

LA LINGÜÍSTICA DE CORPUS (LC) Y SU INCIDENCIA EN LA ENSEÑANZA DE LENGUAS EXTRANJERAS

Chinger Zapata (pp. 53-75)



REVISTA

educare

Órgano Divulgativo de la Subdirección de Investigación y Postgrado del Instituto Pedagógico de Barquisimeto "Luis Beltrán Prieto Figueroa"

BARQUISIMETO – EDO. LARA – VENEZUELA

NUEVA ETAPA

FORMATO ELECTRÓNICO

DEPOSITO

PPI201002LA3674

ISSN: 2244-7296

LEGAL:

VOLUMEN 19 N° 2

MAYO – AGOSTO 2015

LA LINGÜÍSTICA DE CORPUS (LC) Y SU INCIDENCIA EN LA ENSEÑANZA DE LENGUAS EXTRANJERAS

CORPUS LINGUISTICS (CL) AND ITS IMPACT IN FOREIGN LANGUAGE TEACHING

CHINGER ZAPATA*

***UNIVERSIDAD PEDAGÓGICA EXPERIMENTAL LIBERTADOR – INSTITUTO PEDAGÓGICO DE BARQUISIMETO (UPEL - IPB)**

LA LINGÜÍSTICA DE CORPUS (LC) Y SU INCIDENCIA EN LA ENSEÑANZA
DE LENGUAS EXTRANJERAS

*CORPUS LINGUISTICS (CL) AND ITS IMPACT IN FOREIGN LANGUAGE
TEACHING*

REVISIÓN DOCUMENTAL

Chinger Zapata*
UPEL – IPB

Recibido:20/06/2015

Aceptado: 20/08/2015

RESUMEN

La LC ha brindado aportes a distintas ciencias del conocimiento, entre ellas, la didáctica de lenguas extranjeras. No obstante, muchos docentes de inglés en países de habla hispana parecieran no estar muy familiarizados con este tema. En el siguiente estudio de naturaleza documental, se revisa información relacionada a la LC, definición, teorías, métodos, clasificación, orígenes, con el propósito de develar el estado del arte de esta disciplina para poder posteriormente relacionarla con su impacto en la enseñanza de lenguas extranjeras. El estudio revela las ventajas en el diseño de materiales a partir de corpora y de su utilización como herramienta dentro de las estrategias de enseñanza. Se concluye con un énfasis en la importancia de la implementación de los corpora al salón de clases de inglés y con una exhortación a los docentes en la utilización de esta disciplina para optimizar los procesos de aprendizaje de una lengua extranjera.

Descriptor: Lingüística de Corpus, didáctica de lenguas extranjeras, materiales instruccionales

ABSTRACT

Corpus Linguistics (CL) has served many sciences and disciplines in the last years; foreign language teaching is one of them. However, many EFL teachers in Spanish speaking countries seem not to be aware of this fact. This documentary paper first revises the state of the art of this discipline: origins, theories, definitions, methods, classification; second, it sheds light on the impact of CL in the foreign language teaching field. The advantages of designing teaching materials based on corpus are discussed as well as the use of corpora as part of the teaching strategies. The importance of using corpora in the English language classroom is highlighted and teachers are advised to use CL to better foreign language learning processes.

Keywords: Corpus Linguistics, foreign language teaching, teaching materials

* Profesor en Lenguas Extranjeras, Mención Inglés; UPEL-IPB-1998. Magíster en Educación: Mención: EILE; UPEL-IPB-2003. Magíster en Lingüística; UPEL-IPB-2008. Candidato a Doctor del Programa Estudios del Discurso - UCV. Ponente en eventos nacionales e internacionales. Coordinador del Núcleo de Investigación para el Estudio y Enseñanza de Lenguas.

INTRODUCCIÓN

Durante el proceso de enseñanza del inglés como lengua extranjera (EILE), es necesario que los docentes tengan a disposición tantos recursos como les sea posible. Por lo general, los recursos orientados a desarrollar las destrezas receptivas (comprensión lectora y auditiva), están basados en lecturas diversas acerca de tópicos de interés y material auditivo de conversaciones o narraciones. Desafortunadamente, en la mayoría de los casos, tales recursos no constituyen ejemplos verdaderos de las lenguas en sus contextos. En el caso de las lecturas, dichos textos no son sino muestras cuidadosamente diseñadas de una lengua ficticia a través de lo que los expertos denominan “textos simplificados” (Crossley *et al*, 2007; Willis, 1998; Kennedy & Bolitho, 1984). La utilización de este tipo de materiales resulta en muchos casos contraproducente para el aprendiz del inglés como lengua extranjera, ya que, de acuerdo a Crossley *et al* (ob. cit, p.18), este tipo de textos: “...sirven solo para enseñarle a los estudiantes estructuras lingüísticas atípicas y artificiales”¹.

Esta falta de material auténtico y sistemáticamente seleccionado se ha convertido en un obstáculo para que los estudiantes observen y analicen el uso de la lengua en contextos cotidianos durante el proceso de aprendizaje.

Partiendo de la situación antes descrita, se hace necesario que tanto investigadores como docentes recolecten muestras genuinas de lenguaje que puedan ser estudiadas y modeladas por los estudiantes. En este sentido, la LC se presenta como una opción para paliar esta situación. A través de esta disciplina, la lengua puede ser analizada en sus diversas formas, oral, escrita, gestual, entre otras. Por esta razón, la LC ha sido bien recibida a nivel mundial; ella constituye la base para la construcción de diccionarios, atlas lingüísticos, incluso textos de sintaxis, de colocaciones, y cualquier otro tipo de material etnográfico y sociolingüístico.

El impacto y los aportes de la LC en la sociolingüística, la etnografía, los estudios variacionistas y la dialectología, principalmente, han sido inmensos, pero sus contribuciones también alcanzan a disciplinas como el Análisis del Discurso, la Lingüística Computacional y otros campos de investigación emergentes. Es por ello que el ámbito de la

didáctica no queda atrás. Si la LC ha ofrecido información valiosa para la explicación y descripción de la conformación de las lenguas a nivel mundial, entonces es de suponerse que, en el proceso de enseñanza de las mismas, los expertos y docentes hagan uso de esta herramienta para optimizar los procesos de instrucción y desarrollar materiales auténticos que le brinden a los estudiantes muestras verdaderas de la lengua meta. Crossley *et al* (ob. cit.), basados en Larsen-Freeman (2002), señalan que la corriente pedagógica en pro de la enseñanza de lenguas bajo el enfoque comunicativo: "...enfatisa el uso de lenguaje verdadero tanto como sea posible para que los estudiantes puedan estar expuestos a contexto reales y muestras naturales de lenguaje"¹ (p. 17). Estos señalamientos nos conducen a las siguientes interrogantes: ¿Cuál es el estado del arte de la LC como teoría y método para optimizar los procesos de enseñanza y aprendizaje del inglés como lengua extranjera? ¿Cuáles otras aplicaciones tiene la LC en la enseñanza y aprendizaje del inglés como lengua extranjera?

El propósito de esta investigación documental es revisar el panorama general de la LC, sus orígenes, teorías y métodos para, posteriormente, señalar el impacto que, tanto teoría como método, tienen en la didáctica de lenguas extranjeras, específicamente, el inglés.

LA LC

Esta disciplina ha generado, en los últimos años, un debate entre expertos con respecto a si la misma es un referente teórico o metodológico. En esta breve disertación haremos referencia a ambas posturas. Pero antes, iniciaremos con una visión retrospectiva de esta disciplina a manera de contextualización a partir de su objeto de estudio: el corpus, antes de ahondar en sus definiciones, aportes y aplicaciones teóricas y metodológicas.

¿Qué es un corpus?

Para responder esta interrogante, estudiaremos las definiciones de varios autores en orden cronológico. El primero en esta lista es considerado uno de los pioneros y máximos representantes del trabajo con corpora: John Sinclair. Para este autor un corpus es: "Una

colección de fragmentos de lenguaje seleccionados y organizados de acuerdo a criterios lingüísticos explícitos para ser usado como muestras o ejemplos de lenguaje”² (Sinclair, 1996, p. 4). A partir de esta definición, el resto de los autores aportan conceptos que giran en torno a la misma idea planteada por Sinclair.

No obstante, Torruella y Llisterri (1999, p. 7) acotan que el concepto de corpus no debe ya verse como simplemente corpus, fuera del contexto computacional, y agregan que el constructo debe ser “corpus informatizado” para describir la inclusión de las tecnologías en el uso y diseño de corpora. El mismo Sinclair (ob. cit.) ofrece la definición para esta orientación más tecnológica, cuando sostiene que un corpus informatizado es: “Un tipo de corpus que ha sido codificado de manera estandarizada y homogénea para realizar tareas de búsquedas abiertas. Las partes constituyentes de lenguaje están documentadas desde sus orígenes y procedencia”³ (p. 4).

En este mismo orden de ideas, Tognini-Bonelli (2001) sostiene que un corpus es:

Una colección de textos que se presumen representativos de una lengua dada para poder así ser objeto de análisis lingüístico. Por lo general, se asume que el lenguaje que se almacena en un corpus es natural, recolectado de acuerdo a criterios explícitamente diseñados, con propósitos específicos en mente, y bajo la premisa de representar grandes cantidades de lengua de acuerdo a una tipología específica. [...] en líneas generales, el consenso es que un corpus contiene lengua auténtica y natural⁴ (p. 2).

Aunque el concepto de Tognini-Bonelli no revela de manera explícita información relacionada con el uso de las tecnologías, se puede inferir que, para la autora, ésta es indispensable, puesto solo a través de un programa computarizado se puede manejar “grandes cantidades de lengua”, tal y como lo plantean Torruella y Llisterri (ob. cit., p. 7): “...hoy en día la informática facilita tanto la organización y la explotación de grandes cantidades de datos⁵ que sería imposible crear un corpus prescindiendo de este medio o herramienta”.

Otro autor que define corpus dentro de los parámetros informatizados es Hunston (2002), al señalar que un corpus es:

Una colección de textos (un ‘cuerpo’ de lengua) almacenado en una base de datos electrónicos. Los corpora son por lo general grandes cuerpos de textos procesados informáticamente que contienen miles de millones de palabras. Un corpus se diferencia de un archivo en el hecho de que a menudo (pero no

siempre) los textos que han sido seleccionados son representativos de una variedad particular de lengua o género, actuando así como una referencia estándar. Los corpora se usan tanto para análisis cuantitativo como cualitativo⁶ (p. 2-3).

Para cerrar la serie de conceptos, Stubbs (2002) y Meyer (2004) proponen cada uno definiciones que giran más en torno a los propósitos que a la forma como han sido diseñados los corpora. Así tenemos que el primero sostiene que un corpus es:

Una colección de textos diseñado con algún propósito, por lo general, para la enseñanza o la investigación. Un corpus no es algo que el hablante haga o sepa, sino algo que construye un investigador. Es entonces un registro de la actuación, por lo general de muchos usuarios, concebido para su estudio y poder así hacer inferencias acerca del uso típico del lenguaje (Stubbs, 2002, p. 239)⁷.

Para Stubbs, el corpus es, en efecto, una herramienta cuyo propósito, entre otros, es ser explotado en el ámbito de la enseñanza, ya que a través del análisis de su contenido se indaga con respecto al comportamiento natural de las lenguas. El segundo autor, Meyer (2004), con un concepto más sencillo pero igualmente orientado al propósito, declara que: Un corpus puede ser considerado como una colección de textos o partes de textos sobre los cuales se puede llevar a cabo un análisis lingüístico (p. xi)⁸.

Esta serie de conceptos nos hablan, por un lado, de una colección de textos (orales o escritos) como muestra natural de la actuación de los usuarios de una lengua determinada y, por el otro, de cómo esta colección de textos funge como instrumento a través de la cual se puede estudiar, describir y explicar todo lo que acontece en una lengua.

Tipos de corpora

La clasificación de los corpora atiende a criterios diversos y a los propósitos para los cuales son diseñados. Torruella y Llisterri (ob. cit. pp. 8-15) proponen una clasificación que, en principio, se divide en dos grandes grupos: los escritos y los orales. En cada una de estas divisiones, existe una serie de corpora diseñados con propósitos y criterios de selección particulares. La siguiente tabla resume el inventario de los autores.

Tabla. 1. Clasificación de los corpora (torruella y llisterri, 1999).

Escritos		Orales
Según el porcentaje y distribución de los textos	- <i>Corpus grande</i> : alberga una cantidad muy elevada de datos.	- Corpus para la descripción fonética de la lengua : los datos han sido grabados en condiciones acústicas óptimas para su posterior análisis en laboratorio.
	- <i>Corpus equilibrado</i> : contiene variedades de textos con porcentajes de datos equilibrados para cada variedad.	
	- <i>Corpus piramidal</i> : estos corpora tienen varios niveles de textos y porcentajes de datos. A menor variedad de textos por nivel, mayor porcentaje de datos.	
	- <i>Corpus monitor</i> : contiene un número exacto de datos que se actualizan con el tiempo.	
	- <i>Corpus paralelo</i> : lo compone una colección de textos traducidos a dos o más lenguas.	
	- <i>Corpus comparable</i> : se trata de textos del mismo (género) pero escritos en varias lenguas.	
	- <i>Corpus multilingüe</i> : colección de textos de diversos géneros en distintas lenguas.	
Según la especificidad de los textos	- <i>Corpus general</i> : incluye una variedad muy amplia de géneros discursivos.	- Corpus para el desarrollo de sistema en el ámbito de las tecnologías del habla : contienen "...inventarios grabados de unidades de síntesis a partir de los cuales se realiza el paso de una representación ortográfica a una onda sonora..." (p. 14).
	- <i>Corpus especializado</i> : incluye solamente textos especializados de un solo tipo, como los poéticos.	
	- <i>Corpus genérico</i> : incluye únicamente textos de un solo género, incluyendo los que no son especializados.	
	- <i>Corpus canónico</i> : colección de textos de un mismo autor, independientemente del género, como por ejemplo: todas las obras de Shakespeare.	

	<ul style="list-style-type: none"> - <i>Corpus cronológico</i>: incluye datos en secuencia de una época en particular. - <i>Corpus diacrónico</i>: los datos en este corpus contienen textos de la misma lengua en diferentes etapas del tiempo con el propósito de observar su evolución. 	
Según la cantidad de texto que se recoge de cada documento	- <i>Corpus textual</i> : recoge íntegramente todos los textos existentes del tipo de documento que lo constituye.	<p>- Corpus de transcripciones ortográficas de la lengua hablada: contienen grabaciones de conversaciones espontáneas o de los medios de comunicación, incluyéndose también (...) discursos políticos, clases, sermones, etc.” (p.14).</p>
	- <i>Corpus de referencia</i> : a diferencia del textual, este sólo incluye fragmentos de textos de los documentos que lo constituyen.	
	- <i>Corpus léxico</i> : recoge sólo fragmentos de textos de igual tamaño, por lo general, muy pequeños (más pequeños que los que se incluyen en el de referencia).	
Según la codificación y la anotación	<ul style="list-style-type: none"> - <i>Corpus simple</i>: corpus cuyos textos no tienen codificación ni etiquetaje. Los textos están en formato neutros. - <i>Corpus anotado</i>: textos que incluyen codificación y etiquetaje. 	
Según la documentación que acompaña a los textos	- <i>Corpus documentado</i> : corpus cuyos textos incluyen datos de identificación (cabeceras).	
	- <i>Corpus no documentado</i> : los textos no incluyen datos de afiliación o cabecera.	

Como se puede apreciar en la tabla, la lista es amplia. La misma podría incrementarse, ya que con cada investigación podría existir un corpus diferente si se toma en consideración los propósitos del investigador. No obstante, la clasificación de los autores incluye una tipología completa, en el sentido que se ha tratado de incluir los distintos propósitos de investigación a partir de las distintas disciplinas que pueden hacer provecho de esta herramienta.

En esta tipología podemos apreciar cómo en ocasiones algunas clasificaciones constituyen prácticamente lo mismo, por ejemplo, un corpus simple podría al mismo

tiempo ser un corpus monitor por el número reducido de datos que ambos manejan, pero, como acotamos anteriormente, hay una diferencia en cuanto al propósito.

Una vez expuesto este breve preámbulo en torno a los corpora, nos centraremos en la definición y el origen de la LC.

Definición de LC

La definición de este constructo sólo puede hacerse tomando en consideración la orientación filosófica con la que se estudie. Así, si se considera a la LC como una teoría, entonces las definiciones estarán orientadas a señalar los aportes teóricos que desde esta perspectiva se logren. Por su parte, si la misma se define bajo criterios metodológicos, entonces sus definiciones apuntarán a la recolección y procesamiento de datos y al análisis resultante de los mismos.

Definición de la LC como teoría

Para el concepto de LC como teoría, Halliday (1994, p. 24) declara que: "...la lingüística de corpus pone en el mismo plano las actividades de recolección de datos y teorización"⁹, y además sostiene que: "...esto nos está llevando hacia un cambio cualitativo en nuestra comprensión de la lengua"¹⁰.

Este señalamiento revela el carácter formal de la LC en el hecho que no sólo procesa datos, sino también a partir de ellos se construyen hipótesis y postulados que definen el comportamiento de la lengua.

Tognini-Bonelli (ob. cit. p. 1) complementa esta idea cuando declara: "Lo que estamos presenciando es el hecho que la lingüística de corpus se ha transformado en una nueva empresa investigativa y en un nuevo enfoque filosófico para la investigación lingüística"¹¹. Finalmente, esta autora concluye:

Mientras que una metodología puede ser definida como el uso de un conjunto de reglas dadas o porciones de conocimiento en una situación específica, (...) la lingüística de corpus está en posición de definir su propio conjunto de reglas y porciones de conocimientos *antes* de ser aplicados. Esto lleva a los lingüistas a hacer uso de nuevos parámetros para explicar los datos, y esto implica un cambio en lo que puede ser considerado como *la unidad de aceptación* para la descripción lingüística. De este modo, la lingüística de corpus adquiere un

estatus teórico que la coloca en posición de contribuir de manera específica a otras aplicaciones¹² (p. 1).

Para Tognini-Bonelli es notorio el papel teórico de LC en el estudio de las lenguas, por cuanto a través del manejo de conocimiento se pueden formular hipótesis teóricas que pueden ser comprobables luego en la implementación de corpora.

Definición de la LC como metodología

McEnery y Wilson (1996, p. 1) sostienen que la LC trata más con: ...el estudio de la lengua con base en ejemplos de uso en la ‘vida real’ y una metodología, que un aspecto del lenguaje que requiera una explicación o una descripción”¹³.

Asimismo, en un concepto más general pero que sigue por igual la orientación práctica de sus antecesores, Kennedy (1998, p. 1) señala que la LC es: “Una empresa académica que se encarga de la compilación y análisis de corpora”¹⁴. Esta visión metodológica de la LC destaca el carácter pragmático e instrumentalista de la misma como herramienta para el estudio y análisis de las lenguas.

Aunque ambas tendencias de la LC presentan planteamientos interesantes, es prudente destacar que la LC como herramienta metodológica parece dar aportes más significativos en el ámbito de la enseñanza de lenguas. En este sentido, los abanderados de la visión metodológica han dado grandes pasos en relación a su postura. Lo que ellos han propuesto ha surgido luego de señalamientos como los de Halliday (ob. cit.). Tal es el caso de Stubbs (1996), quien dos años después de lo expuesto por Halliday, aclaró que la LC: “...tiene solo unos pequeños esquemas preliminares de teoría con los que se relaciona a los textos individuales con textos de corpora (p. 231)”¹⁵; aclaratoria que fue reforzada posteriormente por las definiciones de McEnery y Wilson y Kennedy ya citadas.

Por su parte, Meyer (2004) sostiene que, a partir del uso que se le ha dado a la LC: ...se hace muy evidente que la lingüística de corpus es más una forma de hacer lingüística que un paradigma aparte dentro de la lingüística”¹⁶ (p. xi).

Ambas posiciones sostienen argumentos serios que le asignan a la LC un estatus diferente. Por lo tanto, el tratamiento, bien sea teórico o metodológico, que cada investigador le dé a la LC dependerá exclusivamente de sus intereses y objeto de estudio.

Enfoque a partir de los datos (Corpus-driven approach - CDA)

En este enfoque, el punto de partida para los lingüistas lo constituye la integridad de la totalidad de los datos; cada descripción de la lengua que se haga debe ser lo suficientemente comprensible y sustentada en la evidencia hallada en el corpus (Tognini-Bonelli, ob. cit., p. 84). Al respecto, Storjohann (2005, p. 5) comenta: “Cada conclusión o señalamiento se hace exclusivamente con base en las observaciones al corpus”¹⁷. Con lo que queda establecido que no se llega al corpus con ideas preconcebidas de lo que sucede en la lengua, sino que, a partir de lo que se observe en ella, se elaboran descripciones que posteriormente podrían convertirse en afirmaciones teóricas. Es decir, los ejemplos extraídos del corpus no se obtienen porque encajan en categorías de análisis pre-existentes, sino que son el producto de patrones recurrentes con los que posteriormente se espera diseñar las categorías de análisis (Baker *et al*, 2006, p. 49).

Tognini-Bonelli (ob. cit., p. 84) concluye que: “La teoría no es independiente de la evidencia”¹⁸; lo cual determina el camino de quienes emplean este enfoque: “...la observación conduce a la hipótesis y ésta a la generalización, por consiguiente, a la unificación de las afirmaciones teóricas”¹⁹ (p. 85).

Enfoque con base en los datos (Corpus-based approach - CBA)

Los investigadores que emplean este enfoque ven al corpus como una herramienta que les permite confirmar señalamientos teóricos. Por tanto, el corpus viene a ser una manera de comprobar que sus intuiciones con respecto al comportamiento de un elemento específico de la lengua son ciertas o falsas, un instrumento para calcular porcentajes de fenómenos lingüísticos y verificar expectativas o extraer ejemplos. En otras palabras, es un material de apoyo adicional (Biber, *et al*, 1998; Tognini-Bonelli, ob.cit.; Storjohann, ob. cit.; Baker *et al*, ob.cit.).

Al parecer, los lingüistas y lo que cada uno de estos enfoques representa no consiguen un punto de acuerdo con respecto al estudio de la lengua. Sin embargo, Xiao (2008a), al igual que Leech (1992) y Meyer (2004), está convencido que: “...la mayoría de los lingüistas está consciente que los dos enfoques son complementarios”²⁰ (p. 987). Ante este planteamiento, los investigadores necesitamos entender que nuestra misión no es hacer

casarnos con un enfoque en particular, sino hacer uso de los recursos con los cuales contamos para una mejor comprensión de los fenómenos de la lengua.

Orígenes de la LC

De acuerdo con Meyer (2008), hablar de LC desde el punto de vista histórico implica, por un lado, hacer referencia a los primeros corpora, así como también los primeros trabajos realizados con en base en ellos; y por el otro, trazar una línea referencial en el tiempo entre lo que este autor considera la era pre-electrónica y la era electrónica. Lo que marca la diferencia entre ambas referencias temporales es la inclusión de la computadora en el diseño y creación de corpora.

La era pre-electrónica

Meyer (2008,p. 1) aclara que el trabajo con corpus en la primera de estas eras:“...consistía en un texto o textos que servían como base para un proyecto particular y tenían que ser analizados manualmente a través de un procedimiento tedioso y prolongado”²¹.

Durante la etapa pre-electrónica, acota Meyer, los cuatro tipos de proyectos diferentes que se pueden destacar son: los realizados para las concordancias bíblicas, los estudios de gramática, los de diccionarios y el proyecto de la primera encuesta para el uso del inglés.

En el proyecto para concordancias bíblicas, el trabajo que destaca, por su elaborado detalle, es el de Alexander Cruden en el siglo XVIII para la versión bíblica King James. La estrategia empleada por Cruden para construir las concordancias fue de organizar alfabéticamente las entradas en listados de papel por separado, de donde la letra C, por citar un ejemplo, contaba con ¡1019 encabezadores! Este corpus de concordancias incluyó aproximadamente unas 2.370.000 palabras en total, número éste que sobrepasa al número de palabras contenidas en la biblia misma. Este arduo trabajo fue realizado en un lapso de dos años, a razón de dieciocho horas diarias (Meyer, 2008). Cabe destacar que el interés de Cruden no fue lingüístico, sino social y religioso, ya que su objetivo fue crear un sistema que le permitiera a los lectores una mejor comprensión del texto bíblico. Podría acortarse

que, a su vez, el propósito de tal trabajo no deja de estar relacionado con el hecho del aprendizaje, en este caso, con el de la comprensión lectora.

En los proyectos para el estudio de gramática se reconocen como pioneros los trabajos de Aristonicus de Alejandría en el Siglo I para el griego, con su obra *Palabras no gramaticales (Ungrammatical Words)*, donde se revisaba las construcciones gramaticales irregulares a partir del corpus de Homero. De la misma manera, en el Siglo IV, se tiene a la gramática de Panini, la cual surgió a partir del corpus del desaparecido idioma de los vedas o sánscrito clásico (Meyer, 2008). Después de estos trabajos, en los siglos siguientes se registraron una serie de propuestas y estudios gramaticales para el inglés como los de Robert Lowth y su obra *Una breve introducción a la gramática del inglés (A Short Introduction of the English Grammar)*. A este ejemplar le siguieron estudios en los Siglos XIX y primera parte del XX, como por ejemplo los de Otto Jespersen, Charles Fries, entre otros.

En el área lexicográfica, los trabajos realizados a partir de corpora le corresponden a John Florio con su Diccionario Inglés-Italiano en 1598, trabajo éste que inspiraría unos años más tarde la clásica obra de Samuel Jhonson y su *Diccionario del Idioma Inglés (Dictionary of the English Language)* en 1775, famoso por la inclusión extensiva de 150.000 citas ilustrativas aproximadamente²².

El último proyecto a destacar en la era pre-electrónica de corpus, es el desarrollado por el *University College London*, a la cabeza de Randolph Quirk a partir de 1959, el cual llevó por título *La Encuesta para el Uso del Inglés* o el SEU (por sus siglas en inglés *Survey for English Usage*²³). Este trabajo fue concebido para subsanar las deficiencias que dejaban las gramáticas de autores como Poutsma, Kruisinga y Jespersen, en relación a la inclusión de registros orales sino exclusivamente escritos (Meyer, 2008). El mismo Quirk (1974, p. 167) señala: “...las grandes gramáticas fallaron [...] su uso generalmente ecléctico de fuentes (escritas) a menudo deja sin una clara explicación la diferencia entre las estructuras normales y las no tan normales, así como las condiciones para selección de las últimas”²⁴.

La era electrónica

La inclusión de la computadora vino a facilitar y optimizar el trabajo con corpus. Al respecto, Llisterri y Torruella (ob. cit., p. 4) señalan: “Muchos trabajos que antes tenían que hacerse a mano, [...] hoy, con la ayuda de la informática, se pueden hacer no solamente con menos tiempo sino también más ordenada y exhaustivamente, o sea, con mayor eficacia y eficiencia”. Con lo que queda claro que la herramienta informática vino a revolucionar el campo de la LC.

Estos mismos autores destacan las múltiples ventajas de los corpora informatizados en diferentes disciplinas relacionadas a los estudios lingüísticos, como la lexicografía y la terminología, la estadística lingüística, la gramática histórica, la fonética y las tecnologías del habla, la didáctica, la sociolingüística, el análisis del discurso y la psicolingüística; así como también otras áreas de las humanidades como la literatura, los estudios históricos, entre otros. Al igual que Stubbs, Llisterri y Torruella también destacan el campo de la didáctica para el uso de los corpora, lo que implica que esta herramienta debería incluirse en la planificación de la enseñanza de lenguas en general.

Hoy día, la LC y la computadora son ambos, disciplina y herramienta inseparables, aunque algunos estudios puedan, por sus propósitos y objetivos, hacer uso de corpora cuyos datos no se procesan electrónicamente. No obstante, en el ámbito de la enseñanza, la inclusión de las herramientas informatizadas para el procesamiento de datos desde los corpora se hace casi imprescindible, ya que con el uso de los corpora se garantiza resultados más generalizables y confiables, especialmente si se trabaja con grandes cantidades de datos.

Entre la segunda etapa del siglo XX y la primera del siglo XXI, los corpora que incluyen al computador como su aliado, de acuerdo con Xiao (2008b), son: el Brown Corpus, del cual se desprende una familia de corpora que siguen el mismo modelo de desarrollo y criterios, con base en las distintas variantes del inglés que datan desde ese mismo año. Así tenemos para el inglés americano, el Brown Corpus (1961) y el Frown Corpus (1991-1992); para el inglés británico existen el LOB Corpus (1961) y el FLOB Corpus (1991-1992); para el inglés hindú, el Kolhapur Corpus (1978); para el inglés

australiano, el ACE Corpus (1986) y, para el inglés neozelandés, el WWC Corpus (1986-1990).

También existen otros corpora con criterios, propósitos y modelos distintos que describen la realidad de cada lengua a nivel nacional. En esta lista encontramos al British National Corpus (BNC), el American National Corpus (ANC), los corpora de referencia del polaco, el corpus nacional del checo, el corpus nacional húngaro, el corpus nacional ruso, el CORIS Corpus del italiano escrito, el corpus nacional helénico, el DWDS Corpus para el alemán, el corpus nacional eslovaco, el corpus moderno del idioma chino, el corpus nacional coreano, entre otros corpora²⁵.

La LC y la enseñanza de lenguas extranjeras

El desarrollo de la LC representa, tal y como se describió en apartados anteriores, un avance significativo como instrumento para el análisis de la lengua. Pero, en relación al ámbito de la enseñanza de lenguas extranjeras, específicamente el inglés, ¿qué aportes puede ofrecer la LC? Al igual que en los estudios acerca de la metodología de la LC (CBA versus CDA), existe un debate entre quienes creen que la LC produce un impacto positivo y significativo en la enseñanza del inglés como lengua extranjera y quienes no lo creen así. Considerando el propósito del presente estudio, desarrollaremos a continuación la utilización de la LC en la enseñanza de lenguas extranjeras.

Impacto de la LC en la EILE

De acuerdo con Römer (2008, p. 112), “El diseño de corpora no se concibió pensando en su aplicabilidad en el ámbito de la didáctica de lenguas”²⁶. Sin embargo, con el paso del tiempo, expertos lingüistas y docentes han descubierto en la LC un herramienta con la cual se puede optimizar los procesos de enseñanza de una lengua extranjera (Davis y Russell-Pinson, 2007; Johansson, 2007; Neff, *et al* 2007; Zinggeler, 2007; Ädel 2006; Braun *et al* 2006; Gavioli 2006; Aston *et al* 2004; Connor y Upton 2004; Granger *et al* 2002; Burnard y McEnery 2000).

Leech (1997) acota que los aportes a la enseñanza de lenguas extranjeras desde la LC se pueden clasificar de dos maneras: aplicaciones indirectas y aplicaciones directas. La

primera se relaciona al hecho de cómo y cuándo utilizar al corpus para el desarrollo de materiales y diseños curriculares; la segunda está más orientada a la utilización del corpus dentro del salón de clases por parte de estudiantes y profesores.

Aplicaciones indirectas de los corpora

La utilización de corpora en el ámbito de la enseñanza de lenguas extranjeras se manifiesta, por un lado, a través del desarrollo y diseño de materiales instruccionales como diccionarios, libros texto de sintaxis, léxico, colocaciones; por el otro, las decisiones acerca del contenido de los programas de estudio con base en muestras de lenguaje (oral o escrito) provenientes de corpora y sobre la base de las evidencias lingüísticas. Como ejemplo de este aporte, en la EILE podemos hacer referencia a la existencia del *Collins COBUILD English Course* (CCEC), el cual es un programa que contiene las palabras y frases más comunes en inglés con sus respectivos significados en contexto. A través del CCEC, también se ha podido revelar aspectos como la estructuración de patrones en el idioma inglés observados en el alto porcentaje de combinaciones de repetición léxica y la inseparable relación entre el léxico y la gramática (Römer, 2008). Con respecto a estos aportes y aplicaciones de los corpora en EILE, el mismo Römer (2008) concluye:

Se puede asumir con toda seguridad que los aprendices podrán desarrollar con facilidad tanto las destrezas receptivas como las productivas cuando se enfrentan con los elementos léxicos más comunes del lenguaje y con los patrones y significados típicos, que cuando lo que se les enseña le da prioridad a palabras y estructuras poco frecuentes que raramente encontrarán en situaciones de la vida real²⁷ (p. 114).

Definitivamente, para el autor, la selección de materiales auténticos para promover y optimizar la enseñanza de la lengua extranjera es un punto primordial. La referencia a vocabulario y patrones comunes con evidencia en los datos de los corpora y, por ende, su preferencia a la hora de emplear algún material didáctico para enseñar inglés, está por encima de los materiales tradicionales, los cuales están basados en muestras ficticias y pre-fabricadas del lenguaje.

Aplicaciones directas de los corpora

La utilización de un corpus dentro del aula de clases para la EILE que provea a los aprendices con muestras del lenguaje genuino, al igual que el desarrollo de materiales didácticos, representa una gran ventaja. En contraposición a la aplicación anterior donde sólo los investigadores y diseñadores de programas tenían acceso al corpus, en la aplicación directa tanto docentes como alumnos pueden utilizar un corpus como parte de la estrategia de enseñanza de algún aspecto particular de la lengua. Esto permite a los estudiantes, por un lado, tener un grado mayor de autonomía en su aprendizaje y, por el otro, tener una apreciación más cercana, de primera mano, sobre cómo funciona un patrón o estructura lingüística o cuál es la posición más común para un determinado vocablo en un contexto escrito u oral. A este tipo de práctica, Johns (1994), citado en Römer (2008.), lo ha denominado “aprendizaje a partir de los datos” (*Data-driven learning:DDL*).El siguiente gráfico resume el papel de la LC en la enseñanza de lenguas.

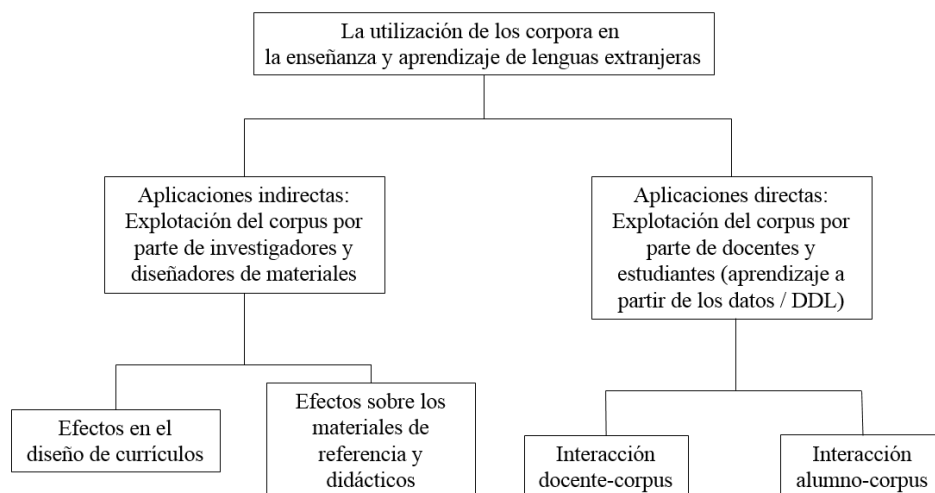


Gráfico 1. Relación de la LC con la enseñanza de lenguas (Römer, 2008)

REFLEXIONES FINALES

Aunque algunos docentes e investigadores han catalogado a la relación entre la LC y la EILE como “un matrimonio forzado”, por no tomar en cuenta algunas consideraciones de tipo pragmático, entre otras, en el hecho didáctico (Véliz Campos, 2008), no cabe la menor

duda de que los aportes desde la LC a la didáctica de las lenguas extranjeras representa un hecho sin precedentes e indetenible, además de efectivo si se maneja de manera apropiada. Evidentemente, la LC no pretende sustituir los métodos tradicionales de enseñanza del inglés; en nuestra opinión personal, ésta constituye un recurso más para optimizar dicho proceso. Además, el docente debe mantener su rol de mediador y supervisar cada actividad dentro del aula, aun cuando el alumno pueda guiar desde un computador parte de su instrucción.

Los resultados de múltiples investigaciones como las citadas en este estudio hasta la fecha, dan cuenta de la efectividad de la implementación de la LC como fuente para el diseño de materiales didácticos y desarrollo de currículo, así como también para la puesta en práctica de ciertas técnicas en las que el uso de corpora es un material fundamental. Pero, la razón principal por la cual se debería utilizar a la LC como un medio para mejorar la enseñanza de inglés en contextos donde esta lengua no es la nativa, es porque, a través de la misma, se puede hacer uso de materiales auténticos que les brinden a los estudiantes un patrón verdadero de la utilización del idioma (de Beaugrande 2001; Kennedy 1992; Römer 2004, 2005; Sinclair 1991). Al respecto, Kennedy (1992,p. 366) acota: "...los ejemplos inventados pueden presentar una versión distorsionada de la cotidianidad o una imagen muy torcida del sistema"²⁸. De la misma manera, en relación a la utilización de ejemplos fabricados, Sinclair (1991: 5) sostiene: "...es un absurdo pensar que los ejemplos inventados pueden en efecto representar al lenguaje mejor que los verdaderos"²⁹.

Estos señalamientos son una prueba contundente de que en la EILE, la utilización de materiales auténticos vía corpora es en definitiva necesaria. Por tanto, exhortamos a los docentes de inglés como lengua extranjera, por un lado, a incorporar dicha herramienta en el salón de clase de inglés para confrontar a los estudiantes con muestras verdaderas del lenguaje y, por el otro, a abocarse a la construcción de sus propios corpora, al diseño de materiales didácticos a partir de esta herramienta y a reorientar los contenidos de los programas de EILE.

REFERENCIAS

Ädel, A. (2006). *Metadiscourse in L1 an L2 English*. Amsterdam: John Benjamins.

- Aston, G., Bernardini, S. y Stewart, D. (eds.). (2004). *Corpora and Language Learners*. Amsterdam: John Benjamins.
- Baker, P., Hardie, A., & McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Biber, D., Conrad, S. y Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Braun, S., Kohn, K. y Mukherjee, J. (eds.). (2006). *Corpus Technology and Language Pedagogy*. Frankfurt: Peter Lang.
- Burnard, L. y McEnery, T. (eds.). (2000). *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt: Peter Lang.
- Connor, U. y Upton, T. A. (eds.). (2004). *Applied Corpus Linguistics: A Multidimensional Perspective*. Amsterdam: Rodopi.
- Crossley, S., McCarthy, P., Louwse, M., & McNamara, D. (2007). A Linguistic Analysis of Simplified and Authentic Texts. *The Modern Language Journal*, 91(i), 15-30.
- Davis, B. y Russell-Pinson, L. (2007). One Corpus, Two Contexts: Intersections of Content-Area Teacher Training and Medical Education. En E. Fitzpatrick. (ed.). *Corpus Linguistics Beyond the World*. [pp. 143-165]. Amsterdam: Rodopi.
- de Beaugrande, R. (2001). Twenty Challenges to Corpus Research. And How to Answer Them. [Documento en línea]. Disponible en: <http://beaugrande.bizland.com/Twenty%20questions%20about%20corpus%re> [Consulta: 2011, Febrero 12].
- Gavioli, L. (2006). *Exploring Corpora for ESP Learning*. Amsterdam: John Benjamins.
- Granger, S., Hung, J. y Petch-Tyson, S. (eds.). (2002). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.
- Halliday, M.A.K. (1994). *An Introduction to Systemic Functional Grammar*. (2ed.). London: Edward Arnold.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Johansson, C. (2007). The Use of Relativizers across Speaker Roles and Gender: Explorations in 19th-century Trials, Drama and Letters. En E. Fitzpatrick. (ed.). *Corpus Linguistics Beyond the World*. [pp. 257-277]. Amsterdam: Rodopi.
- Kennedy, C., & Bolitho, R. (1984). *English for Specific Purposes*. Hong Kong: Macmillan.
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London: Longman.
- Kennedy, G. (1992). Preferred Ways of Putting Things with Implications for Language Teaching. En J. Svartvik. (ed.). *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm*, [pp. 335-373]. Berlin: Mouton de Gruyter.

- Leech, G. (1992). Corpora and Theories of Linguistic Performance. En J. Svartvik. (ed.). *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm*, [pp. 105-122]. Berlin: Mouton de Gruyter.
- Leech, G. (1997). Teaching and Language Corpora: A Convergence. En: A. Wichmann; S. Fligelstone; T. McEnery y G. Knowles. (eds.). *Teaching and Language Corpora*. [pp. 1-23]. London: Longman.
- Meyer, C. (2004). *English Corpus Linguistics: An Introduction*. United Kingdom: Cambridge University Press.
- Meyer, C. (2008). Pre-electronic Corpora. En A. Lüdeling, & M. Kytö. (eds.). *Corpus Linguistics: An International Handbook*. Vol. 1. [pp. 1-14]. Walter de Gruyter GmbH & Co. KG, 10785 Berlin, Germany.
- McEnery, T. y Wilson, A. (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Neff, J., Ballesteros, F., Dafouz, E., Martínez, F., Rica, J-P., Díez, M. y Prieto, R. (2007). A Contrastive Functional Analysis of Errors in Spanish EFL University Writers' Argumentative Texts: Corpus-based Study. En E. Fitzpatrick. (ed.). *Corpus Linguistics Beyond the World*. [pp. 203-225]. Amsterdam: Rodopi.
- Quirk, R. (1974). *The English and the English Language*. London: Edward Arnold.
- Römer, U. (2004). Comparing Real and Ideal Language Learner Input: The Use of an EFL Textbook Corpus in Corpus Linguistics and Language Teaching. En G. Aston, S. Bernardini, y D. Stewart. (eds.). (2004). *Corpora and Language Learners*. [pp. 151-168]. Amsterdam: John Benjamins.
- Römer, U. (2008). Corpora and Language Teaching. En A. Lüdeling, & M. Kytö. (eds.). *Corpus Linguistics: An International Handbook*. Vol. 1. [pp. 112-131]. Walter de Gruyter GmbH & Co. KG, 10785 Berlin, Germany.
- Römer, U. (2005). *Progressives, Patterns, Pedagogy. A Corpus-driven Approach to English Progressive Forms, Functions, Contexts and Didactics*. Amsterdam: John Benjamins.
- Sinclair, J. (1991). *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Sinclair, J. (1996). Preliminary Recommendations on Corpus Typology. Technical report, EAGLES (Expert Advisory Group on Language Engineering Standards). Birmingham: University of Birmingham. [Documento en línea]. Disponible: <http://www.ilc.cnr.it/EAGLES/pub/eagles/corpora/corpus typ.ps.gz> [Consulta: 2011, Mayo 16].
- Storjmann, P. (2005). Corpus-driven vs. Corpus-based Approach to the Study of Relational Patterns. En *Proceedings from the Corpus Linguistics Conference Series, Vol. 1(1)*, (comp.). pp. 1-20. Birmingham: University of Birmingham.
- Stubbs, M. (1996). *Text and Corpus Analysis*. London: Blackwell.
- Stubbs, M. (2002). *Words and Phrases*. (2ed). London: Blackwell.

- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work (Studies in Corpus Linguistics: 6)*, Amsterdam/Atlanta, GA: John Benjamins.
- Torruella, J. & Llisterri, J. (1999). Diseño de corpus textuales y orales. En J. M. Bleca; G, Clavería; C, Sánchez & J. Torruella (comps.). *Filología e informática. Nuevas tecnologías en los estudios filológicos*. pp. 45-77. Editorial Milenio, España.
- Veliz Campos, M. (2008). La Lingüística de Corpus y la enseñanza del inglés (como lengua extranjera): ¿un matrimonio forzado? *Literatura y Lingüística*, 19, 251-263.
- Willis, J. (1998). Concordances in the classroom without a computer: Assembling and exploiting concordances of common words. En B. Tomlinson (Ed.). *Materials Development in Language Teaching*. (pp. 44–66). Cambridge: Cambridge University Press.
- Xiao, R. (2008a). Theory-driven Corpus Research: Using Corpora to Inform Aspect Theory. En A. Lüdeling, & M. Kytö. (eds.). *Corpus Linguistics: An International Handbook*. Vol. 2. [pp. 987-1008]. Walter de Gruyter GmbH & Co. KG, 10785 Berlin, Germany.
- Xiao, R. (2008b). Well-known and Influential Corpora. En A. Lüdeling, & M. Kytö. (eds.). *Corpus Linguistics: An International Handbook*. Vol. 1. [pp. 383-456]. Walter de Gruyter GmbH & Co. KG, 10785 Berlin, Germany.
- Zinggeler, M. (2007). “GRIMMATIK:” German Grammar through the Magic of the Brothers Grimm Fairy Tales and the Online Grimm Corpus. En E. Fitzpatrick. (ed.). *Corpus Linguistics Beyond the World*. [pp. 167-184]. Amsterdam: Rodopi.

NOTAS

1. Original en inglés: “...emphasizes the use of authentic language whenever possible so that students can be introduced to real context and natural examples of language”.
2. Original en inglés: “A collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language”.
3. Original en inglés: “A corpus which is encoded in a standardised and homogeneous way for open-ended retrieval tasks. Its constituent pieces of language are documented as to their origins and provenance”.
4. Original en inglés: A collection of texts assumed to be representative of a given language put together so that it can be used for linguistic analysis. Usually the assumption is that the language stored in a corpus is naturally-occurring, that it is gathered according to explicit design criteria, with a specific purpose in mind, and with a claim to represent larger chunks of language selected according to a specific typology. [...] in general there is consensus that a corpus deals with natural, authentic language

5. El subrayado es del autor del presente estudio.
6. Original en inglés: A collection of texts (a ‘body’ of language) stored in an electronic database. Corpora are usually large bodies of machine-readable text containing thousands or millions of words. A corpus is different from an archive in that often (but not always) the texts have been selected so that they can be said to be representative of a particular language variety or genre, therefore acting as a standard reference. [...] Corpora can be used for both quantitative and qualitative analyses.
7. Original en inglés: A collection of texts, designed for some purpose, usually teaching or research. [...] A corpus is not something that a speaker does or knows, but something constructed by a researcher. It is a record of performance, usually of many different users, and designed to be studied, so that we can make inferences about typical language use.
8. Original en inglés: “A corpus will be considered a collection of texts or parts of texts upon which some general linguistic analysis can be conducted”.
9. Original en inglés: “... corpus linguistics re-unites the activities of data gathering and theorizing...”
10. Original en inglés: “...this is leading to a qualitative change in our understanding of language...”
11. Original en inglés: “What we are witnessing is the fact that corpus linguistics has become a new research enterprise and a new philosophical approach to linguistic enquiry”.
12. Original en inglés: While a methodology can be defined as the use of a given set of rules or pieces of knowledge in a certain situation, (...) corpus linguistics is in a position to define its own set of rules and pieces of knowledge *before* they are applied; this leads the linguist to make use of some new parameters to account for the data, and this entails a change in what can be referred as *the unit of currency* for linguistic description, corpus linguistics has, therefore, a theoretical status and because of this it is in a position to contribute specifically to other applications.
13. Original en inglés: “...the study of language based on examples of ‘real life’ language use and a methodology rather than an aspect of language requiring explanation or description”.
14. Original en inglés: “A scholarly enterprise concerned with the compilation and analysis of corpora”.
15. Original en inglés: “...has as yet only very preliminary outlines of a theory which can relate individual texts to text corpora”.
16. Original en inglés: “...it becomes quite evident that corpus linguistics is more a way of doing linguistics [...] than a separate paradigm within linguistics”.
17. Original en inglés: “Any conclusions or claims are made exclusively on the basis of corpus observations”.

18. Original en inglés: “The theory has no independent existence from the evidence”.
19. Original en inglés: “...observation leads to hypothesis leads to generalization leads to unification in theoretical statement”.
20. Original en inglés: “...the majority of linguists (...) are aware that the two approaches are complementary to each other”.
21. Original en inglés: “...consisted of a text or texts that served as the basis of a particular project, and had to be analyzed through often time-consuming and tedious manual analysis”.
22. Para una mayor comprensión de este trabajo, consultar Meyer (2008).
23. Para una mayor comprensión del SEU, consultar Meyer (2008).
24. Original en inglés: “...the big grammars fell short [...] their generally eclectic use of [written] source materials too often leaves unclear the distinction between normal and relatively abnormal structures and the conditions for selecting the latter...”
25. Para una lista más extensa de los distintos tipos de corpora y su respectiva descripción, consultar Xiao (2008b) y Baker, Hardie & McEnery (2006).
26. Original en inglés: “Corpora was not meant to be used in language teaching”.
27. Original en inglés: It can be safely assumed that learners will find it easier to develop both their receptive and productive skills when they are confronted with the most common lexical items of a language and the patterns and meanings with which they typically occur than when the language teaching input they get gives high priority to infrequent words and structures which the learners will only rarely encounter in real-life situations.
28. Original en inglés: “...invented examples can present a distorted version of typicality or an over-tidy picture of the system”.
29. Original en inglés: “...absurd notion that invented examples can actually represent the language better than real ones”.