

EL CORPUS PARALELO: HERRAMIENTA PARA EL ESTUDIO DE TEXTOS PROCEDENTES DEL INGLÉS MODERNO TEMPRANO Y SUS TRADUCCIONES AL ESPAÑOL

JUDIT MARTÍNEZ MAGAZ
UNIVERSIDAD DE LEÓN

1. Introducción

El objetivo que persigue la investigación que aquí se esboza es la construcción de un corpus paralelo informatizado de textos en las lenguas inglesa y española. Formarán parte de dicho corpus los textos literarios procedentes del período del Inglés Moderno Temprano, período que comúnmente abarca los siglos XVI y XVII. Estos textos incluyen pasajes escritos en las diferentes variedades regionales inglesas o al menos tratan de representar los distintos dialectos que se hablaban en la época. Las traducciones al español de dichos textos son las realizadas en los siglos XX y XXI.

El corpus paralelo permitirá así la realización de un estudio comparativo entre los rasgos dialectales que aparecen en los textos ingleses y las soluciones adoptadas por los traductores al español. En efecto, éste es el objetivo principal que se persigue en la investigación que se realizará durante los próximos años, ya que la construcción de un corpus informatizado de estas características requiere una serie de tareas previas que son las que se han llevado a cabo durante los últimos meses, y que aquí resumimos:

1. Primera etapa (Etapa Teórica): En un primer momento, ha sido necesario realizar el trabajo de índole más teórica: el estudio de los corpus informatizados de textos, y en especial lo que toca a los corpus paralelos entre lenguas distintas. Seguidamente se ha realizado un trabajo teórico y crítico sobre las principales características de la lengua inglesa en el Inglés Moderno Temprano, tanto de la lengua estándar como de las variedades regionales inglesas consideradas no estándar, para posteriormente reconocer el tipo de rasgos que podemos encontrar en los textos literarios.

2. Segunda etapa (Trabajo de Campo): Se procede, por una parte, a la búsqueda de las traducciones al español de los textos para elaborar el catálogo de las mismas, y por otra, al análisis de los pasajes ingleses escritos en dialecto para después introducir todos esos datos en el corpus informatizado. Es en este punto de la investigación en el que nos encontramos actualmente. Ambas tareas siguen caminos completamente distintos, pero en muchas ocasiones, debido a factores como el tiempo o la disponibilidad de los textos, ha sido necesario alternarlas para realizar nuestro trabajo de una manera continua.

En la investigación que perseguimos, Corpus, IMT (Inglés Moderno Temprano), Dialecto y Variación son conceptos clave siempre presentes, y de este modo se convierten en los principales puntos a tratar en la presente comunicación.

2. Los corpus: características generales y tipología

En la actualidad, un corpus lingüístico se define como un conjunto de textos preparado y adaptado para la investigación lingüística. En palabras de A. SÁNCHEZ (2001: 13):

Un corpus lingüístico es un conjunto de datos lingüísticos (pertenecientes al uso oral o escrito de la lengua, o a ambos), sistematizados según determinados criterios, suficientemente extensos en amplitud y profundidad, de manera que sean representativos del total del uso lingüístico o de alguno de sus ámbitos y dispuestos de tal modo que puedan ser procesados mediante ordenador, con el fin de obtener resultados variados y útiles para la descripción y el análisis.

En esta definición encontramos precisamente las tres características más sobresalientes actualmente en todo corpus, señaladas asimismo por muchos otros autores (por ejemplo, en T. McENERY y A. WILSON 1996: 21):

- representatividad
- extensión, pero abarcabilidad

-procesamiento por ordenador

Un corpus debe ser lo suficientemente representativo de la variedad u objeto lingüístico que pretendemos analizar. Para conseguir representatividad, el corpus ha de ser extenso, y buscar variedad de estilos y de registros. Debe ser, además, abarcable, puesto que, de otro modo, la investigación no podría, quizá, llevarse a cabo. La última característica, esto es, el hecho de que los corpus estén informatizados, permite, por un lado, el procesamiento de cantidades ingentes de información impensables sólo cincuenta años atrás, y por otro, combinarlo con un análisis detallado de los textos individuales.

Aún hoy en día, después de varias décadas recorridas desde el nacimiento de los corpus en soporte informático, no existe una clasificación homogénea de los tipos de corpus. Esto se debe «a que los avances informáticos se desarrollan a gran velocidad y permiten la creación de cada vez nuevas combinaciones de conjuntos textuales, facilitando así la aparición de nuevos tipos de corpus» (N. RAMÓN GARCÍA 2000: 72). Precisamente, N. RAMÓN GARCÍA (2000) ha aunado propuestas de diversos autores y ha presentado una clasificación de los tipos de corpus según varios criterios:

1. Según el «origen de las personas que hayan producido los textos del corpus» (N. RAMÓN GARCÍA 2000: 74): normalmente, los textos de un corpus han sido producidos por hablantes nativos de edad adulta de una lengua determinada, pero también podemos encontrar corpus producidos por personas ancianas, niños; por grupos que proceden de una zona geográfica concreta (corpus de variedades regionales), o por grupos de personas de cierto nivel profesional o social (médicos, abogados, etc.).

2. Según el «tratamiento gramatical que haya recibido el corpus» (N. RAMÓN GARCÍA 2000: 73): los corpus pueden ser anotados/etiquetados o no anotados/no etiquetados. Un corpus no anotado se limita a la compilación de los textos sin más. En un corpus anotado, en cambio, las palabras aparecen acompañadas de etiquetas que proporcionan información de diversos tipos (generalmente, información gramatical, pero también extralingüística: autor del texto, fecha de composición, la fuente de la que se ha tomado, etc.).

3. Según el origen de los textos: los corpus pueden ser sincrónicos, aquellos que «recogen textos de un estadio histórico concreto de la lengua y diacrónicos aquellos que recogen textos de diversas etapas en la evolución de la lengua» (N. RAMÓN GARCÍA 2000: 72); existen corpus de textos escritos, orales o mixtos; corpus abiertos (aquellos a los que se añaden constantemente nuevos textos) o cerrados; corpus de textos completos o de fragmentos; corpus generales (pretenden una representación de toda la lengua) o especializados (recogen muestras de un campo de conocimiento en concreto: textos políticos, administrativos, legales, etc.); corpus monolingües, bilingües y multilingües. Dentro de éstos últimos se abren dos grandes grupos: corpus comparables y corpus paralelos. Un corpus comparable se compone de textos originales en una lengua A y originales en lengua B, relacionados por criterios como fecha de composición, género, público al que van dirigidos, etc. Sin embargo, M. BAKER (1995: 234) usa esta etiqueta para referirse a lo que ella llama «a cross between parallel and multilingual corpora»: se trata de dos conjuntos de textos: un corpus formado por textos originales en una lengua, y otro compuesto por textos en esa misma lengua, pero que son traducciones provenientes de otras. En UMIST (Manchester, Reino Unido) y en Lund (Suecia) se han desarrollado proyectos de este tipo de corpus¹.

Entre todos estos tipos de corpus, nos detendremos en los corpus paralelos, ya que son los que interesan en nuestra investigación.

2.1. Los corpus paralelos

Un corpus paralelo es el formado por textos originales en una lengua A y sus traducciones (y supuestas traducciones) en lengua B. Estos textos cubren los mismos significados y tienen idénticas funciones en ambas lenguas. K. AIJMER y B. ALTENBERG (1996: 12)

¹ Para más detalles sobre los ejemplos ofrecidos y sobre la terminología adoptada por M. BAKER (1995), véase N. RAMÓN GARCÍA (2001: 149).

han formulado así las posibilidades de comparación entre lenguas que ofrecen estos tipos de corpus bilingües/multilingües:

- ofrecen nuevas perspectivas de las lenguas a comparar, perspectivas que pasarían inadvertidas con el análisis de un corpus monolingüe;

- se pueden usar para gran variedad de fines comparativos, y así aumentar nuestro conocimiento de las diferencias lingüísticas, culturales y tipológicas, además de características universales;

- ponen de manifiesto las diferencias que hay entre original y traducción; y

- se pueden usar para un gran número de aplicaciones en lexicografía, enseñanza de lenguas y en traducción.

Para conseguir el correcto y fiable desarrollo de un corpus paralelo, es necesario seguir una serie de fases, que resumiremos del siguiente modo:

1. Selección de los textos e informatización de los mismos.

2. Anotación y etiquetación de los textos. Normalmente, a todo texto lo precede un encabezamiento (en inglés, *header*), que contiene información bibliográfica y clasifica el texto según género, autor, fecha de composición, etc.²

3. El siguiente paso es la alineación de los textos: «To *align* a text with a translation of it in another language is [...] to show which of its parts are translated by what parts of the second text» (M. KAY y M. RÖSCHEISEN 1993: 121). Así pues, cada texto original cuenta con su traducción correspondiente. Hoy en día existen diversos programas de alineación automática, entre ellos citaremos *WinAlign* del sistema Trados 5.5, que es uno de los más utilizados. K. HOFLAND y S. JOHANSSON (1998) han desarrollado su propio programa de alineación para el Corpus Paralelo Inglés-Noruego (*The English-Norwegian Parallel Corpus-ENPC*), que es un corpus paralelo «especial», ya que incluye textos originales en inglés y sus traducciones al noruego, y textos originales en noruego y sus traducciones al inglés, compilado en la Universidad de Oslo desde 1994 y ya completado³. En los programas de alineación, la pantalla del ordenador aparece dividida en dos partes: en una de ellas está el texto original, y en la otra su traducción. Una comparación fiable y rigurosa entre las lenguas requiere, además de estos programas, una comprobación manual posterior para corregir cualquier posible error del programa informático. Por otra parte, son necesarias una serie de herramientas informáticas, entre las cuales quizá la más básica es la que se conoce como *browser*. Un *browser* es simplemente un explorador que busca elementos en un corpus (J.EBELING 1998) Así, en la búsqueda directa de una palabra, el *browser* ofrecerá la primera oración donde aparezca esa palabra y su correspondiente oración en el texto paralelo. También nos puede informar de cuántas veces aparece una palabra determinada en el corpus, pero, como señala J. EBELING (1998: 104), las funciones que ofrezca un *browser* dependerán de la flexibilidad de sus opciones de búsqueda. Cuanto más complejas y pulidas sean, mejor será el *browser*. Es obvio que, entre lo que puede localizar, con la adecuada codificación, están los rasgos dialectales, que son los que precisamente interesan en nuestra investigación.

4. Una vez alineados los textos que formarán el corpus paralelo, se procede a la comparación entre lenguas para después ofrecer los resultados.

Los corpus paralelos y su utilidad no han dejado de ser controvertidos, ya que, como ha señalado M. BAKER (1993: 233), en ocasiones, los textos traducidos han sido considerados «as no more than second-hand and distorted versions of "real" texts». En ellos puede haber errores cometidos por la incompetencia o poca atención del traductor, aparte de las lógicas limitaciones lingüísticas al pasar de una lengua a otra (H. PAULUSSEN 1997: 504). De todos modos, bien es cierto que, junto a los corpus comparables, los corpus paralelos se han convertido en una valiosísima herramienta para la comparación entre lenguas.

² Para más información sobre anotación de corpus y sus encabezamientos, véase T. McENERY y A. WILSON 1996: 31-32.

³ Principales compiladores: Stig Johansson, Jarle Ebeling y Signe Oksefjell (véase página web <http://www.hf.uio.no/iba/prosjekt/>).

En efecto, y aun teniendo en cuenta los aspectos negativos que acabamos de comentar, los corpus computerizados ofrecen hoy en día todo tipo de aplicaciones en las distintas ramas y campos lingüísticos. En nuestro caso, nos conciernen los siguientes:

1. Los Estudios de Traducción: los Estudios de Traducción conforman una disciplina académica cuyo objeto de estudio es el fenómeno de la traducción. Los textos traducidos «record genuine communicative events and as such are neither inferior nor superior to other communicative events in any language. They are however different, and the nature of this difference needs to be explored and recorded» (M. BAKER 1993: 234). Los Estudios de Traducción se han visto muy favorecidos por las investigaciones que se llevan a cabo a través de grandes corpus lingüísticos. Como ya «predijo» Mona Baker en 1993:

We are about to witness a turning point in the history of the discipline. I would like to argue that this turning point will come as a direct consequence of access to large corpora of both original and translated texts, and of the development of specific methods and tools for interrogating such corpora in ways which are appropriate to the needs of translation scholars (M. BAKER 1993: 235).

Para M. BAKER (1993: 247), el acceso a corpus computerizados da la oportunidad de investigar aspectos teóricos que de otro modo sería imposible conocer a través de estudios a pequeña escala. Estas cuestiones teóricas incluyen el tamaño y la naturaleza de la unidad de traducción o el tipo de equivalencia conseguida en traducción y a qué nivel. Los corpus paralelos son, obviamente, cruciales para estos fines, además de que ayudan a ofrecer modelos para poner en práctica en traducciones posteriores.

2. Dialectología y variación lingüística: no existen muchos corpus que recojan específicamente variedades dialectales. Sin embargo, éstos son una fuente muy útil para la comparación de variedades diatópicas de una misma lengua, así como para su descripción. En este sentido, esta investigación llenará un vacío en este campo, puesto que, como ya comentamos en la Introducción, los textos literarios seleccionados presentan pasajes escritos en los diferentes dialectos ingleses.

3. Lingüística Diacrónica: ya que los textos literarios que seleccionaremos para nuestro trabajo proceden de los S. XVI y XVII, es obvio que la investigación se encuadra en el marco de la Lingüística Diacrónica. Ésta «es – por excelencia – una lingüística del corpus» (R. CARAVEDO 1999: 49). En la mayoría de los casos, al lingüista diacrónico no le queda más remedio que recurrir, en primera instancia, a los textos para llevar a cabo sus investigaciones. Los textos encontrados de un período histórico anterior o de una lengua muerta forman por sí mismos corpus cerrados que sólo se podrían aumentar con el descubrimiento de otros textos hasta entonces desconocidos (T. MCENERY y A. WILSON 1996: 106). El diseño de un corpus diacrónico presenta más complicaciones que diseñar un corpus sincrónico: «in addition to concerns relating to size and register diversity, there is the added parameter of time that must be adequately represented» (D. BIBER *et al.* 1998: 251). Según M. KYTÖ y M. RISSANEN (1993: 7), para compilar un corpus diacrónico hay que tener en cuenta cuatro aspectos fundamentales: el alcance cronológico (el corpus debe ser representativo de todo el período que pretende cubrir); el alcance regional (la compilación del corpus deberá tener en cuenta las variedades regionales de la lengua); alcance sociolingüístico (los textos procederán de autores que representen distintas edades, sexos, grupos sociales y niveles de educación); y al alcance de género (el corpus debe contener muestras que representen distintos géneros literarios y tipos de texto). Aun teniendo en cuenta estos aspectos, no hay que olvidar que el corpus sólo proporcionará lo que M. RISSANEN (1991: 202) ha llamado «a limited and biased picture of the reality of language». El corpus debe entenderse únicamente como la base para el análisis: es la formación del lingüista la que llevará a obtener la capacidad necesaria para interpretar los datos, unida a todo el bagaje teórico que la Lingüística Diacrónica tiene a su disposición.

3. El período del Inglés Moderno Temprano: la variedad estándar

La etiqueta Inglés Moderno Temprano (IMT) es relativamente reciente y no ha ganado popularidad sin controversia. Muy pocos se han puesto de acuerdo al establecer los límites temporales: R. LASS (1999) lo encuadra entre 1476 y 1776; el Corpus de Helsinki (*The Helsinki Corpus of English Texts*), M. GÖRLACH (1991) y A. C. BAUGH y T. CABLE ([1951] 1993) establecen

unas fechas que abarcan desde 1500 aproximadamente hasta 1700/1710. En nuestro caso particular, y para nuestros fines, hemos establecido los límites entre los S. XVI y XVII, empezando en 1500 y finalizando en 1700, que es cuando daría comienzo la etapa del Inglés Moderno. Pero, ¿porqué esas fechas y porqué el interés en este período?

La corriente humanista comienza a florecer en Inglaterra desde 1485; la educación se convirtió en un fenómeno mucho más extendido (A. C. BAUGH y T. CABLE [1951] 1993: 196); el desarrollo de los medios de comunicación durante el S. XVI favoreció el intercambio de ideas entre distintos países, algo que siempre es beneficioso para el desarrollo de una lengua, puesto que permite su difusión por tierras lejanas, así como la adopción de nuevos términos venidos de fuera (A. C. BAUGH y T. CABLE [1951] 1993: 196); y en 1476 William Caxton instala la primera imprenta en Inglaterra, lo que dio un impulso sin precedentes a la consolidación de la lengua estándar (D. CRYSTAL 1995: 56), que es precisamente el hecho más significativo del período al que nos referimos. Además de imponer y generalizar unas normas ortográficas y de puntuación, la imprenta ofreció más oportunidades para que surgieran nuevos autores y para que sus obras circularan más ampliamente.

El surgimiento de lo que se conoce como variedad estándar ocurre precisamente en IMT. Es éste un concepto de no fácil definición, pero la definición de R. LASS (1987) es la que más adecuada nos parece:

A successful standard must be widely comprehensible, socially highly valued, and «codified» to some extent; that is, if as is usual, control of the standard is a key to social advancement, a mark of having arrived (or being there already), there must be some authoritative consensus, preferably in written form, on what it consists of. The genuinely standard – especially in a class-conscious society – must be distinct from what is «ignorant», «provincial», «lower class», etc. (R. LASS 1987: 65).

En efecto, el término «estándar» hace referencia a una noción y una opción social y no tiene nada que ver con las características internas de esa variedad, al menos en un principio, ya que después se ve profundamente modificada (T. GUZMÁN GONZÁLEZ, comunicación personal). El hecho de que una variedad en concreto sea la elegida para los asuntos políticos y del gobierno, y que sea utilizada en las universidades de prestigio, fomentan su posterior adopción como estándar de esa lengua (R. LASS 1987: 6). La variedad estándar «tiene, por una lado, algo de aceptación generalizada y, por otro, algo de imposición, siendo un modelo que básicamente se adquiere en las escuelas» (M. MELE MARRERO y M^a A. MARTÍN DÍAZ 2001: 574). Cualquier proceso de estandarización comienza por la fonología, y luego por la adopción de una determinada norma escrita (R. LASS 1987: 65-66).

En la lengua inglesa, la variedad dialectal que se erigió como estándar fue la del área de Londres, ya ésta era la capital política, económica y comercial de toda Inglaterra. Además, no podemos olvidar que los centros culturales de Oxford, Cambridge y Canterbury se encuentran bastante cerca de la ciudad (R. LASS 1987: 67). Pero, sobre todo, fue la Cancillería Real (*Royal Chancery*) la que dio el impulso definitivo que la variedad de Londres necesitaba para convertirse en estándar. Desde 1471, los documentos oficiales firmados por el rey Enrique V a través de la Secretaría Real (*Signet Office*), empiezan a circular por todo el país. El gran órgano administrativo nacional era la Cancillería, responsable de los escritos administrativos, de los impuestos, etc (N. F. BLAKE 1996: 177). Los escribas de la Cancillería empezaron entonces a escribir documentos «con una ortografía, un léxico y unos rasgos morfológicos razonablemente homogéneos» (P. TEJADA CALLER 1999: 136). La difusión de esos escritos hizo que los patrones utilizados, con muchas influencias latinas a través del francés, se impusieran sobre cualquier otro modelo, pues ese inglés de la Cancillería era la variedad considerada de mayor prestigio, propia de la corte y de las clases educadas⁴.

En el período que nos ocupa, y desde el momento en que una variedad adquiere el estatus de estándar, suele identificarse con el modo «correcto» de hablar y escribir. John Hart, uno de los más importantes reformadores de la ortografía de aquel período en su *Orthographie* de 1569, explicaba que el habla «educada» de Londres era «that speach which euery

⁴ Aunque la mayor parte de los investigadores coinciden en afirmar que la variedad de la Cancillería fue la que definitivamente se tomó como estándar, recientemente M. BENSKIN (1992) ha arrojado serios argumentos en contra.

reasonable English man, will the nearest he can, frame his tongue therevnto» (citado en R. LASS 1987: 67). Cuando semejantes juicios de valor entran en juego (N. F. BLAKE 1996: 187), la consecuencia inmediata es la infravaloración del resto de variedades, que empiezan a ser asociadas a usos incorrectos e incultos.

3.1. Las variedades dialectales

Cualquier lengua del mundo presenta variedades. Éstas (ya sean de tipo geográfico, social, estilístico, etc.) son algo inherente a las lenguas. El estado natural de una lengua es la variación. Para M. GÖRLACH (1999: 459): «No language in the world is homogeneous, or ever will be». La lengua es siempre objeto de cambio.

Partiendo de este supuesto, podemos distinguir entre dos tipos principales de variedades dentro de una misma lengua:

1. Variedad estándar: como hemos visto anteriormente, el origen del estándar se encuentra en un determinado dialecto que alcanza prestigio principalmente por causas de tipo político.

2. Dialecto:

The difference between dialect and (standard) language appears to reside primarily in a ranking evaluation, the dialect being viewed as subordinate and regionally confined in relation to the (standard) language as a superordinate or overarching language. (F. VAN COETSEM 1990: 16).

La dialectología tradicional ha identificado en término «dialecto» con el tipo de variación diatópica, esto es, teniendo en cuenta el emplazamiento geográfico. Sin embargo, aspectos estilísticos y sociales también son propios de los dialectos (véanse R. LASS 1987; F. VAN COETSEM 1990; J. A. VAN LEUVENSTEDJN y J. B. BERNS, 1990).

Las variedades de tipo social son conocidas como sociolectos: los rasgos lingüísticos varían según la procedencia social del hablante, según el grupo social el que se pertenezca, aunque también entran en juego la edad o el sexo (R. LASS 1987: 237-8). En muchas ocasiones, las variedades dialectales (entendidas como geográficas) y las sociales se solapan y entremezclan.

F. VAN COETSEM (1990: 18) distingue además entre la variedad original de cada individuo (idiolecto), y la variedad propia de una misma comunidad de hablantes (comunilecto: *communilect/comunal variety*).

Para nuestro trabajo hemos elegido estudiar las variedades dialectales regionales que aparecen en los textos literarios. Normalmente, las investigaciones y los trabajos basados en dialectos ingleses tienen en cuenta la diatopía porque «regional features usually belong to an older historical stratum than social ones, and the basis for social variation is generally a regional "given" (if often with some tension between it and a standard which may come from another region». (R. LASS 1987: 216).

El estudio de la variación diatópica en el período del IMT es una tarea no exenta de dificultad, pues los datos y las evidencias con que contamos son realmente limitados: «Our knowledge of regional varieties in the Early Modern period is limited. The information provided by contemporary playwrights, poets and fiction writers may supplement the scarce data provided by orthoepists, grammarians and spelling reformers». (M^a F. GARCÍA-BERMEJO GINER 1999: 250). A esto se añade el hecho de que, en repetidas ocasiones, los rasgos dialectales que aparecen en los textos literarios no son más que alteraciones de la variedad estándar ideadas por el autor para caracterizar a determinados personajes. Aún teniendo en cuenta estas dificultades, la investigación de la variación diatópica en etapas pasadas de la historia de la lengua inglesa es necesaria y valiosa: los dialectos locales suelen conservar rasgos que se pierden en la variedad estándar, porque esos dialectos son menos permeables a las influencias de lenguas del exterior, además de que influyen en gran medida sobre el estándar, en especial en lo que respecta a la pronunciación de palabras de uso cotidiano (G. L. BROOK [1963] 1978: 35-6).

Al comienzo del período que nos ocupa, la lengua inglesa se hablaba en Inglaterra y en ciertas zonas de Escocia, Gales e Irlanda. Las tres variedades dialectales inglesas principales que podremos aislar en los textos serán precisamente las habladas en esas zonas geográficas,

aunque el escocés (*Scots*) es un caso particular y controvertido, ya que muchos autores le otorgan el estatus de lengua aparte y no de dialecto (véanse argumentos en C. BARBER [1976] 1981; M. GÖRLACH 1991).

4. Investigaciones más recientes

4.1. La búsqueda de traducciones

Puesto que el objetivo que perseguimos en nuestro trabajo es la elaboración de un corpus paralelo de textos originales y sus traducciones al español, es obvio que una de las tareas cruciales en el desarrollo de la investigación lo conforma la búsqueda de los textos traducidos a nuestra lengua. Es ésta la tarea que venimos realizando durante los últimos meses, puesto que uno de los objetivos que se persiguen más a corto plazo es la construcción del catálogo de las traducciones.

La localización de los textos, que como ya comentamos al principio de esta comunicación proceden de los S. XX y XXI, supone una búsqueda detallada en los catálogos *on-line* disponibles de todas las bibliotecas posibles, empezando por las universitarias y locales de nuestro país, y siguiendo por las de alcance más grande, esto es, las Bibliotecas Nacionales de España, Argentina, Colombia y Venezuela, la *British Library* (Reino Unido), y en Estados Unidos la *Folger Shakespeare Library* y la Biblioteca del Congreso (*Library of Congress*). Como era de esperar, las traducciones que más abundan son las de las obras teatrales de William Shakespeare. Varias de esas obras presentan variedades dialectales (*Henry V*, *Macbeth*, *The Merry Wives of Windsor*, *King Lear*, etc.) y todas ellas cuentan con numerosas traducciones en lengua española. Además de éstas, otras obras teatrales de autores del período del IMT como Thomas Middleton o Ben Jonson también cuentan con traducciones a nuestra lengua.

Mucho más difícil es encontrar poesía dialectal traducida. Ni siquiera es ésta muy abundante en inglés, y por el momento no hemos encontrado sus traducciones al español. Aún así, y aunque este resultado no es definitivo, debemos tener en cuenta que es más recurrente el dialecto en teatro, puesto que los autores normalmente lo utilizan para la caracterización de ciertos personajes.

En el momento en que el catálogo de las traducciones esté disponible, y los pasajes dialectales completamente identificados, será el momento de pasar todos esos datos a formato informático para desarrollar el corpus paralelo.

4.2. Análisis e identificación del dialecto en los textos

En Inglés Moderno Temprano tenemos que distinguir entre aquellos textos que incluyen pasajes con rasgos dialectales (para indicar la procedencia de un personaje o su nivel cultural) o los textos que están escritos completamente en un determinado dialecto (como ciertos poemas, anécdotas o historias cortas) (M. GÖRLACH 1999: 507). Para los fines que persigue nuestra investigación hemos apartado éstos últimos, ya que, hasta el momento, y mientras no se demuestre lo contrario, no se han encontrado traducciones al español de esa clase de textos. Muy posiblemente no existan, ya que se trata, en la mayoría de los casos, de composiciones muy marginales y que no suponían ningún interés para el mercado de la imprenta y la traducción.

Por tanto, los textos literarios que conformarán el corpus paralelo presentan ciertos pasajes, ya sean éstos más o menos largos en extensión, que incluyen rasgos dialectales.

Entonces, ¿qué es lo que realmente se puede identificar en los textos?. Resumiremos en cinco tipos los rasgos dialectales presentes en los textos:

1. Rasgos gráficos: ciertas grafías han sido alteradas para dar la impresión al lector de que el personaje que habla presenta una pronunciación «desviada» del estándar. Por ejemplo, destacaremos el intercambio de grafías que representan sonidos sonoros por sordos: (1) «*Got pless your house here*» (Sir Hugh Evans en *The Merry Wives of Windsor* – W. Shakespeare).

2. Rasgos morfológicos: flexiones verbales o nominales inaceptables en la variedad estándar: (2) «*I have forgot his name*» (Fluellen en *King Henry V* – W. Shakespeare).

3. Rasgos morfosintácticos: confusiones entre clases gramaticales: principalmente, entre nombres, adjetivos y adverbios: (3) «...*hath been great, reasonable great*» (Fluellen en *King Henry V*-W. Shakespeare).

4. Rasgos sintácticos: Ausencia de concordancia entre sujeto y verbo: (4) «*There is three umpires...*» (Sir Hugh Evans en *The Merry Wives of Windsor* – W. Shakespeare).

5. Rasgos léxicos: vocabulario específico del área geográfica de la que procede el personaje: (5) «*I come, Graymalkin!*» (1st Witch en *Macbeth* – W. Shakespeare).

Este análisis e identificación previa del dialecto en los textos es necesario para después poder establecer comparaciones con las soluciones adoptadas por los traductores al español.

5. Apuntes finales

Durante la presente comunicación el corpus paralelo ha sido presentado como una útil herramienta para el estudio de textos de lenguas distintas. En nuestro caso, éstos proceden de los siglos XVI y XVII, incluyendo pasajes escritos en los distintos dialectos ingleses, y sus traducciones a la lengua española.

Las principales aplicaciones de este tipo de corpus, el período histórico concreto que atañe a nuestra investigación, y cuestiones más prácticas como la identificación del dialecto en los textos literarios o la búsqueda de las traducciones a nuestra lengua son las tareas previas que hemos llevado a cabo durante los últimos meses. En especial, los dos últimos aspectos mencionados serán determinantes para la configuración futura de nuestro corpus, ya que se convierten en los «factores» que lo delimitarán: la disponibilidad de traducciones a nuestra lengua determina el número de textos ingleses en los cuales analizamos los rasgos dialectales, lo que dará como resultado un corpus más o menos amplio.

El estudio de la variación en etapas anteriores de la lengua inglesa es necesario, puesto que ha sido realizado en muy pocas ocasiones. Nuestra investigación pretende llenar ese vacío, además de realizar un estudio comparativo entre las lenguas inglesa y española. El corpus paralelo se diseñará con el fin de obtener resultados fiables durante la comparación de los textos.

El corpus en formato informático estará listo en un futuro muy próximo, adaptado a nuestros fines particulares, esto es, la investigación «diacrónico-dialectal».

Referencias bibliográficas

- AIJMER, KARIN Y ALTENBERG, BENGT, «Introduction», en AIJMER, K., ALTENBERG, B. Y JOHANSSON, M. (eds.), *Languages in Contrast. Papers from a symposium on text-based cross-linguistic studies in Lund, 4-5 March 1994*, Lund, Lund University Press, 1996, 11-16.
- BAKER, MONA, «Corpus Linguistics and Translation Studies. Implications and Applications», en BAKER, M., FRANCIS, G. Y TOGNINI-BONELLI, E. (eds.), *Text and Technology: in honour of John Sinclair*, Amsterdam-Philadelphia, John Benjamins Publishing, 1993, 233-250.
- BAKER, MONA, «Corpora in Translation Studies: An Overview and Some Suggestions for Future Research», *Target*, 7, 1995, 223-243.
- BARBER, CHARLES, *Early Modern English*, London, André Deutsch, [1976] 1981.
- BAUGH, ALBERT C. Y CABLE, THOMAS, *A History of the English Language*, London, Routledge, [1951] 1993.
- BENSKIN, MICHAEL, «Some new perspectives on the origins of standard written English», en VAN LEUVENSTEIJN, J. A. Y BERNS, J. B. (eds.), *Dialect and Standard Language*, Amsterdam, Royal Netherlands Academy of Arts and Sciences, 1990, 71-105.
- BLAKE, NORMAN F., *A History of the English Language*, London, Macmillan Press, 1996.
- BIBER, DOUGLAS, CONRAD, SUSAN Y REPPEN, RANDI, *Corpus Linguistics. Investigating Language Structure and Use*, Cambridge, Cambridge University Press, 1998.
- BROOK, G. L., *English Dialects*, London, André Deutsch, [1963] 1978.
- CARAVEDO, ROCÍO, *Lingüística del Corpus. Cuestiones teórico-metodológicas aplicadas al español*, Salamanca, Ediciones Universidad de Salamanca, 1999.

- VAN COETSEM, FRANS, «The Interaction between Dialect and Standard Language, and the Question of Language Internationalization», en VAN LEUVENSTEIJN, J. A. Y BERNIS, J. B. (eds.), *Dialect and Standard Language*, Amsterdam, Royal Netherlands Academy of Arts and Sciences, 1990, 15-70.
- CRYSTAL, DAVID, *The Cambridge Encyclopedia of the English Language*, Cambridge, Cambridge University Press, 1995.
- EBELING, JARLE, «The Translation Corpus Explorer: A Browser for Parallel Texts», en JOHANSSON, S. Y OKSEFJELL, S. (eds.), *Corpora and Cross-Linguistic Research. Theory, Method and Case Studies*, Amsterdam-Atlanta, GA, Rodopi, 1998, 101-112.
- GARCÍA-BERMEJO GINER, M^a FUENCISLA, «Methods for the Linguistic Analysis of Early Modern English Literary Dialects», en ALONSO, P., GARCÍA-BERMEJO GINER, M^a F., SÁNCHEZ, M^a J. Y MORAN, C., *Teaching and Research in English Language and Linguistics*, León, Celarayn, 1999, 249-266.
- GÖRLACH, MANFRED, *Introduction to Early Modern English*, Cambridge, Cambridge University Press, 1991.
- GÖRLACH, MANFRED, «Regional and Social Variation», en LASS, R. (ed.), *The Cambridge History of the English Language. Vol. III: 1476-1776*, Cambridge, Cambridge University Press, 1999, 459-538.
- HOFLAND, KNUT Y JOHANSSON, STIG, «The Translation Corpus Aligner: A Program for Automatic Alignment of Parallel Texts», en JOHANSSON, S. Y OKSEFJELL, S. (eds.), *Corpora and Cross-Linguistic Research. Theory, Method and Case Studies*, Amsterdam-Atlanta, GA, Rodopi, 1998, 87-100.
- KAY, MARTIN Y RÖSCHEISEN, MARTIN, «Text-translation Alignment», *Computational Linguistics*, 19: 1, 1993, 121-142.
- KYTÖ, MERJA Y RISSANEN, MATTI, «General Introduction», en RISSANEN, M., KYTÖ, M. Y PALANDER-COLLIN, M. (eds.), *Early English in the Computer Age. Explorations through the Helsinki Corpus*, Berlin-New York, Mouton de Gruyter, 1993, pp. 1-17.
- LASS, ROGER, *The Shape of English: Structure and History*, London-Melbourne: J. M. Dent & Sons, 1987.
- LASS, ROGER (ed.), *The Cambridge History of the English Language. Volume III: 1476-1776*, Cambridge, Cambridge University Press, 1999.
- VAN LEUVENSTEIJN, J. ARJAN Y BERNIS, J. B., «Introduction», en VAN LEUVENSTEIJN, J. A. Y BERNIS, J. B. (eds.), *Dialect and Standard Language*, Amsterdam, Royal Netherlands Academy of Arts and Sciences, 1990, pp. 9-14.
- MCENERY, TONY Y WILSON, ANDREW, *Corpus Linguistics*, Edinburgh, Edinburgh University Press, 1996.
- MELE MARRERO, MARGARITA Y MARTÍN DÍAZ, M^a AUXILIADORA, «Formación y desarrollo del Inglés Moderno», en DE LA CRUZ CABANILLAS, I. Y MARTÍN ARISTA, F. J. (eds.), *Lingüística Histórica Inglesa*, Barcelona, Ariel, 2001, 573-596.
- PAULUSSEN, HANS, «Parallel corpora for cognitive contrastive analysis», en LEWANDOWSKA-TOMASZCZYK, B. Y THELEN, M., *Translation and Meaning. Part 4*, Maastricht, Universitaire Pers Maastricht, 1997, 501-507.
- RISSANEN, MATTI, «The diachronic corpus as a window to the history of English», en SVARTVIK, J. (ed.), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82. Stockholm, 4-8 August 1991*, Berlin-New York, Mouton de Gruyter, 1991, 185-205.
- RAMÓN GARCÍA, NOELIA, *Estudio Contrastivo Inglés-Español de la Caracterización de Sustantivos. Una Propuesta de Análisis*, Memoria de Licenciatura Inédita, Universidad de León, 2000.
- RAMÓN GARCÍA, NOELIA, *Estudio Contrastivo Inglés-Español de la Caracterización de Sustantivos*, Tesis Doctoral Inédita, Universidad de León, 2001.
- SÁNCHEZ, AQUILINO, «Investigación y análisis mediante corpus lingüísticos: el poder de atracción de las palabras», en FERNÁNDEZ NISTAL, P. Y BRAVO, J. M. (eds.), *Pathways of Translation Studies*, Valladolid, Centro Buendía – Universidad de Valladolid, 2001, 11-45.
- TEJADA CALLER, PALOMA, *El Cambio Lingüístico. Claves para interpretar la lengua inglesa*, Madrid, Alianza, 1999.

EL CORPUS PARALELO: HERRAMIENTA PARA EL ESTUDIO DE TEXTOS PROCEDENTES DEL INGLÉS MODERNO TEMPRANO Y SUS TRADUCCIONES AL ESPAÑOL

<http://www.hf.uio.no/iba/prosjekt/> : Página web del Corpus Paralelo Inglés-Noruego (*The English-Norwegian Parallel Corpus-ENPC*).