado en un corpus de ejemplos y puede ser aprovechado en la docencia de la interpretación en las instituciones públicas médicas.

(v) Si bien Laviosa, en su caracterización de la última década dentro de los estudios traductológicos basados en corpus, aludía a la existencia de recopilaciones de corpus (como el sitio web de Federico Zanettin<sup>1</sup> o la lista comentada de corpus de Richard Xiao<sup>2</sup>), nosotros hemos observado que en este periodo ha comenzado a darse cierto grado de reflexión sobre los corpus en sí, ejemplificada en trabajos como Ahmad (2008) y Borja (2008) o Zanettin (2012), por citar solo algunos. Esta reflexión surge muy probablemente como resultado de la proliferación de nuevos y muy diferentes corpus, fruto, a su vez, de los avances técnicos que permiten su compilación, almacenamiento, análisis y explotación. La reflexión parece desarrollarse en torno a dos ejes: la clasificación y la evaluación de corpus, observándose paralelismos en otras disciplinas hermanas, como p. ej. Rojo (2010) o Briz y Albelda (2009) dentro de la lingüística del corpus aplicada a los estudios hispánicos. En este volumen, los trabajos de Gorretti Faya, Teresa Tomasziewicz, así como el de Bilbao y Makazaga se insertarían dentro de esta corriente.

(vi) Una última tendencia de la que da testimonio este volumen es la de los corpus en contextos interdisciplinares: Medina y Ramírez han compilado un corpus



de sitios web localizados cuyo análisis puede arrojar conclusiones útiles para otras disciplinas, como el comercio electrónico y el análisis de mercados en este caso. Por otra parte Tomaszkiwicz incluye una interesante reflexión final sobre la posibilidad e incluso la necesidad de analizar un mismo corpus con herramientas de varias disciplinas diferentes a las de la propia traductología.

Finalmente, y antes de pasar a introducir en detalle el contenido de este volumen, quisiéramos llamar la atención precisamente sobre la etiqueta «estudios traductológicos basados en corpus» (v. Sánchez y Lázaro en este volumen). El lector familiarizado con este tipo de investigación advertirá que la denominación «estudios *de traducción* basados en corpus» está más generalizada que la propuesta por Sánchez y Lázaro y empleada aquí. Sin embargo, dado el creciente protagonismo de los estudios de interpretación que recurren a corpus, demostrado aquí con la presencia de tres trabajos, consideramos que la etiqueta propuesta por Sánchez y Lázaro hace más justicia a la realidad investigadora actual dentro de la disciplina.

## 2. La caracterización de la lengua traducida

El estudio de Daniel Ruiz y Chie Motoki, «**La singularidad de la onomatopeya japonesa en la lengua traducida y no traducida**», demuestra el fructífero debate que está surgiendo en torno a la hipótesis de los elementos únicos, que inicialmente fue planteada en estudios con el par de lenguas inglés y finés (Tirkkonen-Condit, 2004), puesta a prueba en un contexto de formación de traductores también en este par de lenguas (Eskola, 2004), y más recientemente en los pares de lenguas inglés-español (Martínez Vilinsky, 2012). Los autores se adentran en un terreno nunca hollado, como es el estudio de la lengua traducida japonesa a partir de textos originales españoles, utilizando la onomatopeya japonesa como indicador. Trabajar en general con corpus comparables del japonés, y en particular con unidades léxicas como la onomatopeya japonesa, plantea una serie de retos metodológicos nada despreciables. Los autores detallan estos retos en su trabajo, que ellos entienden como un primer estudio exploratorio. Sus primeras conclusiones les llevan a acuñar el interesante concepto de «falta de estímulo semántico» (y no tanto de estímulo léxico) en los textos españoles para explicar la menor presencia de onomatopeyas en los textos japoneses traducidos del español (v. pag. 61). Pese a su carácter exploratorio, este trabajo resulta muy original en lo que respecta a las explicaciones propuestas para el fenómeno descrito.

También de la caracterización de la lengua traducida, si bien esta vez del euskera, se ocupa Isabel Etxeberria en su trabajo titulado «**Pana y terciopelo en la literatura vasca traducida y no traducida. Aspectos lingüístico-estilísticos resultantes de la comparación de dos corpus**». Se trata de un trabajo que par-

te de la teoría del polisistema de Even-Zohar (1990) y se lleva a cabo en un corpus monolingüe comparable de prosa vasca, *Ereduzko Prosa Gaur* (<http://www.ehu.es/euskara-orria/euskara/ereduzkoa/>), convenientemente descrito por la autora a partir de la pág. 66, y que le posibilita comparar textos vascos originales y traducidos. Etxeberria busca caracterizar la lengua traducida de un sistema literario «joven, periférico y débil», en el que la traducción desempeña, por tanto, un papel importante, con un «30-35 % de la producción editorial vasca actual» (pág. 64). Su objetivo es identificar «qué rasgos pudieran estar influyendo en el desarrollo léxico, morfosintáctico y estilístico de la lengua vasca» a través de las traducciones que pasan a formar parte del polisistema literario vasco. Para ello, Etxeberria desarrolla una metodología muy ajustada, muy original y, sobre todo, muy claramente expuesta, que comprende (1) la identificación de indicadores organizados en grupos para analizar cuatro factores: precisión semántica, diversidad y riqueza léxicas, usos lingüísticos exclusivos del euskera y binomios *préstamo* vs. vocablo eusquérico), (2) la investigación de dichos indicadores en el corpus de textos traducidos y textos originales y (3) el análisis de los resultados. Entendemos que la metodología que emplea la autora es mixta: En la medida en la que observa los datos de los corpus para formular conclusiones podría calificarse como *corpus-driven* (motivada en corpus); no obstante, en un nivel superior la metodología es *corpus-based* (basada en corpus), puesto que la autora trata de probar la hipótesis formulada a partir de los presupuestos de la teoría del polisistema (v. para la distinción entre ambas metodologías Tognini Bonelli (2001:17), Marco y Van Lawick (2009) o Saldanha (2009), entre otros). En cuanto a las conclusiones, Etxeberria observa un mayor ajuste a la norma en la literatura vasca traducida que en la original, así como una «mayor tendencia a la precisión, el detalle y el matiz» semántico en el euskera traducido (pág. 88), amén de otras cuestiones interesantes que atañen a los textos originales.

Varias contribuciones se ocupan del estudio del estilo traductor y el comportamiento de los traductores literarios, con el apoyo de corpus o de técnicas de la lingüística del corpus. Philippe Humblé, en su inspirador estudio – «**El traductor y el inmigrante. Tres traducciones de Emine Sevgi Özdamar *La vida es un caravasar***»– aplica estas técnicas al estudio de la traducción de novelas escritas por inmigrantes, en las que lo que ha venido denominándose genéricamente «acento» es parte de la identidad del autor, y, por tanto, es objeto de un uso estilístico. Humblé operacionaliza la relación *type/token* para el estudio de la traducción del «acento», y nos permite comprender cómo el análisis cuantitativo abre vías para investigaciones cualitativas posteriores: Los resultados del estudio cuantitativo le permiten establecer unos datos objetivos iniciales y formular sobre ellos hipótesis subsiguientes, impulsando así la espiral del proceso investigador.



Desde un planteamiento metodológico diferente, puesto que trabaja con un corpus de muestras y, por tanto, desde una perspectiva cualitativa, Andreea-Aanca Braescu-Chretariu estudia varias traducciones de un mismo original con el fin de determinar el comportamiento de los traductores literarios y también de evaluar la calidad de las diferentes traducciones. En «**Les expressions idiomatiques dans les versions roumaines du *Rouge et le noir***», Braescu-Chretariu muestra el tratamiento de los fraseologismos por parte de los cinco traductores rumanos que se han enfrentado a la obra sthendaliana a lo largo del último siglo. La autora concluye que la dosificación en ocasiones exagerada de los fraseologismos en las traducciones comporta el riesgo de acercar demasiado el texto origen a la cultura meta, y detecta también dificultades a la hora de tratar en la traducción un rasgo típico de la lengua literaria de Stendhal como es la elipsis en las expresiones idiomáticas.

La contribución de Ariel Novodvorski, «**Estudios de la Traducción basados en Corpus. Aspectos metodológicos para el análisis de estilo**», se centra no tanto en resultados como en cuestiones metodológicas. El autor nos hace partícipes de su inquietud: «Aún se hace necesaria la definición de un cuadro teórico-metodológico más preciso para la investigación de los perfiles estilísticos de la traducción» (p. 118). Como reacción a esta falta de definición, nos revela paso a paso la metodología utilizada en su propio estudio. Los parámetros que Novodvorski operacionaliza («perspectiva», pág.120) para analizar el estilo traductor son el uso de la puntuación, las cursivas y los paratextos, además de la «clavidad» o *keyness* (pág. 118) y la traducción del discurso referido. El autor explica también paso a paso las técnicas y procedimientos utilizados (compilación de los corpus, etiquetados, etc.), resumidos en una relación en la pág. 119. Novodvorski emplea un corpus paralelo bilingüe de tres obras de Ernesto Sábato y sus traducciones al portugués brasileño. Llama la atención la adecuación inteligente de las herramientas informáticas a los fines investigadores, así como el desarrollo de su propio sistema de anotación, que desempeña un papel esencial en el estudio de las variables «discurso referido» y «puntuación».

La aportación de Castillo y Tonin, «Un corpus para la historia de la traducción: *Officina Barezzi* y la edición sinóptica de las traducciones españolas del *Dei delitti e delle Pene* de Cesare Beccaria», resultará de interés para aquellos investigadores de la historia de la traducción, y particularmente «la historia de la mediación cultural entre Italia y España» (p. 140), que deseen observar ejemplos de cómo combinar metodologías motivadas en corpus (*corpus-driven*) con herramientas y técnicas habituales en historia de la traducción, como el estudio de los paratextos. *Officina Barezzi* es un portal que aloja corpus paralelos, concretamente la «edición digital sinóptica de primeras traducciones de obras significativas para la cultura italiana y española» (p. 140). Se trata de un proyecto ambicioso, ejemplo de creación de nuevos recursos libres para la comunidad, en

formato electrónico y accesible, siguiendo el más puro espíritu de las humanidades digitales. Este trabajo incluye interesantes consideraciones sobre decisiones metodológicas importantes, como la selección de las ediciones para la edición sinóptica de las traducciones. Pero sin duda, uno de los puntos destacados de este trabajo reside en cómo las autoras explicitan y ejemplifican los dos posibles acercamientos al estudio comparado del método traductor a partir del material compilado en *Officina Barezzi*: (1) del paratexto al corpus (método deductivo), extrayendo conclusiones de la lectura de los paratextos y buscando ejemplos que confirmen dichas conclusiones, o bien (2) del corpus al paratexto (método inductivo), deduciendo de la comparación de traducciones el método empleado por cada traductor y confirmando las hipótesis en los paratextos.

### 3. Estudios de interpretación y la rama aplicada de los estudios traductológicos basados en corpus

En «**Corpus y formación de intérpretes: un estudio exploratorio**», Araújo y Correia presentan Per-Fide/EPIC, un corpus multimodal de interpretación simultánea bilingüe y bidireccional (EN-en /PT-pt) como extensión del corpus Per-Fide (<http://per-fide.ilch.uminho.pt>), así como los resultados de un análisis exploratorio que da indicios de cómo se puede explotar el corpus con fines investigadores, pero también, y sobre todo, formativos. Las autoras defienden que en didáctica de la interpretación simultánea sería conveniente complementar el paradigma cognitivo con aportaciones de la lingüística contrastiva: A partir de ejemplos extraídos del corpus Per-Fide/EPIC delimitan áreas especialmente problemáticas en interpretación simultánea en general y entre portugués e inglés en particular (relaciones anafóricas, tiempo/aspecto/modo, cohesión interfrástica, papeles semánticos de los argumentos), que podrían ser objeto de estudio en la clase de interpretación simultánea. Con estos análisis contrastivos sobre la producción del intérprete el alumno podría «desarrollar la conciencia metalingüística» (p. **¡Error! Marcador no definido.**) así como para «identificar idiosincrasias de los idiomas» (p. **¡Error! Marcador no definido.**). Creemos que el valor de los testimonios de interpretación que contiene el corpus, que, además de ser auténticos, son multimodales, reside no solo en la autenticidad de los datos, sino en su disponibilidad para ser investigados en trabajo autónomo por parte del alumno.

Este volumen incluye dos trabajos que versan sobre la modalidad de la interpretación en los servicios públicos, también conocida como interpretación social o interpretación comunitaria. Llama la atención el hecho de que ambos incluyen un pequeño debate sobre la (falta de) definición de esta modalidad de interpretación y la mejor manera de etiquetarla, aunque en cada caso se basen en diferentes fuentes (lo que hace aún más interesante una lectura complementaria de ambos trabajos). Concretamente, Santamaría y Fernández achacan a la etiqueta



«interpretación en los servicios públicos» el hecho de que no considera que esta modalidad de interpretación también se lleva a cabo en entidades privadas como ONGs.

En «**El uso de corpus en la formación de intérpretes en los servicios públicos en casos de violencia de género. Aplicaciones pedagógicas e investigación**», Sánchez y Lázaro justifican la necesidad de construir un corpus de conversaciones reales en la lista de características que debe poseer un intérprete que se dedique a la modalidad de interpretación en los servicios públicos. Esta lista, propuesta por Inglis (1984, en Sánchez y Lázaro, p. 186), recoge a su vez los requisitos específicos que esta modalidad exige a los intérpretes. Sánchez y Lázaro refieren las cifras de la violencia de género en España, así como los instrumentos legislativos e institucionales dedicados a paliarla. Asimismo, aluden a la visibilidad de la traducción y la interpretación como *derecho* en la propuesta de directiva COM (2011)/275 y como *necesidad de formación* en los planes de estudios universitarios (poca repercusión en este caso). Las autoras subrayan también la menor presencia de la interpretación dentro de los «estudios traductológicos de corpus» con respecto a la traducción, si bien llaman la atención sobre diferentes y muy diversos ejemplos de aplicación de corpus al estudio de la interpretación: desde corpus de interpretación simultánea para la creación de sistemas de interpretación automática (Ryu et al 2003, en Sánchez y Lázaro, en este volumen) hasta estudios de corte cognitivo (Tagaki et al 2001, Van Besien 1999, también en Sánchez y Lázaro). Finalmente, describen la metodología de su proyecto, incidiendo en la primera fase (recopilación del material) y los problemas que lleva asociados.

Tras repasar las razones de la escasez de estudios basados en corpus en el área de la interpretación comunitaria, Santamaría y Fernández deducen que el propio desconocimiento y falta de regulación de la profesión se refleja en la falta de consenso terminológico para referirse a este tipo de intérpretes. En su contribución, que lleva por título «**Aplicabilidad del uso de corpus a la docencia en interpretación. La herramienta *UniversalDoctor Project* en la formación del intérprete en el ámbito sanitario**», las autoras parten de la idea básica de que la documentación es una necesidad esencial también para los profesionales que trabajan en el ámbito sanitario. A continuación, exploran una herramienta producto de un corpus que puede servir para la fase de documentación y para que los alumnos hagan prácticas de esta modalidad de interpretación. Las autoras hacen interesantes propuestas de mejora de la herramienta, centradas en la necesidad de incorporar elementos que tengan en cuenta condicionantes culturales de la interacción médico-paciente tales como el «historial cultural» y el «grado de cooperación» del paciente.



#### 4. Otros estudios aplicados

Socorro y Cabrera presentan en «**Corpus como base para el desarrollo de material de aprendizaje**» una actividad dentro de una asignatura motivada en el aprendizaje centrado en **competencias**, impulsado en el marco del Espacio Europeo de Educación Superior. Esta actividad constituye una respuesta adecuada a una demanda existente en el entorno social en el que tiene lugar el aprendizaje (la comunidad autónoma de Canarias y su economía basada en el sector terciario), dentro del sector alojativo de la industria del turismo: la traducción de ofertas de alojamientos. En este caso, los alumnos recopilaron un corpus de ofertas alojativas de establecimientos turísticos de todo tipo publicadas en línea y redactadas en español, que, en una fase posterior, se explotó como recurso documental para una actividad de traducción especializada turístico-comercial. Los autores dedican sus esfuerzos a exponer la metodología de la actividad y las dificultades propias de la elaboración de un glosario de estas características (presencia de tecnicismos, dialectalismos, culturemas y préstamos).

El trabajo de Arnáiz, Álvarez y Corell considera también la recopilación de un corpus *ad hoc*, si bien esta vez la finalidad del mismo no es la documentación en el marco de la enseñanza de la traducción, sino en un contexto de traducción profesional. En «**Corpus de textos audiovisuales frente a corpus de textos escritos. La traducción de micropíldoras de aprendizaje**», los autores sopesan los parámetros que han de gobernar la construcción de un corpus para traducir ejemplos de un género audiovisual de muy reciente aparición como es la micropíldora de aprendizaje. El trabajo además contiene interesante información de fondo sobre el contexto en el que nace este género audiovisual y una caracterización del mismo.

En «**Los equivalentes de la informática en los diccionarios generales bilingües: ¿Utopía o realidad?**», Teresa Ortego y Purificación Fernández Nistal utilizan dos corpus generales del español (*CREA* y *Corpus del Español*) y dos corpus especializado (corpus *Open Office* dentro del sistema *OPUS* y *Corpus Tècnic* del Institut Universitari de Lingüística Aplicada) como corpus de referencia para investigar si los equivalentes propuestos en dos de los diccionarios generales bilingües de uso más generalizado en el par de lenguas inglés-español se ajustan a la realidad de uso reflejada en los corpus mencionados. En nuestra opinión, de este estudio pueden sacarse además otro tipo de conclusiones acerca de los propios corpus: Desde otra perspectiva, la metodología empleada por las autoras sirve también para evaluar los recursos. En este sentido resulta particularmente interesante el caso del corpus *OpenOffice* de *OPUS*: Al no poder documentar las autoras en él términos del dominio de la informática que pueden considerarse «clásicos». El esmero con el que las autoras detallan la metodología aplicada en este pequeño estudio permite su replicabilidad en otros contex-



tos, más aún si se tiene en cuenta que el estudio se ha llevado a cabo con recursos disponibles en línea.

Si bien hemos alojado el trabajo Karin Vilar en el bloque de *estudios aplicados*, lo cierto es que lo interpretamos en primer término como una muestra de la relación entre los estudios traductológicos y los estudios contrastivos a través de los corpus. El trabajo «**Análisis cualitativo de textos con un programa CA-QDAS**»<sup>3</sup> tiene como objetivo estudiar las microfunciones del género textual *reseña* y sus realizaciones lingüísticas en alemán y en español. La autora busca explicar unas expectativas lectoras «traicionadas» en el caso de un lector español de una reseña escrita por un autor alemán. Vilar somete los datos –dos conjuntos representativos de reseñas académicas, uno de reseñas escritas en alemán y otro de reseñas escritas en español– a un análisis cualitativo para la identificación de microfunciones. A continuación, la autora compara la representatividad de cada microfunción, así como la variedad y frecuencia de las microfunciones y los recursos lingüísticos asociados a las mismas. El estudio pone de manifiesto que la crítica negativa es un elemento representativo en la reseña alemana, pero no en la española. Vilar también demuestra que, si bien las microfunciones son las mismas en las reseñas de ambos idiomas, el español prefiere ciertas de ellas (p. ej. por la microfunción «algo no es como debe ser»). Además, en los textos alemanes la distribución de las microfunciones parece ser más homogénea. Finalmente, resultan también muy interesantes las reflexiones de la autora acerca de la dificultad de mantener una objetividad total en el análisis.

Incluimos el estudio de Medina y Ramírez, «**Metodología para evaluar la localización web en España. Primeros resultados aplicados a un corpus de sitios web de pymes agroalimentarias andaluzas**», bajo el epígrafe «estudios aplicados» por su orientación a la evaluación de la calidad en localización. En este estudio las autoras presentan resultados parciales de un trabajo más amplio que comprende la recopilación de un corpus compuesto por los sitios web de las pymes andaluzas del sector agroalimentario para, tras la aplicación de un modelo de auditoría, estudiar las características de la localización de las mismas y hacer propuestas de mejora. Este estudio pone de manifiesto la interdisciplinariedad a la que están llamados los estudios de traducción e interpretación, particularmente en este caso la manera en la que los estudios traductológicos basados en corpus y los estudios de mercadotecnia pueden nutrirse mutuamente.

## 5. Creación de recursos

En prácticamente todos los comentarios a los trabajos a los que nos hemos referido hasta ahora hemos aludido a los recursos que se han creado específicamente para llevar a cabo los diferentes estudios, ya se tratara de grandes proyectos o recursos *ad-hoc*. Queremos ahora, no obstante, detenernos mínimamente sobre

algunos trabajos que se centran específicamente en la creación de un recurso de uso general y disponible en línea.

En la contextualización de su trabajo, **Araújo y Correia** presentan el proyecto *PerFide*, que acoge a Per-Fide/EPIC (en el que luego las autoras hacen sus análisis concretos). Destacan por su claridad y su utilidad los detalles en torno a este gran proyecto de compilación que viene en ayuda de la investigación con corpus de la lengua portuguesa, para la que no existen demasiados recursos de este tipo. *Per-Fide* es un corpus multilingüe (PT / ES / RU / IT / FR / EN / DE), descargable y gratuito. Este corpus incluye textos de lo que las autoras denominan «áreas de conocimiento» de la literatura, la religión, lo jurídico-administrativo, la técnica, y el periodismo, permite búsquedas bilingües simultáneas e integra la posibilidad de búsqueda en diccionarios probabilísticos de traducción. En la segunda parte de su trabajo, Araújo y Correia pasan a describir el subcorpus *Per-Fide/EPIC*, un corpus multimodal de interpretaciones simultáneas entre el portugués europeo y el inglés británico de discursos pronunciados en el Parlamento Europeo. En este corpus, las transcripciones de las interpretaciones están alineadas con los discursos originales, y ambos elementos están sincronizados con los vídeos que contienen las grabaciones de los discursos interpretados. Las autoras se refieren a los problemas que planteó el diseño de *Per-Fide/EPIC*, principalmente la imposibilidad de incluir metadatos relevantes sobre los intérpretes, carácter en gran medida ficticio de la oralidad. Además, describen brevemente pero de manera comprensiva el proceso de compilación. Estos últimos detalles resultan, a nuestro juicio, muy interesantes, puesto que el proceso se torna bastante transparente para el lector y puede, por lo tanto, servir de estímulo para la creación de otros recursos de naturaleza similar.

Bilbao y Makazaga, en el capítulo «**EHUskaratuak: un corpus de traducciones académicas en una lengua minoritaria**», presentan el corpus *EHUskaratuak* en su contexto histórico e institucional y los objetivos de los desarrolladores del proyecto. En primer lugar los autores detallan la naturaleza del recurso: Este se nutre de las traducciones del servicio de traducción del servicio de Euskera de la Universidad del País Vasco/EHU en las que se trabaja de forma colaborativa. Se trata de traducciones de manuales universitarios, prosa de divulgación científica y leyes. El corpus se completa con una base de datos terminológica que se publicará en 2014 (*Gaika*) y un servicio de consultas sobre la lengua vasca (*Ehulku*). A continuación, Bilbao y Makazaga presentan los otros corpus existentes de la lengua vasca y los describen con arreglo a una serie de parámetros: su foco de creación, su carácter actual o histórico, la configuración lingüística, su carácter general o especializado, si contienen textos traducidos o no y si están o no lematizados. Por último, los autores describen *EHUskaratuak* con más detalle (es bilingüe paralelo y no tanto multilingüe) y ejemplifican algunos tipos de búsqueda. Como conclusión, Bilbao y Makazaga señalan que el objeti-



vo de los creadores de *EHUskaratuak* es seguir creciendo hasta incluir todas las traducciones académicas coordinadas por el servicio de euskera de la EHU (2007-2015) y subrayan que este corpus cumple los requisitos que exigen los investigadores y los traductores, que son los principales grupos de usuarios del mismo.

## 6. Evaluación y clasificación de corpus

Si bien el trabajo de Bilbao y Makazaga, con su metodología para la clasificación de corpus, podría igualmente incluirse dentro de este apartado. No obstante, ya hicimos referencia al mismo en el apartado «Creación de recursos», por lo que nos centraremos en este punto final en las aportaciones de Goretta Faya y de Teresa Tomasziewicz.

En «**Propuesta de clasificación de corpus**», Faya recopila la más variada terminología utilizada para etiquetar corpus y organiza los conceptos subyacentes en una nueva clasificación en forma de árbol, de carácter integrador y ampliable. Esta propuesta de clasificación resulta útil para describir en profundidad el tipo de corpus que se está utilizando en un nivel determinado, con la ventaja de que los conceptos utilizados son rastreables en la bibliografía relevante y gozan de mucha aceptación en la bibliografía especializada.

Por su parte, Tomasziewicz hace en «**Les corpus dans le cadre de la maîtrise en traduction**» un breve repaso a la noción de corpus, para concluir que el concepto de corpus no es unitario, puesto que se utiliza de manera diferente según las disciplinas (p. 361). Los corpus en traductología son datos en contexto que pueden lanzar preguntas interesantes para más de una disciplina, ya que la traductología está llamada a la interdisciplinariedad, aunque luego la investigación de los fenómenos se haga con herramientas lingüísticas o literarias (v. pág. 369). Resulta muy interesante sus reflexiones acerca de las diferentes maneras de utilizar los corpus por parte de un lingüista y de un traductor (pág. 360), así como sobre la falsa idea de que la noción de corpus es algo nuevo, propio de la investigación contemporánea: La autora achaca esta impresión errónea al hecho de que se confunden los medios tecnológicos disponibles para compilar y analizar los corpus con la idea teórica de corpus en sí (pág. 360). En un breve ejercicio de historiografía, Tomasziewicz explica cómo antes de 1990 en Polonia se investigaba en los TFM de traductología aquellos aspectos que permitían los corpus disponibles (muy restringidos, como detalla la autora), pues la compilación de corpus con los medios disponibles y las condiciones políticas reinantes por entonces era sumamente difícil. Por último, pasa revista a los tipos de corpus que se han utilizado para los TFM del máster de la Universidad Adam Mickiewicz de Poznań.

## 7. Bibliografía

- Ahmad, Khurshid. 2008. Being in Text and Text in Being: Notes on Representative Texts. @ G. Anderman y M. Rogers, eds. *Incorporating Corpora. The linguist and the translator*. Toronto: Multilingual Matters. ISBN 9781853599859, pp. 60-94.
- Borja Albi, Anabel. 2008. Corpora for Translators in Spain. The CDJ-GITRAD Corpus and the GENTT Project. @ G. Anderman, y M. Rogers, eds. *Incorporating Corpora. The linguist and the translator*. Toronto: Multilingual Matters. ISBN 9781853599859, pp. 253-265.
- Eskola, Sari. 2004. Untypical frequencies in translated language. A corpus-based study on a literary corpus of translated and non-translated Finnish. @ A. Mauranen y P. Kujamäki, eds. *Translation universals: Do they exist?* Amsterdam: John Benjamins. ISBN 9789027216540, pp. 83-99.
- Even-Zohar, Itamar. 1990. Polisystem Studies. @ *Poetics Today* 11 / 1.
- Laviosa, Sara. 2014. Corpora and Holistic Cultural Translation. @ M.T. Sánchez Nieto, ed. *Corpus-based Translation and Interpreting Studies: From description to application / Estudios traductológicos basados en corpus: De la descripción a la aplicación*. Berlín: Frank & Timme Verlag für wissenschaftliche Literatur. ISBN 9783732900848, [en preparación].
- Marco, J. 2014. Taking stock: A critical overview of research on (universal) features of translated language. @ Sánchez Nieto, M.T., ed. *Corpus-based Translation and Interpreting Studies: From description to application / Estudios traductológicos basados en corpus: De la descripción a la aplicación*. Berlín: Frank & Timme, Verlag für wissenschaftliche Literatur. ISBN 9783732900848, [en preparación].
- Marco, Josep. y van Lawick, Heike. 2009. Using corpora and retrieval software as a source of materials for the translation classroom. @ A. Beeby, P. Rodríguez-Inés y P. Sánchez-Gijón, eds. *Corpus Use and Translating: Corpus Use for Learning to Translate and Learning Corpus Use to Translate*. Ámsterdam: John Benjamins. ISBN 9789027224262, pp. 9-28.
- Martínez Vilinsky, Barbara. 2012. On the lower frequency of occurrence of Spanish verbal periphrases in translated texts as evidence for the unique items hypothesis. @ *Across Languages and Cultures* 13 / 2: 197-210.
- Olohan, Maeve. 2002. Corpus Linguistics and Translation Studies: Interaction and Reaction. @ *Linguistica Antverpiensia* 1: 419-429.
- Saldanha, Gabriela. 2009. Principles of corpus linguistics and their application to translation studies research. @ *Tradumàtica* 7. Disponible en <<http://www.raco.cat/index.php/Tradumatica/article/view/154828/206722>>. Consultado el 20/9/2014.



- Shlesinger, Miriam. 1998. Corpus-based Interpreting Studies as an Offshoot of Corpus-based Translation Studies. @ *Meta* 43 / 4: 486-493.
- Smith, Nicolas, Sebastian Hoffmann y Paul Rayson. 2008. Corpus Tools and Methods, Today and Tomorrow: Incorporating Linguists' Manual Annotations. @ *Literary and Linguistic Computing* 23 / 2: 163-180.
- Straniero Sergio, Francesco y Caterina Falbo, eds. 2012. *Breaking Ground in Corpus-based Interpreting Studies*, Peter Lang, Berna, etc.
- Sutter, Gert, Patrick Goethals, Thorsten Leuschner & Sonja Vandepitte. 2012. Towards methodologically more rigorous corpus-based translation studies. @ *Across Languages and Cultures* 13 / 2: 137-143.
- Tirkkonen-Condit, Sonja. 2004. Unique items—over-or under-represented in translated language? @ A. Mauranen y P. Kujamäki, eds. *Translation universals: Do they exist?* Amsterdam: John Benjamins. ISBN 9789027216540, pp. 177-184.
- Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Ámsterdam: John Benjamins. ISBN 9789027222763.
- Zanettin, Federico. 2012. *Translation-Driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies*. Manchester: St. Jerome. ISBN: 9781905763290.

---

<sup>1</sup><https://sites.google.com/site/federicozanettinnet/cl-htm#TOC-Translation-driven-Bilingual-and-Multilingual-Corpora>. Fecha de consulta: 16/09/2014.

<sup>2</sup><http://www.lancaster.ac.uk/fass/projects/corpus/cbls/corpora.asp>. Fecha de consulta: 16/09/2014.

<sup>3</sup> CAQDAS: Computer Assisted Qualitative Data Analysis Software.