

La aplicación de un corpus de aprendices en la autocorrección de composiciones escritas

HUI-CHUAN LU

SHENG-YUNG HUNG

National Cheng Kung University

LO HSUEH LU

Providence University

Recibido: 11 abril 2015 / Aceptado: 29 junio 2015

ISSN: 1697-7467

RESUMEN: El presente trabajo evalúa la efectividad de aplicar el Corpus Escrito de Aprendices Taiwaneses de Español en el aprendizaje del idioma español como lengua extranjera, y la autocorrección de la redacción, para evaluar los resultados de las composiciones escritas de los estudiantes, y así entender mejor el valor de la aplicación de dicha herramienta académica. Concluimos que la aplicación del corpus puede facilitar la extensión de la longitud de los textos, evitar los usos inapropiados de la lengua española, y ayudar a corregir los errores de categorías léxicas y de los sustantivos.

Palabras clave: corpus de aprendices, autocorrección, redacción, aplicación.

The Application of a Learners' Corpus for Self-Correction in Written Compositions

ABSTRACT: Previous studies about the application of corpus were limited to the type of native corpus or to English language without analyzing the effectiveness of learning. The present study evaluates the effectiveness of applying the created corpus CEATE (Corpus Escrito de Aprendices Taiwaneses de Español) to foreign language learning and self-correction in writing. From this, we make use of the creation to understand its academic and application value. Our results have indicated that the corpus facilitates the extension of textual length, improves inappropriate usages, and helps participants to correct errors for lexical category and nouns.

Keywords: learners' corpus, self-correction, writing, application.

1. INTRODUCCIÓN

En las últimas dos décadas, la aplicación de corpus ha extendido a casi todas las sub-áreas de la lingüística teórica y aplicada. Los estudios anteriores relacionados con la aplicación de un corpus, generalmente, se limitaban al tipo de corpus nativo y a la lengua inglesa. En otras lenguas no se han realizado muchos estudios en esta área. Además, la mayoría de los estudios anteriores se centran en la aplicabilidad de dicha herramienta a la enseñanza sin evaluar la efectividad de su aplicación. En nuestros estudios anteriores, hemos investigado la aplicación del CREA (Corpus de Referencia del Español Actual) y concluimos que los conocimientos léxicos se obtienen con más facilidad que los gramaticales.

Este trabajo tiene como meta la aplicación del corpus de aprendices CEATE al autoaprendizaje del español como lengua extranjera, para aprovechar su contenido y al mismo tiempo extender su valor académico, después de haber dedicado varios años a su desarrollo. A fin de verificar la efectividad del existente corpus escrito de aprendices CEATE, en su aplicación al autoaprendizaje, se realizó un experimento de evaluación con dos grupos de participantes, el experimental y el de control. Basándonos en los resultados de tres pruebas de composiciones realizadas y aprovechando la herramienta auxiliar –CorpusTool– que facilita hacer anotaciones manuales, estudiamos la diferencia entre los dos grupos de participantes, analizando los errores cometidos y clasificando sus categorías según diferentes niveles lingüísticos.

El contenido de este trabajo se organiza como sigue: La sección 2 se dedica a repasar estudios anteriores, mientras que la metodología se describirá en la sección 3. Luego, los resultados del estudio se señalan en la sección 4. Las limitaciones, las implicaciones pedagógicas y los trabajos futuros se presentan en la sección 5. Finalmente, en la sección 6 se encuentran las conclusiones.

2. ESTUDIOS ANTERIORES

La mayoría de los estudios relacionados con la aplicación de corpus generalmente se asocian con la lengua inglesa (Gilquin, Granger & Paquot (2007), Granger, Kraif, Ponton, Antoniadis & Zampa (2007), Tono (2007), Dalziel & Helm (2008), Mukherjee (2008), Pendar & Chapelle (2008); entre otros). Además, entre los tipos de *córpore* aplicados, los de nativos son más populares que otros tipos de *córpore* y los estudios con respecto a los *córpore* de aprendices son relativamente escasos. Atribuimos esta tendencia a la disponibilidad de *córpore*, porque se ordenan de mayor a menor cantidad. Por un lado, los de inglés, de chino, de alemán, de francés y de italiano; y por el otro, los escritos, orales, históricos, paralelos y los de aprendices, según el cálculo de tipos de *córpore* listados por Lee (2010). Con respecto a los temas de aprendizaje estudiados, los léxicos superan a los demás y los gramaticales varían por sus características específicas. Con respecto a la aplicación de *córpore* paralelos en el aprendizaje, por ejemplo, St. John (2001) indica que tiene una influencia positiva sobre todo para los primerizos. Los *córpore* de aprendices no llaman tanto la atención como los nativos o paralelos. Además, la interfaz, las complicadas anotaciones y las funciones de consulta pueden afectar la voluntad de los usuarios en la aplicación. De ahí que, Pendar & Chapelle (2008) creen que la aplicación de corpus de aprendices requiere un estudio más profundo. Granger et. al. (2007) indican la ayuda de los *córpore* de aprendices en el aprendizaje de la cognición del lenguaje. Gilquin et. al. (2007) enfatizan la importancia de *córpore* de aprendices en la enseñanza de inglés, en contra de los estudios que exclusivamente adoptan los *córpore* nativos. Mukherjee (2008) opina que se debe aprovechar bien el recurso de los *córpore* de aprendices. Dalziel & Helm (2008) apuntan que los *córpore* de aprendices pueden guiar a éstos en el aprendizaje de la lengua mediante autoconsulta. Estos estudios confirman el valor positivo de aplicar estos *córpore* de aprendices en el aprendizaje. Sin embargo, no hay estudios empíricos que puedan ofrecer evidencias concretas para probar su efectividad.

Los estudios sobre la aplicación del corpus en la enseñanza de otras lenguas extranjeras, aparte de los de inglés, son muy limitados. Por ejemplo, Belz & Vyatkina (2008) indican que aunque la aplicación de *córpore* de aprendices llama gradualmente más la atención, aún no sabe entrado en el análisis de su efectividad: por este motivo han realizado, un estudio

sobre la interacción de aprendices alemanes concretamente. Vyatkina (2007) utiliza los textos de un corpus de aprendices de alemán como su fuente para los materiales didácticos. Sin embargo, la aplicación de corpórea españoles sólo se limita a los nativos excluyendo los de aprendices debido a la riqueza de datos y a las poderosas funciones de consulta, por ejemplo, el Corpus de Referencia del Español Actual (CREA) desarrollado por la Real Academia Española (RAE), y el Corpus del Español de Davies. Las obras pertinentes son Chenoll (2008), Pitkowski & Vásquez Gamarra (2008), que limitan a la aplicación de corpus a la enseñanza sin incluir ningún experimento para verificar su efecto. Lu (2007) y Lu, Lin & Pai (2007) investigan el uso del CREA en el aprendizaje de los aprendices de español y concluyen que, para conseguir un mejor resultado de aprendizaje, se debe tener en cuenta las variables tales como el nivel de los participantes y la manipulación de la herramienta. Además, los efectos significativos varían, por ejemplo, se aprende mejor el tema léxico que el gramatical, y las palabras mejor que las frases.

Además de los estudios mencionados anteriormente y debido a la conveniencia de las herramientas del corpus como Sketch Engine, WordSmith, AntConc y ParaConc, los análisis a partir del internet llegan a ser una corriente en el estudio de la lingüística de corpus. No obstante, la abundante cantidad de datos sin registrar su origen ni la calidad puede traer factores incontrolables que afectan los resultados de aprendizaje. Por consiguiente, la aplicación de corpórea a partir de internet requiere más estudios para evaluar su factibilidad.

En resumen, actualmente todavía hacen falta los estudios empíricos de la aplicación de los corpórea de aprendices en el aprendizaje del español. Con el objetivo de avanzar un paso más aprovechando el recurso que tenemos y comparar con los estudios anteriores, vamos a profundizar en la comprensión del corpus de aprendices CEATE mediante el siguiente experimento y análisis.

3. PREGUNTAS DE ESTUDIO Y METODOLOGÍA

3.1. Propósito y pregunta de estudio

El presente trabajo tiene como objetivo estudiar el valor de la aplicación del corpus CEATE, que hemos construido, y centrándose en el autoaprendizaje del español como lengua extranjera. En correspondencia con el propósito de estudio establecido, planteamos la siguiente pregunta de estudio: ¿En qué aspecto mejora el resultado de autocorrección en la redacción la aplicación de un corpus de aprendices?

3.2. Metodología

Para contestar esta pregunta de estudio planteada, realizamos tres pruebas y comparamos las diferencias del progreso entre dos grupos de participantes.

3.2.1. Participantes

Diez estudiantes que habían aprendido español durante 400 horas se dividieron en dos grupos, cinco en el experimental y cinco en el de control, todos participaron en tres pruebas.

Su nivel de español se clasificó según la nota obtenida en la clase de español al final de un año académico. El promedio de ambos grupos es de 83. La nota más baja y la más alta del grupo experimental son 72,55 y 91,57 respectivamente. Mientras que las del grupo de control son 79,99 y 85,65.

3.2.2. Herramienta: Corpus Escrito de Aprendices Taiwanesees de Español, CEATE

En el grupo experimental, se adoptó el corpus escrito de aprendices CEATE (<http://corpora.fld.ncku.edu.tw/>) como instrumento auxiliar, para verificar la efectividad de su aplicación en la autocorrección durante el proceso de redacción en español; estos experimentos se realizaron en un aula donde cada participante pudo usar una computadora con internet para acceder al CEATE. Mientras que al grupo de control, no se le permitió usar ni diccionario ni texto gramatical en un aula sin ordenadores durante las pruebas de redacción.

La herramienta se utilizó cuando los estudiantes tenían dudas con respecto a lo que estaban escribiendo. Ellos podían teclear las palabras clave, seleccionar una de las cuatro funciones principales de consulta (palabras, palabras [POS], palabras [POS]{POS}, lema) para que se apareciera la lista de concordancia que contenía las oraciones con las palabras consultadas. Un paso más avanzado, los estudiantes podían comparar los textos originales y los corregidos contrastando los errores cometidos y las correspondientes correcciones para la autocorrección. Los pasos se señalan en el Diagrama 1.

Diagrama 1: Pasos para utilizar el CEATE

The screenshot shows the CEATE web interface. At the top, there are navigation tabs for '日文', '英文', '俄文', '中文', and '學習網頁'. Below this is a search bar with '帳號: Administrat' and '密碼: *****'. The main content area is titled '台灣西語學習者書面語料庫' and 'Corpus Escrito de Aprendices Taiwanesees de Español'. It features a search form with the following fields: 'ser' in the search box, 'Tipo de consulta: Palabras', 'Categoría gramatical: POS [Adj]', 'Origen de composición: Composición original', 'Hora de aprendizaje: 0 hora - 2500 hora', 'Sexo de aprendiz: No se especifica', 'Departamento de aprendizaje: No se especifica', 'Experiencia relevante de aprendizaje: No se especifica', 'Longitud de composición: 0 palabras - 5000 palabras', 'Tipo textual de composición: No se especifica', 'Tema de composición: No se especifica', and 'Lugar (donde se escribió la composición): No se especifica'. Below the form, it says 'Página 1 2 3 4' and 'En total hay 203 caso(s) almacenados en 147 texto(s) extraídos de un subcorpus de 12062 Palabras.' There is a list of 5 search results, with the first one being: '1. Hubo una empresa que quería firmarme, pero siendo la educación ser tan importante, por eso, refusé la invitación.' At the bottom, there is a comparison table:

Opción consultada	Hubo una empresa que quería firmarme, pero siendo la educación ser tan importante, por eso, refusé la invitación.
Composición corregida	Composición original

3.2.3. Proceso

Las actividades de la evaluación duraron 5 semanas. En la primera semana (iniciada el 5 de marzo del 2009), los participantes de ambos grupos recibieron una lección compuesta de 3 secciones: (1) la enseñanza de seis estrategias para el aprendizaje, (2) la enseñanza de cómo redactar, siguiendo el ejemplo de composición titulada “Mi autobiografía” y (3) la enseñanza sobre cómo utilizar el CEATE.

En la segunda semana (iniciada el 12 de marzo), continuamos la enseñanza de redacción mediante el ejemplo de composición “Viaje/vacaciones”. Luego, en las semanas del 19 y 26 de marzo y la semana del 11 de junio, realizamos tres pruebas (prueba 1, 2 y 3) cuyos títulos fueron “Presentación de mí mismo/a”, “Vacaciones de invierno” y “Recuerdo con mi familia”, respectivamente. Los participantes se separaron en dos aulas para escribir las composiciones con tres temas diferentes correspondientes a los materiales didácticos. El grupo experimental pudo consultar el corpus de aprendices CEATE en las pruebas, mientras que no se permitió el uso de instrumentos auxiliares en el grupo de control. En total, recopilamos 30 composiciones con 4.917 palabras, 10 textos por cada prueba con un promedio de 160 palabras por cada composición.

3.3. Evaluación del progreso de los estudiantes

Los resultados del progreso se evaluaron usando las tres pruebas de composición escrita preparadas por los participantes; identificando los errores cometidos y clasificando sus categorías según diferentes niveles lingüísticos, proceso que se realizó mediante la ayuda del CorpusTool, que facilita las anotaciones manuales, las consultas y el cálculo de diferentes rasgos. En total, realizamos 1.458 anotaciones y clasificamos los rasgos según el uso (uso excesivo, insuficiente o inapropiado), la ortografía y la morfología, la parte de la oración (sustantivo, artículo, adjetivo, pronombre, verbo, adverbio, preposición, conjunción), la sintaxis (orden, funciones gramaticales) y el léxico. Los ejemplos se muestran en las oraciones (1)-(9), los cuales están divididos en las diferentes categorías.

- (1) Uso excesivo:
 - (1a) Oración original:
Vamos a graduar en este semestre y cada persona tiene sus plan.
 - (1b) Oración corregida:
Vamos a graduarnos este semestre y cada persona tiene sus planes.
- (2) Uso insuficiente:
 - (2a) Oración original:
Deseo mis amigos y yo que tenemos buen futuro y trabajo.
 - (2b) Oración corregida:
Deseo que mis amigos y yo tengamos un buen futuro y un buen trabajo.
- (3) Uso inapropiado:
 - (3a) Oración original:
Mis estudiantes y yo practicamos al voleibol en cada martes...
 - (3b) Oración corregida:
Mis compañeros y yo practicamos voleibol todos los martes...

- (4) Forma:
- (4a) Oración original:
...y vamos a competiciones con otros universidades el fin de semana en Taipei.
- (4b) Oración corregida:
...y vamos a competir contra otras universidades el fin de semana en Taipéi.
- (5) Gramática:
- (5a) Oración original:
Vamos a graduuar en este semestre y cada persona tiene sus plan.
- (5b) Oración corregida:
Vamos a graduarnos este semestre y cada persona tiene sus planes.
- (6) Léxica:
- (6a) Oración original:
Mis estudiantes y yo practicamos al voleibol en cada martes,...
- (6b) Oración corregida:
Mis compañeros y yo practicamos voleibol todos los martes...
- (7) Género de sustantivo:
- (7a) Original:
Ella es como mi amigo y charlamos de todas las temas.
- (7b) Corregida:
Ella es como mi amiga y charlamos de todos los temas.
- (8) Número de sustantivo:
- (8a) Original:
Por eso, hago más tiempos con mi mamá y mi hermanoa que mi papá.
- (8b) Corregida:
Por eso, paso más tiempo con mi mamá y mi hermana que con mi papá.
- (9) Orden de sustantivo:
- (9a) Original:
Estoy una estudiante de Cheng Kung Universidad.
- (9b) Corregida:
Soy estudiante de la Universidad de Cheng Kung.

En el cálculo de errores se tiene en cuenta la longitud de los textos escritos para obtener un estándar de comparación. La fórmula para calcular la tasa de errores se señala como sigue: la cantidad de errores dividido por la cantidad de palabras en una composición. Por ejemplo, si un participante comete 46, 27 y 49 errores en las tres pruebas de composición que contienen 134, 115 y 163 palabras, su tasa de errores por composición serían de 34%, 23% y 30% respectivamente. En cuanto al desarrollo del aprendizaje, las tres pruebas se comparan una con otra.

4. RESULTADOS Y DEBATE

En esta sección, vamos a analizar el desarrollo entre los participantes del grupo experimental y los del grupo de control comparando las tasas de errores entre diferentes pruebas

según las categorías lingüísticas clasificadas para las anotaciones del esquema, como se señala en el Cuadro 1.

Cuadro 1
Comparación entre dos grupos

Grupo	Experimental			De control		
	Prueba 1	Prueba 2	Prueba 3	Prueba 1	Prueba 2	Prueba 3
Longitud (palabras)	153,4	159,4	182	182,4	153,4	152,8
Categoría léxica (%)	6,11%	4,40%	5,70%	4,82%	3,74%	6,25%
Categoría gramatical: N (%)	1,36%	1,67%	0,95%	1,76%	2,74%	1,52%

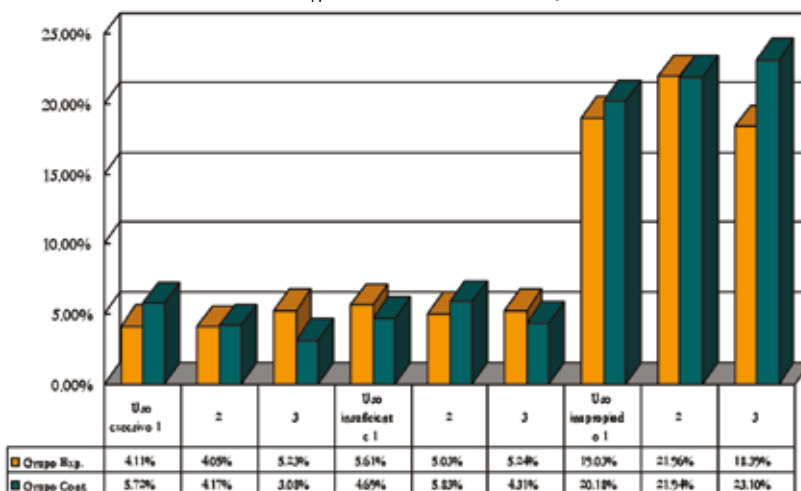
4.1. Longitud

Como se señala en el resultado obtenido según la longitud textual (los promedios de palabras escritas por estudiantes en cada prueba) del Cuadro 1, los promedios de la longitud de las composiciones escritas por el grupo experimental aumentan a lo largo de las tres pruebas (153, 159 y 182 palabras), mientras que los del grupo de control descienden (182, 153 y 152 palabras). Parece que la herramienta ayuda a los participantes a extender la longitud de los textos, es decir, los alumnos escriben más palabras y con más facilidad utilizando el corpus como una herramienta auxiliar, especialmente en el aspecto del contenido del texto.

4.2. Uso

En el Diagrama 2 de esta sección, mostramos los tres usos evaluados.

Diagrama 2: Tres usos analizados



Observamos que el uso inapropiado es el error cometido con mayor frecuencia por los participantes, tanto del grupo experimental como del grupo de control, como se señala en el Diagrama 2: 19,03%, 21,96% y 18,39% son mayores de 5,61%, 5,03% y 5,24% y mayores de 4,11%, 4,05% y 5,23% en las tres pruebas 1, 2 y 3 del grupo experimental y 20,18%, 21,94% y 23,10% superan a 4,69%, 5,83% y 4,31% y a 5,72%, 4,17% y 3,08% en las tres pruebas del grupo de control respectivamente.

Si nos concentramos en la tendencia de los usos inapropiados de los alumnos, nos damos cuenta de que la cifra del grupo experimental (19,03%) sube en la segunda prueba (21,96%) y baja en la tercera prueba (18,39%) mientras que la cifra del grupo de control (20,18%) sube en la segunda prueba (21,94%) y sigue subiendo en la tercera prueba (23,10%). Se indica que los participantes del grupo experimental cometen menos errores en la tercera prueba, pese a la cantidad creciente de los errores cometidos en la segunda prueba, en comparación con el grupo de control. Puede ser que la aplicación del corpus de aprendices facilite la corrección de usos inapropiados después de que los participantes se familiaricen más con la herramienta y la aprovechen como una consulta útil e importante cuando escriben las composiciones.

En los siguientes apartados, se trata el resultado de errores cometidos por los participantes según las categorías lingüísticas desde dos aspectos, léxicos y gramaticales.

4.3. Categoría léxica

Con respecto a los diferentes niveles lingüísticos de los dos grupos, como se muestra en la segunda fila del Cuadro 1, las cifras de ambos grupos bajan en la segunda prueba y suben en la tercera: 6,11%, 4,4% y 5,7% para el grupo experimental y 4,82%, 3,74% y 6,25% para el grupo de control. Sin embargo, los cambios de las cifras indican que se mejora el resultado del aprendizaje con un mayor descenso de los errores en la segunda prueba (pos-prueba) y menor aumento en la tercera prueba para el grupo experimental. Es posible que la herramienta facilite las correcciones de errores léxicos, especialmente después de que los alumnos tienen más experiencia utilizándola. Este resultado es similar al de la aplicación de corpus nativo que se observó en Lu (2007) y Lu et al. (2007).

4.4. Categoría gramatical: Sustantivo

Por último, se presenta el resultado de los errores gramaticales cometidos en los dos grupos, como se señala en la última fila del Cuadro 1. Según el resultado, las cifras de ambos grupos suben en la segunda prueba y bajan en la tercera: 1,36%, 1,67% y 0,95% para el grupo experimental y 1,76%, 2,74% y 1,52% para el grupo de control. Sin embargo, las cifras aún indican que el grupo experimental se comporta mejor al cotejarlo con el grupo de control. Además, si comparamos las categorías léxicas y las gramaticales, observamos que los participantes cometen muchos más errores léxicos que errores de sustantivos. Conectamos esta tendencia con las características comunes de asociación semántica.

Debido al proceso complicado de evaluación, el duro trabajo de hacer anotaciones, y la escasez de alumnos que han estudiado español el suficiente tiempo en un departamento de lenguas extranjeras, el presente estudio se limita a un análisis de datos provenientes de pocos participantes. En el futuro, realizaremos un estudio con más participantes para obtener una generalización más amplia.

5. IMPLICACIONES PEDAGÓGICAS

La aplicación de un corpus de aprendices se distingue de un corpus nativo en muchos aspectos, tanto por su rasgo natural como en sus funciones de consulta, por lo tanto, las características deben destacarse de una manera apropiada. Debemos aprovechar bien las características que posee esta herramienta, las cuales no pueden sustituirse por otras, para conseguir una mejor efectividad en la enseñanza y el aprendizaje (por ejemplo, el contraste entre los textos originales de los aprendices y los de nativos en vez de los textos inventados por profesores; así como los rasgos tales como los usos inapropiados, los errores léxicos y de sustantivos).

Es innegable que la autocorrección durante el proceso de redacción mediante un corpus de aprendices puede causar dificultad cuando los usuarios todavía no son capaces de contrastar las diferencias entre los textos de estudiantes de ELE y los de los nativos. Por lo tanto, es necesario enfatizar en las instrucciones sobre el uso apropiado de la autocorrección antes de comenzar a usar la herramienta. Después de dedicar más esfuerzo y tiempo en la familiarización con la herramienta, un aprendiz activo puede apreciar las triviales distinciones e identificar las posibles causas de errores para mejorar sus conocimientos lingüísticos.

Además, el valor aplicado de la herramienta en cuestión puede ser más accesible si se adopta en la enseñanza. Las explicaciones explícitas y los conceptos lingüísticos proporcionados por los enseñadores respecto a los contrastes entre los errores y las revisiones pueden ser mucho más significativos que las interpretaciones de frases hechas en los textos gramaticales.

6. CONCLUSIONES

En este trabajo, hemos evaluado la aplicación del corpus de aprendices CEATE creado por nosotros a la autocorrección durante el proceso de redacción en español. Ello constituye un paso importante, ya que en estudios anteriores relacionados con la aplicación de corpus, apenas se había realizado muchos estudios en esta área.

Después de analizar el desarrollo entre los participantes del grupo experimental y los del grupo de control, concluimos que la aplicación del corpus puede facilitar la extensión de longitud de los textos, evitar los usos inapropiados y ayudar a corregir los errores de categorías léxicas y de los sustantivos.

Como la aplicación de un corpus de aprendices requiere tiempo y esfuerzo para familiarizarse con la herramienta antes de conseguir un resultado positivo en la autocorrección, son indispensables las instrucciones claras antes de usar la herramienta y las prácticas de consultas. Este paso esencial ayuda a asegurar que todos los usuarios sean capaces de manejar el creado corpus de aprendices para obtener el mejor efecto de aplicación. Con todo esto, se espera que la herramienta mejore el resultado de autocorrección a fin de conseguir un autoaprendizaje más eficaz.

7. AGRADECIMIENTOS

En primer lugar, queremos mostrar nuestro agradecimiento al Consejo Nacional de Ciencia y al Ministro de Educación de Taiwán por su apoyo financiero en el proyecto NSC99-2410-H-006-092-MY2. Así como al desarrollador de la herramienta gratuita –CorpusTool. Además, queremos agradecer a los profesores que nos brindaron su valiosa ayuda en el proceso de recolección de datos para el CEATE. Por último, pero no menos importante, damos las gracias a todos los ayudantes y los participantes por su trabajo en la construcción del CEATE.

8. REFERENCIAS BIBLIOGRÁFICAS

- Belz, J. A. and Vyatkina, N. (2008). “The pedagogical mediation of a developmental learner corpus for classroom-based language instruction”, in *Language Learning & Technology*, 12, 3: 33-52.
- Chenoll, A. (2008). “El e-corpus en la enseñanza de E/LE: del aprendiz al aprendiente”, presentado en el IV Foro de profesores de E/LE. Valencia, Spain.
- Dalziel, F. and Helm, F. (2008). “Exploring modality in a learner corpus of online writing”, in *Linguistic Insights - Studies in Language and Communication*, 74: 287-302.
- Davies, M. (2006). *A frequency dictionary of Spanish: Core vocabulary for learners*. New York, NY: Routledge.
- Gilquin, G., Granger, S. and Paquot, M. (2007). “Learner corpora: The missing link in EAP pedagogy”, in *Journal of English for Academic Purposes*, 6, 4: 319-335.
- Granger, S., Kraif, O., Ponton, C., Antoniadis, G., and Zampa, V. (2007). “Integrating learner corpora and natural language processing: A crucial step towards reconciling technological sophistication and pedagogical effectiveness”, in *ReCALL*, 19, 3: 252-268.
- Lee, D. (2010). *Bookmarks for corpus-based linguists*. <http://www.uow.edu.au/~dlee/CBLLinks.htm>, accessed 21 May, 2012.
- Lu, H.-C. (2007). “Corpus-based study of Spanish teaching in Taiwan: Construction and application of corpus”, in *Foreign Language Studies*: 1-22.
- Lu, H.-C., Lin, L.-T. and Pai, F.-I. (2007). “Corpus-based study of foreign language teaching: Analysis of collocation”, in *Foreign Language Studies*, 6: 39-58.
- Mukherjee, J. (2008). “English corpus linguistics and foreign language research: Line of development and perspectives”, in *ZFF, Zeitschrift für Fremdsprachenforschung*, 19, 1: 31-60.
- O'Donnell, M. (2008). “The UAM CorpusTool: Software for corpus annotation and exploration”, in Bretones Callejas, Carmen M. et al. (eds) *Applied Linguistics Now: Understanding Language and Mind / La Lingüística Aplicada Hoy: Comprendiendo el Lenguaje y la Mente*. Almería: Universidad de Almería, 1433-1447.
- Pendar, N. and Chapelle, C. A. (2008). “Investigating the promise of learner corpora: Methodological issues”, in *CALICO Journal*, 25, 2: 189-206.
- Pitkowski, E. and Vázquez Gamarra, J. (2008). “Los nuevos medios disponibles en Internet para la enseñanza de E/LE: los corpus”, presentado en el CEDELEQ III. Montrel, Quebec.
- St. John, E. (2001). “A case for using a parallel corpus and concordancer for beginners of a foreign language”, in *Language Learning & Technology*, 5, 3: 185-203.
- Tono, Y. (2007). “The roles of oral L2 learner corpora in language teaching: The case of the NICT JLE corpus”, in M. C. Campoy and M. Luzon (eds.), *Spoken Corpora in Applied Linguistics*. Bern: Peter Lang Pub Inc., 163-179.

- Vyatkina, N. A. (2007). *Development of second language pragmatic competence: The data-driven teaching of german modal particles based on a learner corpus*. (Doctoral Dissertation). The Pennsylvania State University: Pennsylvania.
- AntConc, <http://www.antlab.sci.waseda.ac.jp/software.html>
- CEATE, <http://corpora.fild.ncku.edu.tw>
- Corpus del Español, <http://www.corpusdelespanol.org/>
- CREA, <http://corpus.rae.es/creanet.html>
- ParaConc, <http://www.athel.com/para.html>
- Real Academia Española, <http://www.rae.es>
- Sketch Engine, <http://www.sketchengine.co.uk/>
- UAM CorpusTool, <http://www.wagsoft.com/CorpusTool/>
- WordSmith Tools, <http://www.lexically.net/wordsmith/>

