

Al día

María Cecilia Ainciburu

Università degli Studi di Siena

La sección “Al día” contiene una revisión de la bibliografía específica sobre el tema central del artículo que se presenta y se discute en la correspondiente edición de esta Revista. Además, intenta ilustrar los puntos relevantes del debate científico de los últimos cinco años por medio de otras secciones temáticas donde se señalan algunos artículos recientes que pueden interesar a los estudiosos de la lingüística aplicada al español como lengua extranjera.

En este apartado no se pretende recopilar una bibliografía completa sobre el tema, sino seleccionar artículos de interés para el investigador. Nuestros lectores pueden señalar otras referencias bibliográficas que puedan resultarles de interés, escribiendo a ainciburu@unisi.it.

En el caso específico de este número metodológico se ha elegido hacer referencia a los corpus que pueden ser utilizados para describir la interlengua de estudiantes de lenguas extranjeras. La bibliografía general sobre el tema “Interlengua” se encuentra en los artículos publicados en este número.

1. Corpus del español

1. 1 Corpus de producciones orales de estudiantes anglohablantes (también nativos)

Mitchell, Rosamond (2008). SPLLOC : a new database for Spanish second language acquisition research. University of Southampton.

Este proyecto tiene dos objetivos: establecer en pequeña escala pero con alta calidad, una base de datos de alumnos ELE y llevar a cabo un breve programa de actividades sustantivas de investigación en español L2 (segunda lengua). Los datos serán recogidos en el aula los alumnos ingleses de español, desde los principiantes a nivel avanzado. Ha sido diseñado especialmente utilizando tareas de elicitación. Para tener un término de comparación, un grupo de hablantes nativos también realizó las mismas tareas. La base de datos resultante contiene las grabaciones digitales, acompañadas de las transcripciones en CHILDES (Child Idioma Data Exchange System. El material está libremente disponible para su uso a través de un sitio web especialmente creado. La investigación posterior investigará el programa de adquisición de las propiedades morfosintácticas del español: el orden de las palabras, los pronombres clíticos, morfología verbal y las formas interrogativas, proporcionando

una descripción y un análisis de las secuencias de desarrollo de la L2 español de una perspectiva de interfaz.

Tracy-Ventura, Nicole (2008). Spanish Learner Language Oral Corpus project (SPLLOC 1). A New Corpus of Oral L2 Spanish. Actualmente en <http://www.splloc.soton.ac.uk/> y en Talkbank (<http://www.talkbank.org/>).

Los datos de producción oral en español L2 han sido recogidos en el aula de escuelas y universidades de Inglaterra, utilizando una serie de tareas de elicitación diseñadas especialmente. Éstas incluyen la narración, la descripción de imágenes, el debate sobre un determinado tema y la entrevista individual. Hay 20 alumnos (L1 Inglés) en cada uno de 3 niveles: principiantes (novenio año, estudiantes de edades 13-14), intermedio (A2 estudiantes de 17-18 años) y los estudiantes muy avanzados (cuarto año estudiantes de edades 20-21). Dependiendo de su nivel, cada alumno fue grabados en un número variable entre 3 y 5 tareas orales. También se registró la producción de un pequeño número de hablantes nativos en las mismas tareas. La base de datos resultante contiene 290 pistas de audio digital (240 grabaciones - 168.776 palabras, y 50 grabaciones de hablantes nativos - 37238 palabras), las correspondientes transcripciones y archivos XML. Algunos archivos incluyen etiquetado y todos se han transcripto con las normas CHILDES.

Grupo de investigación Woslac (2007) *Proyecto CEDEL2*. Universidad Autónoma de Madrid. En: <http://www.uam.es/woslac/cedel2.htm>

Este proyecto se enmarca dentro de las directrices generales de otros proyectos europeos para la creación de grandes corpus escritos de segundas lenguas. Consta de redacciones de alumnos ELE anglófonos (500 palabras cada una) que fueron realizadas en base a temas predeterminados. Posee datos referidos a diferentes niveles de competencia y se puede acceder a la ficha que identifica al informante.

Koike, Dale (2007). *Spanish Learner Corpus and Exercises (SLCE) project*. Universidad de Texas. En: <http://www.laits.utexas.edu/slce>.

El proyecto tiene como objetivos:

1. proporcionar datos a los alumnos de español y de lingüística aplicada a la adquisición de un segundo idioma con el fin de familiarizarlos con el idioma que se aprende en los diferentes niveles;
2. centrarse interactivamente en las características que pueden ayudar a indicar el nivel de los alumnos;

3 proporcionar información sobre el discurso a través de un trabajo de evaluación de los datos;

4. servir como un corpus de datos de alumnos de español que pueden ser utilizados como base para futuras asignaciones de clase en los Cursos de español y en proyectos de investigación más avanzados de pregrado y postgrado.

1. 2. Corpus de producciones escritas de estudiantes ELE de diferentes lenguas

Cestero Mancera, Ana María; Penadez Martinez, Inmaculada y otros (2001) Corpus para el Análisis de errores de aprendices de E/LE (*CORANE*) Universidad de Alcalá.

Se trata de un corpus de composiciones escritas realizadas por alumnos de los cursos de lengua del año 2000 de la Universidad de Alcalá. Los informantes pertenecen a los cuatro viejos niveles de competencia del Instituto Cervantes, o sea, elemental, intermedio, avanzado y superior. El total de composiciones recogidas, almacenadas y etiquetadas es 1091, de las cuales 1011 realizadas con materiales de apoyo y 80 sin ellos. 43 de ellas han sido realizadas por estudiantes de nivel elemental, 432 pertenecen a alumnos de nivel intermedio, 213 a estudiantes de nivel avanzado y 403 a estudiantes de nivel superior. Hasta el momento, han participado en nuestro corpus 321 estudiantes diferentes: 236 mujeres y 85 hombres, que tienen como lengua materna el japonés (58 estudiantes), el inglés (57), el alemán (36), el francés (29), el sueco (29), el italiano (18), el portugués (18), el japonés (22), el coreano (16), el chino (9), el turco (6), el húngaro (4), el polaco (3), el griego (3), el árabe (4), el ruso (2), el fang (2), el holandés (2), el flamenco (1), el danés (1), el finés (1), el esloveno (1), el pakistaní (1) y el checo (1); además hemos contado con 19 estudiantes bilingües (ewe/inglés, francés/flamenco, francés/inglés, francés/alemán, ruso/polaco, inglés/irlandés, inglés/africano, etc.). Tiene fichas por estudiante y permite trabajar con la producción longitudinal de un mismo informante.

1.3 Corpus de producciones escritas de estudiantes brasileiros de español, inglés y alemán/LE

Tagnin, Stella E. (2002) *USP Multilingual Learner Corpus (MLC)* . Acceso a través del Departamento de Lenguas modernas de la Universidad de San Pablo (Brasil)

El corpus contiene las composiciones en inglés, español y alemán de alumnos de la Universidad de San Pablo (Brasil). Los niveles en español en el Campus (SOC) son Básico, Intermedio y Avanzado. Posteriormente también se ofrecen módulos semestrales de a) Cultura, b), Literatura, y c) de conversación. A diferencia de los corpus de las otras lenguas, el del español se basa en material preparado por sus propios profesores y sólo se basa en el perfil y en un análisis de las necesidades de sus estudiantes. Se matricularon 460 estudiantes, lo que da un total de 920 textos por semestre.

1.4 Corpus de producciones escritas de estudiantes de lenguas extranjeras con diferentes lenguas maternas.

Gilquin, Gaëtanelle, Szilvia Papp y María Belén Diez-Bedmar (Eds.) (2008). *Linking up contrastive and learner corpus research*. Amsterdam: Rodopi

Los trabajos reunidos en este volumen intentan explorar, a través de corpus de datos, la relación entre interlengua contrastiva y análisis de errores. El corpus se aborda desde una perspectiva contrastiva, comparando las producciones de los alumnos con otras pertenecientes a un corpus de nativos o con datos de otros alumnos procedentes de la misma lengua materna. La integración de estos dos marcos, análisis contrastivo y producciones de alumnos permite destacar aspectos cruciales de la producción de éstos, tales como las características del no ser nativos (errores, la sobreexplotación y subutilización, expresiones no idiomáticas), incluyendo las características de la interlengua universal, o cuestiones más generales como la cuestión de la transferencia. Los diez artículos de este volumen abarcan temas que van desde la metodología para la sintaxis (por ejemplo, la colocación de adverbio, postverbal sujetos), a través de las colocaciones) y el discurso (por ejemplo, información de las etiquetas, la elección temática). Entre los idiomas examinados figuran los idiomas inglés, chino, holandés, francés y español. El libro será de interés para una amplia gama de lectores, especialmente los investigadores en adquisición de un segundo idioma y la lingüística contrastiva, pero también para los profesionales que

trabajan en la enseñanza de lenguas extranjeras, tales como profesores de idiomas. El libro incluye los corpus analizados.

2. Corpus de inglés¹

2.1 Corpus de producciones escritas de estudiantes de inglés de diferentes lenguas

AAVV (2008) *Cambridge International Corpus (CIC)*. Cambridge: Cambridge University Press

Es un corpus de más de 30 millones de palabras y en constante actualización. Contiene las producciones de las siguientes certificaciones que se realizan en 190 países y con informantes de 130 lenguas maternas diferentes

KET Key English Test

PET Preliminary English Test

FCE First Certificate in English

CAE Certificate in Advanced English

CPE Certificate of Proficiency in English

BEC Business English Certificate (all levels)

IELTS International English Language Testing System

CELS Certificates in English Language Skills

Al corpus se aplica un sistema de etiquetado de errores y es posible realizar listas y análisis estadísticos mediante el sistema de tratamiento interno que contiene la base de datos. El acceso está limitado a los miembros de Cambridge ESOL y a los autores que trabajan para la Cambridge University Press

¹ La *Revista Nebrija* se ocupa del aprendizaje de todas las lenguas extranjeras. En este número nos interesaba focalizar la atención sobre las bases de datos que permiten realizar Análisis de errores en la producción de estudiantes ELE. De todos modos, hemos querido introducir esta citación porque nos parece que el mismo proyecto podría ser imitado por las Asociaciones certificadoras del español. En la actualidad se puede acceder, realizando una solicitud y pago, a las pruebas escritas de algunos exámenes pero no a los orales.