

Recursos para estudiar el español académico a partir de corpus

María Cecilia Ainciburu

Universidad Antonio de Nebrija.

caincibu@nebrija.es

Cecilia Ainciburu, M. (2012). Recursos para estudiar el español académico a partir de corpus. *Revista Nebrija de Lingüística Aplicada* (2012) 12.

El español académico, ya sea en su versión nativa como no nativa, requiere una descripción teórica y prescriptiva que - por razones de necesidad, en la alfabetización académica y por su condición de nicho económico, en la acción editorial-, tiene buenos y asentados precedentes en los países de lengua española. Nada comparable con la tradición del inglés, como es fácil suponer. Paralelamente, es difícil disponer de material integrado en corpus y mucho menos en corpus no oportunistas, con selección del ámbito de circulación.

En el XX Congreso ASELE realizado en Comillas, la autora de materiales Blanca Aguirre confesaba su perplejidad a la hora de trabajar con modelos textuales de los cuales no existe un corpus representativo. Acceso al video completo en:

<http://www.comunidadcomillas.com/group/xxicongresoaselegrupodembitoacadmico>

Hablaba del Español de los negocios, pero lo mismo puede decirse del Español académico y esa dificultad se potencia aún más cuando el texto meta es de producción oral. Sin embargo, esta es sólo una de las vertientes de un problema que afecta el estudio de los procesos de adquisición de los rasgos típicos de la macro y micro estructura de los géneros de la comunicación académica en ELE. No sólo modelos del español, entonces, sino su relación con la multiplicidad de modelos textuales provenientes de otras lenguas que deben describirse en forma contrastiva.

Por esa razón, evidenciamos la experiencia de la escuela de Valparaíso y del proyecto ADIEU que se presentan como los intentos más articulados, uno en textos producidos por nativos y el otro en su didactización ELE; textos en los que es posible basarse para obtener modelos, realizar análisis e investigación aplicada y organizar itinerarios didácticos específicos por género.

En la sección *Al día* no se realizan análisis generales, sino que se proporcionan recursos. Por esta razón, lo que sigue es la descripción de los recursos provenientes de estas dos experiencias y no su desarrollo histórico o la evaluación de las mismas.

El proyecto de construcción de un corpus de textos propios del ámbito académico en una universidad hispanohablante tiene su arquetipo en la acción de la Escuela Lingüística de la Pontificia Universidad Católica de Valparaíso. Cuenta con un corpus de textos vasto y en constante aumento, de consulta pública y gratuita (como sería deseable en todas las experiencias de este tipo) y con un instrumento de consulta o interfaz de gran utilidad, el llamado "Grial".



The screenshot shows the website 'El Grial' from the 'Escuela Lingüística de Valparaíso' (ELV). The browser address bar shows 'www.elv.d/prontus_linguistica/site/edic/base/port/grial.html'. The page features a navigation menu with 'ELV', 'Postgrados', 'El Grial' (highlighted), 'Revista Signos', 'LECTES', and 'Tutor MUSA'. A search section titled 'Sumérgete en el corazón del Grial' includes 'Herramientas de búsqueda' with buttons for 'Búsqueda Multiniveles' and 'El Manchador'. Below this is 'El Corpus' with a 'Mapa de Corpus' button. A 'Comparación de textos' section is also visible. A text block explains the tool's purpose: '¿Qué es el Grial? Es una interfaz de etiquetaje y consulta computacional de corpus textuales...'. A footer at the bottom provides contact information for ELV.

Esta herramienta integrada permite la búsqueda simple en el corpus previamente etiquetado de formas, lemas y partes de la oración de las que es posible obtener tanto el orden de frecuencia como el contexto. La interfaz permite seleccionar el corpus por modalidad (escrita u oral), registro, géneros, ediciones sucesivas del corpus, temas o -incluso- con el acceso a un determinado documento.

El resultado de esta búsqueda es un documento en el que se muestran y se calculan las frecuencias de aparición de uno o varios rasgos lingüísticos. La herramienta computacional se denomina "Manchador de textos" y proporciona información relevante acerca de la densidad de un rasgo o de la coocurrencia de rasgos en una parte determinada de un texto.

La búsqueda compleja accede a la información sintáctica y de dependencia gramatical relacionada con un cierto tipo de lema (por ejemplo: es adjetivo) o directamente con la etiqueta sintáctico-gramatical. Es necesario, como en la imagen anterior, seleccionar previamente el corpus con el que se desea trabajar.

ELV - El Grial

www.elv.dl/grial2009/busqueda_compleja4.php

el Grial Accede al | Quiénes somos | Qué hacemos | Déjate trabajar

Interfaz de etiquetaje e interrogación de corpus textuales

Búsqueda Simple | **Búsqueda Compleja** | El Manchador | Visualización del Out-Put del análisis

Búsqueda Compleja

Mostrar

Frecuencia Contexto

A=1 Ordenado Completo Parcial 5 Unidades (2) Der

Forma POS EDP ESS

| Secuencia | Token 1 | Token 2 | Token 3 | Token 4 | Token 5 | Token 6 |
|------------|---|---|---|---|---|---|
| Despliegue | <input checked="" type="radio"/> Forma <input type="radio"/> POS <input type="radio"/> EDP <input type="radio"/> ESS | <input checked="" type="radio"/> Forma <input type="radio"/> POS <input type="radio"/> EDP <input type="radio"/> ESS | <input checked="" type="radio"/> Forma <input type="radio"/> POS <input type="radio"/> EDP <input type="radio"/> ESS | <input checked="" type="radio"/> Forma <input type="radio"/> POS <input type="radio"/> EDP <input type="radio"/> ESS | <input checked="" type="radio"/> Forma <input type="radio"/> POS <input type="radio"/> EDP <input type="radio"/> ESS | <input checked="" type="radio"/> Forma <input type="radio"/> POS <input type="radio"/> EDP <input type="radio"/> ESS |
| Forma | <input type="checkbox"/> - |
| Lema | <input type="checkbox"/> - |
| Pos | <input type="checkbox"/> SUST - ADJ ADJ COMP ADJ SUP - |
| Género | <input type="checkbox"/> FEM - MSC AMB NEU - |
| Número | <input type="checkbox"/> SG - PL P1 P2 P3 ABR - |
| Dep Func | <input type="checkbox"/> SV - AUX PREP ENC CNPRE SUJ - |
| ESS | <input type="checkbox"/> PRV(oo) - Vrb Vkt ADV(tra) Adv Nt - |

[Enviar Consulta](#) [Limpiar](#) [Seleccionar Corpus](#)

Dado que en todos los casos debe realizarse una selección previa del corpus, es necesario seleccionar en base a los criterios de la propia investigación aplicados al contenido del mismo. Como se ve en la primera imagen de esta sección Al

día, tal mapa tiene su acceso directo en la portada del Proyecto (pulsante verde). Allí se encuentra una selección de artículos de investigación y de documentos de texto en los que es posible recuperar la información. Es posible que dicha forma no sea la más inmediata (esperaríamos un primer gráfico de acceso), pero las informaciones son complejas y requieren un trabajo de homogenización de las categorías de análisis bastante complejo. La entrada permite recuperar toda la serie de estadísticas por frecuencia completa y normalizada a 1000 palabras, de las etiquetas de un determinado corpus por lo que se presenta como un paso fundamental de la consulta, previa a los dos tipos de búsqueda referidos con anterioridad. Un ejemplo fragmentario de los rasgos que es posible recuperar en el subcorpus Articos-2004.

| Resultados | | |
|--------------------|--------------|--------------------|
| | ARTICOS-2004 | 1996495 |
| | Frecuencia | Normalizada a 1000 |
| | 330913 | 165.7470 |
| <?> <Abbr> N SG | 5044 | 2.5264 |
| <?> <Proper> N SG | 36239 | 18.1513 |
| <?> A COM PL | 1231 | 0.6166 |
| <?> A COM SG | 1203 | 0.6026 |
| <?> A FEM PL | 1457 | 0.7298 |
| <?> A FEM SG | 2909 | 1.4571 |
| <?> A MSC PL | 1854 | 0.9286 |
| <?> A MSC SG | 2145 | 1.0744 |
| <?> ADV | 170 | 0.0851 |
| <?> N COM PL | 1579 | 0.7909 |
| <?> N COM SG | 995 | 0.4984 |
| <?> N FEM PL | 2276 | 1.1400 |
| <?> N FEM SG | 8998 | 4.5069 |
| <?> N MSC PL | 2839 | 1.4220 |
| <?> N MSC SG | 4260 | 2.1337 |
| <?> N PL | 1649 | 0.8259 |
| <?> N SG | 18785 | 9.4090 |
| <?> V CND PL1 | | 0.0000 |
| <?> V CND PL3 | 1 | 0.0005 |
| <?> V CND SG1 | 1 | 0.0005 |
| <?> V CND SG2 | 1 | 0.0005 |
| <?> V CND SG3 | 2 | 0.0010 |
| <?> V GER | 30 | 0.0150 |
| <?> V IND IMPF PL1 | | 0.0000 |
| <?> V IND IMPF PL3 | 8 | 0.0040 |
| <?> V IND IMPF SG1 | 1 | 0.0005 |
| <?> V IND IMPF SG2 | 1 | 0.0005 |
| <?> V IND IMPF SG3 | 4 | 0.0020 |
| <?> V IND PRES PL1 | 14 | 0.0070 |

Como hemos podido ver, aunque solo parcialmente, el corpus de la Escuela de Postgrado en Lingüística de la Pontificia Universidad Católica de Valparaíso es una herramienta muy valiosa y puede aportar datos que la misma construcción del corpus proporciona con criterios de validez y confiabilidad. Sería fundamental que esta experiencia se repitiera en Universidades de otros países latinoamericanos. Eso permitiría acceder a una experiencia más consolidada de producción de textos propios de la comunicación académica en distintas variedades lingüísticas. Lo que destaca en esta experiencia es la posibilidad de consulta rápida y bastante directa porque, si bien no existe un corpus comparable por número de palabras y textos acumulados, el problema de los corpus de español académico de otras universidades reside en su dificultad de consulta en línea. En ese sentido, y por dar un ejemplo, existen noticias de un proyecto similar para el español peninsular realizado bajo el mecenazgo de la Fundación Comillas, pero a la fecha el material no resulta accesible en un portal web.

La experiencia del Proyecto ADIEU, ya referida en uno de los artículos de referencia, es cercana más a la enseñanza de ELE porque parte de un corpus de modelos nativos en las ejemplificaciones didácticas de macro y micro estructura de los géneros textuales escritos (manuales, monografías y ensayos, entre otros), pero recoge un corpus de géneros de los que no había debida documentación (clases magistrales y presentaciones orales). En el primer caso se trata de fragmentos de material publicado y seleccionado con fines didácticos, por lo que resulta de difícil reutilización en el ámbito de la investigación, pero el corpus de clases registradas es único en su género. Además de los productos editoriales ya referidos en el artículo de Llorián (en este número), existe un acceso online a parte del material sobre el cual es posible realizar investigaciones.

Homepage
userpage.fu-berlin.de/vazquez/vazquez/

Graciela Vázquez
ZE Sprachenzentrum
FU Berlin
E-Mail: vazquez@zedat.fu-berlin.de
Impressum

Cursos

Información

TICCAL (Comenius 2.1)
En castellano

Discurso académico y científico

ADIEU

MATERIALES

NOTICIAS

JORNADAS EICLE ENFOQUE INTEGRADO DE CONTENIDOS EN LENGUA EXTRANJERAS
(Enseñanza de asignaturas en castellano)
BERLIN, 21/22/23 de Septiembre de 2006
En la sede del Instituto Cervantes y en la Universidad Libre de Berlín

En la entrada ADIEU se accede a un banco de datos en el que es posible recuperar no solo información bibliográfica, ponencias y materiales sino también las transcripciones de clases magistrales de diferentes áreas universitarias.

HABILIDADES ACADÉMICAS

Autora o autor - ELE - Castellano Primera Lengua - Ponencias - Materiales - Clases magistrales

| CLASES MAGISTRALES | |
|--------------------|----------|
| Arqueología | Práctica |
| Derecho | Práctica |
| Economía | Práctica |
| Lingüística | Práctica |
| Literatura | Práctica |

Participan

- Freie Universität Berlin - ZE Sprachenzentrum
- Universidad de Buenos Aires - Facultad de Filosofía y Letras - Facultad de Ciencias Exactas y Naturales
- Universidad de Córdoba - E.L.E. - Facultad de Lenguas
- Universidad Nacional General San Martín
- Universitat Pompeu Fabra
- Universidad de Tucumán - Filosofía y Letras >>> Carreras

El hecho de tener las transcripciones ya realizadas permite la construcción de un corpus de estudio en el que aplicar rápidamente las herramientas propias de la investigación cuantitativa y de facilitar la cualitativa.

Por ahora no tenemos a disposición un corpus de géneros académicos de producción ELE no nativa. Esa falta de recursos pone en seria dificultad la posibilidad de una investigación extensiva, no reducida a intentos más o menos individuales.