



DESAFÍOS Y PERSPECTIVAS ACTUALES DE LA PSICOLOGÍA EN EL MUNDO DE LA INFANCIA

**REVISIÓN DE ESTUDIOS DE CORPUS Y APLICACIÓN EN DESARROLLO Y
APRENDIZAJE DE LA ESCRITURA
A REVIEW OF CORPORA STUDIES AND USEFULNESS IN WRITING
DEVELOPMENT AND LEARNING**

Esther García-Martín

Jesús Nicasio García Sánchez

Universidad de León

Departamento de Psicología, Sociología y Filosofía, Campus de Vegazana, s/n, 24071 León (España),

*Fecha de recepción: 9 de enero de 2011
Fecha de admisión: 10 de marzo de 2011*

RESUMEN

Se analizan las aplicaciones y utilidad de los estudios de corpus como herramientas metodológicas, y una ilustración con un estudio comparativo de dos tipos de corpus de escritura, uno de desarrollo y otro profesional. Los resultados muestran diferencias estadísticamente significativas en el 61% de los contrastes, con una tendencia similar en los dos corpus, sugiriendo que los patrones evolutivos se van aproximando a los profesionales en la competencia comunicativa escrita. Se valoran los resultados, las implicaciones teóricas y aplicadas en el desarrollo y la educación, y las perspectivas futuras. Durante la realización de este estudio se recibieron ayudas competitivas del proyecto del MICINN (EDU2010-19250) para el trienio 2010-2013; así como del proyecto a Grupos de Excelencia de la JCYL (GR-259; BOCyL 27 de abril de 2009) para el trienio 2009-2011 y con fondos FEDER de la Unión Europea. Ambos concedidos al IP/Director J. N. García.

Palabras clave: estudios de corpus, aplicación en psicología y educación, corpus de desarrollo, corpus profesional, herramienta Seshat.

ABSTRACT

A review of corpora studies and usefulness in writing development and learning. We analyze the usefulness and applications of the corpora studies as methodological tools, and an illustration with a comparative study of two corpora types of writing, one of development and other professional one. Results show significant statistically differences in the 61 % of the contrasts, with a similar tendency in both corpora, suggesting that developmental patterns go closing them to the professional ones in written communicative competence. We value the results, the theoretical and applied implications



REVISIÓN DE ESTUDIOS DE CORPUS Y APLICACIÓN EN DESARROLLO Y APRENDIZAJE DE LA ESCRITURA A REVIEW OF CORPORA STUDIES AND USEFULNESS IN WRITING DEVELOPMENT AND LEARNING

in the development and in the education, and the future perspectives. During this study we received competitive funds from de MICINN project (EDU2010-19250, 2010-2013); besides from the competitive Project for Excellence Groups JCyL (GR-259; BOCyL 27 on April 2009, 2009-2011) and FEDER funds from de European Union. Both awarded to Principal Researcher (J. N. García).

Keywords: corpora studies, psychology and education application, developmental corpus, professional corpus, Seshat tool.

INTRODUCCIÓN

Se analizan las aplicaciones y utilidad de los estudios de corpus como herramientas metodológicas, y una ilustración con un estudio comparativo de dos tipos de corpus de escritura, uno de desarrollo y otro profesional. En primer lugar, los corpus en el contexto de la psicología y la educación permiten analizar diferentes factores y variables psicológicas de la composición escrita, así como la aplicación para estudiar el desarrollo y servir de base para la instrucción de la composición escrita con alumnos con y sin dificultades específicas de aprendizaje, así como para la escritura profesional (estudiantes universitarios, profesionales). En segundo lugar, el estudio comparativo incluye un corpus de desarrollo, con textos manuscritos de alumnos de 8 a 16 años (más de dos millones de palabras); y un corpus profesional inicial procedente del campo de la psicofarmacología (depresión) con más de un millón de palabras. La extracción de indicadores de productividad, coherencia y estructura de los textos, se obtienen aplicando la herramienta informática de análisis automático de textos Seshat (Díez, & García, este congreso).

Han sido diversos los usos en el campo de la educación del estudio de los *corpora*. Se han utilizado principalmente para la enseñanza de segundas lenguas, generalmente en la extracción de datos para la creación de materiales que contengan lengua real y en uso y para el estudio de lenguas en contraste, también con la finalidad de la mejora de la segunda lengua o de la traducción; o se centraban en el léxico. Por esto, nos interesa un corpus con una perspectiva distinta, con la finalidad de que pueda ser utilizado para analizar diferentes factores y variables psicológicas de la composición escrita y para el desarrollo de aplicaciones para el análisis automático de los textos (García-Martín, et al., 2009).

En general, este tipo de estudios de corpus han sido muy importantes entre los lingüistas e investigadores de lenguas (estudios contrastivos y estudios de traducción) en las últimas décadas, ya que permite almacenar grandes cantidades de textos y datos en formato electrónico. Con este proyecto se pretende dar un salto y aplicar esta metodología, que ha sido tan productiva en otros campos, al campo de la psicología para los estudios de escritura dentro de la educación. Además, se están diseñando aplicaciones informáticas que permitirán el análisis automático de los textos. Hay muy pocas aplicaciones en el campo de la psicología educativa y de la educación de este tipo, pero básicamente centrados exclusivamente en el léxico en español (Almela, 2005; Álvarez, Carreiras, & de Vega, 1920a, 1992b; Justicia, 1995), pero no en el discurso y texto escritos, aunque sí en inglés (Biber, Conrad, & Reppen, 1998). Por esto, la aplicación del corpus a este campo de la psicología y la educación y para propósitos tanto del desarrollo como instruccionales será de gran valor, tanto desde el punto de vista teórico como del aplicado.

Es cierto que el trabajo de compilación e informatización de los textos para la construcción del corpus es un trabajo muy tedioso pero una vez están los textos agrupados en el ordenador pueden reutilizarse tantas veces como sea deseado o necesario para análisis de tipos diferentes y con distintas perspectivas para el tipo de lengua o textos que contenga el corpus. A pesar de todo el trabajo realizado desde 1994 y la gran cantidad de datos registrados por el equipo de investigación, todos los textos y análisis estaban en formato de papel. Además, somos conscientes de la existen-



DESAFÍOS Y PERSPECTIVAS ACTUALES DE LA PSICOLOGÍA EN EL MUNDO DE LA INFANCIA

cia de otros corpus similares en español basados en el léxico y vocabulario que aparece en los libros de texto infantiles y cuyo propósito era la extracción de datos de tipo psicolingüístico como frecuencia, longitud, frecuencia silábica posicional (Álvarez, Carreiras, & de Vega, 1992a, 1992b), que permitieron la creación de tareas evaluativas para diferentes propósitos, pero, principalmente, para evaluar la ruta fonológica (por ejemplo listas de pseudo-palabras con frecuencias diferentes, longitud o frecuencia silábica), o de la ruta ortográfica (por ejemplo homófonos, palabras altamente frecuentes, palabras con ortografía arbitraria, en cuanto a la frecuencia, longitud). Otro corpus, que contiene textos escritos por niños, se compiló teniendo en cuenta este propósito (Justicia, 1995). Hay otros ejemplos de corpus similares que no han sido publicados, como el construido por el equipo de investigación de Jiménez en 1997, pero ninguno de ellos de naturaleza evolutiva o educacional ni tampoco con las características que se pretenden en nuestro proyecto (ver por ejemplo Almela, 2005). En ninguno de los casos mencionados se consideró el corpus como una herramienta metodológica y el foco del análisis ha sido léxico, quedándose al nivel de la palabra y sus características desde el punto de vista psicológico. Ninguno de ellos tenía el propósito final de ser utilizados para análisis, para el estudio de la composición escrita y tampoco para el estudio de, además del ámbito léxico, de aspectos estructurales como pueden ser la coherencia referencial o relacional o la productividad. Esta última se ha estudiado solo en parte en inglés con el uso de una herramienta implementada por Biber, Conrad, & Reppen (1998), cuyo interés principal era el análisis de datos lingüísticos de forma empírica a partir de corpus (*Corpus-based studies*). Gracias a estudios de este tipo podemos ahora aplicar el uso del corpus a nuestras necesidades particulares, centrándonos en la complejidad del discurso escrito en alumnos con y sin dificultades de aprendizaje (Gregg, Coleman, Stennett, & Davis, 2002). Teniendo en cuenta las discusiones existentes a cerca del tamaño de corpus en lingüística de corpus, se podría decir que la mayoría de los corpus utilizados para aspectos psicolingüísticos son más pequeños de lo que se desearía. Ver por ejemplo el estudio de Justicia (1995) con 2166 casos.

Es teóricamente relevante el encontrar patrones específicos y del desarrollo acerca de dificultades y habilidades en la composición escrita de los alumnos a la hora de establecer dificultades comunes en alumnos con y sin dificultades de aprendizaje en la escritura. Es también relevante en el contexto escolar el modo en que se podría ayudar a padres y profesores en el proceso educativo, lo que conllevaría a una mejora en los logros y en la auto-regulación del aprendizaje de los alumnos con dificultades de aprendizaje, así como del resto de alumnos, si sabemos las diferentes variables y factores psicológicos involucrados en la escritura, tanto el aprender a escribir como en otros campos, el aprender mediante la escritura.

Un corpus es una colección de textos, almacenados en formato electrónico, seleccionados y organizados de acuerdo con criterios explícitos y para propósitos lingüísticos. Sin embargo, la idea de que tenga una finalidad para el estudio lingüístico ha ido cambiando y se ha ido expandiendo a otros campos como el estudio de las lenguas en contraste para traducción, para enseñanza de segundas lenguas o como en nuestro caso para *estudios de tipo evolutivo y educativo*.

Los elementos básicos de una definición de corpus sobre los que hay un consenso, no tienen la misma importancia para cada uno de los autores. Una de las mayores discusiones gira en torno al tamaño de la muestra, del corpus y la representatividad del corpus. La otra es más de tipo teórico acerca de la posición del corpus como disciplina teórica, aplicada o metodología. Para nuestros propósitos particulares nos quedaremos con el *corpus como metodología* ya que para nosotros es una *herramienta útil para el estudio de los aspectos psicológicos entorno a la escritura*, a partir de datos empíricos y fiables de textos compilados por el equipo de investigación sobre composición escrita de alumnos desde los 8 a los 16 años de edad, y no tendrá la finalidad de puro análisis lingüístico. Pero volviendo a la discusión acerca del tamaño adecuado o deseado del corpus y la repre-



REVISIÓN DE ESTUDIOS DE CORPUS Y APLICACIÓN EN DESARROLLO Y APRENDIZAJE DE LA ESCRITURA A REVIEW OF CORPORA STUDIES AND USEFULNESS IN WRITING DEVELOPMENT AND LEARNING

sentatividad, a pesar de que los autores no tengan demasiado claro el tamaño, sí están de acuerdo en la idea de que el *corpus ha de ser representativo* (Biber, 1993). La idea de representatividad es central y esencial al trabajo de corpus ya que guía y estructura la construcción y análisis de los corpus. Las consideraciones generales a tener en cuenta a la hora de construir un corpus representativo tienen que ver con el muestreo, el tamaño y la distribución de la población (Biber, 1993). Sin embargo, como se ha mencionado, es más importante que un corpus sea representativo que contenga un número muy elevado de textos, así como la inclusión de una gran variedad de aspectos de la población estudiada, es decir, la variedad de tipos textuales son más importantes. Por definición, según Biber (1993) cualquier selección de textos es una muestra, si se ha extraído de una población total y teniendo en cuenta los límites de las jerarquías. Pero para poder tener una buena muestra, esta ha de ser obtenida de manera aleatoria (*random sampling; stratified random sampling*). Biber (1993) propuso un modelo de muestreo jerárquico que muestra cómo se deben seleccionar datos de cada estrato y cómo decidir, por tanto, en la composición final de la muestra, para los casos de poblaciones heterogéneas. Por ejemplo, identifican, sucesivamente, seis estratos: (i) canal primario, sea escrito u oral, (ii) formato, esté publicado o no, (iii) entorno o marco, sea institucional, público, o privado y personal, (iv) destinatario, referido a pluralidad, presencia, interactividad y conocimientos compartidos, (v) exponente, incluyendo variación demográfica y reconocimiento, (vi) veracidad, es decir, que sea informativo veraz, (vii) propósito, para persuadir, entretener, informar, instruir, explicar, describir, y (viii) temas. Esto puede aplicarse a otros campos, así podemos decir que *estas características principales o ventajas nos permiten la aplicación del corpus a los estudios de composición escrita, en los campos de psicología del desarrollo y de la educación*.

Una ilustración de interés son las *aplicaciones de los corpus de investigación*. La idea de las aplicaciones es muy importante para este estudio, ya que la finalidad del uso del corpus que aquí se pretende diseñar y construir, además de *análisis de tipo psicológico y educativo*, va destinado a la creación de una aplicación de análisis automático de textos. En la mayoría de los casos se ha prestado atención a la creación de conocimiento pero no de tecnología en los campos de las humanidades o las ciencias sociales, por eso creemos conveniente que es un buen momento de dar este salto. Para la creación de aplicaciones se necesitan principios operacionales explícitos, una metodología basada en corpus para la obtención de datos, buenos sistemas de clasificaciones y estrategias basadas o centradas en problemas. Los resultados obtenidos del corpus o de los análisis realizados a partir de corpus deberán estar caracterizados por dos conceptos importantes: *utilidad* y la *usabilidad* (*usefulness and usability*) de la información o de los resultados. Es decir, que los resultados de esta investigación puedan ser puestos en uso (*útiles*) y al mismo tiempo, prácticos y convenientes al uso (*usabilidad*), aunque no es fácil alcanzar ambas al mismo tiempo.

Entre las aplicaciones del corpus está la *construcción de herramientas o programas informáticos*. El desarrollo de este tipo de aplicaciones en los últimos años ha crecido y se aplican a muchos campos pero hay dos que son relevantes y conviene destacar. En primer lugar, el *diseño de documentos escritos en el entorno laboral* y en segundo lugar los *avances en herramientas de escritura*.

El primero, el *diseño de documentos escritos* pretende describir el modo en que los escritores profesionales utilizan sus competencias (cualidades y conocimientos); mejorar el impacto comunicativo de dichos documentos y crear modelos expertos (ERN-LWE, 2007). La *escritura profesional* está caracterizada por la diversidad y variedad de medios utilizados y esto dificulta la construcción de herramientas útiles y utilizables que puedan cubrir todas las necesidades. La diversidad mencionada cubre todo tipo de géneros como artículos científicos, informes, formularios, panfletos de instrucciones, memorias, etc., y cada uno de los géneros, a su vez, incluye diferentes tipos textuales como, por ejemplo, informes de laboratorio, estudios clínicos, estudios experimentales o estudios de caso dentro de los artículos científicos, lo cual hace la situación aún más difícil y compleja. La variedad de soportes pueden incluir papel, formato electrónico, hipertexto, etc. Además, se ha de



DESAFÍOS Y PERSPECTIVAS ACTUALES DE LA PSICOLOGÍA EN EL MUNDO DE LA INFANCIA

tener en cuenta al lector o receptor de los textos ya que no es lo mismo si se escribe para expertos del campo, para el público general o para un cliente. Es importante tener en cuenta el *impacto que la escritura tiene en los usuarios*, ya que esto condiciona la manera de escribir y las elecciones que el escritor debe tomar a la hora de escribir un texto o estrategias empleadas. Además de todo esto, la información en el texto ha de organizarse de tal modo que pueda ser comunicada de manera eficiente (Alamargot, Terrier, & Cellier, 2007).

El segundo, los *avances tecnológicos en herramientas de escritura* se centran en el desarrollo de aplicaciones estandarizadas para el estudio de la escritura o para la mejora de los procesos de la escritura. En los últimos años, se han desarrollado programas con fines diferentes como pueden ser para registrar la escritura en formato digital, para poder analizar los procesos de escritura, o grabar el tecleo ("*Keylogging*"), sistemas de dictado y reconocimiento de voz, y otros sistemas que pueden encontrarse recogidos en Olive y Levy (2002) o en la revisión de MacArthur, Graham y Fitzgerald (2006, pp. 357-448). Un ejemplo a destacar es el "*ScriptLog*" desarrollado para registrar la escritura de forma digital por Stromqvist et al., (2006).

Ilustración de los estudios de corpus: comparación de un corpus de desarrollo y otro profesional

Con el fin de ilustrar las posibilidades y utilidad de los estudios de corpus en psicología y educación, se presenta la construcción y comparación de dos corpus de textos escritos informatizados, uno de desarrollo y otro profesional. Presentamos, en primer lugar, la descripción del corpus de desarrollo y después la descripción de un corpus de tipo profesional. Y, seguidamente, tras la aplicación de la herramienta Seshat de análisis automático de textos, se describen los indicadores en los que encontramos diferencias estadísticamente significativas, así como una valoración sobre las mismas.

MÉTODO

Corpus de desarrollo

El corpus de desarrollo contiene textos de un total de más de 8.000 alumnos españoles, entre 8 y 16 años de edad, con y sin dificultades de aprendizaje, para el estudio del desarrollo de la composición escrita entre edad y tipologías de alumnos, desde la perspectiva del desarrollo y comparativa; teniendo en cuenta, no solo los niveles de escritura (calidad y productividad) y logros generales, sino, además, en función de otras variables psicológicas (procesos y factores diferentes: planificación, reflexividad, motivación y auto-eficacia, auto-conocimiento y auto-regulación), y variables como habilidades sociales o el papel de la práctica del profesor en la enseñanza de la escritura que modula los logros de los alumnos. Los textos han sido recogidos en muestreos de diferentes investigaciones del equipo, en relación a diferentes proyectos, con el tema de la composición escrita como central o común para el estudio de su evaluación e intervención en alumnos con y sin dificultades de aprendizaje (cfr., García-Martín, et al., 2009). Se han escaneado, informatizado y compilado cerca de 20.000 textos manuscritos por los alumnos, desde el año 1994, y se siguen añadiendo textos de estudios en marcha. Cada texto tiene una media de 100 palabras, lo que nos da un total de unos 2 millones de palabras, lo cual justifica de sobra la cantidad y representatividad de un corpus de este tipo, que no es general, sino específico, de una población limitada. El consenso general es de un millón de palabras más o menos para poder asegurar la representatividad en los *corpora* para fines específicos o que tratan con lo que llamamos sub-lengua en lugar de corpus general (McEnery, Xiao, & Tono, 2006).

Además, se han implementado dos tipos de medidas a lo largo del análisis de los textos compilados. Una de ellas, con criterios subjetivos del lector, y otra, con criterios más formales basados



REVISIÓN DE ESTUDIOS DE CORPUS Y APLICACIÓN EN DESARROLLO Y APRENDIZAJE DE LA ESCRITURA A REVIEW OF CORPORA STUDIES AND USEFULNESS IN WRITING DEVELOPMENT AND LEARNING

en el texto. En los casos en que se tomaron medidas basadas en el texto, se aplicaron medidas de productividad o generación de información. Este tipo de evaluación está relacionada con la cantidad de textos producidos (número de párrafos, frases, verbos, número total de palabras en la redacción, entre otros). También, se han tomado medidas de organización de la información, coherencia relacional y referencial, así como en estructura. La evaluación del texto basada en el lector, como la estructura, coherencia y calidad siguen los criterios descritos por Spencer y Fitzgerald (1993), pero con ligeras variaciones para hacerlos más apropiados para los textos de comparación y contraste, ya que es el tipo que se analizó para este propósito particular.

Corpus profesional

Se construyó un corpus profesional inicial a partir de revistas científicas de psicofarmacología. Se ha conseguido un total de poco más de un millón de palabras, al ser esta cifra el umbral deseado, se han tenido que incluir todos los textos encontrados sin necesidad de hacer ningún tipo de muestreo. Por lo que la población y la muestra coinciden en este caso. El corpus 1 contiene la colección de textos físicos a partir de los cuales será extraída la información deseada. Contamos con un total de 159 textos y 1.485.811 palabras.

Herramienta informática

Para la aplicación de los indicadores y medidas del producto textual, a cada uno de los dos corpus, se utilizó el análisis automatizado de textos psicolingüísticos que realiza la herramienta informática en fase de construcción avanzada por otro técnico contratado y actual miembro del equipo, denominada Seshat (Díez, & García, este congreso). Esta herramienta produce un análisis automatizado en muy pocos segundos de miles de textos. En este momento analiza los aspectos objetivos, y está en desarrollo la obtención de los indicadores subjetivos o cualitativos. Igualmente, las medidas utilizadas en la evaluación de los textos se realizan mediante la evaluación del producto textual (MacArthur, Graham, & Fitzgerald, 2006; Olive, Kellogg & Piolat, 2002). Hasta la construcción de esta herramienta informática, la corrección de los textos se hacía de forma manual, lo que implicaba un consumo considerable de tiempo.

RESULTADOS

De la comparación de los indicadores, medidas y variables obtenidas de los dos corpus (de desarrollo y profesional), a partir de la aplicación de la herramienta informática Seshat se consideraron 18 variables del producto textual (evaluación basada en el texto), referidas a generación de información, organización de las ideas y coherencia (relacional y referencial). El número total de indicadores extraídos de la aplicación de la herramienta informática Seshat, es considerable, con medio millón para el corpus de desarrollo y tres veces más para el profesional, por lo que se han pasado los indicadores directos a porcentajes con el fin de hacer una primera aproximación general que permita su comparación.

Observando los resultados, podemos hacernos una idea de la proximidad de los porcentajes en ambos corpus, lo que indica un patrón similar que, para el caso del corpus de desarrollo, supone una conquista progresiva a lo largo de toda la escolaridad en cuanto a la adquisición de la competencia comunicativa escrita.

Se han encontrado diferencias estadísticamente significativas en el 61% de los indicadores contrastados. Pero podemos ver que hay diferencias en un mayor tipo de indicadores en cuanto a generación de ideas que en organización de ideas y coherencia. Así, en la tabla 1 de porcentajes de generación de ideas, podemos ver que hay diferencias estadísticamente significativas en los indicadores siguientes: número de párrafos, número de unidades textuales, número de palabras de contenido,



DESAFÍOS Y PERSPECTIVAS ACTUALES DE LA PSICOLOGÍA EN EL MUNDO DE LA INFANCIA

número de palabras funcionales, número de determinantes y en el total de productividad. En cambio, si tenemos en cuenta la organización de ideas y la coherencia no se han encontrado diferencias en todos los indicadores, pero sí en los siguientes: indicadores lexicales, coherencia referencial, conectores discursivos, coherencia relacional y coherencia total. En el 39% de casos restantes no se han encontrado diferencias estadísticamente significativas.

DISCUSIÓN Y CONCLUSIONES

En síntesis, podemos afirmar que se encuentran diferencias estadísticamente significativas en generación de información o productividad, en organización de la información y coherencia; si bien, el patrón de ambos corpus es similar, reflejando el patrón evolutivo y educativo, que, para el caso del corpus de desarrollo está perfeccionándose, y para el caso del corpus profesional, se supone conquistado y, en cierto sentido canónico. Practicados diversos análisis, se obtienen diferencias estadísticamente significativas en el 61% de los contrastes, aunque *las tendencias de los dos tipos de corpus en todas estas medidas es similar, lo que indica que, en esencia, los patrones evolutivos se van aproximando a los patrones profesionales en la competencia comunicativa escrita*.

Como se presentó en los resultados, el mayor número de diferencias se encontró en las variables de generación de ideas. Esto podría deberse a que la estructura, desde temprana edad, ya está desarrollada y por lo tanto es la misma en el corpus de desarrollo y en el profesional. Esto nos muestra que, quizás, hubiera sido necesario un análisis adicional de contenido semántico, del proceso de la escritura, o de otro tipo, a partir de los cuales se habrían encontrado más diferencias. Es decir, que el análisis cuantitativo sería conveniente analizarlo más en detalle y contrastarlo con otro de tipo cualitativo. Ya que las diferencias entre el corpus de desarrollo y profesional posiblemente serían mucho mayores si analizásemos los contenidos semánticos de los textos en detalle, además del número de frecuencias en indicadores diversos. El grado de especificidad de las palabras variaría posiblemente, ya que esperaríamos encontrar, en el corpus de desarrollo, un alto número de palabras de contenido pertenecientes al léxico de lengua general, mientras que en el profesional predominarían los términos pertenecientes al vocabulario de los lenguajes de especialidad (LSP) (Bergenholtz, & Nielsen, 2006).

Por último, hay que enfatizar la necesidad de añadir, al análisis cuantitativo, un análisis cualitativo de tipo semántico o de contenido. También, sería interesante estudiar el proceso *on line* desplegado en la escritura, y no solo el producto, como estamos realizando y desarrollando (Álvarez, García, & Robledo, 2011, este congreso).

REFERENCIAS

- Alamargot, D. Terrier, P., & Cellier, J. M., (2007). *Improving the Production and Understanding of Written Documents in the Workplace*. Amsterdam: Elsevier Science & Technology Books.
- Almela, R. (coord.) (2005). *Frecuencias del español: Diccionario y Estudios Léxicos y Morfológicos*. Madrid: Universitas.
- Álvarez, C. J., Carreiras, M., & de Vega, M. (1992a). Estudio estadístico de la ortografía castellana (1): La frecuencia silábica. *Cognitiva*, 4, 75-105.
- Álvarez, C. J., Carreiras, M., & de Vega, M. (1992b). Estudio estadístico de la ortografía castellana (1): Frecuencia de bigramas. *Cognitiva*, 4, 107-125.
- Bergenholtz, H., & Nielsen, S. (2006). "Subject-field components as integrated parts of LSP dictionaries", in *Terminology Journal*. Amsterdam/Philadelphia: John Benjamins, vol 12 (2): 281-303.
- Biber, D. (1993). 'Representativeness in corpus design'. *Literary and Linguistic Computing*, 8/4, 243-257.

**REVISIÓN DE ESTUDIOS DE CORPUS Y APLICACIÓN EN DESARROLLO Y APRENDIZAJE DE LA ESCRITURA
A REVIEW OF CORPORA STUDIES AND USEFULNESS IN WRITING DEVELOPMENT AND LEARNING**

- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and uses*. Cambridge, UK: Cambridge University Press.
- García-Martín, E., & García, J. N. (2011, marzo). Corpora studies in educational sciences. A review. *Actas VI Congreso Internacional de Psicología y Educación*. Valladolid: COP.
- García-Martín, E., & García, J. N. (2011, marzo). Corpus usefulness for the learning of writing and development. *Actas VI Congreso Internacional de Psicología y Educación*. Valladolid: COP.
- García-Martín, E., García, J. N., Pacheco, D. I., & Díez, C. (2009). Design of an Open Corpus and Computer Tool for Writing Development and Instruction among Students 8 to 16 years old, with and without Learning Disabilities. *Internacional Journal of Developmental and Educational Psychology*, XXI (1), 2, 107-116.
- Gregg, N., Coleman, C., Stennett, R. B., & Davis, M. (2002). Discourse complexity of college writers with and without disabilities: A multidimensional analysis. *Journal of Learning Disabilities*, 35 (1), 23-38, 56.
- Justicia, F. (1995). *El desarrollo del vocabulario. Diccionario de frecuencias*. Granada: Servicio de Publicaciones de la Universidad.
- MacArthur, C. A., Graham, S., & Fitzgerald, J. (eds) (2006). *Handbook of Writing Research*. New York/ London: The Guildford Press.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: an advanced resource book*. Oxon; New York : Routledge.
- Olive T., & Levy C. M, (eds) (2002). *Contemporary tools and Techniques for Studying writing*. Dordrecht/ Boston/ London: Kluwer Academic Publishers.
- Olive, T., Kellogg, R. T., & Piolat, A. (2002). The triple task technique for studying the process of writing. In T. Olive & C.M. Levy (Eds.), *Contemporary tools and techniques for studying writing (pp. 31-61)*. Dordrecht, A.A.: Kluwer Academic Publishers.
- Spencer, S. L., & Fitzgerald, J. (1993). Validity and Structure, Coherence, and Quality Measures in Writing. *Journal of Reading Behavior*, 25(2), 209-231.
- Stromqvist, E. (2006). <http://www.scriptlog.net/> ; Visited August 2008.
- The European Research Network on Learning to Write Effectively (ERN-LWE), (2007). COST Action IS0703. http://www.cost.esf.org/index.php?id=233&action_number=IS0703 Links to Download MoU: Memorandum of Understanding. Visited August 2008.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.