

Propuesta de clasificación de corpus textuales

[Proposal of textual corpora classification]

GORETTI FAYA ORNIA

Universidad de Oviedo

gorettifayaornia@gmail.com

Resumen

El empleo de corpus en diversas disciplinas ha ocasionado una falta de consenso en una doble vía. Por un lado, en la terminología empleada, y por otro, en la clasificación de corpus.

Es conveniente contar con una clasificación de corpus clara, organizada e integradora, ya que permite observar las características de cada tipo así como su relación con los demás. En consecuencia, facilita la selección del tipo de corpus más adecuado para un estudio concreto, muestra otras alternativas (otros tipos de corpus), y permite ubicar el trabajo dentro de un marco general de corpus lingüísticos.

En esta comunicación, pretendemos dar a conocer nuestra propia propuesta de clasificación de corpus. Hemos tratado de que sea lo suficientemente amplia y abarque los intereses de las disciplinas que trabajan con más frecuencia con corpus lingüísticos (la Lingüística Contrastiva y los Estudios de Traducción).

Abstract

Using corpora for different subjects has resulted in a lack of agreement in two ways. First, with respect to the terminology used, and secondly, regarding the way corpora are classified.

A clear, organised and integrated classification is advisable, since it shows the features of each type of corpus as well as its relationships with other corpora. Consequently, such a classification (a) makes it easier to select the most suitable type of corpus for a specific study, (b) reveals other alternatives (other types of corpora) and (c) allows for a study in a general frame of linguistic corpora.

With this paper, we aim to show our own proposal for classifying corpora. We have attempted to make our classification wide enough so as to cover the interests of those disciplines that most often work with linguistic corpora (i.e. Contrastive Linguistics and Translation Studies).

Palabras clave

Clasificación de corpus, propuesta de clasificación, corpus en la Lingüística Contrastiva y Estudios de Traducción, falta de consenso en terminología y clasificación

Keywords

Corpora classification, classification proposal, corpora in Contrastive Linguistics and Translation Studies, lack of agreement about terminology and classification



Esta obra se publica con una licencia Creative Commons **Reconocimiento – NoComercial – SinObraDerivada (by-nc-nd)**: No se permite un uso comercial de la obra original ni la generación de obras derivadas.

1. Introducción

En este trabajo, daremos a conocer nuestra propia propuesta de clasificación de corpus. Su elaboración se debe a las necesidades surgidas en otro trabajo de investigación que estábamos llevando a cabo.

Dicho estudio consistía en un análisis contrastivo del género del folleto médico en inglés y en español. En él trabajábamos con dos corpus de folletos. Uno estaba formado por documentos traducidos a inglés en España y otro por documentos originales redactados en lengua original (inglés y español).

Nos resultaba preciso ubicar en una clasificación general estos dos corpus, para así observar las relaciones entre ellos. Sin embargo, las clasificaciones existentes no cubrían todas nuestras necesidades.

2. Terminología empleada en el presente trabajo

Veamos a continuación el uso de la terminología en el presente trabajo. Ordenamos los términos alfabéticamente.

Corpus alineado

Es un corpus en el que coexisten originales con sus correspondientes traducciones y que ha sido sometido a un proceso de alineamiento (Rabadán y Fernández Nistal 2002: 51). El proceso de alineación se puede llevar a cabo tanto a nivel de frase como de palabra (P. Baker 2006: 9).

Corpus canónico

Término utilizado principalmente por Torruella y Llisterri (1999: 55) y Corpus Pastor (2001: 158) en sus tipologías para referirse a un corpus que contiene las obras completas de un determinado autor.

Corpus comparable

Es uno de los términos en donde encontramos mayores discrepancias e incoherencias terminológicas entre la Lingüística Contrastiva y los Estudios de Traducción. Para la primera (Teubert 1996, Peters, Picchi y Biagini 1996, McEnery y Wilson 1996, Johansson y Hasselgard 1999), se trata de textos originales en dos o más idiomas, según la época de composición, la categoría de texto, el público al que se dirigen, etc. Son por tanto multilingües, y como consecuencia, facilitan las comparaciones entre lenguas así como las investigaciones cruzadas. En los Estudios de Traducción, el término suele hacer referencia a dos colecciones de textos en la misma lengua. Por un lado, un conjunto de textos originales en una sola lengua. Por otro, un grupo de textos traducidos a esa misma lengua a partir de otras lenguas diferentes (sin ser traducciones de los primeros) (Granger 2003: 20).

En nuestra opinión, los corpus comparables pueden ser tanto monolingües como multilingües. Se distinguen tres categorías. En primer lugar, textos originales en una o varias lenguas. En segundo lugar, textos traducidos en una o en varias lenguas. En tercer lugar, una combinación de textos originales y traducidos en una o varias lenguas (sin ser estas traducciones de los primeros). Como vemos, la condición esencial de este grupo es que no pueden estar integrados por originales con sus correspondientes traducciones.

Corpus cronológico

Se trata de un corpus que incluye textos producidos durante un determinado periodo de tiempo (Torruella y Llisterra 1999: 55; Corpus Pastor 2001: 158).

Corpus de muestras

Se compone de partes de un texto seleccionadas según determinados principios concierne al tamaño, ubicación de la muestra en texto completo y método de selección (Laviosa 1997: 34).

Corpus de referencia

Este tipo de corpus trata de representar la naturaleza general de una lengua. Por ello, consta de una amplia variedad de textos que incluyen diversos géneros y fuentes (P. Baker 2006: 137-138). Sinclair (1996: 10) define del siguiente modo este tipo de corpus en el informe EAGLES:

A reference corpus is one designed to provide comprehensive information about a language. It aims to be large enough to represent all the relevant varieties of the language, and the characteristic vocabulary, so that it can be used as a basis for reliable grammars, dictionaries, thesauri and other language reference material. (Sinclair 1996: 10)

Corpus de textos completos

Un corpus de textos completos contiene textos íntegros (Laviosa 1997: 291).

Corpus diacrónico

Corpus que trata de ser representativo de una lengua o de una variedad en un determinado periodo de tiempo, de modo que permita estudiar la evolución lingüística (P. Baker 2006: 56). Son por tanto aquellos cuyos textos no están restringidos a un periodo de tiempo (Laviosa 1997: 35).

Los términos «corpus de referencia» y «corpus diacrónico» suelen tener el mismo referente, ya que para que un corpus sea representativo de la lengua en su conjunto (corpus de referencia) no debe ceñirse a un periodo concreto (rasgo típico de los corpus diacrónicos).



Corpus electrónico

Se trata de un tipo de corpus que se encuentra en soporte electrónico para poder trabajar con él de manera más rápida y eficaz (Rabadán y Fernández Nistal 2002: 46).

Corpus escrito

El término hace referencia a un corpus que contiene únicamente textos producidos o publicados en formato escrito (P. Baker 2006: 171). Puede encontrarse en papel o en soporte electrónico (ya sea escaneado o mecanografiado).

Corpus especializado

Aquellos que están compuestos por textos que no pertenecen a la lengua general, sino que se centran en un ámbito determinado, ya sea una temática concreta, un campo del saber en particular, un determinado autor, cierta variante dialectal, un periodo de tiempo concreto, etc. (Torruella y Llisterri 1999: 55, Corpas Pastor 2002: 158, y Rabadán y Fernández Nistal 2002: 50).

Corpus generalista

Son representativos de la lengua común e incluyen una gran variedad de textos producidos en situaciones comunicativas cotidianas (Torruella y Llisterri 1999: 55, Corpas Pastor 2001: 158).

Corpus genérico

Corpus representativo de un determinado género (Torruella y Llisterri 1999: 55, Corpas Pastor 2001: 158).

Corpus informatizado

Se trata de «una recopilación de textos seleccionados según criterios lingüísticos, codificados de modo estándar y homogéneo, con la finalidad de poder ser tratados mediante procesos informáticos y destinados a reflejar el comportamiento de una o más lenguas» (Torruella y Llisterri 1999: 51).

Corpus mixto

Son aquellos corpus compuestos tanto por textos completos como por muestras (Laviosa 1997: 291).

Corpus monolingüe

Estos corpus están integrados por textos escritos en una sola lengua, y por ello, son siempre comparables. El motivo de que los corpus monolingües no puedan ser paralelos se debe a que, como veremos más adelante, los corpus paralelos constan de originales con sus correspondientes traducciones, es decir, están redactados en al menos dos lenguas (Granger 2003: 20).

Corpus monitorizado

Se trata de un corpus con un volumen textual constante (como sucede en los corpus estáticos), pero actualizado: se van añadiendo nuevos elementos a la vez que se eliminan cantidades equivalentes de material antiguo (Torruella y Llisterrí 1999: 54, *Corpas Pastor* 2001: 158). Dichos textos «are scanned on continuing basis, ‘filtered’ to extract data for database, but not permanently archived» (Atkins et al 1992: 5).

Con ellos podemos «monitorizar» la evolución de una lengua. Este tipo de corpus solo puede encontrarse en los corpus diacrónicos. El motivo es que estos últimos no están ligados a un periodo restringido y son, por tanto, los únicos que permitirán la inclusión de nuevos textos.

Corpus multilingüe

Los corpus multilingües son simplemente aquellos que contienen textos en más de un idioma (P. Baker 2006: 119).

Corpus oral

Según Baker et al (2006: 148), estos corpus constan de discursos orales transcritos. Su elaboración resulta compleja debido a las repeticiones, pseudocomienzos de frases, dubitaciones, vocalizaciones e interrupciones de los discursos espontáneos. Se confunde frecuentemente con los corpus de habla (speech corpus). Estos son corpus integrados por grabaciones realizadas en un estudio, cuya finalidad principal es estudiar diferentes rasgos fonéticos.

Sin embargo, discrepamos con Baker en la definición de este tipo de corpus. En nuestro trabajo, emplearemos el término «corpus oral» para referirnos a discursos orales grabados (y no a discursos orales transcritos). Nuestro uso del término se asemeja por tanto a lo que el autor denomina «corpus de habla» (speech corpus). En nuestra opinión, la transcripción de los discursos orales (lo que él denomina «speech corpus»), se incluye dentro de los corpus escritos.

Corpus paralelo

Este término hace referencia a un conjunto de textos originales con sus correspondientes traducciones (Brown y Cocke 1988, Gale y Church 1991, Marinai et al 1991, Baker 1993 y 1995, Malmkjaer 1998, Kenny 2001, Rabadán y Fernández Nistal 2002, Altenberg y Granger 2002, Olohan 2004, Borja Albi 2007, entre otros). Los textos integrantes podrán estar alineados.

Corpus publicado

Consta de textos que se han publicado y están disponibles para su venta.

Corpus sincrónico



Contiene textos producidos en un periodo limitado de tiempo (Laviosa 2007: 35).

3. Propuestas de clasificación existentes

Se han publicado numerosos trabajos sobre corpus que han tratado directa o indirectamente cuestiones de etiquetado y clasificación de corpus. Debido a su relevancia para trabajos posteriores, destacamos especialmente las obras de Laviosa (1997), Johansson (1998), Torruella y Llisterri (1999), Corpas Pastor 2001, Granger (2003), Olohan (2004), Paul Baker (2006), Borja Albi (2007), Cabré, M. T. (2007), Briz y Albelda (2009), Zanettin (2012), entre otros. Las principales diferencias que se detectan entre las propuestas de clasificación atañen a los criterios seleccionados (Pearson 1998).

A pesar de su importante contribución al mundo académico, hemos podido observar que estas propuestas presentan ciertas limitaciones, entre las que destacamos las siguientes. En primer lugar, algunas clasificaciones propuestas no abarcan todos los tipos de texto, sino que se centran en un único tipo (por ejemplo, textos traducidos, textos monolingües, discursos orales, etc.). En segundo lugar, algunas de las clasificaciones resultan demasiado amplias y sería conveniente completarlas con otros subgrupos o atendiendo a otros criterios. Por último, no todas se representan en forma de árbol, por lo que la relación entre los distintos tipos de corpus no queda reflejada con claridad.

4. Nuestra propuesta de clasificación

En este apartado procedemos a presentar nuestra propuesta de clasificación. Consideramos que es suficientemente amplia, ya que abarca los intereses tanto de la Lingüística Contrastiva como de los Estudios de Traducción. Además, los tipos de corpus que se distinguen en la clasificación no son categorías independientes, sino que la representación en forma de árbol permite observar la relación entre ellos.

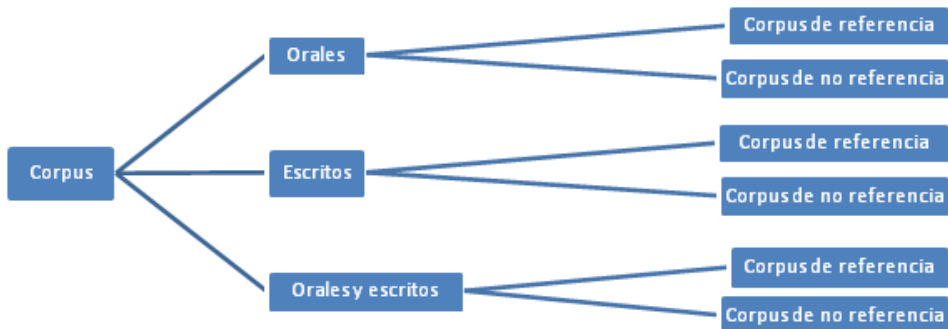
Nuestra clasificación consta de dos fases. La primera atañe a aspectos formales, concretamente al medio de difusión, al grado de amplitud del análisis, al momento temporal, a la posibilidad de adición, a la extensión de los textos, a su publicación y a la temática tratada. La segunda se centra en los siguientes aspectos lingüísticos: número de lenguas, composición de originales y traducciones y relación entre ellos, la dirección de la traducción y, por último, la alineación. A continuación, explicamos con más detalle cada uno de los pasos de nuestra clasificación.

La primera división efectuada corresponde al medio de difusión; esto es, si el corpus contiene textos escritos u orales. Asimismo, añadimos una tercera categoría para aquellos corpus en los que se combinan ambos.

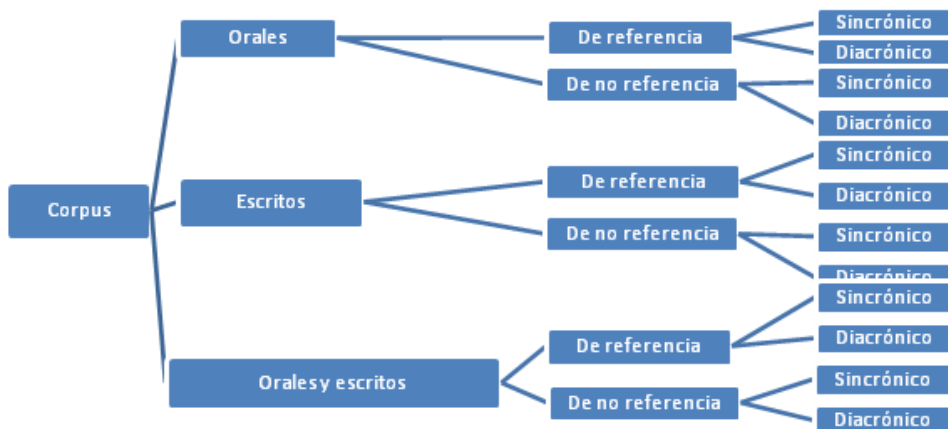


Los textos pertenecientes al ámbito escrito pueden estar en soporte electrónico o impreso. Dentro del ámbito oral existen variadas posibilidades de clasificación, debido principalmente al desarrollo de los materiales multimedia. Por citar algunos ejemplos, los textos que componen el corpus podrían clasificarse según (a) el soporte en el que se encuentran (CD, cinta de video o audio, reproductor MP3, medio electrónico, etc.), (b) el formato de grabación (.mp3, .mp4, .wma, etc.), (c) el momento de emisión (directo o diferido), (d) el canal de difusión (radio, televisión, rueda de prensa, Internet, videoconferencia, etc.) e incluso (e) si los textos que lo integran se han producido con preparación previa (conferencias, comunicados informativos, etc.) o si por el contrario son improvisados (debate, rueda de prensa, etc.). No detallaremos estos tipos en nuestra clasificación, pero si en un trabajo de investigación resultara necesario indicar esta información, creemos que este sería el momento adecuado.

En este punto podrían distinguirse dos tipos de corpus según el grado de amplitud o restricción del análisis: corpus de referencia y corpus de no referencia. Entendemos los primeros como aquellos que tratan de ser representativos de una lengua.

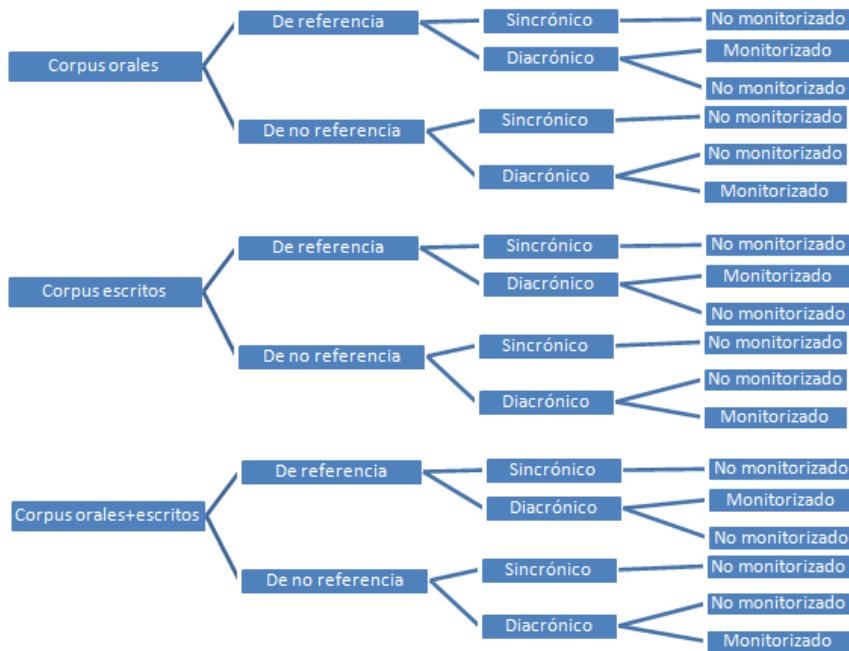


A su vez, atendiendo al momento temporal que se analiza, estos corpus pueden ser corpus sincrónicos (si los textos están restringidos a un momento concreto) o diacrónicos (si no están restringidos a un determinado momento temporal).

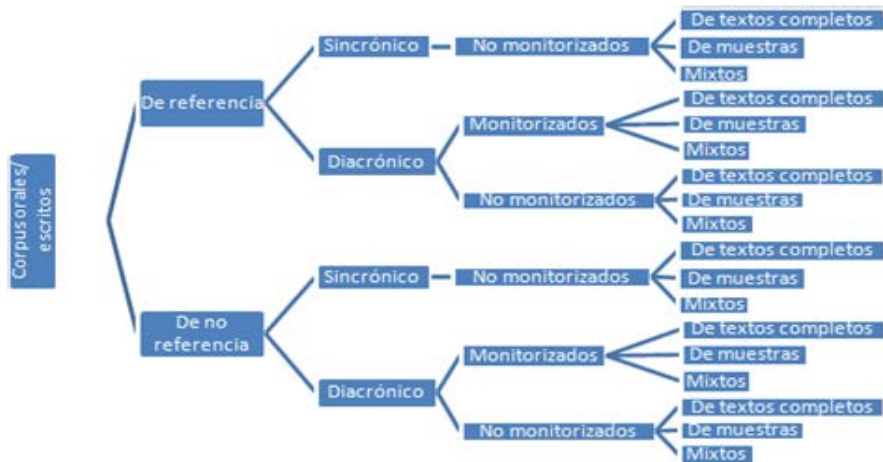


Si fuera preciso especificar si el corpus es informatizado (esto es tratado para que pueda ser utilizado por programas informáticos), consideramos que debería indicarse a continuación. Sin embargo, no lo reflejamos en nuestro esquema para no extenderlo demasiado de forma innecesaria, dado que entendemos que hoy en día la mayor parte de los corpus son informatizados.

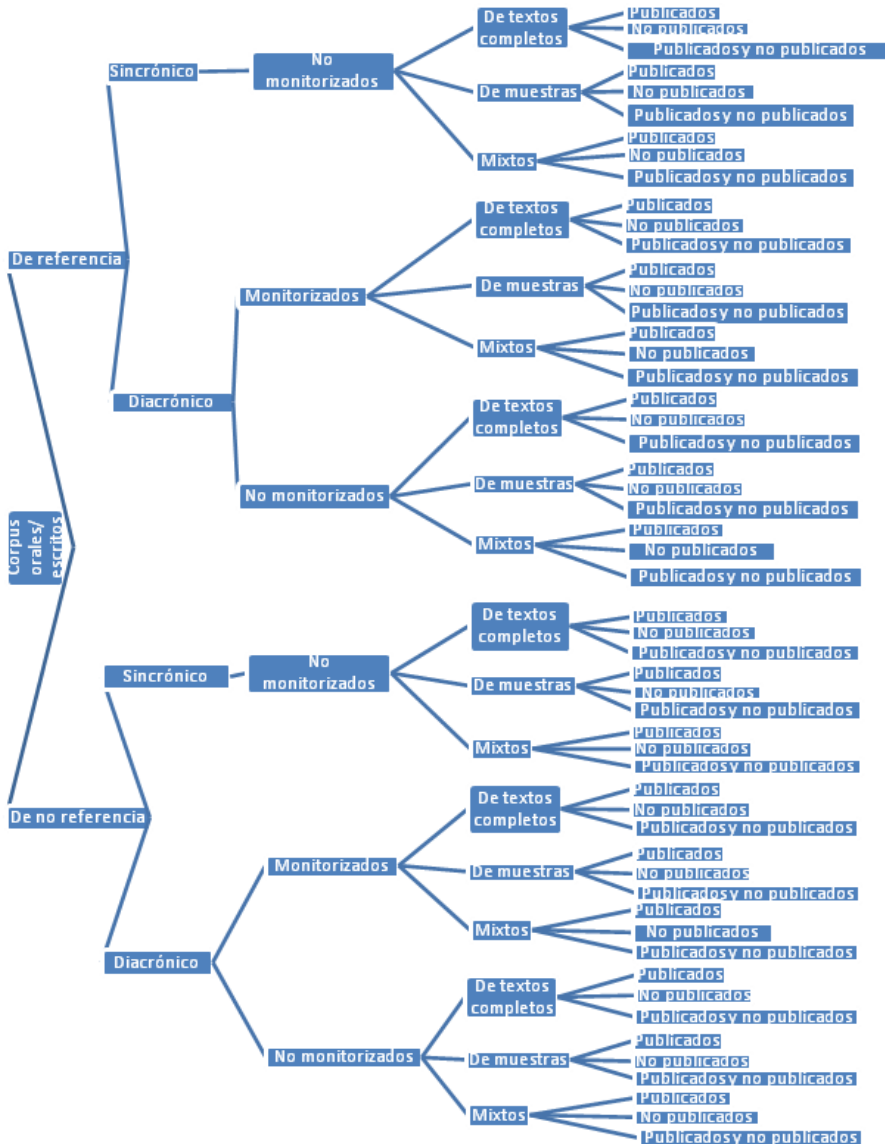
El siguiente nivel corresponde a las posibilidades de actualización y aumento de información de un corpus. Incluimos aquí el término «monitorizado», entendiéndolo como «aquellos corpus compuestos por textos escaneados y filtrados frecuentemente para que los datos estén actualizados». De este modo, los corpus diacrónicos pueden dividirse en «monitorizados» o «no monitorizados», ya que al no ceñirse a un periodo restringido, podrán admitir la incorporación constante de nuevos textos. Por el contrario, los corpus sincrónicos siempre serán «no monitorizados», ya que estudian una determinada época con principio y fin.



El siguiente paso de la clasificación atañe a la extensión de los documentos que componen el corpus. Siguiendo la clasificación de Laviosa (1997) en este punto, cabe distinguir entre (a) corpus de textos completos; (b) corpus compuesto por muestras y (c) corpus mixtos. Téngase en cuenta que en este esquema y en los siguientes, ya no desglosaremos «corpus orales», «corpus escritos» y «corpus orales y escritos», ya que contienen las mismas divisiones y ocuparía demasiado espacio indicarlo por triplicado. En su lugar, nos referiremos a ellos de forma simplificada como «corpus orales/escritos».



A continuación, atendemos a la publicación de los textos, es decir, si el corpus consta de textos publicados, no publicados o una combinación de ambos.



El siguiente estadio de la clasificación, el último de la primera fase, atañe al grado de especialización de los documentos, es decir, si pertenecen a la lengua general (corpus generalistas) o a la lengua especializada (corpus especializados). En base al criterio que se adopte, existen diversas posibilidades para sub-



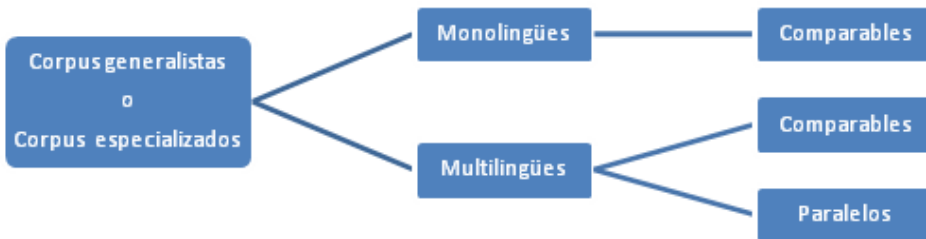
dividir los corpus especializados. Siguiendo la terminología de autores como Laviosa (1997) o P. Baker (2006), podemos citar como ejemplo los corpus genéricos (un determinado género), los canónicos (un determinado autor), los cronológicos (según periodo de tiempo, y siguiendo la terminología de Corpas Pastor), etc. Podrían distinguirse también otros corpus especializados, como los dialectales (atendiendo al dialecto), los históricos (en base al momento histórico al que hagan referencia), los nacionales/regionales (atendiendo al ámbito), los pedagógicos/formación (si tienen función didáctica), los de segunda generación (creados con posterioridad a 1990; también llamados «megacorpus» por su gran tamaño) o los corpus médicos, jurídicos, económicos, etc. (si se atiende a la temática). De este modo, la primera fase de nuestra clasificación quedaría como se indica a continuación. Téngase en cuenta que con el fin de simplificar el esquema no desglosamos de forma independiente los corpus de referencia y los de no referencia.



lenguas que intervienen en el corpus. Consecuentemente distinguiremos entre monolingües y multilingües.

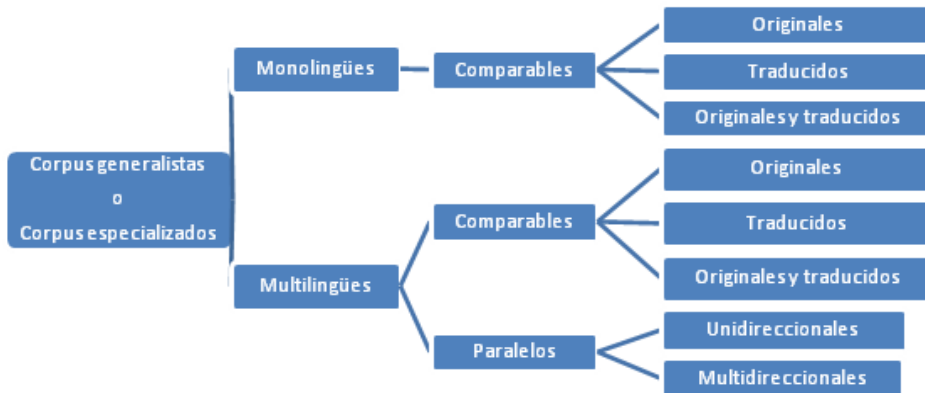


Si nos centramos en los corpus multilingües, podemos distinguir dos tipos: comparables y paralelos. La condición esencial de un corpus comparable es que no esté integrado por originales con sus correspondientes traducciones. Por el contrario, un corpus paralelo es un conjunto de textos originales con sus correspondientes traducciones.

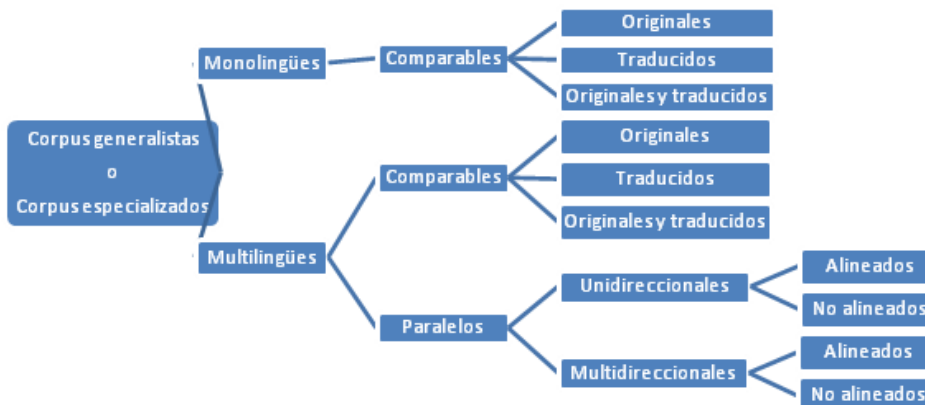


Por tanto, los corpus comparables pueden estar compuestos por diferentes conjuntos: (a) textos originales, ya sea en una misma lengua o en varias lenguas; (b) textos traducidos en varias lenguas; o bien (c) una combinación de textos originales y traducidos, sin ser estos traducción de los primeros.

En los corpus paralelos no es necesario especificar si los textos son originales o traducciones, puesto que el término «paralelo» ya implica que está compuesto por originales y sus correspondientes traducciones. Estos corpus se dividirán en «unidireccionales» o «multidireccionales» (en donde se incluyen los «bidireccionales»). Entendemos por «unidireccional» aquellos corpus que incluyen originales con sus correspondientes traducciones (esto es, paralelo), y por tanto multilingües, en los que la traducción se ha llevado en una única dirección (es decir, siempre con un mismo idioma origen). En este sentido, un corpus multidireccional es aquel que incluye originales con sus correspondientes traducciones (esto es, paralelo), y por tanto multilingües, en los que la traducción se ha llevado en más de una dirección.



Además, los corpus paralelos pueden estar alineados o no alineados. Dicha alineación puede ser a nivel de palabra, de frase, de párrafos, etc. Esta sería la última división de la segunda fase, y por tanto, el fin de nuestra propuesta de clasificación.



Todas las categorías que contengan textos traducidos pueden completarse con las alternativas que propone Laviosa (1997). En primer lugar, según esta autora, desde el punto de vista de la lengua origen, pueden ser «corpus con una única lengua origen» o «corpus con varias lenguas origen». La categoría de «corpus con dos lenguas origen» la hemos cambiado a «corpus con varias lenguas origen», por ser una categoría más amplia. En segundo lugar, según esta autora, atendiendo al modo de traducción (escrito u oral), pueden distinguirse «corpus con un solo modo de traducción» o «corpus con dos modos de traducción». En tercer lugar, según el método de traducción (es decir, si se trata de una traducción humana –ya sea asistida o no- o traducción automática), podrán ser «corpus con un solo método de traducción» o «corpus con varios métodos de traducción». La categoría «corpus con dos métodos de traducción» de Laviosa se

corresponde con lo que nosotros denominamos «corpus con varios métodos de traducción».

En este punto también podrían efectuarse otras divisiones de rango inferior. Por ejemplo, las que atañen a la dirección de la lengua del traductor (lengua materna, lengua extranjera o estado mixto de lenguas meta) así como a la categoría del traductor (estudiante o profesional). En nuestro esquema no reflejaremos todas estas categorías, ya que las consideramos secundarias y creemos que su presencia complicaría demasiado la clasificación. No obstante, si fuera necesario especificar alguno de estos aspectos en un determinado trabajo de investigación, en nuestra opinión, este sería el lugar apropiado.

5. Conclusiones

En nuestra opinión, nuestra propuesta de clasificación de corpus es amplia e integradora. No se centra en una única disciplina, sino que combina los intereses de varias, principalmente los de la Lingüística Contrastiva y los Estudios de Traducción (por ser dos de las disciplinas que emplean corpus con más frecuencia).

Además, nuestra clasificación es flexible porque permite la incorporación de subtipos de corpus, en caso de que en un proyecto concreto sea preciso atender a criterios más específicos que los que hemos expuesto.

Por último, consideramos que la representación en forma de árbol es sistemática, fácil de entender y permite observar con claridad el lugar que ocupa un determinado corpus en la clasificación general, así como su relación con otros tipos.

6. Bibliografía

- Altenberg Bengt y Sylviane Granger. 2001. The grammatical and lexical patterning of make in native and non-native student writing. @ *Applied Linguistics* 22/2: 173-194
- Atkins, Sue, Jeremy Clear y Nicholas Ostler. 1992. Corpus design criteria. Literary and Linguistic Computing. @ *Journal of the Association for Literary and Linguistic Computing* 7/1: 1-16.
- Baker, Mona. 1993. Corpus Linguistics and Translation Studies — Implications and Applications. @ M. Baker, G. Francis y E. Tognini-Bonelli, eds. *Text and Technology: In honour of John Sinclair*. ISBN 9789027285874, pp. 233-252.
- Baker, Paul. 2006. *Using Corpora in Discourse Analysis*. Londres y Nueva York: Continuum. ISBN 9780826477255.



- Borja Albi, Anabel. 2007. *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*. Clevedon: Series Translating Europe. Multilingual Matters. ISBN 1853599859, pp. 243-265.
- Briz Gómez, Antonio y Marta Albelda Marco. 2009. *Estado actual de los corpus de lengua española hablada y escrita. I+D*. @ Instituto Cervantes. *El español en el mundo*. Madrid: Instituto Cervantes. ISBN 9788434020368, pp. 165-225.
- Brown, Peter, John Cocke et al. 1988. A Statistical Approach to Language Translation. @ *Proceedings of the 12th International Conference on Computational Linguistics*: Budapest.
- Cabré, María Teresa. 2007. *Constituir un corpus de textos de especialidad: condiciones y posibilidades*. @ M. Ballard y C. Pineira-Tresmontant, ed. *Les corpus en linguistique et en traductologie*. Arras: Artois Presses Université. ISBN 9782848320632, pp. 89-106.
- Gale, William y Kenneth W. Church. 1991. Identifying Word Correspondences in Parallel Text. @ P. Price, ed. *Proc. of Fourth Darpa Workshop on Speech and Natural Language*. (del 19 al 22 de febrero). Asilomar (California): DARPA, pp 152-157.
- Corpas Pastor, Gloria. 2001. Compilación de un corpus ad hoc para la enseñanza de la traducción inversa especializada. @ *TRANS. Revista de Traductología* 5: 155-184.
- Granger, Sylviane, Jacques Lerot y Stephanie Petch-Tyson. 2007. *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Beijing: Foreign Language Teaching and Research Press. ISBN 9789042010468
- Johansson, Stig. 1998. On the role of corpora in cross-linguistic research. @ S. Johansson y S. Oksefjell, eds. *Corpora and cross-linguistic research: Theory, method, and case studies*. Ámsterdam: Rodopi B.V. ISBN 9042002913, pp. 3-24.
- Johansson Stig y Hilde Hasselgård. 1999. Corpora and cross-linguistic research in the Nordic countries. @ Granger et al, eds. *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Beijing: Foreign Language Teaching and Research Press. 9789042010468, pp. 145-162.
- Kenny, Dorothy. 2001. *Lexis and Creativity in Translation. A Corpus-based Approach*. Manchester: St. Jerome
- Laviosa, Sara. 1997. How Comparable Can 'Comparable Corpora' Be? @ *Target* 9/2: 289-319.
- Laviosa, Sara. 2007. Studying Anglicisms with Comparable and Parallel Corpora. @ W. Vandeweghe, S. Vandepitte y M. van de Velde, eds. *The Study of Language and Translation*. ISBN 9789027226815, pp: 123-136.
- Malmkjaer, Kirsten et al. 1998. *Translation & Language Teaching. Language Teaching & Translation*. Manchester: UK. St. Jerome Publishing.

- Marinai, Elisabetta, Carol Peters y Eugenio Picchi. 1991. Bilingual Reference Corpora: A System for Parallel Text Retrieval. @ *Using Corpora*, Proc. of 7th Annual Conference of the UW Centre for the New OED and Text Research. UK: OUP, pp 63-70.
- McEnery, Tony y Andrew Wilson. 1996. *Corpus Linguistics*. Edinburgh: Edinburgh University Press, ISBN 0748608087.
- Olohan, Maeve. 2004. *Introducing Corpora in Translation Studies*. Londres y Nueva York: Routledge. ISBN 9780415268844.
- Peters, Carol, Eugenio Picchi y Lisa Biagini. 1996. Parallel and Comparable Bilingual Corpora in Language Teaching and Learning. @ S. Botley et al., eds. *Proceedings of Teaching and Language Corpora*. Lancaster: Lancaster University. ISBN 186220 013, pp. 68-82.
- Rabadán Rosa y Purificación Fernández Nistal. 2002. *La traducción inglés-español: fundamentos, herramientas, aplicaciones*. León/Valladolid: Universidad de León/ITBYTE. ISBN 8497730100.
- Sinclair, John. 1996. *Preliminary Recommendations on Corpus Typology*. EAGLES. <http://www.ilc.cnr.it/EAGLES/corpusstyp/corpusstyp.html>
- Teubert, Wolfgang. 1996. Comparable or parallel corpora? @ *Corpus to corpus: a study of translation equivalence: International Journal of lexicography*. London: Oxford University Press.
- Torruella, Joan y Joaquim Llisterra. 1999. *Diseño de corpus textuales y orales*. @ J. M. Blecua et al., eds. *Filología e informática. Nuevas tecnologías en los estudios filológicos*. Barcelona: Editorial Milenio. ISBN 978-8489790414, pp. 45-77.
- Zanettin, Federico. 2012. *Translation-Driven Corpora. Corpus Resources for Descriptive and Applied Translation Studies*. Manchester: St Jerome. ISBN 978-1905763290.

