

## **La enseñanza y el aprendizaje de los conectores contraargumentativos en español como lengua extranjera. El uso del CORPUS de alumno**

### **Learning and teaching of connectors of contrargumentation in the Spanish language. The use of the student corpus**

*Weronika Górska*

Instytut Językoznawstwa, Uniwersytet im. Adama Mickiewicza  
ul. Międzychodzka 5, 60-371 Poznań

weragorska@poczta.onet.pl

#### **Abstract**

The present article is a part of M.A. thesis completed and defended in the Institute of Linguistics at Adam Mickiewicz University in Poznań (June 2006). The main focus of the work is the presentation of the application of analysis of the student corpus in the examination of the mastering of the connectors of contrargumentation. The base for the work is being formed by a corpus of 300 pages. The material has been created by the author who would get information via forum La ruta de la lengua Española, where information are being transferred between Spanish people – future teachers of their native language, and the students of Spanish from abroad. The text contains introduction to the linguistics of the corpus, methodology of its creation as well as analysis and comparison to the standard corpus of the Spanish language.

#### **Streszczenie po polsku**

Niniejszy artykuł jest częścią pracy magisterskiej napisanej pod kierunkiem dr Agaty Wolarskiej w Instytucie Językoznawstwa i obronionej w roku 2006. Praca ta powstała dzięki badaniom, które przeprowadziłam w ramach programu SOCRATES-ERASMUS na Universitat de Barcelona od maja do września 2005 roku.

Zbiegają się w niej różne zagadnienia z takich dziedzin jak pragmatyka (analiza dyskursu), lingwistyka komputerowa (językoznawstwo korpusowe) oraz nowe technologie w nauczaniu języków obcych (przede wszystkim wykorzystanie wiadomości elektronicznych przesyłanych przez forum).

Celem pracy było sprawdzenie hipotezy czy można się nauczyć łączników kontrargumentacji za pomocą inputu, który uczniowie (z poziomu średniozaawansowanego lub zaawansowanego) otrzymują dzięki wymianom maili z hiszpańskimi studentami. Bazę badań stanowił CORPUS<sup>1</sup> utworzony z maili które, wymieniali między sobą uczniowie języka hiszpańskiego jako obcego z Wiednia i z Tel Awiwu oraz studenci przedmiotu przygotowującego do nauczania tego języka z Universitat de Barcelona. CORPUS obejmujący około 300 stron został następnie poddany analizie leksykometrycznej w programie

---

<sup>1</sup> Nazwa ta używana jest tutaj na określenie korpusu stworzonego przez autorkę.

HYPERBASE. Badania zostały przeprowadzone dzięki uprzejmości wykładowców z Universitat de Barcelona.

Badania wykazały, iż uczniowie języka hiszpańskiego używają przede wszystkim podstawowego łącznika PERO (ale) w odróżnieniu od rodzimych użytkowników tego języka, w których wiadomościach znaleziono występujący z dużą frekwencją konektor AUNQUE (mimo że) oraz inne konektory. Wykazano także, iż nie wzrasta użycie łączników bardziej zaawansowanych i trudniejszych pod względem zarówno gramatycznym, jak i logicznym. Niniejszy artykuł opiera się na części pracy dotyczącej wykorzystania korpusu uczniowskiego w badaniach z zakresu dydaktyki języka hiszpańskiego.<sup>2</sup>

Lingwistyka korpusu jest dyscypliną stosunkowo nową, polem w którym językoznawca wraz z informatykiem łączy swoje wysiłki w celu lepszego zrozumienia ludzkiego języka. Początki jej sięgają roku 1967, w którym to H. Kucera wraz z N. Francis opublikowali *Computational Analysis of Present-Day American English*. Pierwszy korpus był kompilacją języka amerykańskiego, zawierającą około miliona słów pochodzących z różnych źródeł. Dzięki rozwojowi tej dziedziny możliwe są obecnie badania językoznawcze, literaturoznawcze (np. analizujące styl określonego autora), a także badania z zakresu traduktologii, socjologii czy politologii. Glottodydaktyka jest kolejną dziedziną, która korzysta z osiągnięć lingwistyki korpusu. W Internecie są dostępne kompilacje tekstów literackich, słowniki, programy do tworzenia ćwiczeń, encyklopedie i wiele innych pożytecznych pomocy dydaktycznych powstałych dzięki badaniom opartym na korpusach.

Aktualnie za korpus uznaje się zbiór wielu tekstów autentycznych zapisany w formie elektronicznej. Teksty te wybiera się według różnych kryteriów, ale zawsze biorąc pod uwagę ich reprezentatywność. Najważniejszą cechą korpusu jest jego forma elektroniczna, która pozwala na późniejszą jego obróbkę za pomocą różnego rodzaju programów komputerowych. Dzięki formatowi elektronicznemu wyciągnięcie wniosków statystycznych wymagające przeszukania olbrzymiej ilości danych staje się stosunkowo proste. Istnieje wiele typów korpusów. CORPUS jest po części korpusem uczniowskim – stosunkowo mało znanym i niewiele jeszcze używanym w dydaktyce języków obcych. Według moich informacji nie istnieją dostępne korpusy uczniowskie w języku hiszpańskim. CORPUS stanowi próbę wypełnienia tej luki.

CORPUS jest zbiorem tekstów pisanych (wiadomości elektronicznych), jest jednojęzyczny, synchroniczny (wiadomości wysyłane były od lutego do czerwca 2005 roku) i oznaczony (wprowadzono 2 rodzaje znaków, &&&1&&& dla oznaczenia części CORPUSU i \$ dla oznaczenia każdego paragrafu). Jeśli chodzi o ograniczenia geograficzne – CORPUS zawiera przede wszystkim wiadomości pisane przez mieszkańców Izraela, Austrii i Hiszpanii. Pojawiają się także wiadomości napisane przez Polaków, Rosjan i Amerykanów.

CORPUS zawiera różne odmiany języka hiszpańskiego oraz wiele próbek języka uczniowskiego, który w mniej lub bardziej zbliżonym stopniu przypomina język standardowy. W CORPUSIE zastosowano nowatorski system oznaczania paragrafów pozwalający na natychmiastowe odnalezienie nadawcy oraz daty wysłania wiadomości.

Korpus można stworzyć w różnorodny sposób: przepisując tekst wydrukowany, skanując go, kopiując z dyskiety (CD-ROMU itp.), lub wykorzystując teksty znajdujące się w Internecie. Tworząc CORPUS posłużono się właśnie tą ostatnią metodą.

CORPUS na CD-ROM dostępny jest w *Bibliotece Instytutu Językoznawstwa*. W niniejszym artykule prezentuje się szczegółowo metodologię oraz analizę CORPUSU, a także porównanie uzyskanych wyników z oficjalnym korpusem języka hiszpańskiego CREA.

## 1 Introducción

El presente artículo forma parte de la tesina de licenciatura defendida en 2006 en la Universidad de Adam Mickiewicz en Poznań. Dicha tesina es fruto de una investigación llevada a cabo en la Universitat de Barcelona.

Su objetivo era comprobar cómo los nativos y los estudiantes de español como lengua extranjera (más adelante llamado ELE) usan los conectores contraargumentativos en los mensajes electrónicos que intercambiaron entre sí, y además, cómo se enseña estas piezas en los manuales.

---

<sup>2</sup> W dalszej części pracy sprawdzam jak przedstawiane są konektory kontrargumentacji w 10 podręcznikach do nauki języka hiszpańskiego oraz proponuję plan lekcji wyjaśniającej różnicę pomiędzy PERO (ale) i AUNQUE (pomimo że).

El interés por los conectores contraargumentativos se despertó durante la estancia en Barcelona cuando la autora observó que muchos estudiantes de ELE no sabían usarlos bien y no entendían las diferencias existentes entre ellos. Ahora bien, éstos son los elementos imprescindibles en la conversación, y además, como dice J. Portolés, han sido pobremente descritas en las gramáticas y los diccionarios de español. Por esto, los profesores del ELE muy a menudo sólo pueden contar con su propia intuición lingüística a la hora de explicar el uso de los conectores contraargumentativos a sus alumnos.

La aspiración de la tesina fue comprobar si los estudiantes de ELE pueden entender la noción de los conectores contraargumentativos implícitamente, mediante el intercambio de correo electrónico con los nativos. Buscaba la respuesta a la pregunta si los estudiantes usaban los conectores más difíciles lógicamente y gramaticalmente que el muy popular PERO y si el uso de PERO y AUNQUE, como conectores prototípicos, cambiaba con el tiempo.

Dadas las características del medio, se suponía que aparecerían los conectores contraargumentativos usados también en la lengua oral y se intentó comprobar si su uso por parte de los nativos tenía como consecuencia el enriquecimiento del vocabulario de otra variedad que el estándar por parte de los estudiantes de ELE.

Este trabajo abarca tres campos de lingüística: la pragmática de la que se eligieron como objeto de investigación los conectores contraargumentativos; la lingüística computacional cuyas herramientas sirvieron para realizar el análisis lexicométrico (en este caso mediante el programa HYPERBASE) y la didáctica de las lenguas extranjeras con su corriente más reciente: el uso de las nuevas tecnologías en la enseñanza de ELE. De este último campo elegí como fuente de materiales de análisis los mensajes electrónicos que intercambiaron durante medio año los estudiantes de ELE de Viena y de Tel Aviv con los estudiantes de la asignatura llamada: Cuestiones de ELE (una de las asignaturas optativas para los estudiantes de Hispánicas en la Universitat de Barcelona), mediante un foro de una página Web titulada: *La ruta de la lengua española* (más adelante llamado FORO). Este intercambio se realizó en el marco del programa coordinado por M. Cruz Piñol, R. Sitman, V. Berger (la parte lingüística) y el creador de la página Web en la que se halla el FORO – J. García Garrido. En los años 2005–2006 empezó la quinta edición de este proyecto.

Los mensajes enviados por 35 tutores del año 2005 constituyeron la base del CORPUS<sup>3</sup> que a continuación fue tratado con el programa HYPERBASE, con el objetivo de llevar a cabo el análisis lexicométrico.

El presente artículo se refiere sólo a esta parte de la tesina que habla del uso de corpus de alumno en las investigaciones en el campo de ELE.<sup>4</sup>

En cuanto al sistema de citación, dado el gran número de recursos en línea, se indica siempre el número de cada página Web citada que ésta tiene en la bibliografía. Así, se pone este número al lado de la Web indicada, por ejemplo: “disponible en: <[www.rae.es](http://www.rae.es)> (número).”

## 2 La lingüística computacional

La lingüística computacional es una disciplina nueva en la que los lingüistas y los informáticos unen sus esfuerzos para estudiar el lenguaje humano. Hay muchas áreas de las que se ocupa esta disciplina, por ejemplo: traducción automática, diseño de analizadores sintácticos, de etiquetadores o lematizadores y también diseño y aprovechamiento lingüístico de corpus.

La rama de la lingüística computacional que se ocupa de los corpus se llama lingüística de corpus. Iniciada en 1967 por dos científicos (H. Kucera y N. Francis<sup>5</sup>) con la publicación de *Computational Analysis of Present-Day American English*, hoy en día se utiliza mucho en diferentes disciplinas, enseñanza de las lenguas entre otras. El primer corpus fue

<sup>3</sup> Así se va a llamar en este trabajo el corpus creado por la autora.

<sup>4</sup> Los demás tratados en mi tesina son: el análisis de los manuales (en cuanto al uso de los conectores contraargumentativos) y una propuesta de clase sobre PERO y AUNQUE como los conectores introductorios de los argumentos fuertes y débiles.

<sup>5</sup> Kucera, Francis (1967).

una compilación de inglés americano de aproximadamente un millón de palabras seleccionadas de una amplia variedad de fuentes.

Un método de análisis de texto que estudia la frecuencia, repetición y distribución de las palabras que lo integran se llama la lexicometría. Es la estadística textual que permite establecer las relaciones existentes entre las unidades léxicas.

## **2.1 La definición de corpus**

Un corpus lingüístico es un conjunto, normalmente muy amplio, de ejemplos reales de uso de una lengua. Es cualquier conjunto de textos agrupados y clasificados según determinarios criterios (tamaño, contenido, tipo de ordenación) que se halla en soporte informático. J. Torruella y J. Llisterri lo definen como: (...) *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.*<sup>6</sup>

Un corpus es sólo una muestra de la lengua. Sin embargo, debería ser una muestra representativa para poder sacar las conclusiones generales sobre el uso del lenguaje. Como dicen los autores de *Diseño de corpus textuales y orales: Un corpus siempre está construido a base de muestras con la intención de que de su observación se puedan extraer generalizaciones sobre la lengua (...). El corpus tiene que mostrar a pequeña escala cómo funciona una lengua natural.*<sup>7</sup>

Los autores describen en su artículo la manera de crear un corpus para que sea una muestra representativa.

Gracias al hecho de estar en soporte electrónico, el corpus permite extraer conclusiones estadísticas y realizar consultas automáticas. Hay que darse cuenta que el ordenador permite el tratamiento de cantidades de texto que un hombre nunca sería capaz de realizar.

J. Waliński<sup>8</sup> cita a J. Sinclair y a W. McEnery, quienes establecen rasgos característicos que debería cumplir un conjunto de textos para ser llamado corpus. Según el primero son: la cantidad de textos, la calidad, la simplicidad (en el formato ASCII, ahora más bien se usa Unicode) y la documentación. El segundo habla de la representatividad, el tamaño cerrado, el formato electrónico y la correspondencia con el estándar que tiene mucho que ver con dos primeros puntos.

### **2.1.1 Tipos de corpus**

En cuanto a la tipología de los corpus se puede dividirlos en muchos tipos según diferentes criterios. Por ejemplo:

- basados en textos escritos/hablados; hay dos tipos de corpus hablado: los que se conservan en el formato sonoro y aquellos en los que se ha transcrito una muestra oral, llamados también textuales;
- corpus monolingües/multilingües/pararelos;
- corpus sincrónicos/diacrónicos;
- corpus simples/codificados (por ejemplo etiquetado con diferentes categorías gramaticales);
- corpus generales/especializados (corpus de alumno constituye una subcategoría de éste).

Hay más clasificaciones de corpus, lo cual se debe a la multitud de criterios, según los que se las puede establecer. Se podría diferenciar los corpus según:<sup>9</sup> el porcentaje y la distribución de los diferentes tipos de texto; la especificidad de los textos; la cantidad de texto que se recoge en cada documento; la codificación y la anotación.

---

<sup>6</sup> Torruella, Llisterri (1999: 51).

<sup>7</sup> *Ibidem* (63 y 43).

<sup>8</sup> Waliński (2005: 27-28).

<sup>9</sup> Véase más detalles en Torruella, Llisterri (op.cit.: 53-56) y en Waliński (op.cit.27).

## 2.1.2 Ejemplos de corpus en español<sup>10</sup>

Hay que mencionar que los corpus orales tienen sus propios criterios por tener el origen diferente. Sin embargo, se renuncia a su descripción en el presente trabajo, dado el hecho de que el CORPUS es un corpus basado en textos escritos.

Como un corpus lingüístico puede emplearse cualquier texto o grabación en formato electrónico. Por ejemplo los diarios en línea, la antología de textos, los repertorios de libros en soporte electrónico o diferentes repertorios de textos.<sup>11</sup>

Ejemplos de corpus existentes de español disponibles en la red:

- CREA y CORDE<sup>12</sup>
- SOL<sup>13</sup>
- Corpus del Español<sup>14</sup>
- The Spanish FrameNet Project<sup>15</sup>
- Base de Datos Sintácticos del Español Actual (BDS)<sup>16</sup>
- VARILEX<sup>17</sup>
- AGLÉ<sup>18</sup>
- Corpus oral<sup>19</sup>

## 2.2 Maneras de crear y tratar el corpus

Hay cuatro maneras de obtener un corpus: manual (desde el teclado), utilizando el escáner y la tecnología OCR, desde el CD-ROM y buscando los textos que ya están en Internet.

En cuanto a la manera de tratar el corpus, todo depende de lo que se quiere obtener. Ya los programas tipo Word teóricamente permiten extraer por ejemplo las apariciones de una palabra concreta (mediante la orden *buscar*). S. Sotelo Docío y M<sup>a</sup> N. Vázquez González subrayan la utilidad de las bases de datos y hojas de cálculo para llevar a cabo una recopilación de datos ordenada, su manipulación, la extracción de una serie de conclusiones y su presentación.<sup>20</sup> Claro está que éstos no son unos programas preparados para trabajar con el corpus. Entre los diferentes programas creados para la manipulación de textos en formato electrónico cabe destacar TACT, WordSmith, WordCruncher, Lexa. Los otros sobre los que se puede consultar el Internet son por ejemplo: Leximappe, Unitex, AntConc, Morfix-NLP, Natural Language Toolkit. Muchos de ellos gratuitos y compatibles también con el sistema operativo Linux.<sup>21</sup>

En el presente trabajo se utilizó el programa HYPERBASE, creado inicialmente para la lengua francesa.

Los diferentes programas varían entre sí. Algunos menos sofisticados permiten solamente la extracción de las concordancias, las frecuencias, etc. Otros permiten la etiquetación de las palabras, lo que añade la información explícita de diferente tipo: morfosintáctica, fonética, semántica. Esto permite afinar las búsquedas posteriores.

<sup>10</sup> En Lewandowska-Tomaszczyk (2005) se hallan los ejemplos de los corpus en diferentes lenguas.

<sup>11</sup> Unos ejemplos concretos: la antología sobre *La conquista de América* disponible en: <<http://www.aselered.org/>> (7) o el repertorio de libros en soporte electrónico *Biblioteca Virtual Miguel de Cervantes*, disponible en: <<http://www.cervantes.es/>> (32).

<sup>12</sup> Disponible en: <<http://corpus.rae.es/cordenet.html>> (4) y <<http://corpus.rae.es/creanet.html>> (5).

<sup>13</sup> Disponible en: <<http://spraakbanken.gu.se/lb/konk/rom2/>> (14).

<sup>14</sup> Disponible en: <<http://www.corpusdelespanol.org>> (15).

<sup>15</sup> Disponible en: <<http://gemini.uab.es/SFN/>> (16).

<sup>16</sup> Disponible en: <<http://www.bds.usc.es/>> (17).

<sup>17</sup> Disponible en: <<http://www.lenguaje.com/glosario/glosario.php>> (18).

<sup>18</sup> Disponible en: <<http://cvc.cervantes.es/obref/agle/>> (12).

<sup>19</sup> Disponible en: <[http://www.cervantes.es/seg\\_nivel/lect\\_ens/oesi/coral\\_01.htm](http://www.cervantes.es/seg_nivel/lect_ens/oesi/coral_01.htm)> (13).

<sup>20</sup> Véase Sotelo Docío y Vázquez González (1998).

<sup>21</sup> Más informaciones sobre estos programas en <<http://www.smess.egss.ulg.ac.be/lejeune/logiciels/opencqdas.html>> (8).

### 2.3 El corpus de alumno

El corpus de este tipo está basado en una recopilación de los textos escritos o dichos por los alumnos de segunda lengua. En la lingüística aplicada constituyen una novedad. La mayoría de los corpus existentes de este tipo se creó hace pocos años.

Para crear un corpus de alumno hay que seguir unos criterios que Agnieszka Leńko-Szymańska<sup>22</sup> presenta en un cuadro. Son los criterios conectados con: la lengua – el tipo de texto, la forma de texto, el estilo, el registro, el tema; la tarea – el tipo, el tiempo para realizarla, el acceso a los materiales complementarios (diccionario, etc.); el alumnado – la edad, el sexo, la formación, la lengua materna; el contexto de aprendizaje de la lengua extranjera, el nivel, la experiencia en el uso de la lengua extranjera, el conocimiento de otras lenguas extranjeras.

El corpus de alumno tiene dos aplicaciones primordiales: sirve como material para el análisis de errores para los científicos o como texto – base para la preparación de los ejercicios.

El análisis del corpus de alumno puede también servir para mejorar su estilo. La autora cita un programa, creado por P. Nation de Nueva Zelanda, que permite establecer *lexical frequency profile*. Este indicador muestra la variedad de las palabras usadas por el estudiante (si usa más o menos complicadas, la riqueza de vocabulario, etc.).

Un empleo del corpus de alumno que seguramente se va a extender es éste de la adquisición de segundas lenguas. Como dice la autora, dados los pocos corpus existentes de este tipo: *no son todavía utilizados con frecuencia en las investigaciones sobre la adquisición de la segunda lengua*.<sup>23</sup>

Con el corpus de alumno hay un problema técnico. A saber, todo el trabajo de marcar el texto (introducir algún tipo de etiqueta, por ejemplo de partes de habla) tiene que ser realizado manualmente. Existen los editores de errores que ayudan un poco en este penoso trabajo; no obstante, es el hombre, no una máquina, quien tiene que vigilar todo el proceso.

No hay muchos ejemplos de este tipo de corpus. La autora presenta 7 proyectos: todos ingleses, excepto uno francés. Por ejemplo: International Corpus of Learner English (ICLE), Hong Kong University of Science and Technology Learner Corpus (HKUST), PELCRA Learner Corpus (creado por la Universidad de Łódź), Cambridge Learner Corpus (CLC) y Longman Learners' Corpus (LLC). La mayoría de estos corpus no está al alcance del público. Solamente ICLE es una base de datos a la que se puede parcialmente acceder desde Internet o compararla en el CD (junto con el buscador de concordancias).

Según los datos de A. Leńko-Szymańska no existen los corpus de alumno en otros idiomas. El CORPUS constituye una pequeña muestra que rellena este hueco existente en la lengua española.

### 3 Metodología de la creación del CORPUS

El corpus analizado en este trabajo se compone de los mensajes electrónicos que intercambiaron los estudiantes de Cuestiones de ELE de Universitat de Barcelona con los estudiantes de ELE de Tel Aviv y de Viena en el año 2005. Todos los mensajes fueron enviados mediante un FORO<sup>24</sup> de *La ruta de la lengua española*. Por tanto este CORPUS es un corpus especializado (una mezcla de correo enviado por los alumnos de ELE – esta parte constituye un corpus de alumno – y correo enviado por los nativos).

En el CORPUS se prescindió de indicar la procedencia exacta de cada texto. Se optó por dar sólo la referencia general<sup>25</sup> en lugar de indicar la referencia de cada mensaje, ya que desde la primera web se puede acceder a cada uno de los mensajes.

La finalidad del CORPUS es el análisis de los conectores contraargumentativos.

En cuanto a los límites son siguientes:

1. temporales – es un corpus sincrónico, limitado en el tiempo (desde el febrero hasta el julio de 2005);

<sup>22</sup> Leńko-Szymańska (2005: 204).

<sup>23</sup> Ibidem (202); traducción de WG.

<sup>24</sup> FORO de la *La ruta de la lengua española*, disponible en: <<http://larutadelalengua.com/foro/index.php?c=10>> (fuente)

<sup>25</sup> Ibidem

2. geográficos – se limita a los habitantes de Israel, Austria y España. Aparecen también los mensajes de otras nacionalidades como Estados Unidos, Polonia, Rusia;
3. lingüísticos – se presentan diferentes variedades de español y la interlengua de los estudiantes que más o menos, según el caso, se acerca al español estándar. Aparecen también las palabras en alemán o inglés. Un mensaje entero en alemán fue eliminado del CORPUS.

El CORPUS está basado en textos escritos, es monolingüe, sincrónico, como ya se ha mencionado y simple. Para extraer la información necesaria no hacía falta etiquetarlo. Esta etapa sería imprescindible por ejemplo en el caso de análisis de errores en las frases concesivas con aunque. En este caso se debería etiquetar el modo subjuntivo e indicativo para poder tratar el CORPUS mediante un programa informático.

El CORPUS fue preparado de la siguiente manera: se copió todos los mensajes al editor de texto – WORD. Se eliminó los mensajes que erróneamente aparecieron dos veces. En esta etapa cada mensaje se presentaba así:

Hola! 

Me llamo Irene. Soy una chica de 22 años que vive en Barcelona.

Estudio cuarto de Filología Hispánica; me interesa, sobre todo, la literatura del Renacimiento y del Barroco. Hablo varios idiomas: catalán, español, francés e inglés, aunque este último debería refrescarlo porque, al no practicarlo con frecuencia, se me está olvidando.

En un futuro, me gustaría irme a vivir fuera de España y dar clases de español o de literatura española.

Leer, lógicamente, es mi mayor afición, pero no es la única. También adoro viajar, bailar, pasear por el centro de la ciudad, escuchar música y cantar....ah! y beber café mientras hablo durante horas con un amigo!

A continuación se quitó todas las mayúsculas (dejando éstas de nombres propios) y los espacios entre párrafos y líneas.<sup>26</sup> Además se quitó todos los cuadros, dibujos, etc. dejando sólo el texto del mensaje. Se quitó también todos los temas, dado que no contenían ninguna información relevante para el análisis posterior. En esta etapa cada mensaje se presentaba así:

Publicado: Vie Feb 25, 2005 2:36 pm

hola!

me llamo Irene. soy una chica de 22 años que vive en Barcelona.

estudio cuarto de filología hispánica; me interesa, sobre todo, la literatura del renacimiento y del barroco. hablo varios idiomas: catalán, español, francés e inglés, aunque este último debería refrescarlo porque, al no practicarlo con frecuencia, se me está olvidando.

en un futuro, me gustaría irme a vivir fuera de España y dar clases de español o de literatura española.

leer, lógicamente, es mi mayor afición, pero no es la única. también adoro viajar, bailar, pasear por el centro de la ciudad, escuchar música y cantar....ah! y beber café mientras hablo durante horas con un amigo!

A continuación, se atribuyó a cada mensaje un número especial que está visible en los resultados del análisis al principio de cada ocurrencia.<sup>27</sup> El sistema de cifras que fue inventado por la autora, permite no sólo saber a primera vista por quien está escrito cada mensaje, sino también cuándo exactamente fue enviado. En esta etapa cada mensaje se presentaba así:

\$1252

TUTOR 1 Irene PONS JULIÁ

Publicado: Vie Feb 25, 2005 2:36 pm

hola! me llamo Irene. soy una chica de 22 años que vive en Barcelona.

estudio cuarto de filología hispánica; me interesa, sobre todo, la literatura del renacimiento y del barroco. hablo varios idiomas: catalán, español, francés e inglés, aunque este último debería refrescarlo porque, al no practicarlo con frecuencia, se me está olvidando.

en un futuro, me gustaría irme a vivir fuera de España y dar clases de español o de literatura española.

<sup>26</sup> Fueron acciones requeridas para que el programa de análisis HYPERBASE funcionara bien.

<sup>27</sup> Ocurrencia – cada aparición de una palabra concreta en un corpus.

leer, lógicamente, es mi mayor afición, pero no es la única. también adoro viajar, bailar, pasear por el centro de la ciudad, escuchar música y cantar....ah! y beber café mientras hablo durante horas con un amigo!

Después el CORPUS fue dividido en 8 partes:

1. Los mensajes de todos los tutores para los estudiantes de Tel Aviv.
2. Los mensajes de todos los tutores para los estudiantes de Viena.
3. Los mensajes de febrero y marzo de todos los estudiantes israelíes.
4. Los mensajes de abril de todos los estudiantes israelíes.
5. Los mensajes de mayo y junio de todos los estudiantes israelíes.
6. Los mensajes de febrero y marzo de todos los estudiantes austríacos.
7. Los mensajes de abril de todos los estudiantes austríacos.
8. Los mensajes de mayo y junio de todos los estudiantes austríacos.

Se dividió los mensajes de estudiantes en los meses (febrero y marzo como los meses en los que el FORO empezó a funcionar, abril – el mes de medio y mayo con junio – dos meses finales) para poder comprobar si el uso de los conectores crece con el paso de tiempo.

En esta etapa cada primer mensaje de cada parte del CORPUS se presentaba de la siguiente manera:

&&&1&&&

\$1225

TUTOR 1 Irene PONS JULIÁ

Publicado: Vie Feb 25, 2005 2:36 pm

hola!

me llamo Irene. soy una chica de 22 años que vive en Barcelona.

estudio cuarto de filología hispánica; me interesa, sobre todo, la literatura del renacimiento y del barroco. hablo varios idiomas: catalán, español, francés e inglés, aunque este último debería refrescarlo porque, al no practicarlo con frecuencia, se me está olvidando.

en un futuro, me gustaría irme a vivir fuera de España y dar clases de español o de literatura española.

leer, lógicamente, es mi mayor afición, pero no es la única. también adoro viajar, bailar, pasear por el centro de la ciudad, escuchar música y cantar....ah! y beber café mientras hablo durante horas con un amigo!

El símbolo &&&1&&& significa una parte del CORPUS. Es el símbolo requerido por el programa HYPERBASE, como también lo es el símbolo \$ que significa un párrafo<sup>28</sup> del CORPUS. El número que le sigue es el mencionado arriba, establecido por la autora para mejor orientación en el CORPUS. Hay que leerlo de la siguiente manera (aunque se lo explica en el orden desde la izquierda hacia derecha, se aconseja empezar desde el fin y establecer al principio la fecha):

11225 – mensaje de un tutor

- el primer número es el número de la parte concreta del CORPUS, en este caso 1 – los mensajes de todos los tutores para los estudiantes de Tel Aviv;
- el segundo número es el número personal de cada tutor, en este caso 1 – Irene;
- los demás números se refieren a la fecha en la que fue enviado este mensaje (mes, día), en este caso 225, es decir: el veinticinco de febrero.

810512 – mensaje de un estudiante

- el primer número es el número de la parte concreta del CORPUS, en este caso 8 – los mensajes de mayo y junio de todos los estudiantes austríacos;
- el segundo número (o en algunos casos dos o más siguientes números) es el número personal de cada estudiante,<sup>29</sup> en este caso 10 – Orit;
- los demás números se refieren a la fecha<sup>30</sup> en la que fue enviado este mensaje (mes, día), en este caso 512, es decir: el doce de mayo.

<sup>28</sup> Cada mensaje constituye un párrafo.

<sup>29</sup> Es el mismo número que posee su tutor. Si un tutor tiene dos estudiantes se da al primer estudiante su número y al otro el mismo número con O, por ejemplo Irene tiene dos estudiantes: Daniel con el número 1 y Orit con el número 10. Su tercer estudiante por consiguiente tendría un número con dos O – 100.

A continuación se presenta la lista de todos los tutores con sus estudiantes y números personales.

LOS TUTORES DE TEL AVIV:

1. Irene<sup>31</sup>: Daniel (1) y Orit (10)
2. Raquel<sup>32</sup>: Morik (2) y Aviva (20)
3. Nathan<sup>33</sup>: Yarra (3) y Mira (30)
4. Soña: Eli (4)
5. Marta: Shany (5) y Edith (50)
6. Carmen: Sholmo (6) y Tal (60)
7. M<sup>a</sup>Teresa: Rachel (7)
8. Encarna: Kerren (8) y Lili (80)
9. Raul: Adi (9) y Inbal<sup>34</sup>(90)
10. Diana: Shira (100)<sup>35</sup>
11. Jordi: Niv (11) y Adi (110)
12. Sara: Cuinat (12) y Naomi (120)
13. Wera: Dorina (13) y Olga (130)
14. Aga: Yula (14)
15. Francesca: Gabriela (15) y Caramit (150)
16. Sergi: Yana (16) y Johnatan (170)
17. Sonia: Meron (17)
18. Fatima: Limor (18) y Dima (180)
19. Selima: Eran (19)

LOS TUTORES DE VIENA:

1. Marta: Amal (1) y Claudia (10)
2. David: Eva (2) y Cristina (20)
3. Raquel: Conny (3) y Ricardo (30)
4. Juan Miguel: Melanie (4) y Myram (40)
5. Ana: Mirela (5) y Roland (50) y Theresa (500)
6. Barbara: Sabina (6) y Sandra (60)
7. Blanca: Nicole (7) y Miguel (70) y Alice (700)
8. Yurena: Michi (8)
9. Cristina: Andreas (9)
10. Marta: Christine (10)<sup>36</sup> y Stephanie (100)
11. Yolanda: Melanie (11) y Mathias (110) y Julia (1100)
12. Anna: Vanessa (12)<sup>37</sup> y Barbara (120)
- 13.<sup>38</sup>

---

<sup>30</sup> A veces algunas personas enviaban dos mensajes el mismo día. En este caso se los trata como un mensaje, es decir: están en el mismo párrafo en el CORPUS.

<sup>31</sup> Se anota solamente los nombres, siendo los apellidos irrelevantes para el análisis.

<sup>32</sup> Esta tutora recibió un mensaje erróneamente de una Cristina. Dado que fue error no se la adjunta a la lista.

<sup>33</sup> Este tutor de origen estadounidense fue uno de los estudiantes del programa Socrates – Erasmus que participó en el proyecto de FORO de la *Ruta de la lengua española*.

<sup>34</sup> Este estudiante como también: Johnatan, Meron, Limor y Dima envió sólo un mensaje.

<sup>35</sup> Se atribuyó este número a Shira dado que el 10 ya lo tiene Orit.

<sup>36</sup> En este grupo de estudiantes hay dos que tienen el mismo número – 10. Se trata de Claudia y Christine. Se optó por dejarles el mismo número, ya que ninguna de las dos chicas escribió el mismo día, así que de todas maneras no se puede equivocar a estas dos estudiantes. Para que el lector sepa de quien es un mensaje concreto se presenta aquí todos los mensajes de las dos chicas (sólo de las partes 7 y 8 porque en la parte 6 Claudia no escribió nada). En la parte 7: Claudia – 10430; Christine – 10410, 10417, 10428. En la parte 8: Claudia – 10513, 10520, 10529; Christine – 10501, 10512, 10519, 10526.

<sup>37</sup> Esta tutora recibió también un mensaje en alemán de una tal Verena. No obstante, dado que Verena no escribió ningún mensaje más, se tachó a esta chica de la lista.

14. Arlena: Theresa (14) y Stephanie (140) y Claudia (1400)
15. Laura: Sarah (15) y Simon (150)
16. María: Pía (16) y Tabea (160)

### **3.1 El uso del programa HYPERBASE**

Una vez preparado el CORPUS, se lo puso en el programa HYPERBASE. A continuación el programa llevó a la autora por todo el proceso de preparación del CORPUS para su posterior tratamiento.<sup>39</sup> Acabada la parte técnica, se pudo ya buscar las ocurrencias de cada palabra y en caso de PERO y AUNQUE observar el comportamiento de estas voces en cada parte del CORPUS en un gráfico. Los resultados de estas investigaciones se presenta más abajo.

### **3.2 Nivel de español representado por los alumnos participantes en el proyecto del Foro de *La ruta de la lengua española*<sup>40</sup>**

De las presentaciones de los alumnos participantes en el FORO de La ruta de la lengua española resulta que tienen un nivel muy diferenciado: desde los que estudiaron un o dos años hasta los que además hicieron varios cursos en España y visitaron a menudo este país. Se escribió a las dos profesoras coordinadoras, con el objetivo de calificar el nivel de sus alumnos. Según las respuestas que se recibió, se puede constatar que los estudiantes de Tel Aviv eran de nivel avanzado y de nivel superior. Los de nivel avanzado cursaban el segundo año de español y los de superior el tercer año o más. Los semestres constan de 14 semanas cada uno; cada semana los estudiantes del nivel avanzado estudian 4 horas. Los de superior estudian 2 horas semanales. En general los cursos son de 120 horas anuales – 45 minutos cada “hora”.

En cuanto a los estudiantes de Viena, no se recibió ninguna respuesta concreta de la parte de la coordinadora. V. Berger constató que generalmente los estudiantes que habían participado en este FORO, eran del primer nivel obligatorio en el que se intentaba lograr un nivel homogéneo. Además añadió que no se podía hablar allí de primer u otro año porque la universidad en Viena no funcionaba de esta manera (no explicó cómo funcionaba esta Universidad). Algunos de los estudiantes de Viena han hecho un curso de principiantes en la Universidad. En cuanto al horario, los estudiantes austríacos quedan dos veces a la semana para una clase tradicional de 90 minutos. Tienen entonces el mismo horario que sus compañeros de Tel Aviv del curso avanzado.

Dado que no se dispone de más información, resulta imposible establecer algún tipo de relación entre los años de aprendizaje y el nivel de español, en este caso especialmente el uso de los conectores contraargumentativos que constituyen el objeto de análisis del CORPUS.

## **4 El análisis del CORPUS**

El análisis del CORPUS tiene como objetivo encontrar las respuestas a las siguientes preguntas: ¿Cuáles conectores usan los nativos y cuáles eligen los estudiantes de ELE? ¿Aumenta el uso de los conectores en los mensajes de los estudiantes? ¿Empiezan los alumnos a usar los conectores más complicados gramaticalmente?<sup>41</sup>

Encontrando las respuestas a estas dos preguntas se suponía corroborar o negar la hipótesis que para usar los conectores contraargumentativos propiamente no es suficiente el INPUT que se recibe gracias al intercambio de los mensajes con los nativos.

---

<sup>38</sup> El tutor con este número no existe, dado que hubo un error a la hora de organizar el FORO.

<sup>39</sup> Se prescinde de la descripción de esta etapa, dado que es sólo una cuestión técnica, sin importancia para este trabajo.

<sup>40</sup> Se debería tomar en consideración el nivel de los alumnos para poder relacionarlo con el conocimiento que tienen sobre los marcadores contraargumentativos. Como se explicó en la parte teórica, no se puede esperar el uso de estas piezas en el nivel principiante.

<sup>41</sup> El CORPUS constituye un campo muy amplio para un investigador: se podría por ejemplo comprobar si gracias a las explicaciones de los tutores los estudiantes eliminan los errores en el uso de los conectores, empezando por la ortografía: mas/más. El CORPUS puede servir a otros científicos para hacer el análisis de errores de cualquier tipo.

Cuadro 1: Ocurrencias de los conectores contraargumentativos en el CORPUS<sup>42</sup>

| PARTES           | tutores | tutores | alumnos | alumnos | alumnos | alumnos | alumnos | alumnos | TOTAL       |
|------------------|---------|---------|---------|---------|---------|---------|---------|---------|-------------|
|                  | Israel  | Viena   | Israel  | Israel  | Israel  | Viena   | Viena   | Viena   |             |
|                  | I       | II      | III     | IV      | V       | VI      | VII     | VIII    |             |
| POR EL CONTRARIO | 1       | 0       | 0       | 0       | 0       | 0       | 0       | 0       | <b>1</b>    |
| SI BIEN          | 1       | 0       | 0       | 0       | 0       | 0       | 0       | 0       | <b>1</b>    |
| PESE A           | 0       | 1       | 0       | 0       | 0       | 0       | 0       | 0       | <b>1</b>    |
| CON TODO         | 0       | 1       | 0       | 0       | 0       | 0       | 0       | 0       | <b>1</b>    |
| ESO SÍ           | 2       | 1       | 0       | 0       | 0       | 0       | 0       | 0       | <b>3</b>    |
| AHORA BIEN       | 4       | 0       | 0       | 0       | 0       | 0       | 0       | 0       | <b>4</b>    |
| AUN ASÍ          | 2       | 2       | 0       | 0       | 0       | 0       | 0       | 0       | <b>4</b>    |
| A PESAR DE QUE   | 2       | 2       | 0       | 0       | 0       | 0       | 1       | 0       | <b>5</b>    |
| DE TODAS MANERAS | 3       | 1       | 0       | 1       | 0       | 0       | 1       | 0       | <b>6</b>    |
| AL CONTRA-RIO    | 2       | 4       | 0       | 0       | 0       | 0       | 0       | 0       | <b>6</b>    |
| DE TODOS MODOS   | 6       | 2       | 0       | 0       | 0       | 0       | 0       | 0       | <b>8</b>    |
| NO OBSTANTE      | 2       | 5       | 0       | 0       | 0       | 0       | 1       | 0       | <b>8</b>    |
| MAS              | 6       | 1       | 1       | 2       | 0       | 0       | 0       | 0       | <b>10</b>   |
| SIN EMBARGO      | 11      | 4       | 0       | 1       | 0       | 0       | 0       | 2       | <b>18</b>   |
| EN CAMBIO        | 8       | 10      | 1       | 0       | 0       | 0       | 0       | 0       | <b>19</b>   |
| DE TODAS FORMAS  | 19      | 11      | 0       | 0       | 0       | 1       | 0       | 0       | <b>31</b>   |
| AUNQUE           | 82      | 71      | 3       | 3       | 1       | 0       | 3       | 4       | <b>167</b>  |
| PERO             | 315     | 331     | 108     | 69      | 66      | 68      | 92      | 98      | <b>1147</b> |

<sup>42</sup> Este cuadro presenta las ocurrencias de cada parte del CORPUS, empezando de la primera hasta la última. Cada cifra indica cuántas veces aparece una palabra concreta en una parte concreta del CORPUS. Por ejemplo: POR EL CONTRARIO sólo aparece una vez en la primera parte, es decir que esta voz fue usada sólo una vez por algún tutor.

Antes de analizar las ocurrencias de los conectores contraargumentativos se eliminó de ellas los casos de “más” (escrito a menudo sin acento) comparativo como también la expresión “con todo lo que...” que el programa HYPERBASE calificó erróneamente como un conector contraargumentativo. Como se explicó más arriba el CORPUS está dividido en ocho partes: dos constituyen los mensajes enviados por los tutores y seis siguientes se refieren a los mensajes escritos por los estudiantes.

Como se puede observar del análisis cuantitativo, el conector más usado por los tutores es el conector adversativo PERO. Le sigue el conector concesivo AUNQUE. Los demás conectores, aunque son usados por los tutores, lo son mucho menos que los dos conectores mencionados arriba. En cuanto a los estudiantes, se limitan al conector PERO. AUNQUE aparece sólo 14 veces. En el CORPUS hay cuatro hápax<sup>43</sup>: si bien, por el contrario, pese a, con todo.

El programa HYPERBASE permite el análisis en forma de un gráfico que suministra una información muy relevante.<sup>44</sup> A saber: el programa, mediante unos cálculos estadísticos, preve cuántas veces debería aparecer teóricamente cada palabra del texto en cada una de las partes. Es decir, si PERO aparece 1147 veces en el total del CORPUS, calcula cuántas veces debería aparecer en cada una de las 8 partes, proporcionalmente a la extensión de cada una de estas partes. Después, lo compara con el número de apariciones reales de esa palabra y calcula la diferencia.

Todo esto se traduce a cifras estadísticas, es decir, no son cifras reales. Así pues, "0" significa que el cálculo teórico de las veces que debería aparecer una palabra y las veces que aparece en realidad coinciden o son muy aproximadas. Si una palabra aparece más veces de lo calculado en una parte se considera que es específicamente positiva o sobreutilizada en esa parte en concreto; si aparece menos veces, se considera que es específicamente negativa o infrautilizada en esa parte en concreto. Esto es lo que se conoce como cálculo de especificidades.

Ahora bien, para que dichas especificidades sean relevantes, deben ser superiores a "2" o inferiores a "-2". Por tanto, todas las especificidades que sean 0, 1, -1 no se tendrán en cuenta en el análisis porque se considera que aunque la cifra teórica y la real no coincidan, la diferencia entre ellas no resulta significativa.

En el gráfico aparecen tres columnas: la columna "Texte" indica la parte del corpus; la columna "fréq." el número de apariciones de una palabra en cada parte y la columna "écart" el grado de especificidad.<sup>45</sup>

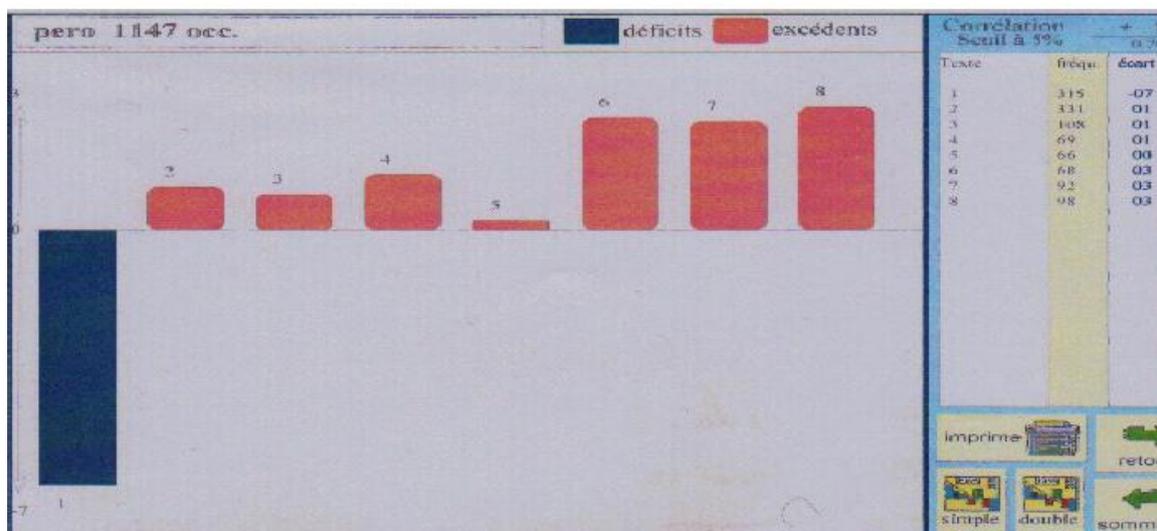


Gráfico 1: Ocurrencias de PERO

<sup>43</sup> Hápax es una palabra que aparece sólo una vez en un corpus.

<sup>44</sup> Desgraciadamente sólo se ha podido analizar PERO y AUNQUE mediante los gráficos, dado que los demás conectores están tratados como expresiones y el programa HYPERBASE contiene la opción de los gráficos sólo para las palabras.

<sup>45</sup> Explicación recibida de la parte de M. Nofre.

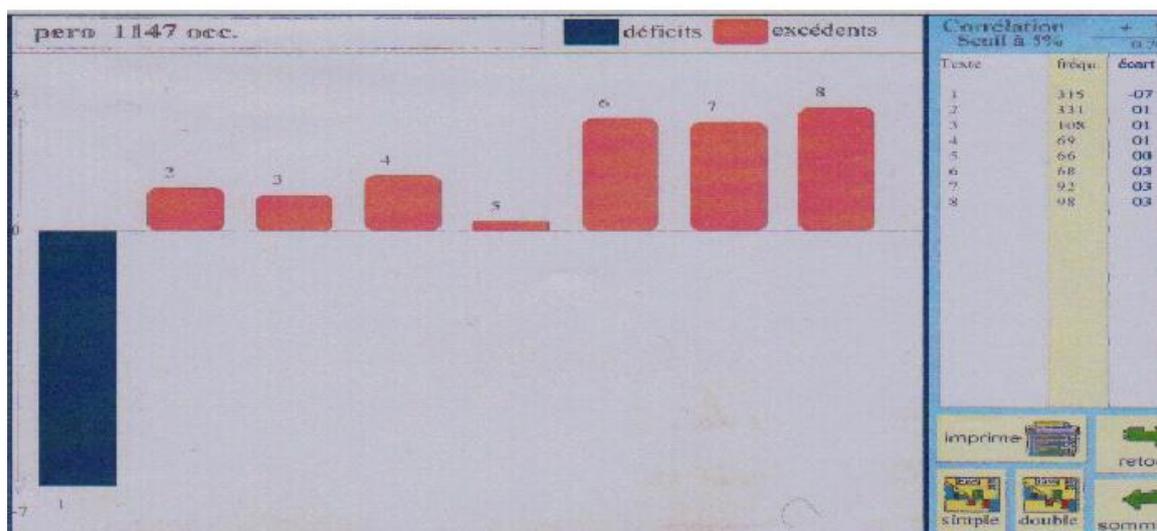


Gráfico 2: Ocurrencias de AUNQUE

El color azul que se puede observar en el gráfico, significa que la palabra PERO es infrautilizada y el color rojo significa que la misma palabra aparece demasiadas veces en el CORPUS. Dejando al lado las partes: 2-5 (dado que "écart" está entre 0 - -1), se puede constatar que los tutores del primer grupo utilizan poco el conector PERO, mientras que los estudiantes austriacos lo utilizan demasiado. Aquí se ve que hay una gran diferencia entre los alumnos de Israel y los de Viena, siendo los primeros los que utilizan PERO mucho menos que los segundos. Además en la tercera etapa de intercambio de los mensajes disminuye el número de ocurrencias de este vocablo. No es el caso del segundo grupo, en el que el uso de PERO crece con el tiempo.

En cuanto a AUNQUE los resultados de análisis no son de extrañar. Los tutores usan este vocablo muchas veces (comparando con todo el CORPUS incluso demasiadas) y los estudiantes lo usan insuficientemente comparando con todo el CORPUS. Ahora bien, si miramos el uso de esta voz durante los diferentes meses, vemos que en el caso de los estudiantes de Viena el uso crece y en el caso de sus compañeros de Tel Aviv el uso disminuye.

Aunque del gráfico resulta que los estudiantes austriacos empiezan a usar más el conector AUNQUE no se puede decir que sea un cambio relevante porque en cifras son solo 7 casos entre todos los 32 estudiantes.

En definitiva, se constató que a pesar del INPUT que reciben los estudiantes de ELE mediante los mensajes intercambiados con los nativos, que usan diferentes conectores contraargumentativos, éste no es suficiente para aprender a usar estos vocablos. Por consiguiente, se corroboró la hipótesis admitida antes de hacer el análisis del CORPUS.

En consecuencia se puede constatar que hay que explicar explícitamente las diferencias en el uso de los conectores contraargumentativos a los estudiantes de ELE.

Uno se podría preguntar si el aprendizaje de los conectores es posible mediante el INPUT que se recibe en este tipo de actividades. Según M. Cruz Piñol<sup>46</sup> mucho depende de tipo de alumnado, lo cual no tiene que ser necesariamente relacionado con su origen, sino más bien con su carácter. Un alumno minucioso y capaz de analizar sus propios errores podría aprender más mediante un intercambio de los mensajes electrónicos.

En este tipo de análisis no se puede olvidar del problema de nivel de los alumnos. Los alumnos que participaron en el FORO tienen un nivel muy diferenciado y en muchos casos imposible de determinar. Por consiguiente, lo único que se puede constatar con seguridad, es que hasta los estudiantes que estudiaron español como mínimo 2 o 3 años, no recurren por lo general a otros

<sup>46</sup> Informaciones adquiridas en la conversación privada que la autora tuvo con M. Cruz Piñol.

contraargumentativos que PERO<sup>47</sup>. El intercambio de mensajes con los nativos que sí usan estos marcadores no es suficiente para aprender a aplicar estas palabras.

Al final, hay que comparar las ocurrencias de CORPUS con un corpus más objetivo y de peso más científico, como el Corpus de Referencia del Español Actual (CREA). Los resultados de esta búsqueda se presentan en el siguiente cuadro. Hay tres columnas. En la primera aparecen todas las ocurrencias de un conector escrito con minúscula, en la segunda todas las ocurrencias de la misma palabra escrita con mayúscula y en la tercera el total. Se presenta de la misma manera todos los conectores. Se tuvo que presentarlo de este modo puesto que en el CREA se distingue las minúsculas de las mayúsculas. Las ocurrencias se refieren a todos los textos (en el CREA hay la posibilidad de elegir un campo especial, según el tema de texto, la variedad de español en el que está escrito o el medio en el que se encuentra) y a todas las variedades de español. En el presente cuadro se puede observar que en los textos actuales aparecen todos los conectores que están presentes en el CORPUS (no se hizo el análisis de otros conectores porque se quiso comparar sólo los que aparecen en el CORPUS).

*Cuadro 2: Ocurrencias de los conectores contraargumentativos en el CREA*

| CONECTORES EN MINÚSCULA | CONECTORES EN MAYÚSCULA | TOTAL          | NOMBRE DEL CONECTOR |
|-------------------------|-------------------------|----------------|---------------------|
| 187                     | 71                      | <b>258</b>     | ANTES BIEN          |
| 825                     | 772                     | <b>1597</b>    | AUN ASÍ             |
| 858                     | 1088                    | <b>1946</b>    | DE TODAS MANERAS    |
| 236                     | 1765                    | <b>2001</b>    | DE TODOS MODOS      |
| 739                     | 1583                    | <b>2322</b>    | DE TODAS FORMAS     |
| 123                     | 3167                    | <b>3290</b>    | AHORA BIEN          |
| 3503                    | 1701                    | <b>5204</b>    | ESO SÍ              |
| 5363                    | 2071                    | <b>7434</b>    | A PESAR DE QUE      |
| 4720                    | 3014                    | <b>7734</b>    | POR EL CONTRARIO    |
| 5259                    | 2614                    | <b>7873</b>    | MAS                 |
| 6183                    | 2066                    | <b>8249</b>    | CON TODO            |
| 8344                    | 3543                    | <b>11887</b>   | SI BIEN             |
| 10052                   | 4345                    | <b>14397</b>   | PESE A              |
| 9983                    | 5254                    | <b>15237</b>   | EN CAMBIO           |
| 27633                   | 37592                   | <b>65225</b>   | SIN EMBARGO         |
| 92923                   | 23565                   | <b>116488</b>  | AUNQUE              |
| 28547                   | 128547                  | <b>157094</b>  | NO OBSTANTE         |
| 325079                  | 1065479                 | <b>1390558</b> | PERO                |

Se pensó que sería interesante ver los resultados de ambos análisis a la vez. Se los presenta en el siguiente cuadro. En la primera columna aparecen todos los conectores desde los menos hasta

<sup>47</sup> Por razones de espacio no se hizo el análisis más detallado (cual estudiante exactamente usa cuales conectores). Es un posible campo de futura investigación.

los más usados en el CORPUS. En la cuarta columna aparecen todos los conectores desde los menos hasta los más usados en el CREA. La segunda y la tercera columna representan las ocurrencias de estos vocablos en cifras en cada corpus respectivamente.

Cuadro 3: Comparación de los marcadores contraargumentativos en el CORPUS y el CREA

| CORPUS                |             | CREA           |                       |
|-----------------------|-------------|----------------|-----------------------|
| POR EL CONTRARIO      | 1           | 258            | ANTES BIEN            |
| SI BIEN               | 1           | 1597           | AUN ASÍ               |
| PESE A                | 1           | 1946           | DE TODAS MANERAS      |
| CON TODO              | 1           | 2001           | DE TODOS MODOS        |
| ESO SÍ                | 3           | 2322           | DE TODAS FORMAS       |
| <b>AHORA BIEN</b>     | <b>4</b>    | <b>3290</b>    | <b>AHORA BIEN</b>     |
| AUN ASÍ               | 4           | 5204           | ESO SÍ                |
| <b>A PESAR DE QUE</b> | <b>5</b>    | <b>7434</b>    | <b>A PESAR DE QUE</b> |
| DE TODAS MANERAS      | 6           | 7734           | POR EL CONTRARIO      |
| AL CONTRARIO          | 6           | 7873           | MAS                   |
| DE TODOS MODOS        | 8           | 8249           | CON TODO              |
| NO OBSTANTE           | 8           | 11887          | SI BIEN               |
| MAS                   | 10          | 14397          | PESE A                |
| <b>SIN EMBARGO</b>    | <b>18</b>   | <b>15237</b>   | <b>EN CAMBIO</b>      |
| <b>EN CAMBIO</b>      | <b>19</b>   | <b>65225</b>   | <b>SIN EMBARGO</b>    |
| DE TODAS FORMAS       | 31          | 116488         | AUNQUE                |
| AUNQUE                | 167         | 157094         | NO OBSTANTE           |
| <b>PERO</b>           | <b>1147</b> | <b>1390558</b> | <b>PERO</b>           |

Si se analiza las diferencias entre el CORPUS con el CREA, se ve que hay tres marcadores del CORPUS, cuyas posiciones coinciden exactamente con el CREA (**ahora bien, a pesar de que, pero**) y además hay tres que se encuentran muy cerca en la escala de su uso (*aunque, sin embargo, en cambio*).

Este cuadro se preparó de la misma manera que el cuadro 4. En la primera columna aparecen todos los conectores del CORPUS, desde los menos, hasta los más usados en la parte de los tutores. En la tercera columna aparecen todos los conectores del CORPUS desde los menos hasta los más usados en la parte de los alumnos. En la quinta columna aparecen todos los conectores desde los menos hasta los más usados en el CREA. La segunda, la cuarta y la sexta columna representan las ocurrencias de estas palabras en cifras en cada corpus o su parte respectivamente.

Como era de esperar, hay más coincidencias con el CREA en la parte de los tutores que en la parte de los alumnos. Con el color rojo se marca estos conectores que coinciden en la parte de los tutores con el CREA pero que no se encuentran en las mismas posiciones, si los comparamos con la parte de alumnos. Según se observa, el contenido de los mensajes escritos por los tutores se asemeja más a las producciones escritas u orales presentes en el CREA que éste producido por alumnos.

Se puede entonces arriesgar una tesis que el CORPUS al menos en algunos casos, refleja bastante bien el uso de la lengua española (en cuanto a los conectores contraargumentativos). Aquí hay que tomar en consideración las diferencias tipológicas entre los corpus. El CREA es un corpus muy grande, de español actual que abarca muchas disciplinas. El CORPUS es muy pequeño en comparación con el CREA y además, es un corpus específico que representa en su gran parte la interlengua de los que están todavía estudiando español mientras que el CREA es un corpus representativo de la lengua española. El CORPUS no puede serlo, dado el origen de los textos que lo constituyen. Sin embargo, gracias a la comparación presentada más arriba, se puede ver que algunos conectores tienden a aparecer con la misma frecuencia. En cualquier corpus

comparado con el CREA de la misma manera, cuantos más conectores estuvieron en las mismas posiciones, más se acercaría este corpus a un corpus representativo.

*Cuadro 4: Comparación de los marcadores contraargumentativos presentes en la parte de tutores del CORPUS y en la parte de alumnos del CORPUS con el CREA*

| TUTORES               |     | ALUMNOS          |     | CREA                  |         |
|-----------------------|-----|------------------|-----|-----------------------|---------|
| POR EL CONTRARIO      | 1   | POR EL CONTRARIO | 0   | ANTES BIEN            | 258     |
| SI BIEN               | 1   | SI BIEN          | 0   | AUN ASÍ               | 1597    |
| PESE A                | 1   | PESE A           | 0   | DE TODAS MANERAS      | 1946    |
| CON TODO              | 1   | CON TODO         | 0   | DE TODOS MODOS        | 2001    |
| ESO SÍ                | 3   | ESO SÍ           | 0   | DE TODAS FORMAS       | 2322    |
| AHORA BIEN            | 4   | AHORA BIEN       | 0   | AHORA BIEN            | 3290    |
| AUN ASÍ               | 4   | AUN ASÍ          | 0   | ESO SÍ                | 5204    |
| <b>A PESAR DE QUE</b> | 4   | AL CONTRARIO     | 0   | <b>A PESAR DE QUE</b> | 7434    |
| DE TODAS MANERAS      | 4   | DE TODOS MODOS   | 0   | POR EL CONTRARIO      | 7734    |
| AL CONTRARIO          | 6   | A PESAR DE QUE   | 1   | MAS                   | 7873    |
| <b>NO OBSTANTE</b>    | 7   | NO OBSTANTE      | 1   | CON TODO              | 8249    |
| MAS                   | 7   | EN CAMBIO        | 1   | SI BIEN               | 11887   |
| DE TODOS MODOS        | 8   | DE TODAS FORMAS  | 1   | PESE A                | 14397   |
| <b>SIN EMBARGO</b>    | 15  | DE TODAS MANERAS | 2   | <b>EN CAMBIO</b>      | 15237   |
| <b>EN CAMBIO</b>      | 18  | MAS              | 3   | <b>SIN EMBARGO</b>    | 65225   |
| DE TODAS FORMAS       | 30  | SIN EMBARGO      | 3   | <b>AUNQUE</b>         | 116488  |
| <b>AUNQUE</b>         | 153 | <b>AUNQUE</b>    | 14  | NO OBSTANTE           | 157094  |
| PERO                  | 646 | PERO             | 501 | PERO                  | 1390558 |

## 5 A modo de conclusión

El presente trabajo tuvo como objetivo comprobar si los estudiantes de ELE eran capaces de concebir la noción de los conectores contraargumentativos implícitamente. Gracias al análisis del corpus de alumno, se pudo observar cómo los estudiantes de ELE usaban estos elementos. Se creó el CORPUS de los mensajes electrónicos para ver si los estudiantes cambiaban su manera de escribir gracias al input que recibían de la parte de los tutores.

Dada la especificidad del medio, se creía poder encontrar en los mensajes los conectores, no sólo propios del lenguaje escrito, sino también los que usan los nativos en las conversaciones cotidianas. Se suponía, que gracias a escribir en un foro, que constituye un espacio abierto donde los estudiantes pueden hablar libremente sobre los temas que les interesan, teniendo al alcance todos los recursos electrónicos de apoyo, los estudiantes se sentirían más libres para improvisar y buscar diferentes modos de expresarse. Se esperaba que imitaran las construcciones que veían en los mensajes de sus tutores, que empezaran a utilizar los recursos de contraargumentación más complicados que el conector PERO. Se quería comprobar también si intentarían manejar otra variedad del estándar: el español conversacional.<sup>48</sup>

Después de analizar el CORPUS, se puede constatar que los estudiantes generalmente no recurren a los conectores contraargumentativos más complicados, aunque estén en los niveles en los que ya deberían hacerlo. Gracias a la comparación del CORPUS con el CREA, se pudo observar que los nativos también recurren por lo general a los marcadores adversativos aunque usen también los marcadores concesivos. Por esto, se cree que hay que presentar todos tipos de conectores a los estudiantes de español. La comparación del CORPUS con CREA muestra evidentemente que los estudiantes van a encontrarse con diferentes contraargumentativos y que es necesaria la explicación explícita de estas piezas.

<sup>48</sup> Hay que recordar que en los mensajes de tutores aparecieron los conectores típicos para esta variedad.

Enseñar una lengua como lengua extranjera debería significar abrir puertas y ventanas a otros mundos más o menos lejanos: los de las personas que hablan esa lengua como propia. Porque las lenguas son instrumentos de representación del mundo y de las personas que las habla, que las utilizan para construirse a sí mismas en relación con las demás en la sociedad. Las lenguas son parte integrante de las culturas de los pueblos y, por lo tanto, acceder al conocimiento de una lengua extranjera es una manera de acceder a otras formas de pensar, de concebir el mundo y las relaciones humanas; es una forma también, de entender que, si bien el lenguaje es algo universal que hermana a todos los componentes de especie humana, las lenguas concretas que se usan en cada comunidad están impregnadas de las características culturales propias de cada grupo de hablantes. En un mundo en que las relaciones entre hablantes de diferentes procedencias lingüísticas son cada vez más habituales, se hace más necesario y deseable el conocimiento de varias lenguas.<sup>49</sup>

Saber la lengua quiere decir saber comportarse en diferentes tipos de situaciones. Los nativos, hablando con los estudiantes de ELE, pueden juzgarlos mal como personas por cometer errores pragmáticos. Pasa muy a menudo que los estudiantes extranjeros parecen muy bruscos o muy tímidos. Esto se debe al grado insuficiente del conocimiento de la lengua oral que es como la llave que abre las puertas a la sociedad en la que se encuentra el alumno. Resulta que el conocimiento con el que un estudiante extranjero llega a un país hispanohablante muchas veces no es suficiente para poder comunicarse con los nativos. Como (...) *los nativos suelen aceptar más fácilmente las erratas de pronunciación y de sintaxis que las violaciones de las reglas conversacionales, las cuales se pueden interpretar como mala educación*<sup>50</sup> los profesores deberían tener más en cuenta diferentes variedades del español a la hora de enseñarlo.

Con esto, hay que subrayar la importancia de intercalar el español coloquial en el aula de ELE. Los conectores analizados en el presente trabajo forman obviamente sólo una parte de conocimiento que debería adquirir el estudiante.

En el presente artículo se intentó también demostrar la utilidad de servirse de los corpus de alumno en las investigaciones lingüísticas<sup>51</sup> y de usar el correo electrónico en la enseñanza de ELE. Aunque éste no constituya un medio suficiente para el estudio de los conectores contraargumentativos, su utilidad en las clases de ELE es indudable.

---

<sup>49</sup> Ruiz Bikandi, Tusón Valls (2003: 7).

<sup>50</sup> Tiina-Leena (2000: 622).

<sup>51</sup> Se pone a disposición de todos los interesados el CORPUS bajo la condición de tener el previo consentimiento de la autora. Se puede contactar con ella vía Internet: weragorska@poczta.onet.pl.

## Bibliografía

### FUENTES:

García Garrido, J.: *Foro de la Ruta de la lengua española* (Universidad de Salamanca, España) 2001-2002 Última actualización: 06/04/06 [07.05.2006] disponible en: <<http://www.larutadelalengua.com/foro/>>

### ESTUDIOS:

- Anscombe, J.-C., Ducrot, O. 1994. *La argumentación en la lengua*. Madrid: Gredos.
- Becerra Hiraldo, J.M<sup>a</sup>, Barros García, P., Martínez González, A., Molina Redondo, J. A. de (eds.) 1999. *La enseñanza de segundas lenguas*. Granada: Universidad de Granada.
- Bécue Bertaut, M. 1999. Análisis estadístico de textos. En: Blecua, J.M., Clavería, G., Sánchez, C., Torruella, J. (eds.) *Filologías e informática Nuevas tecnologías en los estudios filológicos* Barcelona: Universitat Autònoma de Barcelona, p: 81-109.
- Cano López, R. 2000. Aplicaciones del ordenador personal en la enseñanza de lenguas. En: Moya Corral, J. A. (ed.) *Adquisición y enseñanza de la Lengua Española. Actas de las IV Jornadas sobre la enseñanza de la Lengua Española*. Granada: Centro de Formación Continua de la Universidad de Granada, p: 97-101.
- Ciganda Canals, Q. 2002. *Las estructuras concesivas con aunque y la enseñanza de E/LE*. Directora: Estrella Montolio. Barcelona: Universitat de Barcelona. Departamento de Didáctica de la Lengua y la Literatura.
- Consejo de Europa 2003. *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: Ministerio de Educación, Cultura y Deporte, Anaya.
- Consuela Pérez, L. 2002. El uso del correo electrónico (asincrónico) y de las salas electrónicas de conversación (sincrónico) en la clase de español como lengua extranjera. En: Gimeno Sanz, A. M. (coord.) *Tecnologías de la información y de las comunicaciones en la enseñanza de ELE Actas del XII congreso internacional de ASELE*. Valencia: Universidad Politécnica de Valencia, p: 477-484.
- Cruz Piñol, M. 2002. *Enseñar español en la era de Internet*. Barcelona: OCTADERO.
- Garrido Rodríguez, M<sup>a</sup> del Camino 2000. Los conectores pragmáticos en la enseñanza de ELE: argumentación u relevancia (I). En: Franco, M., Soler, C., Cos, J. de, Rivas, M., Ruiz, F. *Nuevas perspectivas en la enseñanza del español como lengua extranjera*. Cádiz: Universidad de Cádiz, p: 323-329.
- Goźdz-Roszkowski, S., Witzczak-Plisiecek, I. 2005. Korpusy a języki specjalistyczne. En: Lewandowska-Tomaszczyk, B.(red.) *Podstawy językoznawstwa korpusowego*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, p: 174-200.
- Kucera, H., Francis, W.N. 1967. *Computational Analysis of Present-Day American English*. Rhode Island: Brown University Press.
- Lence Guilbert, M<sup>a</sup> A. 2002. Práctica de lengua escrita sobre aspectos socioculturales pro correo electrónico. En: Gimeno Sanz, A. M. (coord.) *Tecnologías de la información y de las comunicaciones en la enseñanza de ELE Actas del XII congreso internacional de ASELE*, Valencia: Universidad Politécnica de Valencia, p: 655-665.
- Leńko-Szymańska, A. 2005. Korpusy Uczniowskie. En: Lewandowska-Tomaszczyk, B.(red.) *Podstawy językoznawstwa korpusowego*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, p: 201-220.
- Leńko-Szymańska, A. 2005. Korpusy w nauczaniu języków obcych. En: Lewandowska-Tomaszczyk, B.(red.) *Podstawy językoznawstwa korpusowego*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, p: 221-239.
- Llisterri, J. 1998. Nuevas tecnologías y enseñanza del español como lengua extranjera. En: Moreno F., Gil M., Alonso K. *La enseñanza del español como lengua extranjera: del pasado al futuro*. Madrid: Universidad de Alcalá, p: 45-69.
- López Alonso, C., Séré, A. (eds.) 2003. *Nuevos géneros discursivos: los textos electrónicos*. Madrid: Biblioteca Nueva.
- Martín Zorraquino, M<sup>a</sup> A. 1994. Los marcadores del discurso en la enseñanza del español como lengua extranjera (aspectos gramaticales y cuestiones pragmáticas). En *Enseñanza de II Lenguas*. Granada: Universidad de Granada.
- Portolés, J. 1998. *Marcadores del discurso*. Barcelona: Ariel.
- Sanz Iglesias, J. 2002. Correo electrónico, atención y elaboración de tareas escritas en estudiantes de E/LE: estudio de caso. En: Gimeno Sanz, A. M. (coord.) *Tecnologías de la información y de las comunicaciones en la enseñanza de ELE Actas del XII congreso internacional de ASELE*. Valencia: Universidad Politécnica de Valencia, p: 65-72.
- Soler-Espiauba, D. 2002. Yo emilio, tú emilias... todos emiliamos ( El Correo Electrónico en el aula de E/LE). En: Gimeno Sanz, A. M. (coord.) *Tecnologías de la información y de las comunicaciones en la enseñanza de ELE Actas del XII congreso internacional de ASELE*. Valencia: Universidad Politécnica de Valencia, p: 73-84.
- Sotelo Docío, S., Vázquez González, M<sup>a</sup> N. 1998. *Herramientas informáticas de aplicación lingüística*; "Interlingüística", núm.7: 223-228.
- Ruiz Bikandi, U., Tusón Valls, A. 2003. *La enseñanza del español como lengua extranjera*; "Textos de Didáctica de la Lengua y la Literatura", núm. 34: 7-9.
- Tiina-Leena, S. 2000. Estrategias conversacionales en los manuales de ELE y su aplicación en el aula. En: Franco, M., Soler, C., Cos, J. de, Rivas, M., Ruiz, F. *Nuevas perspectivas en la enseñanza del español como lengua extranjera*. Cádiz: Universidad de Cádiz, pág.621-627.
- Trenchs, M. 1998. E-mails a una mestra. Correu electrònic i aprenentatge de llengües. Lleida: Pagès Editors.
- Torruella, J., Llisterri, J. 1999. Diseño de corpus textuales y orales. En: Blecua, J.M., Clavería, G., Sánchez, C., Torruella, J. (eds.) *Filologías e informática Nuevas tecnologías en los estudios filológicos*. Barcelona: Universitat de Autònoma de Barcelona, p: 45-77.

Waliński, J. 2005. Typologia korpusów oraz warsztat informatyczny lingwistyki korpusowej. En: Lewandowska-Tomaszczyk, B.(red.) Podstawy językoznawstwa korpusowego. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, p: 27-41.

### **Recursos en línea:**

1. Utah Consular Corps. 2003 Última actualización: 14/09/2004 [13.04.2006], disponible en: <<http://www.utahconsuls.org/>>
2. La Asociación para la Enseñanza del Español como Lengua Extranjera: Página Principal. (España) [13.05.2006], disponible en: <<http://www.aselered.org/>>
3. Real Academia Española: Página Principal (España) [09.05.2006], disponible en: <<http://www.rae.es>>
4. Real Academia Española: CORDE (España) [09.05.2006], disponible en: <<http://corpus.rae.es/cordenet.html>>
5. Real Academia Española: CREA (España) [09.05.2006], disponible en: <<http://corpus.rae.es/creanet.html>>
6. Asociación cultural Antonio de Lebrija: La página del idioma español el castellano.org (España) 1996 [20.01.2005], disponible en: <<http://www.elcastellano.org/ns/edicion/2004/septiembre/lectura.html>>
7. Antonio Millán J.: Vocabulario de ordenadores e Internet. (España) Última actualización: 17/01/2005 [19.01.2005], disponible en: <[http://www.jamillan.com/v\\_index.htm](http://www.jamillan.com/v_index.htm)>
8. Lejeune Ch. : Des outils libres [10.05.2006], disponible en: <<http://www.smess.egss.ulg.ac.be/lejeune/logiciels/opencaqdas.html>>
9. Instituto Cervantes: Página Principal. (España) 1997 [13.04.2005], disponible en: <<http://www.cervantes.es>>
10. Centro Virtual Cervantes: Diccionarios, léxicos y glosarios. (España) [13.05.2006], disponible en: <[http://cvc.cervantes.es/oteador/default.asp?l=2&id\\_rama=88&ct=catalogo88](http://cvc.cervantes.es/oteador/default.asp?l=2&id_rama=88&ct=catalogo88)>
11. Instituto Cervantes: Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza. (España) [25.05.2006], disponible en: <<http://cvc.cervantes.es/obref/marco/>>
12. Fernández Ramírez S.: Archivo Gramatical de la Lengua Española (Instituto Cervantes, España) [13.05.2006], disponible en: <<http://cvc.cervantes.es/obref/agle>>
13. Instituto Cervantes: Integrated Reference Corpora for Spoken Romance Languages. (España) [13.05.2006], disponible en: <[http://www.cervantes.es/seg\\_nivel/lect\\_ens/oesi/coral\\_01.htm](http://www.cervantes.es/seg_nivel/lect_ens/oesi/coral_01.htm)>
14. David Mighetto: SOL – Spanish Online 1998 Última actualización: 5/12/2001 [13.05.2006], disponible en: <<http://spraakbanken.gu.se/lb/konk/rom2/>>
15. Mark Davies: Corpus del Español. 2001-2002 [10.05.2006], disponible en: <<http://www.corpusdelespanol.org>>
16. Department of Education of Spain: The Spanish FrameNet Project Última actualización: 04/28/2005 [10.05.2006], disponible en: <<http://gemini.uab.es/SFN/>>
17. Universidad de Santiago: Base de Datos Sintácticos del Español Actual (BDS) (España) Última actualización: 16/10/01 [10.05.2006], disponible en: <<http://www.bds.usc.es/>>
18. SIGNUM Cía. Ltda.: VARILEX (Ecuador) 1999 [10.05.2006], disponible en: <<http://www.lenguaje.c>>