

CORPUS BILINGÜES Y LAS TRES “CES”: CONTRASTE, COMPARACIÓN Y CONTACTO LINGÜÍSTICOS

M^a. Belén LABRADOR DE LA CRUZ.

Universidad de León

Hoy en día resulta indiscutible afirmar que los grandes corpus computarizados están en boga y ya no nos sorprende ni la cantidad de corpus que han construido o están construyendo distintas universidades, editoriales y otros organismos (unas simples horas de navegación por Internet nos lo confirman) ni el tamaño de dichos corpus: *Thirty years ago, when this research started, it was considered impossible to process texts of several million words in length. Twenty years ago it was considered quite possible but still lunatic. Today it is very popular.* (Sinclair, 1991:1).

¿Qué son en definitiva los corpus y cuál es su significación hoy en día? Originariamente, la palabra *corpus*¹ designaba una colección de escritos, generalmente del mismo autor. No debemos olvidar que la palabra *corpus* procede del latín y significa cuerpo, o sea que se refiere a un cuerpo textual. La edición de 1992 de la RAE lo define como *conjunto lo más extenso y ordenado posible de datos o textos científicos, literarios, etc. que pueden servir de base a una investigación*. Para Sánchez...*cabe definir un corpus lingüístico como una colección de textos anexados uno a otro de forma secuencial* (Sánchez, 1995:40). Leech define corpus como *a sufficiently large body of naturally occurring data of the language to be investigated* (Leech, 1991:8).

Estas definiciones señalan como propiedades inherentes de los corpus su finalidad investigadora, su amplia extensión y su orden. Para explicitar un poco más las definiciones de estos autores, que consideran el corpus en un sentido amplio de la palabra, añadiremos que a lo que nos solemos referir hoy en día al utilizar el término *corpus*, es a un conjunto de textos,

¹ Las palabras inglesas procedentes del latín que terminan en *-us* tienen tres posibles plurales; algunas palabras forman su plural sólo en *-i* (*stimulus/-i*), otras sólo en *-es* (*campus/-es*), algunas pueden optar por cualquiera de los dos finales (*fungus/-es/-i*) y por último están las palabras *genus-genera* y *corpus*, que tiene dos posibles terminaciones: *corpora* y *corpuses*. El primero parece ser el elegido en toda la literatura sobre temas lingüísticos y a veces se introduce en la lengua española como anglicismo. Hasta tal punto es común este término, que a veces se cree erróneamente que es una forma de singular por lo que uno puede incluso encontrarse con la forma *corporae*, un ejemplo de hipercorrección. Sin embargo, en castellano, si una palabra es polisílaba, termina en *-s* y no es aguda, el plural no se marca. No se distinguen, pues, el singular y el plural, salvo por la concordancia con los artículos y adjetivos como, por ejemplo, la tesis/las tesis, el virus/los virus o el corpus/los corpus. Hacemos esta distinción porque en el mundo académico se escucha con demasiada frecuencia la palabra *corpora* en español por influencia negativa del inglés.

orales y/o escritos, de fuentes variadas, producidos por un elevado número de hablantes y escritores diferentes, que están procesados en ordenador y que pueden analizarse automáticamente de distintas formas.

Partiendo del hecho de que los corpus son herramientas útiles para los estudios lingüísticos de corte descriptivista, es necesario tener en cuenta que son instrumentos y fuentes de datos textuales que resultan tanto o más eficaces cuanto más se adapten a los objetivos de un tipo concreto de investigación, por lo que a continuación resumimos la variedad de corpus que existen.

Podemos establecer diferentes tipologías de corpus según distintos criterios. En cuanto a la extensión o posibilidades de extensión de los corpus y el tamaño de sus componentes, se pueden dividir en:

- corpus *limitados*: no se añaden nuevos textos al conjunto inicial; o corpus *ilimitados*: son dinámicos, permanecen abiertos, por lo que su extensión es variable, va aumentando.
- corpus *cerrados* ("*sample*"): están formados por pequeños textos tomados como muestras según criterios muy fijos de procedencia; o corpus abiertos ("*monitor*"), de gran tamaño y crecientes, es decir, se van añadiendo nuevos textos; pueden mostrar la paulatina evolución de la lengua mediante la recogida de datos que evidencian nuevos usos.
- corpus de *textos enteros*: formados por textos completos, cualquiera que sea su extensión; o corpus de *extractos*: más equilibrados y variados, por tomar fragmentos de un mayor número de textos diversos.

Por otro lado, según el criterio de selección, los corpus pueden ser *aleatorios*, formados por textos elegidos al azar; o corpus *seleccionados*, constituidos por textos escogidos cuidadosamente para lograr equilibrio y variedad.

En lo que respecta a su soporte material, existen corpus *transportables* en medios como CD-Rom, disquettes, cintas magnéticas, etc. o corpus *online* en el ordenador o una red (adaptado de Roberts, 1997).

Por último y en base al tipo de lenguaje de los textos que componen el corpus, cabe recordar la clasificación que realiza Baker (1995):

- lengua *general* versus dominio *restringido*
- lengua *escrita* versus lengua *hablada*
- estado *sincrónico* versus estado *diacrónico*
- distintos tipos de procedencia (autores y/o hablantes) y de género (artículos de periódico, ficción, científico, literario, etc.)
- según zonas geográficas (p. ej. inglés británico, inglés americano, etc.)
- *monolingüe* versus *bilingüe/multilingüe*.

Estos últimos corpus, los bilingües o multilingües, son los más apropiados para la investigación interlingüística por el hecho de describir más de una lengua. Baker añade a esta tipología un grado de precisión mayor al definir las clases de corpus que son o deberían ser utilizados en el marco de los estudios de traducción. Habla de dos tipos de corpus existentes y un tercero cuyo término es acuñado por ella misma como resultado de una propuesta personal.:

- Corpus *paralelos*: están compuestos por textos de la lengua origen y sus correspondientes traducciones en la lengua meta.
- Corpus *multilingües*: Se trata de un conjunto de textos de distintas lenguas, todos ellos texto-origen, que se han creado en la misma o distinta institución pero en base a los mismos criterios de diseño.
- Por último, los corpus *comparables* son el tipo de corpus por el que aboga Baker. Estarían formados por dos subcorpus, uno de textos originales y otro de traducciones, ambos en la misma lengua y ambos de un género, variedad lingüística y época similares y de longitud aproximada (Baker, 1995: 233-5).

Cada uno de estos tipos de corpus es más adecuado para un objetivo concreto. Así como los paralelos son muy útiles para la crítica de traducciones y para observar las características del proceso traductor, los comparables contribuyen a resaltar los rasgos característicos de los textos traducidos en comparación con los originales, y los multilingües son los que mejor sirven para estudiar la forma natural de decir las cosas en cada lengua y qué tipo de construcciones sintácticas, expresiones y vocabulario resultan más idiomáticos para expresar un determinado significado.

Nosotros añadiríamos una distinción entre los corpus multilingües y los bilingües. Los primeros se compondrían de aquellos textos escritos simultáneamente en varias lenguas distintas sin que ninguno constituya en realidad una traducción de otro. Este fenómeno, cada vez más extendido a causa de las instituciones y relaciones internacionales, ha dado lugar a una disciplina muy reciente perteneciente al ámbito de los estudios de traducción: es la llamada *Parallel Writing* o escritura paralela.

Los segundos, si bien han de estar compuestos necesariamente por textos originales en dos lenguas, no tienen que ser textos equivalentes escritos simultáneamente y bajo las mismas circunstancias. Nos referimos a la yuxtaposición de dos corpus monolingües, de dos lenguas diferentes, cuyas características respecto al tamaño global, al género, registro y variantes diatópica, diastrática y diacrónica de las dos lenguas sean lo más similares posible, pero sin tratarse necesariamente de los mismos textos.

De esta forma, los corpus bilingües son los más apropiados para observar y comparar pares de lenguas, exentas en gran medida de las influencias externas como calcos y extranjerismos que tienden a introducirse, en mayor cantidad que la habitual, en las traducciones.

La ventaja concreta de la que se benefician, en comparación con los corpus multilingües, es la mayor facilidad que tienen de formarse a partir de la incesante y potencialmente infinita producción textual de las lenguas objeto de estudio (exceptuando estudios diacrónicos, lenguas muertas u otros casos en los que existen restricciones particulares). En cambio, los corpus multilingües se hallarían condicionados, en cuanto a su fuente de datos, por la cantidad de escritura paralela que se produjese.

Los fines y métodos de estudio serían muy diferentes en ambos: Así como los corpus multilingües son de mayor utilidad en los estudios de traducción

y su uso consiste en establecer binomios textuales a partir de la correspondencia de líneas en ambos subcorpus y su consiguiente análisis comparativo, los corpus bilingües son los más adecuados en el ámbito de los estudios contrastivos y el método de análisis consiste en buscar apariciones de determinados *input*: palabras, combinaciones de palabras, estructuras sintácticas, etc., observarlas en sus contextos naturales, obtener índices de frecuencia, listados de colocaciones más habituales y otros datos estadísticos y a partir de ellos realizar estudios descriptivos y comparativos entre ambas lenguas.

Una vez situados los corpus bilingües dentro de una tipología general de los corpus lingüísticos, pasemos a esbozar unas breves ideas sobre la disciplina lingüística cuya existencia justifica la creación y utilidad de los corpus bilingües: los estudios contrastivos.

Como introducción me parece adecuado comenzar por una distinción de tipo terminológico y conceptual. Cuando hablamos de *estudios contrastivos*, nos referimos a una disciplina que estudia simultáneamente dos lenguas con el fin de establecer una comparación entre ellas que permita observar cuáles son sus semejanzas y sus diferencias, y así facilitar la comunicación interlingüística e intercultural.

Hablamos pues de *contraste*, *comparación* y *contacto* entre lenguas, pero estas tres palabras están cargadas de connotaciones con matices diferentes que conviene discernir. Observemos qué significan o, más bien, cómo se utilizan estas tres palabras en el ámbito interlingüístico.

Atendamos en primer lugar a la noción de *contacto*: Dos lenguas se ponen en contacto, con la posibilidad de influencia mutua, por medio de dos vías.

A) La primera es la vía *geográfica*. Dos comunidades lingüísticas comparten un mismo espacio o conviven en terrenos fronterizos, por lo que pueden producirse situaciones de bilingüismo, alternancia de códigos ("*code-switching*"), diglosia, formación de *pidgins*, *sabires* o criollos ("*creoles*") como *lingua franca*, etc.

Se denomina diglosia a la situación bilingüe en la que una de las dos lenguas presenta un estatuto socio-político inferior. Si una es la empleada por el gobierno, los medios de comunicación, el sistema educativo, los actos religiosos..., la otra es la variante no prestigiosa utilizada con la familia, los amigos, en la calle, etc. Cada una de las lenguas se utiliza con propósitos distintos en contextos comunicativos diferentes.

La alternancia de códigos se refiere a un cambio de una lengua a otra, producido por un hablante, en una conversación. Un hablante puede hacer una pregunta en una lengua y su interlocutor puede contestar en otra lengua distinta, o un hablante puede iniciar un discurso en una lengua y finalizarlo en otra, cambiando de lengua incluso en mitad de una oración (esto suele suceder cuando los interlocutores por lo menos comprenden esas lenguas, aunque no sean capaces de hablarlas).

Lingua franca o lengua auxiliar es cualquier lengua que se utiliza para la comunicación entre diferentes grupos de personas que hablan distintos

idiomas; puede ser la lengua nativa de uno de los grupos (p. e. inglés como lengua internacional) o una lengua simplificada formada a partir de otras, como es el caso de los *pidgins*. El término *lingua franca* procede de la lengua heterogénea, basada en el italiano central y con distintos elementos de otras lenguas románicas, que se hablaba en los puertos mediterráneos desde la Edad Media hasta el siglo XIX.

El *pidgin* es una lengua secundaria, nacida del contacto entre varias lenguas, generalmente una europea y otra de una antigua colonia, con el fin de permitir la comunicación entre distintas comunidades intelingüísticas. Se compone de reglas gramaticales simples y vocabulario reducido, mezcla de las dos lenguas. El léxico y la gramática pueden expandirse hasta tal punto que a veces estos *pidgins* se convierten en lenguas oficiales, p. e. el Tok Pisin de Nueva Guinea y el Krio de Sierra Leona.

Cuando esta mezcla de lenguas no se ha desarrollado como sistema completo y su gramática está mal caracterizada y su léxico es pobre, se trata de un *sabir*. Se diferencian, por un lado, de los *pidgins*, en que éstos son sistemas completos utilizados como segundas lenguas, mientras que los *sabires* no constituyen sistemas completos y, por otro lado, de los *criollos*, en que, éstos nacen como *sabires* o *pidgins* y se convierten en las lenguas maternas de comunidades culturales; los *sabires* nunca llegan a convertirse en lenguas maternas. El *chinook*, por ejemplo, en Estados Unidos es un caso de *sabir* basado en el chino y con elementos ingleses.

Los *criollos*, constituyen el estadio posterior de esa mezcla de lenguas que ha evolucionado hasta desarrollar reglas gramaticales y de formación de nuevo vocabulario propias. Los *criollos* son ya adquiridos por los niños como lenguas maternas, por lo que se convierten en primeras lenguas, a diferencia de los *sabires* y *pidgins*, utilizados como segundas lenguas. Hay *criollos* franceses en Haití, en la Martinica, en Guadalupe, *criollos* ingleses en Jamaica, Hawái y en los EE.UU. ("*gullah*"), *criollos* portugueses y neerlandeses. El origen mixto de los *criollos* los diferencia de los dialectos de una lengua y su estatuto sociocultural los opone a la lengua misma.

B) La segunda vía de contacto es la **cultural**, es decir, a través de los medios de comunicación, traducciones, enseñanza de lenguas extranjeras, contacto personal de hablantes individuales, en vez de comunidades enteras, que aprenden otro idioma o se desplazan a zonas geográficas propias de otras comunidades lingüísticas: *The coexistence of languages is not merely a problem of geography; languages do not come into contact only across their geographic borders. Nor is it only a problem of social bilingualism or diglossia. In our present world this coexistence is more and more expressed by means of the learning and teaching of second languages and the increasingly common activities of translation and interpretation connected with all kinds of communication media and cultural transmission. Although acquiring a second language or translating can initially be considered as individual activities, the massive scale on which they are nowadays carried out invests them with a clear social importance* (Hernández-Sacristán, 1994: 20).

En un sentido amplio todos estos casos son de lenguas en contacto. El contacto puede producirse mediante estas dos vías; también se mantiene entre comunidades enteras o en la mente de sujetos individuales bilingües... sin embargo existe una diferencia radical en dos enfoques distintos de estudio.

Por un lado, se puede centrar la atención en las influencias de una lengua sobre otra, los calcos, los préstamos, los casos de bilingüismo, alternancia de códigos, diglosia e incluso los casos extremos de surgimiento de nuevas lenguas, híbridas, producto de la fusión de otras ya existentes. La palabra contacto alude de forma especial a todas estas consecuencias de la relación directa de lenguas.

Por otro lado, la palabra contraste se refiere específicamente a la comparación de dos lenguas como sistemas diferenciados, delimitados y claramente independientes. Y es éste el enfoque específico de los estudios contrastivos.

Aunque ambos enfoques son diferentes y a nosotros nos interesa el segundo a propósito de este estudio, se debe tener en cuenta que existen grandes semejanzas entre las estructuras de los lenguajes *pidgin* y criollos por un lado, y las interlenguas de los estudiantes de lenguas extranjeras por otro. Al fin y al cabo, los estudios contrastivos se hallan intrínsecamente relacionados con el análisis y prevención de errores, los cuales conciernen de forma directa al intelenguaje. La investigación se realiza sobre dos sistemas lingüísticos diferentes; sin embargo, las aplicaciones atañen el aprendizaje de uno a partir de otro y a través de un sistema aproximativo: *An approximative system is the deviant linguistic system actually employed by the learner attempting to utilize the target language* (Nemser, 1988: 2).

Este contacto de las lenguas origen y meta en el aprendizaje de lenguas extranjeras se produce en la mente del estudiante de forma individual y se trata de un *sistema desviado*, se analiza, corrige y acerca paulatinamente a la lengua meta, mientras que el contacto entendido en el sentido de las influencias y mezclas de unas lenguas con otras abarca comunidades lingüísticas y en ocasiones modifica los sistemas hasta tal punto que se produce una evolución diferente y paralela a la del *sistema madre* o incluso se origina un sistema lingüístico nuevo.

En cuanto al término *comparación*, se podría pensar en hablar de lingüística comparada o comparativa, para referirnos a la lingüística contrastiva, pero hay que diferenciar de nuevo entre dos perspectivas a las que alude este adjetivo de *comparado*.

La *lingüística comparada* se ha utilizado con frecuencia como sinónimo de *lingüística histórica*, el enfoque lingüístico que prevalecía en el siglo XIX. Este tipo de aproximación a las lenguas se centra en el estudio de los cambios intralingüísticos a través del tiempo y de las relaciones de parentesco entre lenguas diferentes. Comparando formas anteriores y posteriores de una lengua y de varias, se ha podido demostrar cómo algunas lenguas están emparentadas (p.e. todas las lenguas procedentes del indoeuropeo).

También se han podido reconstruir formas antiguas de las que no quedaba prueba escrita alguna.

Este punto de vista es el que se decanta por el estudio de la evolución de lenguas y su clasificación en familias y en tipologías sintácticas, según el orden de palabras (p. e. lenguas SVO (Subjctct-Verb-Objctc), como el inglés, SOV como el turco, o VSO como el galés) y morfológicas, según la formación de las palabras (p. e. lenguas analíticas, como el chino, aglutinantes, como el turco, sintéticas como el latín, polisintéticas, como el esquimal, e infijas como el árabe o el hebreo).

Distinguimos por tanto entre comparación genética, histórica y diacrónica por un lado, y comparación contrastiva, descriptiva y sincrónica por el otro. Aunque de nuevo nos decantamos por este segundo tipo de comparación en un estudio contrastivo, merece la pena tener presente que los resultados de la comparación de dos lenguas dependerán en gran medida de cuál sea ese par de lenguas elegido, de cuál sea su relación de parentesco y su pasado histórico.

En cuanto al concepto de *contraste*, observemos cuáles son las nomenclaturas compuestas por este término. Tenemos *estudios contrastivos*, *lingüística contrastiva* o *de contrastes* o *análisis contrastivo*. Todos ellos son términos utilizados actualmente, pero quizá el de estudios contrastivos abarca un campo más amplio, menos restringido por la lingüística y que implica una mayor independencia, si cabe, dentro de ser una rama interdisciplinar. La definimos como rama interdisciplinar por sus múltiples afinidades con otras disciplinas. Además, este último es un término paralelo a *estudios de traducción*, una ciencia paralela.

Por el hecho de estudiar elementos pertenecientes a la lengua, los estudios contrastivos se hallan vinculados inevitablemente a la lingüística, y por trabajar con parejas de lenguas y relaciones interlingüísticas, están emparentados indiscutiblemente con la traducción y con los estudios de lenguas en contacto.

De igual modo, aunque en menor grado, los estudios contrastivos comparten ciertos aspectos con: la semiótica, al tratar sobre sistemas de comunicación; la sociología, en el estudio del uso real de ambas lenguas, las características de los hablantes y las circunstancias en que se produce la comunicación; la psicología, especialmente en cuanto se refiere a los mecanismos de aprendizaje de la segunda lengua, o al proceso de trasvase de la información; la filosofía; la antropología; los estudios culturales, etc.

Después de este breve repaso nominativo y conceptual, en el que hemos aludido a las denominaciones y las nociones, podemos deducir que los estudios contrastivos son también estudios lingüísticos comparados y estudios de lenguas en contacto, pero tanto la lingüística comparada como las lenguas en contacto comprenden campos de acción específicos al margen de la materia que abarcan en particular los estudios contrastivos.

Parece difícil delimitar el ámbito de los estudios contrastivos, tanto más cuanto que lazos muy estrechos le unen a otras disciplinas afines dentro del

campo de las ciencias humanas, como acabamos de ver. En especial, un motivo de controversia es la relación de interdependencia de los estudios contrastivos, la traducción y la lingüística, relación que intentamos resumir a continuación.

Observemos dónde se integran los estudios contrastivos con respecto a la traducción dentro de una clasificación lingüística. Las opiniones son diversas:

- Halliday incluye los estudios contrastivos dentro del ámbito de la lingüística, y habla también de los estudios de traducción, que integra en los estudios contrastivos: *The theory and method for comparing the working of different languages in known either as 'comparative descriptive linguistics' or as 'contrastive linguistics'. Since translation can be regarded as a special case of this kind of comparison, comparative descriptive linguistics includes the theory of translation* (Halliday et al., 1965: 112 citado en Kühlwein et al., 1981: 7).
- Otra opinión relevante es la de James, el cual engloba la traducción, el análisis contrastivo y el de errores en una rama de la lingüística llamada *Interlanguage Study: There are thus three branches of two-valued (2 languages are involved) interlingual linguistics: translation theory -which is concerned with the processes of text conversion; error analysis and contrastive analysis- these last two having as object of enquiry the means whereby a monolingual learns to be bilingual* (James, 1980: 4) y más adelante afirma que los estudios interlingüales son *...a hybrid discipline, constituted not only of linguistics but also of psychology and sociology* (James, 1980: 6).
- La visión de Rabadán, también desde la perspectiva de la teoría de la traducción, es parecida; habla de tres disciplinas diferenciadas, aunque con terrenos fronterizos: *Del mismo modo que la lingüística tiene puntos de contacto con la Translémica, con terrenos fronterizos a veces compartidos, a veces en mutuo usufructo, pero no comparte ni fines ni método, el análisis contrastivo puede aportar datos de interés para la traducción. Se trata de tres campos cuyas fronteras son muy difusas, especialmente en el caso de los dos últimos, pero que responden a necesidades distintas, persiguen metas diferentes, y por tanto, hacen uso de marcos metodológicos propios de cada uno* (Rabadán, 1991: 43).

Hemos visto que la traducción y los estudios contrastivos se hallan estrechamente vinculados, pero, ¿de qué forma?, ¿cómo colaboran mutuamente? Para contestar a esta pregunta vamos a formular otra: ¿Cuál es el porqué o el para qué de los estudios contrastivos?

Se ha hablado extensamente en su utilidad como herramienta pedagógica, para planificar la enseñanza de segundas lenguas, para predecir las áreas problemáticas y por consiguiente prevenir los errores. James afirma que: *If one could juxtapose the structures of the mother tongue against those of the target language, course designers (and teachers and learners) would be better able to plan their learning and teaching; better able to*

foresee difficulty and consequently better able to husband resources and direct learning and teaching effort (James, 1980: iii).

Un segundo fin de los estudios contrastivos es el de conducir a reflexiones sobre el lenguaje en general, partiendo de la descripción de lenguas particulares y siguiendo un método inductivo: *Pourquoi la linguistique contrastive? Tout d'abord parce que la comparaison de deux langues met à jour des phénomènes qui concernent le langage en général* (Guillemin-Flescher, 1992:1).

Un tercer objetivo muy importante de los estudios contrastivos es su aplicación a la traducción. Al igual que sucede con la didáctica de las lenguas extranjeras, se fomenta la prevención y corrección de errores y el incremento de dominio en ambas lenguas y destreza en la transferencia, si bien se distinguen en la dirección hacia la que se hallan orientadas.

El rasgo de direccionalidad distingue a los estudios de traducción (dirección desde la lengua extranjera, o lengua *origen*, a la lengua nativa, o lengua *meta*) del aprendizaje de una segunda lengua (dirección desde la lengua nativa a la lengua extranjera). Los estudios contrastivos se limitan a describir y constatar, de entrada sin priorizar ninguna dirección, para que ambas aplicaciones sean factibles.

Intentan mantenerse lo más asépticos o neutrales posibles, sin rasgo de direccionalidad y centrándose en la lengua como producto, para que a partir de esas dos lenguas producto analizadas, la enseñanza de lenguas pueda observar el proceso de interlengua y la traducción pueda obtener datos del habla real y así marcar directrices sobre el proceso de traducción. Dicho de otro modo, los estudios contrastivos ofrecen una fuente importante de material didáctico, no sólo enfocado hacia el aprendizaje de lenguas extranjeras y la predicción de errores, sino también de cara al aprendizaje de la traducción.

Al parecer, los estudios contrastivos conforman un área intesdisciplinar con haces de relaciones que la unen a muchas otras ramas del saber y que se caracteriza por el hecho de trabajar a doble banda; con dos lenguas, dos culturas, y codo a codo con la traducción en una rama bipartita de la lingüística. Se trata, pues, de una disciplina interlingüística, intercultural e interdisciplinar.

Así pues, como tal tipo de disciplina necesita de datos empíricos bilingües como material para la labor de comparación y contraste, nada mejor que los corpus bilingües para proporcionárselos. Intentaremos justificar finalmente la utilización de los corpus bilingües y su relevancia en el ámbito de los estudios contrastivos.

Uno de los precursores de la lingüística de corpus fue Bloomfield, que utilizó corpus recopilados manualmente para la reflexión y el análisis lingüístico. Bloomfield y el resto de los descriptivistas americanos defendían la objetividad y la necesidad de contrastar las suposiciones con datos empíricos para poder llegar a extraer conclusiones fiables sobre las lenguas.

Todos ellos basaron su investigación en el análisis de un corpus de enunciados emitidos de forma natural por los hablantes nativos de una lengua. Pero se acepta sólo lo que los hablantes nativos dicen en su lengua. Pero se acepta sólo lo que los hablantes nativos dicen en su lengua, no lo que dicen sobre ella: *Accept everything a native speaker says in his language and nothing he says about it... The folklorist may be interested in Englishmen's beliefs about English; the linguist must concentrate rather on how Englishmen speak when they are not thinking about their language.* (Sampson, 1980: 64, 66).

Se trata de descubrir y describir regularidades y patrones formales en el corpus y, en el caso de un estudio interlingüístico o contrastivo, de descubrir y describir las semejanzas y diferencias entre ambas lenguas. Como afirma Baker, la forma de comparar el uso real de dos lenguas es utilizar material auténtico producido por los hablantes nativos, en cantidades significativas, es decir, grandes corpus textuales (Baker, 1995: 223).

Podemos resumir las ventajas de utilizar los corpus bilingües como metodología en los estudios contrastivos de la siguiente forma:

Observación frente a introspección: 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 emitir juicios sobre el lenguaje. La intuición no es, según nuestro punto de vista, un criterio irrefutable a la hora de juzgar la gramaticalidad de un hecho de habla. 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 tomadas de un contexto real comunicativo.

Descriptivismo frente al prescriptivismo: o lo que es lo mismo, norma frente a uso. Respecto a esta dicotomía, la lingüística de corpus se decanta claramente por el uso, el uso real que los hablantes nativos hacen de su lengua. Se prefiere describir objetivamente a dictar normas que regulen el habla; se observa cómo se expresan los mensajes en lugar de afirmar cómo se han de decir. Es el uso el que afecta 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.

Tipicidad, aceptabilidad y frecuencia. Los lingüistas, en principio, tienen dos opciones: trabajar con ejemplos posibles, generados por reglas gramaticales, pero que probablemente ni se han utilizado ni se utilizarán jamás en una situación comunicativa real, o tomar ejemplos de la realidad, aunque muchos no reflejen lo que es propio, incluso correcto, del idioma; una tercera alternativa consiste en seleccionar, de un corpus real, lo que es típico y común de una lengua, según un criterio de frecuencia.

El estudio del uso real de un idioma (que es observable, mientras que la competencia no lo es) es el que, una vez analizado, nos lleva a inducir pautas generales de comportamiento existentes en el ámbito de la competencia

lingüística y aquellos rasgos más frecuentes serán los más significativos del idioma, si bien algunos elementos poco comunes pueden resultar en ocasiones significativos. Gracias a los corpus tenemos a nuestra disposición no sólo material auténtico y aceptable (sea o no gramatical), sino también aquello que es típico de una lengua concreta según un criterio de frecuencia.

Capacidad cuantitativa y cualitativa del corpus. El lingüista, por muy experto que sea, tiene sólo un conocimiento parcial de la lengua que habla. Es difícil que se le ocurran todo tipo de ejemplos posibles o significados posibles de una palabra mediante mera introspección. Una gran cantidad de ejemplos no aparece cuando los buscamos conscientemente, y en cambio afloran, sin ningún esfuerzo, en situaciones comunicativas reales. En realidad, por muchos ejemplos que puedan ocurrírsele, nunca abarcará la extensión de un corpus extenso y tampoco puede intuir la frecuencia de palabras determinadas, más que muy vagamente. Así, un corpus nos ofrece tanto información cuantitativa -frecuencia en cifras o porcentajes de una palabra, como cualitativa- todas las acepciones y usos de una determinada palabra.

Material auténtico. Ejemplos apropiados: Otra razón para justificar el uso de los corpus es el hecho de que los lingüistas tienden a buscar ejemplos que justifiquen sus hipótesis, lo cual es aparentemente fácil pero no siempre representa la realidad pues pueden ser excepciones o ejemplos que no son relevantes ante el conjunto de otros ejemplos que podrían incluso justificar lo contrario. Muchos de estos ejemplos resultan forzados, artificiales o poco naturales (del tipo de: "Juan es alto" o "María come peras").

Un corpus proporciona material auténtico, ejemplos genuinos que constituyen la verdadera lengua viva que es objeto de estudio. Por otro lado, aquellos que defienden la invención de ejemplos mediante intuición también pueden afirmar que algunos ejemplos son correctos si se emiten situaciones o contextos muy específicos, pero, de hecho, no se producen nunca o casi nunca. Un corpus, por el contrario, muestra de forma objetiva aquella que de verdad sucede en el habla real y si es lo bastante representativo nos muestra no sólo lo que es posible, sino lo que es común, lo que es frecuente en el habla espontánea, carente de condicionamientos metalingüísticos.

Datos estadísticos. Respecto a la frecuencia, los lingüistas, sin más ayuda que su propia intuición, pueden sugerir si una palabra es mucho más común que otra sólo en casos en los que salta a la vista, y un corpus sin embargo puede ofrecer cifras exactas sobre la frecuencia de uso de todas las palabras que lo componen. Gracias a los avances tecnológicos puede obtenerse una gran cantidad de información en un mínimo de tiempo, lo cual facilita y agiliza enormemente la labor del lingüista.

Dadas todas estas ventajas consideramos justificado el hecho de basar las investigaciones que versen sobre estudios contrastivos de lenguas en corpus de datos, en hechos y cifras reales que demuestren la veracidad de las hipótesis sostenidas, de modo que tras la descripción y el contraste de los usos y estructuras de ambas lenguas objeto de comparación pueda darse un mayor entendimiento de las dos lenguas y éste a su vez posibilite tanto

un análisis de errores y una mayor destreza de la segunda lengua por parte de los estudiantes de lenguas extranjeras como una mayor comprensión de la lengua origen y una mejor expresión de la lengua materna por parte de los traductores.

El propósito de este artículo ha sido doble: en primer lugar, situar los estudios contrastivos dentro del ámbito general del saber y en relación con disciplinas afines, así como definir someramente una serie de variables y conceptos de suma importancia en lo que se refiere a cualquier tipo de investigación interlingüística. En segundo lugar se ha intentado defender una metodología concreta a seguir en el campo de los estudios contrastivos.

Por último y a modo de conclusión, simplemente resaltar como método de análisis de los estudios contrastivos, la lingüística de corpus y ampliar el conjunto interrelacionado de las "ces" de las que hablábamos al principio, de modo que al contraste, comparación y contacto lingüístico se unan las utilidades de los corpus bilingües.

BIBLIOGRAFÍA

- AIJMER, K & ALTENBERG, B. 1991. *English Corpus Linguistics*. London: Longman.
- BAKER, M. 1995. "Corpora in Translation Studies: An Overview and Some Suggestions for Future Research" en *Target* 7: 2, 223-243. Amsterdam: John Benjamins.
- GUILLEMIN-FLESCHER, J. 1992. *Linguistique Contrastive et Traduction*. Paris: Orphys.
- HERNÁNDEZ, C. 1994. *Aspects of Linguistic Contrast and Translation*. Germany: Peter Lang.
- JAMES, C. 1980. *Contrastive Analysis*. Singapore: Longman.
- KÜHLWEIN, W & THOME, G. et al. 1981. *Kontrastive Linguistik und Übersetzungswissenschaft*. München: Wilhelm Fink Verlag.
- LEECH, G. 1991. "The State of the Art in Corpus Linguistics" en Aijmer, K. pp. 8-30.
- NEHLS, D. 1988. *Interlanguage Studies*. Heidelberg: Julius Groos Verlag.
- NEMSER, W. 1988. "Approximative Systems of Foreign Language Learners" en Nehls, D. pp. 1-10.
- RABADÁN, R. 1991. *Equivalencia y traducción*. León: Universidad de León
- ROBERTS, R. 1997. Ponencia del VI Curso Superior de Traducción Inglés/Español. Itbyte. Universidad de Valladolid.
- SÁNCHEZ, A. 1995. *Corpus lingüístico del español contemporáneo*. Madrid: SGEL.
- SAMPSON, G. 1980. *Schools of Linguistics*. London: Hutchinson.
- SCHUMANN, J. H. 1978. *The Pidginization Process: A Model for Second Language Acquisition*. Rowley MA: Newbury House.
- SINCLAIR, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.