

ESTUDIO DEL USO DEL ARTÍCULO A PARTIR DE UN CORPUS PARALELO DE APRENDICES, CPATEI¹

Lu Hui-Chuan & Lu Lo Hsueh
National Cheng Kung University & Providence University

Resumen: El principal objetivo de nuestro trabajo es averiguar los usos del artículo por parte de alumnos taiwaneses que estudian español como segunda lengua extranjera e inglés como primera lengua extranjera, aplicando herramientas de corpus desarrolladas para facilitar la eficacia en el análisis sistemático del lenguaje. El presente estudio se basa en datos extraídos del Corpus Paralelo de Aprendices Taiwaneses de Español e Inglés (CPATEI) que construimos desde 2006. Basándonos en los estudios previos, utilizamos varias herramientas auxiliares con la intención de facilitar, semi-automáticamente, la anotación y el procesamiento de los usos del artículo. Además, examinamos errores como el uso excesivo, insuficiente e inapropiado del artículo, así como las características de los participantes. También nos dedicamos a contrastar los datos, tanto de aprendices como de nativos. Esperamos que los resultados de la investigación puedan proporcionar referencias útiles para el desarrollo de estrategias didácticas en el aprendizaje del español como lengua extranjera.

Palabras clave: corpus de los aprendices, corpus paralelo, artículo, segunda lengua, tercera lengua.

1. INTRODUCCIÓN

Los estudios relacionados con los usos del artículo han llamado bastante la atención en el área de la lingüística aplicada. Además, es un punto gramatical difícil de asimilar para la mayoría de los alumnos extranjeros en el proceso de aprendizaje del inglés o el español, y generalmente se relaciona con la complejidad contrastiva entre la lengua meta y la lengua materna. Como los estudios previos dejaron problemas no resueltos, esperamos un método más sistemático para alcanzar un aprendizaje eficaz. Mientras tanto, el florecimiento de la lingüística de corpus nos arroja una nueva luz, gracias a su poderosa función de proporcionar anotaciones y consultas que sirvan como ejemplo. Aprovechando los instrumentos basados en corpus que ya existen, el presente trabajo se dedica a investigar los posibles factores relacionados entre el español y el inglés, y sus resultados contribuirán al aprendizaje de los artículos del español de estudiantes taiwaneses que aprenden inglés como primera lengua extranjera y español como segunda. La investigación se basa en la creación de un Corpus Paralelo de Aprendices Taiwaneses de Español e Inglés, (CPATEI), de manera que esperamos utilizar la tecnología como apoyo en este proceso de enseñanza-aprendizaje.

La organización de este trabajo consiste en presentar la introducción en la sección 1, los estudios previos en la sección 2, la pregunta generadora de la investigación, los datos, la metodología y los resultados en la sección 3, y por último la conclusión en la sección 4.

¹ Este presente trabajo fue presentado oralmente en el VII Congreso Internacional de la Asociación Asiática de Hispanistas que tuvo lugar en Beijing, China.

2. ESTUDIOS PREVIOS

Entre otros estudios anteriores, Leung (2005: 39), propone que la transferencia de la adquisición de la tercera lengua proviene de dos orígenes: de la gramática de interlengua de la primera y la de la segunda lengua. Además, De Angelis (2007: 19), actualiza su anterior estudio realizado en el año 2005, indicando que los factores relacionados con la influencia cross-lingüística y con la transferencia de conocimiento de lenguas pueden incluir distancia entre lenguas, habilidad en la lengua meta, orden de adquisición de diferentes lenguas, etc. Ringbom (2007:117), también estudia los factores de distancia entre cada lengua y sus niveles. En el presente estudio nos dedicamos a construir un corpus paralelo como primer paso, que incluye inglés y español (CPATEI)², con el fin de investigar los temas multilingües. En el anterior estudio de Lu & Lu (2010:15), también basado en CPATEI, nos dimos cuenta de que ni la relación entre los usos incorrectos de las conjunciones españolas y los errores en inglés, ni la relación entre los errores en español y el nivel de español de los sujetos llegan a una diferencia significativa. Además, llegamos a la conclusión de que los usos insuficientes de conjunciones en español y en inglés se pueden atribuir al poco uso que se les da en chino, (lengua materna de los aprendices taiwaneses) en contraste con los resultados de Field y Yip (1992:26) y Granger y Tyson (1996:20).

Por otro lado, Fernández López (1990:111) muestra que el total de errores acerca del artículo que presentan los estudiantes de su investigación es de 161. Entre ellos, los errores relacionados con la omisión del artículo llegan a 142, mientras que la incorrecta elección entre el artículo determinado e indeterminado es de 19. En cuanto a los cuatro grupos de estudiantes, la escala de errores cometidos es la siguiente: los japoneses, los alemanes, los franceses y los árabes. Las posibles causas que se pueden deducir son: en primer lugar, que en el caso de los estudiantes japoneses, como no existe artículo determinado en su lengua materna, lo sustituyen por el determinante demostrativo; y en segundo lugar, que en los demás grupos los errores se cometen debido a falsas analogías entre la lengua materna y la lengua meta. También hemos repasado la investigación de Goitia Pastor (2007:411), que se basa en la observación de los errores cometidos en el uso del artículo en un total de 110 exámenes de estudiantes estadounidenses. Ella indica que el uso del artículo definido resulta más difícil.

Acerca de los estudios relacionados con los aprendices de español taiwaneses, Lu (1998:519) estudia el uso del artículo en el español de los estudiantes taiwaneses en 136 tareas traducidas del inglés al español. Los resultados señalan que el chino, como lengua materna, afecta al uso correcto del artículo determinado. Asimismo, Lin (2005:274) indica que los errores en la omisión del artículo se comenten por transferencia de la lengua materna. Según ella, al no haber artículo en chino, los estudiantes taiwaneses presentan problemas en el aprendizaje del sistema lingüístico. Por último, Lu *et al.* (2009:1) se concentran en estudiar la relación entre la presencia y la ausencia del artículo definido, analizando los datos extraídos del Corpus Oral de Aprendices Taiwaneses. Los resultados muestran que los aprendices taiwaneses usan los artículos definidos más correctamente en inglés que en español. Además, postula la Hipótesis Paralela para referirse a las relaciones significativas señaladas en el cuadro de la nota al pie³.

² Hemos recopilado 6.185 palabras en las composiciones españolas y 6.057 palabras en sus correspondientes traducciones al inglés en el año 2009.

³ Relación entre diferentes usos del artículo:

Correlación de español/inglés	Artículo definido	∅
Artículo definido	0,469* (correlación positiva 0,010) Ing: 84,38% Esp: 71,88%	No significativo Ing: 78,91% Esp: 53,13%
∅	0,509* (correlación negativa 0,005) Ing: 45,31% Esp: 45,31%	X

3. ESTUDIO DE LOS USOS DEL ARTÍCULO

3.1. Preguntas generadoras del estudio

Planteamos las siguientes preguntas:

1. ¿Cuál será la distribución y la inclinación de los artículos que usan los aprendices taiwaneses?
2. ¿Cuáles son los posibles factores que afectan al uso de los artículos españoles por parte de los aprendices taiwaneses?
 - a. ¿Cómo se relacionan los artículos españoles y los ingleses que usan los aprendices taiwaneses?
 - b. ¿Cómo se relacionan los artículos españoles que usan los aprendices participantes en este estudio con su nivel de español?

3.2. Textos recopilados

En el presente estudio se realiza un análisis de 3.721 palabras de 33 composiciones en español y 3.520 palabras de los mismos 33 textos traducidos al inglés, que son extraídos del Corpus Paralelo de Aprendices Taiwaneses de Español e Inglés (*CPATEI*) que empezó a construirse en el año 2006⁴. Los participantes provienen del segundo curso del Departamento de Español de la Universidad Providence, cuya lengua materna es el chino mandarín y que aprenden inglés como primera lengua extranjera y español como segunda.

Antes de recoger los datos, pedimos a los participantes que rellenen la información personal y firmen una carta de consentimiento. Por un lado, con los datos personales, podemos hacer análisis más profundos relacionados con las características de los aprendices; por el otro, las cartas de consentimiento firmadas nos conceden el derecho de poder incluir sus textos escritos en nuestro corpus paralelo de aprendices de forma que los investigadores puedan estudiar los textos legalmente siempre y cuando sea con propósito de investigación. Luego de una semana, los alumnos contaron con 45 minutos para escribir una composición en español y luego traducirla al inglés sin poder usar ningún diccionario ni instrumento auxiliar. Después, todos los participantes tomaron el *Wisconsin Placement Test for Spanish* para obtener un criterio más justo al clasificar el nivel de español de los alumnos. La distribución de las características de los datos recogidos se señala en el Cuadro 1.

Cuadro 1. Características de los datos extraídos.

Datos recogidos en 2009	Alumnos (m, f)	Horas de aprendizaje	Notas del Wisconsin Placement test (media)	Palabras
PU2	33 (1, 32)	1,024	371-588 (445)	esp: 3,721 ing: 3,520

⁴ *CPATEI*-2009 se caracteriza por los siguientes rasgos: En primer lugar, los textos son paralelos en dos lenguas extranjeras. En segundo lugar, los textos recogidos obedecen a un mismo criterio: el mismo tema y las mismas tareas. Además, el nivel de español de los participantes se califica por una prueba general, con el fin de obtener un estándar consistente de alumnos de diferentes sistemas de educación. Para más detalles sobre el corpus creado, *CPATEI*, véase Lu & Lu 2010.

3.3. Metodología

Las composiciones españolas y las traducciones inglesas recolectadas son corregidas por los hablantes nativos de español y de inglés respectivamente. Después, los cuatro tipos de datos son POS etiquetados⁵ y enumerados paralelamente para análisis posteriores.

Las herramientas auxiliares utilizadas para el presente estudio incluyen: *Corpus Tool de UAM*⁶, *WordSmith*⁷, *ParaConc*⁸ y *Colocación de Español*⁹. El uso de cada una depende de la función requerida en nuestra necesidad de llevar a cabo un proceso analítico de calidad, como: etiquetación semi-automática, anotación de usos, elementos paralelos, palabras colocadas, etc. Por ejemplo, utilizamos el *Corpus Tool de UAM* para anotar los artículos usados por los aprendices. En el esquema que organizamos, dividimos los artículos en dos grupos: los correctos y los incorrectos. Por un lado, los usos correctos se clasifican en definidos e indefinidos, en combinación con sus características de género y número (el, la, los, las, un, una, unos, unas). Por otro lado, anotamos los usos incorrectos según tres categorías de errores cometidos por los alumnos (uso excesivo, uso insuficiente y uso inapropiado).

Calculamos los usos correctos e incorrectos, los porcentajes y examinamos las relaciones significativas entre las variables estudiadas, utilizando el método estadístico de correlación de SPSS (versión 15). Para hacer referencias a los usos nativos, consultamos dos corpus, el *Corpus Paralelo de Español, Inglés y Chino (CPEIC)*¹⁰ y el *Corpus de Español*¹¹.

3.4. Resultados y debate

En esta sección, tratamos de contestar las preguntas de investigación analizando los resultados consultados con diferentes instrumentos auxiliares. En primer lugar, señalamos la distribución de los artículos usados en los textos en español, tanto en la versión escrita por los alumnos, como en la corregida por los hablantes nativos. Podemos ver los resultados en los Cuadros 2 y 3.

Cuadro 2. Distribución de artículos usados en los textos españoles.

Versión\Art	el	la	los	las	un	una
Original	37	203	36	95	37	14
Revisada	50	210	46	111	22	27

Cuadro 3. Distribución de artículos usados en los textos ingleses.

Versión\Art	the	a	an
Original	185	73	1
Revisada	179	87	3

⁵ Utilizamos el Tree Tagger de University of Stuttgart programado por Helmut Schmid. El tema estudiado en este trabajo se concentra en ART en español y DT en inglés.

⁶ *Corpus Tool de UAM* ha sido programado por Michael O'Donnell.

⁷ *WordSmith* (versión 5) de Mike Scott se ha usado para consultar los artículos usados en los textos estudiados. Los elementos combinados en los sintagmas nominales incluyen NC (nombre común) y NP (nombre propio) en español y NN (nombre común singular), NNS (nombre común plural), NP (nombre propio singular) y NPS (nombre propio plural) en inglés.

⁸ Michael Barlow es el programador de *ParaConc*. Por un lado se ha utilizado para contrastar las frases paralelas entre los textos de los aprendices y los revisados, y por otro lado para contrastar las frases paralelas entre los textos españoles y los ingleses.

⁹ *Colocación de Español* se ha desarrollado por Nacional Cheng Kung University, Taiwán. No sólo se presta atención al valor de probabilidad de χ^2 sino también consideramos la cantidad de aparición (≥ 5)

¹⁰ Es un corpus que se ha construido por Nacional Cheng Kung University y se dedica a recopilar datos paralelos de hablantes nativos de las diferentes lenguas. Por la escasez de datos paralelos disponibles, los datos recopilados son textos de la *Biblia*.

¹¹ Las consultas de este corpus nos provee de un panorama amplio de los usos más frecuentes en la lengua española desde una perspectiva nativa.

Observamos que, en comparación con la versión corregida por un hablante nativo, el Cuadro 2 muestra que los alumnos usaron menos artículos definidos en español que los nativos, mientras que el Cuadro 3 muestra que se usaron más en los casos de artículos definidos y menos en los casos de artículos indefinidos ingleses que los nativos. Si profundizamos la investigación en los sustantivos colocados después de los artículos ingleses, nos damos cuenta de que se escalonan en ambas versiones, la versión estudiantil y la nativa: los sustantivos comunes singulares >(se usan con más frecuencia que) los sustantivos comunes plurales > los sustantivos propios singulares > los sustantivos propios plurales. Los usos de artículos se dividieron en correctos e incorrectos, los resultados se señalan en el Cuadro 4.

Cuadro 4. Comparación entre los usos correctos e incorrectos.

Usos\Lengua	Español	Inglés
Usos correctos	408	299
Usos incorrectos	135	81
Porcentaje de corrección	75,14%	78,68%

En el Cuadro 4 se nota que los aprendices alcanzan un nivel de corrección más alto en los usos de artículos ingleses que en los españoles. Como se muestra en el Cuadro 5, se estudiaron los posibles factores, analizando la correlación entre los usos de artículos españoles, los ingleses y el nivel de español.

Cuadro 5. Correlación entre diferentes factores.

Valor de P: corrección/error	Usos de artículos españoles	Usos de artículos ingleses
Nivel de español	0,081/0,993	0,384/0,489
Usos de artículos españoles		0,151/0,206

En el Cuadro 5 observamos que todos los valores de P (Probabilidad) de correlación entre los factores examinados superan 0,05 sin importar si se evalúan los usos correctos o incorrectos del artículo. Es decir, ni el nivel de español, ni el uso de artículos ingleses se relacionan significativamente con el uso de los artículos españoles. Como los factores estudiados no determinan los usos de los aprendices, vamos a estudiar los usos del artículo con más detalle según sus características, los resultados se señalan en el Cuadro 6.

Cuadro 6. Distribución de usos correctos e incorrectos.

Usos\Artículos	el	la	los	las	un	una
Correctos	39	158	26	78	46	12
%	75,48%		68,42%		78,38%	
	72,88%				78,38%	
Incorrectos	29	35	22	26	7	9
%	24,52%		31,58%		21,62%	
	27,19%				21,62%	

En el Cuadro 6 notamos que los aprendices manejan los artículos indefinidos mejor que los definidos, y que entre los definidos, usan las formas singulares mejor que las plurales.

A continuación, vamos a clasificar los errores cometidos por los aprendices para obtener una idea sobre la inclinación de los usos. Los dividimos en los usos excesivos, los insuficientes y los inapropiados, como se ve en el Cuadro 7.

Cuadro 7. Errores clasificados según los usos.

Uso\Artículo	el	la	los	las	un	una	total
Excesivo	5	8	3	2	2	1	21
Insuficiente	19	23	19	23	4	1	89
Inapropiado	5	4	0	1	1	7	18
Incorrecto	29	35	22	26	7	9	128

Según el Cuadro 7, las ocurrencias de los errores se escalonan de más a menos frecuentes en el siguiente orden: insuficiente > excesivo > inapropiado. A pesar de no obtener un resultado significativo deducido del Cuadro 5, averiguaremos los contrastes entre las dos lenguas mediante las oraciones paralelas en español e inglés, a fin de conseguir una conclusión más sistematizada, como se señala en el Cuadro 8.

Cuadro 8. Artículos usados por los aprendices en las oraciones paralelas de CPATEI (del español al inglés).

Art	Español	Inglés
Definido	el	sin traducirse: 72,97%; the+N: 24,32%
	la	sin traducirse: 76,35%; the+N: 20,69%
	los	sin traducirse: 79,41%; the+N: 8,82%
	las	sin traducirse: 94,74%; the+N: 4,21%
Indefinido	un	sin traducirse: 67,57%; a, an+N(sing): 29,73%
	una	a, an+N(sing): 50%; sin traducirse: 42,86%

Cuadro 9. Artículos usados por los aprendices en las oraciones paralelas de CPATEI (del inglés al español).

Art	Inglés	Español
Definido	the	sin traducirse: 72,28%; el, la: 23,91%; los: 3,26%; un, una: 0,54%
Indefinido	a, an	sin traducirse: 74,32%; un, una: 16,22%; el, la: 5,41%; los: 2,70%

Observando los artículos usados por los aprendices traduciendo del español al inglés (Cuadro 8) y del inglés al español (Cuadro 9), nos damos cuenta de que si consultamos las oraciones paralelas, los artículos se omiten con mucha frecuencia, aún sin importar la lengua en la cual se haga el análisis. Esto quiere decir que el uso de los artículos por parte de los alumnos en las dos lenguas se procesa a través de sistemas distintos, o se ve afectado por su lengua nativa, en la que no existen los artículos definidos. Si quisiéramos saber las diferencias entre los aprendices y los nativos, estudiaríamos los contrastes observados en ambos casos, analizando los datos provenientes del *Corpus Paralelo de Español, Inglés y Chino (CPEIC)*, como se muestran en los Cuadros 10 y 11.

Cuadro 10. Artículos usados por los hablantes nativos en las oraciones paralelas de *CPEIC*.

Art	Español	Inglés
Definido	el	the+N(sing): 74,5%; sin traducirse: 16%
	la	the+N(sing): 81,5%; sin traducirse: 17,5%
	los	the+N(plur): 54%; sin traducirse: 17,5%; the+N(sing): 17%
	las	the+N(plur): 59%; sin traducirse: 20,5%; the+N(sing): 16,5%
Indefinido	un	a, an+N(sing): 85%; sin traducirse: 9%

Cuadro 11. Artículos usados por los hablantes nativos en las oraciones paralelas de *CPEIC*.

Art	Inglés	Español
Definido	the+N(sing)	el, la: 57,14%; sin traducirse: 12,61%; los, las: 5,6%; un, una: 0,84%
	the+N(plur)	los, las: 94,34%; sin traducirse: 5,66%
Indefinido	a, an	un, una: 56%; sin traducirse: 38%; el, la: 1%

En los Cuadros 10 y 11, podemos observar la inclinación presente en el corpus nativo, *CPEIC*, entre español e inglés; los casos omitidos no ocurren tanto como se observa en los usos por parte de los aprendices, es decir, en los datos de los hablantes nativos, los artículos tienen una inclinación más constante entre los artículos definidos y los indefinidos, mientras que en comparación con los resultados extraídos de los Cuadros 8 y 9, nos percatamos de que el grado de similitud entre las dos lenguas es mayor que el que hay entre los artículos usados por los aprendices.

Por último, analizamos los elementos colocados más frecuentemente, comparando el uso que le dan los aprendices y el que le dan los hablantes nativos, con la intención de obtener una mejoría en el proceso de enseñanza. En primer lugar, consultamos la herramienta de *Colocación de Español* para comparar las palabras combinatorias usadas por los aprendices y por los correctores nativos, como se muestra en el Cuadro 12.

Cuadro 12. Comparación de elementos colocados usados por los aprendices y los corregidos por nativos.

Texto	Ejemplos
Original	la vida (31), la noche (13), la comida (6)
Revisado	los estudiantes (8), la vida (29), la universidad (15), la noche (14), la Universidad (6), la comida (7), la gente (8), la residencia (5), las clases (6), la clase (8)

En el Cuadro 12, nos damos cuenta de que las colocaciones usadas por los nativos que corrigen los textos originales (salidas del método estadístico de χ^2), son muchas más que las usadas por los aprendices, por ejemplo, “los estudiantes, la universidad, la gente, la residencia, las clases, la clase”.

En segundo lugar, consultamos las colocaciones de alta frecuencia en el *Corpus del Español*, programado por Mark Davies, para obtener una lista general de las combinaciones “artículo+sustantivo” como punto de referencia para la enseñanza o el aprendizaje. Los resultados se señalan en el Cuadro 13.

Cuadro 13. Combinaciones de “artículo+sustantivo” de alta frecuencia consultadas con *Corpus del Español*¹².

Artículos	10 ejemplos de alta frecuencia
Definidos	el (2.360.220) el rey (55.474), el mundo (26.292), el señor (17.091), el tiempo (16.546), el día (15.801), el sol (14.977), el cielo (14.537), el hombre (13.942), el alma (13.892), el cuerpo (13.219)
	la (2.921.455) la tierra (40.256), la vida (33.861), la ciudad (25.234), la muerte (23.417), la mano (22.172), la casa (19.692), la gente (17.230), la noche (17.154), la puerta (16.761), la ley (16.361)
	los (1.525.636) los ojos (22.955), los hombres (20.938), los indios (17.717), los españoles (11.025), los reyes (9.858), los pueblos (8.522), los pies (7.753), los moros (7.039), los niños (5.980), los medios (5.966)
	las (927.222) las cosas (29.554), las manos (15.503), las leyes (8.635), las armas (8.222), las mujeres (7.936), las palabras (6.540), las personas (5.934), las casas (5.742), las obras (5.703), las partes (5.339)
Indefinidos	un (571.332) un hombre (14.135), un día (8.055), un momento (4.860), un año (4.717), un tiempo (3.709), un punto (3.286), un lado (3.245), un pueblo (3.106), un libro (2.604), un lugar (2.598)
	una (441.154) una vez (10.161), una cosa (7.243), una mujer (6.386), una parte (5.854), una especie (3.474), una serie (3.460), una persona (3.185), una noche (2.818), una casa (2.793), una carta (2.769)

Las listas señaladas en el Cuadro 13 nos ofrecen una idea básica sobre las combinaciones “artículo+sustantivo” a las que debe prestarse más atención en la enseñanza. El siguiente paso será estudiarlas en diversos contextos para mejorar el entendimiento de sus funciones discursivas, como parte de trabajos futuros.

4. CONCLUSIÓN

Basándonos en el *Corpus Paralelo de Aprendices Taiwaneses de Español e Inglés (CPATEI)* y el *Corpus Paralelo de Español, Inglés y Chino (CPEIC)*, este trabajo ha estudiado los usos del artículo por parte de alumnos taiwaneses, aplicando las herramientas existentes de corpus tales como: *UAM Corpus Tool*, *WordSmith*, *ParaConc*, *Colocación de Español*, *Corpus del Español*. Todas ellas han sido desarrolladas para facilitar el análisis sistemático del lenguaje.

Para contestar las preguntas planteadas, llegamos a las siguientes conclusiones: En primer lugar, ni el nivel de español ni el uso de artículos ingleses se relacionan significativamente con el uso de los artículos españoles. En segundo lugar, los aprendices alcanzan un nivel de corrección más alto en el uso de artículos ingleses que en el de los españoles, a pesar de que pertenecen

¹² Davies, Mark. (2002-) *Corpus del Español* (100 millones de palabras, siglo XIII - siglo XX). Disponible en: <http://www.corpusdelespanol.org>.

al departamento de español. Además, los aprendices manejan los artículos indefinidos mejor que los definidos y usan las formas singulares mejor que las plurales dentro de los definidos. En tercer lugar, la frecuencia de los errores se escalona de alto a bajo en el siguiente orden: insuficiente > excesivo > inapropiado. Por último, al contrastar los textos paralelos descubrimos que los aprendices adoptan diferentes estrategias (presencia o ausencia del artículo) para traducir del español al inglés, de manera más obvia que la que se observa en los nativos.

Las conclusiones presentadas anteriormente pueden estudiarse con más profundidad en el futuro, para así mejorar los aportes del tema presentado. Con respecto a las limitaciones, una de ellas se debe a la poca cantidad de datos que se pueden recolectar en el futuro. Con todo esto, a partir del corpus, esperamos que este estudio provea una perspectiva diferente en la investigación de ELE.

AGRADECIMIENTOS

Queríamos aprovechar esta oportunidad para agradecer la ayuda financiera del Landmark Project (R021) patrocinado por Nacional Cheng Kung University. De igual manera, nuestro agradecimiento a *Corpus Tool de UAM*, por permitirnos utilizarlo de forma gratuita.

En cuanto a la construcción del *Corpus Paralelo de Aprendices Taiwanese de Español e Inglés (CPATEI)*, el desarrollo de las herramientas auxiliares, la anotación y el análisis de los datos, la revisión de las composiciones en español, y las traducciones en inglés, queremos dar nuestro sincero agradecimiento a los ayudantes Ting-Xuan Wang, Ciro Chang, Belisa Liu, Emilia Chang, Blanca Ye, Serafina Chang y Emilia Lin, de igual manera que a los hablantes nativos Cindy M. Barquero, María H. Contreras y Derek Murphy.

REFERENCIAS

- Davies, M. (2002-). Corpus del español (100 millones de palabras, siglo XIII - siglo XX). Disponible en <http://www.corpusdelespanol.org>. [fecha acceso: 28.7.2009]
- De Angelis, G. (2005). "Interlanguage transfer of function words", *Language Learning* 99/3: 379-414.
- De Angelis, G. (2007). *Third or Additional Language Acquisition*. Clevedon: Multilingual Matters.
- Fernández López, S. (1990). "El uso del 'artículo' en aprendices de español lengua extranjera", en Actas del Primer Congreso nacional de ASELE. Granada, Universidad de Granada. 109-118.
- Field, Y. y Yip, L. M. O. (1992). "A comparison of internal cohesive conjunction in the English essay writing of Cantonese speakers and native speakers of English", *RELJ Journal* 23/1: 15-28.
- Goitia Pastor, L. (2007). "Un estudio del uso del artículo definido por parte de estudiantes estadounidenses de español como lengua extranjera (E/LE) mediante un inventario de frases correctas e incorrectas", *Interlingüística* 17: 409-418.
- Granger, S. y S. Tyson. (1996). "Connectors usage in the English essay writing of native and non-native EFL speakers of English", *World Englishes* 15/1: 17-27.
- Leung, Y.-K. I. (2005). "L2 vs. L3 initial state: A comparative study of the acquisition of French DPs by Vietnamese monolinguals and Cantonese-English bilinguals", *Bilingualism, Language and Cognition* 8: 39-61.
- Lin, T.-J. (2005). *La Adquisición y el Uso del Artículo por Alumnos Chinos*. Madrid: RedELE.

- Lu, H.-C. (1998). "El uso del artículo en el español: errores e implicaciones pedagógicas", en K. Alonso, F. M. Fernández y M. G. Bürmann (eds.) *Actas del VIII Congreso Nacional de ASELE*. Alcalá de Henares: Universidad de Alcalá, 519-525.
- Lu, H.-C., C.-Y. Chang y L. H. Lu. (2009). "CPATEI: A parallel Spanish-English learner corpus with mutual reference analysis between multiple inter-languages", resumen aceptado por AACL2009, Alberta, Canada.
- Lu, H.-C. & H. L. Lu. (2010). "Parallel corpus-based study of conjunctions", ponencia presentada en el International Conference on New Trends in Contemporary European Literature, Culture and Language, Taipei, Taiwan.
- Ringbom, H. (2007). *Cross-linguistics Similarity in Foreign Language Learning*. Clevedon: Multilingual Matters.

CORPORA Y HERRAMIENTAS

- Corpus del Español, Mark Davies
<http://www.corpusdelespanol.org/x.asp>
- ParaConc, Michael Barlow
<http://paraconc.com/>
- Tree Tagger, University of Stuttgart
<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>
- UAM Corpus Tool, Michael O'Donnell
<http://www.wagsoft.com/CorpusTool/>
- WordSmith Tool v.5, Mike Scott
<http://www.lexically.net/wordsmith/>