



## **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**

**Esther García-Martín, Jesús-Nicasio García, Deilis I Pacheco, & Carmen Díez**  
University of León

### ABSTRACT

We aim, in this study, to create an open monolingual corpus in Spanish for the study of the development of writing composition between typologies of students with and without LD between 8 and 16 years of age, both from a developmental and comparative approach. We pursue the building of a computerized open corpus containing written language texts by Spanish students. These texts have been compiled by the research team since 1994 in which thousands of students have been studied (at least 7232) and between 14000 and 20000 texts have been gathered and will keep on growing. We also attempt to build a computer tool that will allow automatic texts analysis and evaluation and correction of written texts from a developmental and instructional psychological perspective to be applied in the educational context and used by teachers in schools to automate the correction of texts in order to evaluate productivity, quality, structure and coherence of the written composition. During this study, a research grant was received by the regional government of Castilla and Leon (JCyL). During the development of this research we received funds from the DGI-MEC, research competitive project SEJ200-66898-EDUC (2007-2010), and with FEDER funds from the European Union, awarded to the Principal Researcher (PR) (J N García), Chairman of the research team GR259 recognized as Excellence Research Group of Castilla y León.

### **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.**

For the realization of this research we pursue the building of a computerized open corpus, containing the texts written by students between 8 to 16 years of age and collected by our research team since 1994. This open corpus which will keep on growing will allow us to create a tool for the automatic evaluation and analysis of those texts, focusing on the text based assessment (productivity, structure and coherence), and reader-based assessment (quality).

Although similar tools exist in other languages, like the one in English implemented by Biber (Biber, Conrad, & Reppen, 1998) allowing the extraction of markers of relational and referential coherence, obtaining very interesting comparative analysis and results in students with and without LD through its



## DESIGN OF AN OPEN CORPUS AND COMPUTER TOOL FOR WRITING DEVELOPMENT...

application (Gregg et al., 2002, 2006).

For our specific purposes, this will allow us to do developmental and comparative studies, and also the dissemination and selling of the tool, aiming at its generalization among teachers in their classrooms, making the computerized analysis of texts easier, as well as the productivity, the quality, the structure and the coherence, and considering the level, the age, etc., implementing programs and computerized instructional sequences and strategies, according to the level of each student in each moment.

It seems appropriate to begin by defining what a corpus is as a whole. As there are different approaches to the definition of a corpus, we will describe a set of common characteristics that seem to appear in most definitions. Although there have been many different definitions of what a corpus is since the beginning of the discipline in the early 1960's, when what Laviosa (2002: 5) calls the first generation of one-million word computer-readable corpora were created in the area of corpus linguistics, a term that first appeared in the 1980's (McEnery, Xiao and Tono, 2006: 3), it seems more useful to build a definition of corpus by gathering its most important features, which at the same time are the characteristics most authors will agree on, even if they do not agree on specific characteristics. We shall begin by stating that a corpus is a collection of texts normally in electronic form, selected and organized according to explicit selection criteria and for linguistic purposes. We will add to this basic definition by identifying a set of common features or characteristics found in many corpus definitions.

For example, Tognini-Bonelli (2001: 2) affirms that:

“A corpus can be defined as a collection of texts assumed to be representative of a given language put together so that it can be used for linguistic analysis [...] language stored in a corpus is naturally-occurring, it is gathered according to explicit design criteria, with a specific purpose in mind, and with a claim to represent larger chunks of language selected according to a specific typology. In general there is consensus that a corpus deals with natural, authentic language.”

As we can see, the basic elements of this definition are: collection of texts, representativeness, used for linguistic analysis, naturally occurring authentic language, explicit design criteria, specific purpose, specific typology. All these definitions fall within the scope of Corpus Linguistics or Applied Linguistics. Corpora can be said to have three major characteristics or advantages that can be applied to the study of written composition in our case. First, we are dealing with an empirical approach that focuses on authentic, natural language use, and as a result with a communicative approach to language. Second, what is highlighted is the importance of context and function in relation to meaning. Third, it is clear that corpus applications rely on new technologies and computer science, as it is very difficult to build a corpus without the aid of computers, it is thanks to this use of technology that the corpus has become increasingly popular among scholars as a methodological tool that can be applied to different areas of research as it allows for the storage of great quantities of texts and data retrieval for analysis in applied linguistics and now also in the field of developmental and educational psychology.

We shall not forget to mention McEnery et al. (2006), where the main characteristics of and approaches to corpus linguistics are grouped in (i) a definition of a corpus; (ii) a justification of the use of computers for the study of language; (iii) an explanation of a scientific approach to empirical data (based on real evidence), rather than an intuitive one based on assumptions; (iv) a discussion of the controversy over whether corpus linguistics is a methodology (from ancient Greek, meaning *path*) or a theory in itself (scientific discipline). In a previous and general definition by McEnery and Wilson (1996) a corpus is simply said to refer to any collection of texts, as long as it is more than one, but today a corpus includes more specific traits, including sampling and representativeness, finite size -very large in general, although there are corpora that keep constantly increasing size, machine-readable form, and used for linguistic analysis. For our own particular purposes we will take the corpus methodology and use it as a tool for the study of psychological aspects, relying on the empirical data collected by the research team on the written compositions of school children and adolescents from 8 to 16 years old and not linguistic analysis alone. It is very important though to keep always in mind the idea of representativeness (Biber, 1993) which is central in corpus research because it guides and structures the



## CONTEXTOS EDUCATIVOS ESCOLARES: FAMILIA, EDUCACIÓN Y DESARROLLO

procedures in building and analyzing corpora. Summing up, if we review the available literature in corpora we can see how most authors agree on defining a corpus as a collection of computerized texts (texts stored in electronic format with data that can be extracted using corpus retrieval software), selected according to specific criteria, for language analysis or the study of language/s, using authentic language, containing representative samples of a specific language or languages (see Table 1).

**Table 1 Some of the main characteristics normally found in a corpus definition.**

---

<i>A corpus is a collection of texts</i>	, whether they are written or spoken, full texts or parts of texts.
<i>A corpus has a communicative aspect</i>	, that is, it is real language use, authentic pieces of texts with a communicative purpose within a society or group of people, not artificial made up texts. (Bowker and Pearson 2002)
<i>A corpus can normally be found in electronic format,</i>	the texts are stored in a computer.
<i>A corpus is compiled following a specific set of criteria</i>	, and the texts are selected according to these selection criteria.
<i>A corpus is said to be representative</i>	of a specific language, sub-language, text-type, genre... (Biber 1993)
<i>A corpus should have a purpose.</i>	It is built with a purpose in mind, and it is defined according to its purpose (Hunston 2002)

---

There are many different typologies and classifications of corpora, some are original, others are modifications of the better known typologies. While we can find general typologies published and/or used in different works, created with different criteria in mind, there is no consensus agreement on a typology system. To avoid confusion in the terminology used, we find it more appropriate to present here a general classification based on the different aspects of the corpus, instead of just borrowing a specific system. The essential elements or criteria to be taken into account will be presented in sets of binary, opposed elements, or, as McEnery et al. (2006: 59) put it, according to their 'potential use'. A corpus can be general purpose (the best known in English is the British National Corpus: BNC) vs. special purpose; written vs. spoken; monolingual vs. multilingual; original language vs. translated language; synchronic vs. diachronic (the best known is the Helsinki Diachronic Corpus of English texts); open vs. closed vs. monitor (that is, a corpus that is increased on regular basis. An example of this is the Corpus de Referencia del Español Actual (CREA), from the Spanish Royal Academy of Language); learner (L2) vs. native or developmental corpus (L1) (containing child language acquisition texts in a mother tongue as it is the case in this project) vs. adult native or near-native user. We will not expand more on these as this is not the aim of our research. Further examples and explanations of existing corpora can be found in Meyer (2002).

Indeed, the type of corpus one compiles depends on the purpose of the corpus. It is therefore appropriate at this point to discuss the applied possibilities that corpora offers for research in general and more specifically to our own purposes.

In most cases, the centre of attention has been focused on the creation of knowledge, description and data analysis and not in the creation of technology. Contrastive corpus-based analysis should produced results that can be characterized by two important concepts: usefulness and usability of data (Foraker design, 1999-2008). The results obtained from this type of investigation can be put to use

**DESIGN OF AN OPEN CORPUS AND COMPUTER TOOL FOR WRITING DEVELOPMENT...**

(useful), and can also be convenient and practicable for use (usable), although it is not always possible to achieve both characteristics at the same time. This is what imposes restrictions at the applied stage.

But it is important to unveil what a corpus can do, what corpora are used for (Hunston 2002: 3, 13) and why we use corpora. A store of used language, that is a corpus, is not useful in itself; it has to be accessible by readily available software, and analyzed or compared in order to yield any results. In the same way that a language can contain an "open repertory of rich guidelines for constructing meaningful discourse events" (de Beaugrande, 2001: 11), a corpus is a potential source of data in context that allows researchers to find information that would not be available without the use of corpora and to use that information for different purposes such as the creation of dictionaries or glossaries in lexicography or terminology; creation of real language teaching material in native or foreign language teaching, the development of writing applications or aided translation software, and many others already available. With the appropriate data-retrieval software the possibilities of corpora keep on growing. Having presented a brief overview on corpora, its most common definitions, types and applications we shall proceed to place present research within the frame of educational psychology. The present research is part of a larger project that intends to develop writing composition from the psychological and instructional points of view. The aims are on the one hand the implementation of evaluative tools/instruments of the processes and psychological factors involved in writing composition, and on the other, the design and contrast of interventions for the optimization of these factors among students with and without LD (García, et al., 2009, Graham & Perin, 2007a, 2007b, McArthur, et al., 2006). In spite of all the work done since 1994 and the great amount of data collected by our research team, everything was still in paper format. Apart from this, we are also aware of similar corpus in Spain based on lexis and vocabulary that appears in children textbooks, and which aim was to extract relevant psycholinguistic data such as frequency, length, syllable-frequency, and so on (Álvarez, Carreiras, & de Vega, 1920a, 1992b), which allowed the creation of evaluative tasks for diverse purposes, but mainly to evaluate the phonological path (i.e., lists of pseudo-words with different frequency, length or syllable-frequency) or the orthographic path (i.e. homophones, highly frequent words, words with an arbitrary orthography in terms of frequency, length). Another corpus was compiled according to this purpose. It contained texts originally written by children (Justicia, 1995, Justicia, Santiago, Palma, Huertas, & Gutiérrez, 1996). There are a few other examples of similar corpora which have not been published such as the one build by the Jimenez research team in 1997, but none of the evolutive and educational nature nor with the characteristics pursued by the present Project (Almela, 2005). In none of the above cases has the corpus been considered as a methodological tool and the focus of the analysis has been mainly lexical, working at word level and its characteristics from a psycholinguistic point of view. None of them had, neither, the final purpose of being used for the analysis, and study of writing composition, nor for the focus of, not only, lexical aspects, but also, structural aspects, such as relational and referential coherence or productivity. The later have been studied only in part in English by the use of the well-known tool implemented by Biber, Conrad & Reppen (1998), with its main focus on analyzing linguistic data empirically from corpora (Corpus-based studies). Thanks studies of this type we can now apply the use of corpora to our particular needs, focusing on the complexity of written discourse in students with and without LD (Gregg, Coleman, Stennett, & Davis, 2002). Apart from this, and considering the present discussions about corpus size in the area of corpus linguistics, the area of study where corpora have been mostly developed, we could say that most corpora, used for the study of psychological or psycholinguistic aspects, are, in fact, smaller than what would be desirable to find. See for example a study by Justicia et al., (1996) which includes 2166 cases.



## CONTEXTOS EDUCATIVOS ESCOLARES: FAMILIA, EDUCACIÓN Y DESARROLLO

## METHOD

*Participants*

A total of over 7232 Spanish students, between 8 and 16 years of age, with and without LD took part in this research. They were randomly assigned to either experimental or control conditions to study the development of writing composition between age and typologies of students, both from a developmental and comparative approach; considering, not only the writing levels (quality and productivity) and the general achievement of the students, but also, as a function of other writing psychological variables (processes and different factors: planning, reflexivity, motivation and self-efficacy, self-knowledge and self-regulation), and variables such as social skills or the role of the teacher's practice in the teaching of the writing, and which modulate the achievement of the students.

All the texts have been compiled by different researchers of the team in relation to different projects all having written composition as a common core for the investigation of writing, evaluation and intervention in students with and without LD (Arias & García, 2007, Fidalgo & García, 2008, García, 2002; García & Arias, 2004; García & de Caso, 2002a; 2004; 2006a, 2006b, 2006c, 2007, García & Fidalgo, 2003; 2004; 2006, 2008a, 2008b; García & Marbán, 2003; García, et al., 2009, Torrance, Fidalgo, & García, 2007).

*Measures*

Between 14000 and 20000 written texts, have been collected, since 1994, in paper format, plus the texts that will be collected for future studies. Each texts contains about 100 words which will give us a total of about 2 million words, which will be enough to justify the representativeness of the type of language being studied, as it is a general consensus that over one million words is a good number to assure representativeness in corpora for specific purposes, that is dealing with a sub-language as opposed to general language corpora (Bowker and Pearson 2002; McEnery, Xiao and Tono 2006; Pearson 1998). Moreover, mainly, two types of measures have been implemented throughout the analysis of the texts collected: one according to a subjective reader based criteria, and also by a more formal text-based criteria. Whenever text-based measures were taken productivity measures were used. This type of assessment is concerned with the quantity of texts produced (number of paragraphs, sentences, verbs, total number of words in the essay...). Relational and referential coherence measures have also been taken. Textual length and density have been measured, too. The reader based assessment such as structure, coherence and quality, followed the criteria described by Spencer and Fitzgerald (1993), but with slight variation to make them appropriate for a comparative-contrast expository text as was the type of text analyzed for this particular purpose. Apart from quantitative measures, qualitative ones have also been used to asses written essays using a 6-point scale from 1 (difficult to understand) to 6 (excellent), the essay structure was assessed on a four point scale from 1 (unstructured) to 4 (well structured), and coherence was also assessed on a four point scale, from 1 (incoherent) to 4 (very coherent). These measures take into account the way the text is perceived by the reader as being coherent or not, of good or bad quality and well or badly structured.

*Procedure*

The scanning and proof reading of texts is a time consuming task, but the results then will be great. Moreover, normally in corpus building, the selection of criteria, finding, and sampling of the texts is one of the more time-consuming tasks, but as we already count with a big amount of textual material, we can proceed to the next step which is the scanning or typing of texts into a computer so that they are in electronic format. Firstly, we will have to scan or type all the texts into a computer so that we could have all the information in electronic format for its storage and the retrieval of information. Proof read is needed in order to eliminate typing or scanning errors. The texts will be codified following Biber's



## DESIGN OF AN OPEN CORPUS AND COMPUTER TOOL FOR WRITING DEVELOPMENT...

model. And contextual information about the text arranged according to the different variables of the students (age, gender, intellectual level, writing level, school year, typology...). Once all the texts have been introduced into the computer, all types of formatting will have to be cleared out and converted to some type of computer language such as html or xml depending on the program used to run the texts and from which information will be retrieved in order to analyze the empirical data. This part has not been fully developed yet, but if the texts are going to be processed by programs such as WordSmith Tools TM (2008), which allows the search by word and extraction of frequency lists or concordances (KWIC, Key Word in Context, useful to find patterns for example), all the texts would have to be in .txt format so that the program can run them correctly (for further information see references).

## RESULTS

### *Corpus design*

It will be a special purpose corpus geared towards a specific type of study. It is therefore domain or genre specific, and represents a particular sub-language. It will include texts written by students. Other corpora have been previously compiled using student's texts but with a different focus of analysis and in other languages, as it is the case of the ICLE (International Corpus of Learner English) project coordinated by Sylviane Granger (2003) which aimed at studying the "interlanguage of the foreign language learner" and other corpora compiled in Asia (Pravec, N.A., 2002). In our case, the corpus will be used for the study of writing compositions from a psychological point of view in students with and without LD, and the focus will be on textual writing in their mother tongue.

It will contain written language and will be monolingual containing original Spanish language. It will be open as texts will be added constantly. It could also be considered a "developmental corpus", as it contains child language acquisition texts in their mother tongue.

### *Steps carried out in the compilation of corpus*

For corpus building criteria, there are different approaches or models that we could follow (see Table 2). One of the best known in corpus linguistics is the work of McEnery et al. (2006). But in this case, we have found more appropriate and useful to follow Bowker and Pearson's guide (2002), it focuses on the use of corpora in language for specialized purposes (LSP), but describes in a practical and operational way the steps to be followed in this area of research; other proposals either deal mainly with general language corpora, or are not as specific, easy to use and detailed as this one (see also Pearson 1998).

Normally, the first step in corpus design is the identification of selection criteria for the text to include in the corpus, in our case we do not need to do this because the texts have already been compiled since the beginning in 1994. Thousands of students have been studied between 8 and 16 years of age, which has originated an amount of 14000-20000 texts total until now.

Following Bowker and Pearson's guide (2002), we would have to delimit the *time frame* for the selection of texts. As it has been agreed to create an open corpus that will keep on growing, the time frame will be large, as from 1994 to present and it will be constantly updated, and new texts will be included as new studies are carried out within the research team.

Secondly, we could decide on the *genres* or text types to be included in the corpus. Because of the research that has been done up to now, at the moment there are four basic genres to be included: description, narration, and free composition or creative writing and comparison and contrast texts.

Finally, due to the actual state of the texts, they will be in *handwritten format* and only we are only dealing with *written language*.

For these reasons, the *aim* at this stage of our research is to build an open special purpose monolingual student corpus in Spanish. The *purpose* of the corpus will be the automatic analysis and eva-



## CONTEXTOS EDUCATIVOS ESCOLARES: FAMILIA, EDUCACIÓN Y DESARROLLO

luation of these texts from the psychological perspective of quality and productivity comparison as well as the general student's performance in relation to other psychological variables such as planning, reflexivity, motivation and self-efficacy, self-concept and self-regulation.

**Table 2 Initial selection criteria for student corpus building. Developed, adapted and applied according to our research needs from Bowker and Pearson (2002).**

<i>Criteria and categories</i>	<i>Initial selection criteria</i>
Date	From 1994 to present and growing
Gender or text type	Student written compositions: compare and contrast, descriptive, argumentative, creative writing compositions...
Mode	Written language
Medium	Handwritten texts
Aim	Build open special purpose monolingual student/developmental corpus in Spanish
Purpose of corpora	Writing applications for developmental and instructional purposes
Size	Start at 14000-20000 texts (1400000 words)
Linguistic variety	Child and adolescent Castillian Spanish (from 8 to 16 years old at school level)

The matter of corpus *size* has been broadly discussed in the field of Corpus Linguistics. Although there has been an important debate on this issue, and it is generally believed that the larger the corpus the better (Sinclair 1991: 18), for specialized fields there is a consensus on a million words (Pearson 1998: 56), being enough to achieve representativeness. Representativeness is considered more important than size itself (Pearson 1998: 51). Moreover, the actual size of a corpus depends to some extent on the type of applications to be developed and the scope of the subject field being studied (Biber 1993). This clearly justifies the corpus here being built, considering that we have a minimum of 14000 texts compiled around different schools in the region and each of them has around 100 words, some of them bigger, will have at least 1400000 words to start with .

## DISCUSSION

This type of corpus studies have been very popular among linguistic researchers in the past decades as it provides the possibility of storing big amounts of data in an electronic form, but most of the studies stop at the descriptive level. With this study we will be not only applying the corpus methodology to another discipline, but also bridging the gap that exists between theory and application that in other areas such as translation studies is being demanded (see Rabadán 2008). Hardly any applications have been created in the area of developmental psychology and education, except for those mentioned above (Almela, 2005; Álvarez, Carreiras, & de Vega, 1992a, 1992b; Biber, Conrad, & Reppen, 1998; Justicia, 1995, Justicia, Santiago, Palma, Huertas, & Gutiérrez, 1996). For this reason the application of the use of corpora to this particular area of psychology and education and for developmental and instructional purposes will be of great value from a theoretical and specially applied point of view. Corpus work is very tedious in its first stages, but once all the texts have been introduced into the computer, they can be used once and again for different studies and with different purposes of analysis concerning the particular language or sublanguage included.

In this way, the corpus we aim to build will be used for two main purposes, on the one hand, for the empirical study of real data, and on the other hand, for applied purposes (see Rabadán, Labrador and Ramón 2004), to build a computer tool that will allow us an automatic analysis of the texts as well



## DESIGN OF AN OPEN CORPUS AND COMPUTER TOOL FOR WRITING DEVELOPMENT...

as their evaluation and correction from a psychological perspective. It could be applied to education and used by teachers in schools and high school levels to automate the correction of texts in order to evaluate productivity, quality, structure and coherence of the written composition. It could also be used by teachers in order to assess the student's progress and for the evaluation of texts. Finding developmental and specific patterns about difficulties and abilities in the students' written compositions is theoretically relevant in order to establish the common core of difficulties in students with and without LD in writing. It is also relevant in the school context in the way that we could help parents and teachers in the educative process, which will, consequently, follow an improvement in the students with LD achievement and self-comfort, as well as the rest of students if we know the different variables and psychological factors involved in writing, both in learning how to write and in other areas by learning by writing.

### *Future perspectives*

Other studies could be carried out with the use of corpora and its application into the social sciences, and, in our particular case, for the study of writing composition among children and adolescents ranging between 8-16 years of age at school levels, focusing on the developmental and instructional points of view in the area of educational psychology. On the one hand, developmental comparative studies could be carried out taking into consideration the difference of age of the students, contrasting different typologies of students, backgrounds and different levels. On the other hand, psycho-social and cultural studies could be developed, by the comparison of structural patterns or topic of the written compositions done by the students, and so on. Furthermore, other possible studies could derive from the corpus such as the comparative study of writing between native Spanish students and learners of Spanish as a foreign language, both in the writing process and product evaluation and assessment.

## REFERENCIAS

- Almela, R. (coord.) (2005). *Frecuencias del español: Diccionario y Estudios Léxicos y Morfológicos*. Madrid: Universitat.
- 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.
- Arias-Gundín, O., & García, J. N. (2007). Eficacia de la instrucción en los aspectos mecánicos y/o semánticos de la revisión textual. *Análisis y Modificación de Conducta*, 33(147), 5-30.
- Biber, D. (1993). 'Representativeness in corpus design'. *Literary and Linguistic Computing*, 8/4, 243-257.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and uses*. Cambridge, UK: Cambridge University Press.
- Bowker, L., & Pearson, J. (2002). *Working with Specialized Language: A Practical Guide to Using Corpora*. London; New York: Routledge.
- de Beaugrande, R. (2001). 'Large Corpora, Small Corpora and the Learning of "Language"' in M Ghadessy, A. Henry, and R. L. Roseberry (eds.), *Small Corpus Studies and ELT: Theory and Practice*, pp. 3- 27. Amsterdam: John Benjamins Publishing Company.
- Fidalgo, R., & García, J. N. (2008). El desarrollo de la competencia escrita a través de una enseñanza metacognitiva de la escritura. *Cultura y Educación*, 20(3), 307-328.
- Foraker Design (1999-2008). Foraker Design. <http://www.foraker.com/> Links to Usability Testing & Analysis. Visited August 2008.
- García, J. N., & Arias-Gundín, O. (2004). Intervención en estrategias de revisión del mensaje escrito. *Psicothema*, 16 (2), 194-202.
- García, J. N., & de Caso, A. M. (2002). ¿Es posible mejorar la composición en alumnos con dificultad?



## CONTEXTOS EDUCATIVOS ESCOLARES: FAMILIA, EDUCACIÓN Y DESARROLLO

- des de aprendizaje y/o bajo rendimiento sin que cambie la reflexividad hacia la escritura? *Psicothema*, 14 (2), 456-462.
- García, J. N., & de Caso, A. M. (2004). Effects of a motivational intervention for improving the writing of children with learning disabilities. *Learning Disabilities Quarterly*, 27 (3), 141-159.
- García, J. N., & de Caso, A. M. M. (2006a). Changes in writing self-efficacy and writing products and processes through specific training in the self-efficacy beliefs of students the learning disabled. *Learning Disabilities. A Contemporary Journal*, 4 (2), 1-27.
- García, J. N., & de Caso, A. M. M. (2006b). Comparison of the effects on writing attitudes and writing self-efficacy of three different training programs in students with learning disabilities. *International Journal of Educational Research*, 43, 272-289.
- García, J. N., & de Caso, A. M. (2007). Effectiveness of an Improvement Writing Program According to Students' Reflexivity Levels. *The Spanish Journal of Psychology*, 10 (2), 303-313.
- García, J. N., & Fidalgo, R. (2003). Desarrollo de la conciencia de los procesos cognitivos de la escritura, mecánicos frente a sustantivos y otros en alumnos de 8 a 16 años. *Psicothema*, 15, 41-48
- García, J. N., & Fidalgo, R. (2008a). Changes in the calibration of writing self-efficacy in students with learning disabilities by gender. *The Spanish Journal of Psychology*, 11 (2), 444-432.
- García, J. N., & Fidalgo, R. (2008b). The Orchestration of Writing Processes and Writing Products: A comparison of 6<sup>th</sup> Grade Students With and Without Learning Disabilities. *Learning Disabilities. A Contemporary Journal*, 6 (2), 77-98.
- García, J. N., & Fidalgo, R. (2008c). Fostering the self-regulation of the recursive thinking involved in composition writing. En A. Valle, J. C. Núñez, R. González-Cabanach, J. A. González-Pienda & S. Rodríguez (Eds.), *Handbook of Instructional Resources & Applications* (pp. 171-185). New York, NY: Nova Science Publishers.
- García, J. N., & Marbán, J. M. (2003). El proceso de composición escrita es alumnos con DA y/o BR. estudio instruccional con énfasis en la planificación. *Infancia y Aprendizaje. Journal for the Study of Education and Development*, 26 (1), 97-113.
- García, J. N., Rodríguez, C., Pacheco, D. I., Díez, C. (2009). Influencia del esfuerzo cognitivo y variables relacionadas con el TDAH en el proceso y producto de la composición escrita. Un estudio experimental. *Estudios de Psicología*, 30 (1), 31-50.
- Graham, S., & Perin, D. (2007a). A Meta-analysis of writing instruction for adolescents students. *Journal of Educational Psychology*, 99, 445-476.
- Graham, S., & Perin, D. (2007b). *Writing next. Effective strategies to improve writing of adolescents in middle and high schools*. New York: Alliance for Excellent Education.
- Granger, S. (2003). *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. S. Granger, J. Lerot and S. Petch-Tyson (eds). Amsterdam : Rodopi.
- 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.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge : Cambridge University Press.
- Justicia, F. (1995). *El desarrollo del vocabulario. Diccionario de frecuencias*. Granada: Servicio de Publicaciones de la Universidad.
- Laviosa, S. (2002). *Corpus-based translation Studies. Theory, Findings, Applications*. Amsterdam : Rodopi.
- 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.
- Meyer, C. F. (2002). *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.
- Pearson, J. 1998. *Terms in Context*. Amsterdam : John Benjamins.



## DESIGN OF AN OPEN CORPUS AND COMPUTER TOOL FOR WRITING DEVELOPMENT...

- Pravec, N.A., (2002). "Survey of Learner Corpora" *ICAME JOURNAL*, 26. April.
- Rabadán, R. , Labrador, B., & Ramón, N. (2004). 'English-Spanish Corpus-Based Contrastive Analysis: Translational Applications from the ACTRES Project.' In B. Lewandowska-Tomaszczyk (ed). *Practical Applications in Language Corpora (PALC 2003)* (pp. 141-152). Frankfurt-am-Main: Peter Lang
- Rabadán, R. (2008). 'Refining the Idea of "Applied Extensions" in A. Pym, M. Shlesinger and D. Simeoni (eds) *Beyond Descriptive Translation Studies. Investigations in homage to Gideon Toury* (pp. 103 – 117). Amsterdam/Philadelphia: John Benjamins.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Torrance, M., Fidalgo, R., & García, J. N. (2007). The teachability and effectiveness of strategies for cognitive self-regulation in sixth grade writers. *Learning and Instruction*, 17(3), 265-285.
- WordSmith Tools (2008). <http://www.lexically.net/wordsmith/version4/guides/Spanshtutorial4WordSmithsite.pdf> , Visited August.

*Fecha de recepción: 28 febrero 2009*

*Fecha de admisión: 19 marzo 2009*