La utilización de la lingüística de corpus en la enseñanza de español: usos de *pero* y *sino*

Iandra Maria da Silva
(CEFET-MG/UFSC)

1. CONSIDERACIONES INICIALES

Últimamente se discute mucho la importancia y el potencial de los corpora lingüísticos para el área de la enseñanza/aprendizaje de lenguas y, según nuestra observación, se verifica el creciente número de trabajos, bajo esta perspectiva. Sin embargo, la gran mayoría de estos estudios pertenecen al campo del inglés.

Lo que nos anima a realizar esta comunicación es abordar la utilización de corpora direccionados a la lengua española con el objetivo de presentar y ofrecer sugerencias de tratamiento de elementos gramaticales de difícil acceso para alumnos de ELE.

La lingüística de corpus (LC) trabaja con este objetivo, y por lo tanto, nos proponemos discutir las aportaciones que la LC puede traer a la clase de español como lengua extranjera, no solamente al profesor al preparar sus clases sino también a los aprendices como herramienta capaz de reunir ejemplos de uso auténtico de lenguaje, proporcionar independencia y estimular la reflexión de los alumnos con respecto a la lengua que están aprendiendo.

Ilustramos este estudio presentando las etapas de una actividad realizada con alumnos de español en el estudio de las diferencias de uso de *pero* y *sino* basada en el corpus. La propuesta se desarrolla con un grupo de estudiantes de español, de nivel intermedio, en un curso de lenguas ofrecido por la Universidad Federal de Santa Catarina, durante el año de 2005.

Durante el trabajo que realicé como profesora de este grupo, me llamaban la atención algunas de las dudas y dificultades específicas manifestadas por los aprendices en relación con algunos puntos gramaticales y también la escasez de informaciones en los libros didácticos y en las gramáticas de referencia. Dicha situación motivó la investigación que tiene por objetivo describir y reflexionar, a través de textos reales de la lengua española, los usos de las conjunciones especificadas.

La investigación basada en corpora tiene en la lingüística de corpus el fundamento teórico que orienta al análisis de información frecuencial, con una visión general de acuerdo con los postulados de Sinclair, 1991; y Sánchez, 1995, y Berber Sardinha, 2000, 2004.

De esta manera, este trabajo contribuye para que la LC pueda expandir sus aplicaciones a la enseñanza de español y pueda ofrecer ventajas al desarrollo de

actividades a partir de ejemplos concretos que permiten rastrear problemas gramaticales y proporcionar material auténtico para examinarlos.

Para el desarrollo del presente estudio, utilizamos una muestra de datos de periódicos electrónicos de 20 países de lengua hispánica recolectados durante el año 2005, más precisamente entre los meses de enero a junio. La verificación de los datos es realizada con la ayuda del programa WordSmith (Mike Scott, 1998). Los resultados indican factores que influyen en la elección de las conjunciones de acuerdo con empleos de mayor frecuencia.

2. PRESUPUESTOS GENERALES DEL CUADRO TEÓRICO

El objetivo de esta sección es describir el aparato teórico que justifica este estudio. Para eso, presentamos algunas consideraciones teóricas sobre la Lingüística de Corpus, la definición de *corpus* y discusiones sobre uso de concordancias y las herramientas disponibles para analizarlas.

El trabajo propuesto está fundamentado en el área de investigación basada en un corpus direccionado a la enseñanza y aprendizaje del español. La cuestión principal se refiere a la utilización de conceptos y aportes metodológicos de la lingüística de corpus como el soporte teórico para este estudio en el tratamiento de los usos de *pero* y *sino*, permitiendo hacer una descripción de la lengua basada en padrones y informaciones de frecuencia.

2.1 Lingüística de corpus

La metodología de análisis lingüística denominada *lingüística de corpus* puede ser definida como el estudio del lenguaje basado en ejemplos auténticos de uso, colectados en situaciones de vida real (McEnery y Wilson, 1996).

El desarrollo tecnológico impulsa cada vez más las investigaciones en esta área dentro del ámbito de la lingüística aplicada. Con la utilización de corpora varios investigadores presentan propuestas de estudio utilizando la LC en la enseñanza de lenguas extranjeras, principalmente del inglés.

En la gran mayoría, los trabajos desarrollados a nivel nacional, se usa el aparato teórico de la LC para preparación y utilización de material didáctico y descripción de corpus de aprendices.

Entre la diversidad de nuevos trabajos, citamos algunos estudios de «dissertação de mestrado» como la de Oliveira Souza, 2003; Balbás, 2003; Conde, 2002, y Jacobi 2001. Estos aportes utilizan el corpus para desarrollar sus investigaciones caracterizadas por una visión del lenguaje como proceso natural de construcción de sentido, de carácter eminentemente empírico.

De acuerdo con McEnery y Wilson (1996: 84), la condición empírica del lenguaje «es denominada por la observación de datos que ocurren naturalmente» y el corpus es un medio que posibilita la compilación de estos datos. En esta perspectiva, lo que caracteriza el paradigma empiricista en estudios del lenguaje es la naturaleza, el origen de los datos utilizados en las pesquisas.

De acuerdo con Berber Sardinha, 2004:

> A Lingüística de Corpus se ocupa da coleta de dados e exploração de corpora, ou conjuntos de dados lingüísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio do computador (Berber Sardinha, 2004: 3).

Son muchos los estudiosos que presentan definiciones sobre corpus. Según Sinclair «un *corpus* es una colección de textos naturales elegidos para caracterizar un estado o variedad de la lengua» (Sinclair, 1991: 171).

Para Sánchez 1995, «un *corpus* es un conjunto de datos lingüísticos que pueden pertenecer al uso oral o escrito de la lengua, sistematizados según determinados criterios extensos en amplitud y profundidad, de manera que sean representativos en la totalidad del uso lingüístico o de alguno de sus ámbitos, dispuestos de tal modo que puedan ser procesados por un ordenador, con la finalidad de propiciar resultados varios y útiles para la descripción y análisis» (Sánchez, 1995: 8-9).

La lingüística de corpus, por lo tanto, es una área que se basa en una visión probabilística del lenguaje y se dedica a la exploración de la lengua a través de evidencias empíricas extraídas por medio del ordenador (Berber Sardinha, 2000).

Para que se haga la exploración de la lengua son necesarios algunos criterios para la formación del corpus: I) el corpus debe estar compuesto por textos auténticos en lengua natural, II) el contenido debe ser seleccionado a fin de garantizar que el corpus tenga sus características propias y III) el corpus debe ser representativo.

Refiriéndose a ese tercer ítem, la representatividad, tenemos la cuestión de la extensión del *corpus*. De acuerdo con Berber Sardinha (2000: 344), la representatividad de un *corpus* depende da las formas y del sentido:

> A freqüência das formas em si não é suficiente porque mesmo palavras de alta freqüência possuem vários sentidos. Assim, uma freqüência alta pode «esconder» vários sentidos, os quais separados teriam baixa freqüência. Para que seja representativo, um corpus deve conter o maior número possível de sentidos de cada forma.

De acuerdo con Sinclair, 1991, y Halliday, 1991 y 1992, la lengua es un sistema probabilístico en que determinados trazos lingüísticos como los lexicales, estructurales, pragmáticos o discursivos son más frecuentes que otros. Por eso,

para que haya representatividad, un corpus debe ser lo más grande posible, pues con un *corpus* de gran extensión, mayor será la probabilidad de ocurrencias de los rasgos lingüísticos que se quieran investigar.

En otras palabras, la LC presupone que los trazos lingüísticos no ocurren con la misma frecuencia y que su variación no ocurre de forma aleatoria (Biber *et al,* 1998). Esta noción nos remite al hecho de que el hablante de una determinada lengua tiene a disposición una gran cantidad de expresiones que pueden asociarse dentro de rasgos lingüísticos específicos.

2.2 Uso de concordancias

La utilización de los corpora en la enseñanza de LEs tiene relación con el Aprendizaje Movido a Datos *(Data-Driven Learning),* creado por Johns 1986 para la enseñanza de la gramática inglesa a partir del análisis de concordancias.

Según Berber Sardinha, 2000, el aprendizaje movido a datos proporciona tres ventajas a los alumnos: I) el desarrollo de la habilidad de identificar regularidades y hacer generalizaciones para explicarlas; II) el profesor deja de ser la figura central y pasa a ser el coordinador de la investigación; III) la enseñanza de la gramática no se fundamenta en la transmisión de reglas sino en el descubrimiento del propio alumno.

El aprendizaje movido a datos ubica al aprendiz en el papel de descubridor e investigador en el proceso de manejo del corpus y estimula al alumno el desarrollo de estrategias de aprendizaje. Este tipo de actividad le permite al alumno que encuentre sus propias respuestas a través de la observación y análisis de las concordancias obtenidas por medio del ordenador.

Una concordancia[1] puede ser obtenida usando un programa concordanceador, como el Microconcord (Tim Johns, 1986), disponible gratuitamente en la web[2] y el WordSmith (Mike Scott, 1998).

El programa operacional WordSmith, por ejemplo, dispone de herramientas para analizar las construcciones seleccionadas, permite manipular datos, posibilita visualizarlos, indica la frecuencia de las palabras, las concordancias, entre otros recursos útiles en el trabajo con base en corpus.

Es posible, a través del WordSmith, seleccionar todos los archivos disponibles para el análisis de una sola vez y dar el comando de busca de la palabra en análisis. El programa presenta la cantidad de datos encontrados en el total de los archivos analizados, produciendo listas de oraciones, generalmente incompletas, cuyos

---

1 Lista de oraciones de una determinada palabra presentada en el contexto de un corpus.

2 Disponible en http://www.lexically.net/downloads/_freebies/mconcord.zip.

párrafos pueden ser ampliados ofreciendo informaciones sobre el tipo de palabra que acompaña la palabra central y evidencia determinadas combinaciones.

Es un recurso eficiente para aislar palabras y expresiones lingüísticas de un corpus y posibilita el análisis de la lengua en su uso auténtico.

La importancia lingüística de tales concordancias es observada a partir de un modelo compuesto por variables a fin de probar los factores que influyen en las elecciones lingüísticas. Según Sankoff (1988: 984) estas elecciones pueden ser influidas por distintos factores tales como los rasgos fonológicos, el contexto sintáctico, la función discursiva, el estilo, la situación interactiva, entre otros factores.

Estas concordancias pueden ser utilizadas en la enseñanza de la lengua de dos formas. En una de ellas el alumno tiene acceso al ordenador, al corpus y al programa de concordancias para contestar a las preguntas y hacer exámenes específicos. En la otra, el profesor es el mediador, él edita las concordancias y las transforma en actividades didácticas para las clases de lengua extranjera.

3. METODOLOGÍA

3.1 El escenario de la investigación

El trabajo con los conectores *pero* y *sino* a través de la utilización de la lingüística de corpus como herramienta de análisis se desarrolla con un grupo de aprendices de español de nivel intermedio del curso extracurricular de la Universidade Federal de Santa Catarina (UFSC). Seleccionamos este grupo porque fuimos su profesora de lengua española durante el año de 2005.

El grupo de alumnos que participa de este estudio suma un total de ocho aprendices de diferentes carreras y también de antiguos alumnos de la UFSC. Este trabajo se inició durante el mes de setiembre de 2005 con los primeros contactos en clase.

El material didáctico utilizado con este grupo, de autoría de Maria Cabral Bruno y de Maria Angélica Mendoza, es el libro *Hacia el español, curso de lengua y cultura hispánica* (1998), publicado por la editorial Saraiva.

La dificultad presentada con el uso de los elementos *pero* en contraste con *sino,* llevaron a la utilización de textos periodísticos con el objetivo de hacer un análisis de los usos concretos de estas conjunciones.

3.2 Selección y análisis del corpora

Los datos utilizados en este trabajo son de textos periodísticos seleccionados entre los meses de enero a junio de 2005 y posee una extensión aproximada de 2983 palabras.

El criterio elegido para la elección fue, primero, la disponibilidad en medio electrónico del material recolectado. El *corpus* utilizado en este trabajo está compuesto por periódicos de veinte países hispánicos a fin de garantizar la representatividad de este tipo de texto. Después de la recolección de textos de la prensa escrita electrónica, ese material fue organizado en archivos específicos. Presentamos, abajo, los periódicos y sus respectivos países.

Argentina: *La Nación, Clarín, Diario Los Andes, El Día*.

Bolivia: *La Razón, Los Tiempos*.

Chile: *El Mercurio, Diario Crónica, El Diario Austral*.

Colombia: *El Mundo, El Tiempo, El Espectador*.

Costa Rica: *Prensa Libre*.

Cuba: *La Nueva Cuba, Digital Granma*.

El Salvador: *El Diario de Hoy, La Prensa Gráfica*.

España: *ABC, El Mundo, El Periódico, El Correo Digital*.

Ecuador: *El Universo, Diario de Hoy, El Comercio*.

Guatemala: *Prensa Libre, La Hora*.

Honduras: *Tiempo, La Prensa, La Tribuna*.

México: *La Crónica, El Imparcial, El Universal*.

Nicaragua: *La Prensa*.

Panamá: *Critica en Línea, El Siglo, La Prensa*.

Paraguay: *Última Hora, La Nación, ABC*.

Perú: *La República, El Peruano, Expreso*.

Puerto Rico: *El Nuevo Día, Univisión*.

Republica Dominicana: *El Caribe*.

Uruguay: *El Observador, Últimas Noticias*.

Venezuela: *El Universal, El Carabobeño, La Hora*.

Con la ayuda del programa computacional para análisis de corpora WordSmith (Mike Scout, 1998), disponemos de herramientas para examinar los usos de las conjunciones en las construcciones seleccionadas. El programa permite manipular datos, visualizar las sentencias, indicar la frecuencia de las palabras, así como realiza otras tareas útiles para el lingüista.

4. LA ACTIVIDAD DIDÁCTICA

La actividad sugerida que mostramos a continuación fue desarrollada con un grupo de estudiantes de nivel intermedio de la lengua española. El interés que impulsiona este trabajo surge de la comprobación de que los aprendices presentan dificultades en la utilización de los conectores *pero* y *sino*.

A partir del manual *Hacia el español* (1998), nivel avanzado, fueron desarrolladas actividades con el uso de algunos operadores como recurso para la cohesión textual, tales como: *aunque, aun + gerundio, pero, sino, sin embargo* y *mientras*.

De acuerdo con este manual los nexos *sino* y *pero* tienen la siguiente función:

> Pero: Introduce una oposición que limita la idea anteriormente expresada, le añade un nuevo aspecto, sin excluirla. Pero tiene como sinónimo el mas, pero éste, de uso arcaico.
>
> Sino: En general se usa tras una negación para sustituir el elemento negado por otro de la misma categoría. Su diferencia en relación al pero es justamente la exclusión –y no la limitación o restricción– de lo dicho anteriormente (Bruno et al., 1998: 23- 4).

Tras realizar el estudio del contenido planteado por el libro didáctico, propusimos el análisis de una selección de oraciones tomadas de textos periodísticos.

Después de verificar que los estudiantes tuvieron problemas con usos de las conjunciones *sino* y *pero,* decidimos ampliar estos análisis. En un primer momento, presentamos y discutimos los análisis de los alumnos. Estas discusiones se refieren al conocimiento y las suposiciones con respecto a los usos de las conjunciones.

En un segundo momento, realizamos el análisis con la ayuda de los ordenadores en el laboratorio de lenguas de la institución, mediante la utilización de la herramienta WordSmith, permitiendo la selección, ordenamiento y cálculo de los textos almacenados electrónicamente.

Durante la realización de esta tarea los alumnos se comunicaban en español y cambiaban informaciones con sus compañeros probando sus suposiciones con respecto de los usos de las conjunciones.

El papel del profesor en este momento es de fundamental importancia. Dependiendo del grado de conocimiento de los alumnos y de su experiencia, la mayoría de las veces, logran realizar las tareas de forma autónoma, conduciendo sus análisis a partir de las instrucciones del profesor.

Los estudiantes descubren al sistematizar los usos de *pero* y *sino* a partir de la observación y análisis de concordancias una idea de cómo estos nexos son utilizados produciendo coherencia en los textos.

Veamos, a título de ilustración, algunos ejemplos de concordancias extraídas del corpus con *pero* y *sino:*

N	Concordance
1.	el 5 al 14 de agosto en Argentina. **Pero** para muchos conocedores del bal
2.	taba entre la Calle 50 y la Vía España, **pero** un perdigonazo la bajó. Un calor
3.	o conoce a alguien que desee hacerlo, **pero** no sabe dónde ir, visite la página
4.	n base a un inexistente apoyo político. ¿**Pero** es que ya no existe gente honra
5.	a ni la vuelta atrás a la vida que amaba. **Pero** sí ama a su hija Gertie (Raquel C
6.	ó a vulnerar la autonomía universitaria, **pero** les urgió a resolver responsablem
7.	creo que van a ser juegos bien difíciles, **pero** vamos con bastante preparación y

N	Concordance
1.	ente a 1492, pero no para conquistar, **sino** para confiscar la tierra no ya pro
2.	nos vamos a unir a los magistrados, **sino** que se garantice lo que preceptú
3.	sólo lo referido a la ley de narcotráfico, **sino** también la forma en la cual se d
4.	históricos, van a salir, no en grupos, **sino** en bandadas. Porque ninguno d
5.	Al llegar a la escuela no hacemos **sino** continuar obedientes al grito de l
6.	exigen a sí mismos como jugadores **sino** que la fanaticada les exige gana
7.	no a través de referendos nacionales, **sino** de un único referendo europeo c

Por su disposición, las concordancias proveen información sobre el tipo de palabras que acompañan la palabra central y evidencian la recurrencia de determinadas combinaciones, o fraseologías dando la posibilidad de distinguir las funciones en cada caso de uso y facilitando la comparación entre estos usos en los contextos situacionales.

5. CONSIDERACIONES FINALES

El crecimiento del área de la lingüística de corpus es evidente en los últimos años y crece cada vez más en la medida en que más investigadores descubren en el corpus la herramienta que posibilita el trabajo práctico con el lenguaje y la fuente de informaciones sobre ella, posibilitando esclarecimientos sobre la lengua

de una forma única (Fillmore, 1992, ápud Berber Sardinha, 2000), de que ningún otro modo lo habría descubierto.

La realización de este trabajo fue motivada por la necesidad de poder ofrecer a mis alumnos de lengua española una alternativa diferente para el aprendizaje de determinados fenómenos lingüísticos.

La disponibilidad de ejemplos de uso auténtico del lenguaje y la oportunidad de exponer a los aprendices de LEs a las distintas situaciones reales de comunicación son algunas de las posibilidades que ofrece la LC.

Este tipo de análisis puede utilizarse para la elaboración de otro tipo de actividades que consisten en la observación y análisis de un gran número de ejemplos reales de la lengua que pueden ser extraídos de diversas maneras, de textos periodísticos, disponibles en la Internet; de un gran número de autores de varias lenguas, disponible en la Red en direcciones como el Proyecto Gutenberg[3]; de hablantes nativos en situaciones auténticas de comunicación; de aprendices de la lengua española; de textos literarios; entre otros.

El trabajo propuesto con uso de corpora en la enseñanza de lengua es relevante, pues pone de relieve las dificultades específicas manifestadas por los estudiantes de un determinado contenido, permitiendo la selección de temas y ampliando las posibilidades de uso de la lengua. El uso del *corpus* también se muestra eficaz ampliando las informaciones que muchas veces se revelan insuficientes en los diccionarios y gramáticas.

Otros aspectos que logramos con este tipo de trabajo se refiere a la estimulación de la reflexión sobre la lengua, a la resolución de problemas y al intercambio de experiencias.

Este enfoque de enseñanza basado en la utilización de corpora propicia la concienciación lingüística y la autonomía, permitiendo el contacto con la lengua en uso y creando naturalmente un ambiente favorable a la enseñanza comunicativa de lengua extranjera y cambiando los papeles tradicionalmente asumidos por profesor y alumnos en clase.

BIBLIOGRAFÍA

Balbás, Marcial Soto, 2003, *Análise de erros, baseada na Lingüística de Corpus, da escrita de aprendizes brasileiros universitários de Espanhol como Língua Estrangeira,* LAEL, Pontifícia Universidade Católica de São Paulo, Dissertação de Mestrado.

Berber Sardinha, A.P., 2000, «Lingüística de Corpus: Histórico e Problemática», *Delta,* 16/2: 323-367.

[3] Disponible en http://promo.net/pg/.

— 2004, *Lingüística de Corpus*. São Paulo, Manole.
Biber, Douglas; Conrad, Susan y Reppen, Randi, 1998, *Corpus Linguistics: investigating language structure and use*, New York, Cambridge University Press.
Bruno, Fátima Cabral y Mendoza, María Angélica, 1998, *Hacia el Español. Curso de lengua y cultura hispánica,* São Paulo, Editora Saraiva.
Conde, Helena Maria de Andrade, 2002, *Escolhas léxico-gramaticais em composições de alunos avançados de inglês originários de instituições de ensino bilíngües e monolíngües - um estudo multidimensional baseado em corpus,* LAEL, Pontifícia Universidade Católica de São Paulo, Dissertação de Mestrado.
Halliday, M.A.K., 1991, «Corpus studies and probabilistic grammar», en: K. Aijmer & B. Altenberg (Org.), *English Corpus Linguistics: Studies in honour of Jan Svartvik,* London, Longman. pp. 30-43.
— 1992, «Language as system and language as instance: The corpus as a theoretical construct», en: Svartvik, J., (ed.), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82,* Berlim, De Gruyter, pp. 61-78.
Jacobi, Claudia, 2002, «Computadores, corpora y la enseñanza de español en cursos de letras», *Anuario brasileño de estudios hispánicos,* 12: 29-44.
— 2001*, Lingüística de Corpus e ensino de espanhol a brasileiros: descrição de padrões e preparação de atividades didáticas (decir/hablar; mismo; mientras/ en cuanto/aunque),* LAEL, Pontifícia Universidade Católica de São Paulo, Dissertação de Mestrado.
Johns, T., 1986, «Microconcord: A language learner's research tool», *System:* 14 (2): 151-62.
Mcenery, T., & Wilson, A., 1996, *Corpus Linguistics,* Edinburgh*,* Edinburgh University Press.
Oliveira Souza, Cláudia Maria de, 2003, *Adverbiais conjuntivos na escrita de alunos brasileiros de inglês como língua estrangeira: uma pesquisa baseada em corpora*, Lael, Pontifícia Universidade Católica de São Paulo, Dissertação de Mestrado.
Sánchez, A., 1995, «Definición e historia de los corpus», en: Sánchez, A., *et al*. (ed.), *CUMBRE - Corpus linguístico del español contemporáneo,* Madrid, SGEL.
Sankoff, David, 1988, «Variable rules». en: Ammon, U.; Dttmar, N., y Matteir, K.J., (eds.) *Sociolinguistics: an international handbook of the science of language and society,* New York, Walter de Gruyter, pp. 984-998.
Scott, Mike & Oxford University Press, 1998, *WordSmith tools manual version 30*.
Sinclair, J., 1991, *Corpus, Concordance, Collocatio,* Oxford, Oxford University Press.