

LA LINGÜÍSTICA DE CORPUS EN UN ESTUDIO CONTRASTIVO INGLÉS-ESPAÑOL

M^a Belén Labrador de la Cruz
Universidad de León

Probablemente sea una intencionalidad práctica y didáctica lo que ha promovido la iniciación y el desarrollo de los estudios contrastivos entre lenguas; en el siglo XVIII, James Pickbourne, el primero en utilizar el término “contraste” en este sentido, creyó que sería útil, con fines pedagógicos, contrastar el verbo inglés con el de otras lenguas (Krezswoski, 1990:2). Podría argüirse también que es la reflexión teórico-descriptiva la que después conlleva aplicaciones prácticas; se trata de dos caras de la misma moneda: la práctica de la traducción, el análisis del error y la enseñanza de segundas lenguas son a un tiempo causa y consecuencia de las descripciones comparativas de lenguas.

En el momento en el que existe una comparación, partimos de una base en principio paradójica, pues aquello que es idéntico no es susceptible de comparación mientras que lo diferente es por definición no comparable. ¿De qué tipo son entonces los objetos de nuestro estudio, tales que permiten ser sometidos a comparación? Sólo en el caso de que exista una misma base semántica, universal, que se materialice a través de distintos recursos formales en dos lenguas diferentes podemos obtener objetos que son similares semántica y/o pragmáticamente y cuyas diferencias estriban únicamente a nivel fonológico, morfológico-sintáctico y léxico.

Con el término universal no pretendemos abarcar exactamente una estructura profunda global e idéntica basada en principios gramaticales universales en el sentido chomskiano, sino que admitimos, como los representantes del distribucionalismo americano, en su versión más moderada, la importancia de la cultura y la visión del mundo de los hablantes de una lengua, y observamos hasta qué punto éstas se hallan relacionadas con el lenguaje. El hecho de considerar el significado cognitivo de forma global nos remite al concepto de *gestalt*. Según Lakoff, “*thought, perception, the emotions, cognitive processing, motor activity and language are all organized in terms of the same kind of structures*”, las llamadas *gestalts* (en Krezswoski, 1990:215). De esta forma y por mucho que varíen el conocimiento del mundo, la forma de ver las cosas y las costumbres de las distintas comunidades culturales y, aunque existen vacíos conceptuales y zonas de inequivalencia interlingüística, las lenguas, por el mero hecho de ser tipos de lenguaje humanos y naturales, constan de una serie de características comunes, o funciones básicas, desde un enfoque funcionalista, como el de Halliday, que pueden tomarse como *tertium comparationis* en una comparación interlingüística.

Una vez seleccionado el objeto de estudio y al buscar áreas de convergencia y divergencia entre los dos elementos de la comparación, nos cruzamos de nuevo con una paradoja. Aunque a primera vista se pudiese creer que la dificultad, en un trasvase aceptable, semántico-pragmático y la diferencia formal son directamente proporcionales, el hecho es que el grado de dificultad es mayor entre elementos que guardan cierta semejanza que entre aquéllos que no tienen ningún punto en común (James, 1980:188-9).

Por este motivo, al analizar la transferencia lingüística nos centraremos en aquellos rasgos diferentes y, especialmente, en los matices distintos de aquellos rasgos que aparentemente se asemejan.

Con el fin de identificar los puntos conflictivos que pueden dar origen a problemas y errores tanto en las traducciones como en el aprendizaje del inglés o el español como lengua extranjera, la opción más adecuada parece ser tomar la lingüística de corpus como método de análisis.

Basarse en la observación de datos, en la evidencia textual, aporta muchas más ventajas que tomar la intuición como único criterio para discernir fenómenos lingüísticos y pronunciar aseveraciones sobre el lenguaje. Los defensores de la introspección pueden argüir que no todo lo que se dice es correcto, por lo que no todas las oraciones recogidas en un corpus son gramaticales. Sin embargo, en general, sí serán significativas y válidas para la comunicación pues han sido extraídas de un contexto real eminentemente comunicativo. Por otro lado, existen oraciones gramaticalmente correctas que no pueden considerarse aceptables. La recursividad elevada al infinito es en teoría gramatical, pero ¿dónde se establecen los límites de la aceptabilidad?

Siempre que se produce un cambio, éste sucede primero en el *habla* y comienza siendo individual; si es aceptado y repetido por otros se convierte en social hasta que es reconocido ampliamente por una mayoría y entonces entra a engrosar las filas de la *lengua*. Es el uso el que afecta a la norma aunque ésta intente regular el uso y lo consiga, en parte. La lengua es lo suficientemente inestable para estar viva y lo suficientemente estable para que la comunicación sea posible.

Es el estudio del uso real de un idioma, que es observable, el que una vez analizado nos lleva a inducir pautas generales de comportamiento existentes en el ámbito de la *competencia* lingüística.

La observación y análisis del uso del inglés y el español, en este caso, se establece mediante un estudio descriptivo, descriptivo en una doble vertiente significativa -por un lado, como punto medio en el continuum teórico-descriptivo-aplicado y, por otro, en contraposición con el término prescriptivo, el cual implica la primacía de la norma sobre el uso.

Quizá sea cierto que cualquier hablante nativo de una lengua sepa intuitivamente si se puede o no decir una expresión en particular, tal y como afirma Chomsky; no obstante, esa tendencia innata puede hallarse condicionada

por factores sociales, geográficos, generacionales, etc., por lo que varios nativos pueden discrepar respecto a algunos casos conflictivos. Un corpus suficientemente representativo nos muestra en cambio, no sólo lo que es posible, sino lo que es común, lo que es frecuente en el habla espontánea, que carece de condicionamientos metalingüísticos. Además es difícil que a un lingüista se le ocurran todo tipo de ejemplos posibles mediante mera introspección. Una gran cantidad de ejemplos no aparece cuando los buscamos conscientemente, y en cambio afloran espontáneamente, sin ningún esfuerzo, en situaciones comunicativas reales.

Por este tipo de ventajas, incluida la fácil y rápida extracción de datos estadísticos que ofrecen los corpus computerizados, nos pareció muy conveniente decantarnos por este tipo de análisis.

Haciendo un breve repaso a la historia de la lingüística de corpus, se observa la rápida evolución y notable mejoría de los corpus. La primera generación se inaugura, en 1959, con el SEU (Survey of the English Usage), recopilado manualmente por Quirk y compuesto por 5000 palabras en sus inicios y, con el Brown Corpus, de la Universidad de Brown. La segunda generación, constituida por los corpus de Longman/Lancaster y el de la Universidad de Birmingham, se centra en el aumento del tamaño, como objetivo principal. Y la tercera generación advierte que el tamaño del corpus no es lo único importante, por lo que aunque en el de Oxford sigue prevaleciendo la cantidad, los de Brown y SEU tienden a ser muy representativos y el de Birmingham aboga por un término medio (Leech 1991:8-30).

Este último, denominado *Bank of English*, constituye el proyecto Cobuild, empresa conjunta de la editorial CollinsHarper y la Universidad de Birmingham. Hoy en día el corpus global de datos asciende a un total de 320 millones de palabras, que proceden de dieciséis fuentes distintas: revistas, periódicos, programas de radio, textos efímeros de diversa índole, textos orales coloquiales transcritos y una miscelánea de libros. El tipo de información al que se puede acceder es el siguiente: líneas de concordancia, índices de frecuencia y distribución de las palabras, nodos y sus colocaciones, etc. De este modo se pueden buscar sintagmas oracionales, combinaciones de palabras y de clases de palabras, series de inflexiones, etc. que aparecen en su propio contexto lingüístico, el cual facilita la observación de patrones formales asociados a significados diferentes.

El estado-de-la cuestión en el terreno español se encuentra algo menos avanzado y gran parte de los corpus que se han comenzado a compilar recientemente se halla aún en estado de desarrollo. La mayoría se centra en un área especializada (ej. corpus textual del español periodístico), una edad determinada (ej. corpus de habla infantil CSIC-UNED) o una zona geográfica (como el ALMECOR, corpus oral del habla rural de la Alpujarra Almeriense). Entre los

generales cabe destacar el proyecto CREA, de la Real Academia Española y el CUMBRE, de la editorial SGEL (Informe del Instituto Cervantes).

Las líneas de investigación, en el terreno de los estudios contrastivos, abiertas por la creación de amplios corpus representativos parecen muy prometedoras y pueden repercutir de forma importante en el ámbito de la traducción y en los campos editorial, escolar y universitario.

Referencias bibliográficas

Corpus escritos y orales disponibles y en desarrollo en España (1996), Alcalá de Henares: Instituto Cervantes.

JAMES, C. (1980): *Contrastive Analysis*. Singapore: Longman.

KRZESZOWSKI, T.(1990): *Contrasting Languages*. Berlin: M. Gruyter.

LEECH, G. (1991): "The State of the Art in Corpus Linguistics" en *English Corpus Linguistics*. London: Longman, pp. 8-30.

SÁNCHEZ, A. (1995): *Cumbre. Corpus lingüístico del español contemporáneo*. Madrid: SGEL.

SINCLAIR, J. (1991): *Corpus, Concordance, Collocation*. Hong Kong: Oxford University Press.