

# WRITTEN LEARNER CORPORA BY SPANISH STUDENTS OF ENGLISH: AN OVERVIEW

María Belén Díez Bedmar  
*Universidad de Jaén*

## 1. Introduction

Long before the arrival of technology and the possibility to store learners' data in electronic format rather than in shoe boxes, i.e. the pre-Computer Learner Corpora era, the oral and written production by (Spanish) students of English as a second or foreign language had been compiled and analysed for various research purposes, as can be seen in conference proceedings, international journals, PhD dissertations, etc.

However, Granger's first papers on Computer Learner Corpora (CLC) in the 90s (cf. Granger, 1993; 1994; 1998, etc.) led to an ever-growing body of publications which uses a wide range of CLC with data by learners from various mother tongues learning diverse target languages. Among the most frequent research questions addressed are the description of the students' second/foreign language acquisition process, their interlanguage description, their proficiency level, or the design and piloting of teaching materials to enhance their language learning process (cf. among others, Granger, 1998; Lewandowska-Tomaszczyk and Melia, 2000; Granger, Hung and Petch-Tyson, 2002; Gilquin, Papp and Díez-Bedmar, 2008; Meunier and Granger, 2008, etc.).

As a consequence of such a substantial number of CLC and publications, some researchers have reviewed and classified them according to various criteria (cf. Pravec, 2002; Tono, 2003; Myles, 2005; Schiftner, 2008). Nevertheless, and despite its growing number, the learner corpora compiled in Spain with the oral and/or written production by Spanish learners of English have not yet undergone the same process.

Thus, this paper focuses on the main written CLC (or those having an oral and a written component) compiled with data by Spanish learners of English as a foreign language. With that objective in mind, the main CLC compiled by Research Groups with a teaching purpose, i.e. in order to analyse or describe the students' proficiency level, their interlanguage and/or create teaching materials to meet the students' needs, will be described.<sup>1</sup> It is important to notice here that only those learner corpora which may be considered as 'more typical' learner corpora, rather than the 'peripheral types' (Nesselhauf, 2004: 128) will be considered. For this reason, projects such as the *INTELeNG Project* which keeps the students' errors in a *Microsoft Access database* (cf. Mendikoetxea, Murcia and Rollinson, 2006; Mendikoetxea, Murcia and Rollinson, in press) are not included.

The methodologies used by these Research Groups to conduct the analyses of learners' language in some of their publications will be highlighted. Among them, the most frequent ones are Computer-aided Error Analysis, CEA, (Dagneaux, Dennes and Granger, 1998), Contrastive Interlanguage Analysis, CIA, (Granger, 1996) and the Integrated Contrastive Model (Granger, 1996; Gilquin, 2000/2001).

---

<sup>1</sup> Therefore, the wide range of computer learner corpora compiled by individual researchers to conduct analyses on different aspects of the students' written production in the foreign language will not be considered here.

Since the compilation and methodology used to conduct analyses of CLC depend on the research questions sought, the results obtained so far tackle diverse aspects of the students' production in the foreign language. Consequently, the overall picture of Spanish students' written command in English is somehow patchy and difficult to visualise. For this reason, this paper also provides an overview of the main research topics covered when analysing the seven learner corpora described, in an attempt to grasp a better understanding of what has been done so far regarding written CLC by Spanish students of English.

## 2. Main computer learner corpora in Spain

In the following sections the learner corpora and corpus-based publications by the members of seven Research Groups, in alphabetical order, will be described.<sup>2</sup>

### 2.1. ENWIL

*ENWIL*, *English Written Interlanguage*, was a project by researchers at the *Universidad de Alcalá* and *CENUA*, *Center for Norteamerican Studies*, which provided information on the students' production in the foreign language conducting CIAs and CEAs in a corpus of essays written by first-year students of English Philology (Valero Garcés, Mancho Barés, Flys Junquera and Cerdá Redondo, 2000a: 1855).<sup>3</sup>

Previous to the creation of the error taxonomy used for a CEA, some problematic aspects in the writing of the students in the learner corpus were highlighted qualitatively. Among them, errors related to cohesion, the use of coordination and subordination, paragraph writing and punctuation were reported (Flys, Valero, Mancho and Cerdá, 1999).

Then, a descriptive taxonomy, divided into five main categories (morphological and syntactic, lexical, discourse, spelling and punctuation), was designed (cf. Valero, Mancho, Flys and Cerdá, 2000a: 1854), and a piece of software was created to help the error-tagging of the students' texts. Similarly to the *UCL Error Editor* (Hutchinson, 1996), a tool bar with the error taxonomy facilitates the insertion of tags. This annotation can then be retrieved in the form of quantitative reports which indicate the number and the type of errors in each of the five categories (Mancho, Valero, Flys and Cerdá, 2001: 422).

The data in the learner corpus was studied by means of a CEA and a CIA to analyse the students' production before and after they had received some training in writing in the foreign language in the first year (Valero, Mancho, Flys and Cerdá, 2000b).

As a result of their investigation, a resource book addressed to Spanish students of English was published (Valero Garcés, Mancho Barés, Flys Junquera and Cerdá Redondo, 2003). Among the contents of the book, a section entitled 'Writing Effective Texts' suggests clues on how to write effective pieces of academic writing. Then, students are presented with for self didactic units based on the results of the learner corpus study and a glossary with metalinguistic terms used throughout the book.

---

<sup>2</sup> Due to space limitations, only some publications by the members of each Research Group will be mentioned.

<sup>3</sup> Even though the compilation of the production by this group of students in successive years was planned (Valero, Mancho, Flys and Cerdá, 2000: 1854), only the production by first-year students has been analysed.

## 2.2. Grupo de Investigación de Adquisición de Lenguas (GRAEL)

Together with the Research in English Applied Linguistics (REAL) Research Group at the University of the Basque Country (see Section 2.3. below), the *Grupo de Investigación en Adquisición de Lenguas* (GRAL) is also interested in analysing the role played by the age at which bilingual students begin their instruction in English as well as the hours of English classes received.

The progressive implementation of a new Education Law in Spain, by means of which students begin their instruction in the foreign language at the age of 8 (instead of at the age of 11), made it possible to compile the production by two groups of students: i) students who began studying English at the age of 8 (early onset time); and ii) those who began their English classes at the age of 11.

As described by Muñoz (2006a), the data for the *Barcelona Age Factor (BAF)* project was compiled using various questionnaires and tests (composition writing among them) from the production by 2063 bilingual students (Spanish/Catalan) and it was collected in four data samplings, i.e. after 200, 416, 726 and 926 instruction hours, even though only the first group could be tracked over the four samplings. As a result, two learner corpora have been compiled. The first one is the *BAF corpus*, containing all the above-mentioned data, and the second one the *Barcelona English Language Corpus (BELC)*, which originates from the *BAF project* but only includes the data by the subjects who could be tracked longitudinally over a period of seven years.<sup>4</sup>

This wealth of data has allowed an important number of studies, which have revealed measures to quantify the students' use of the foreign language (Celaya Villanueva, Pérez-Vidal and Torras Cherta, 2000/2001), and longitudinal or cross-sectional analyses of the oral and written data, as can be seen in the web page of the research group.<sup>5</sup> For instance, Celaya and Torras (2001) conducted a CEA to analyse the role of L1 influence on vocabulary, and Celaya Villanueva (2006) conducted a 7-year longitudinal analysis of the role played by lexical transfer in the written production in the foreign language by 16 low proficiency learners.

Among the publications by the Research Group, a book with the main results of the project stands out (Muñoz, 2006b). The chapter by Torras, Navés, Celaya and Pérez-Vidal (2006) provides a summary of the results of the four main studies conducted on the effect of age on the development of written competence by the students in the corpus. By means of measures related to fluency, accuracy, lexical and grammatical complexity, these studies analysed (i) short and mid-term effects of an early start, (ii) the comparison of the rate of acquisition between learners with different instructional time but the same age, (iii) patterns of development in writing, and (iv) the effect of age of onset in the long run.

Another chapter focuses on the oral and written production in the foreign language by two groups of students at the end of their secondary education with the same hours of instruction, but having begun them at different ages (Miralpeix, 2006).

---

<sup>4</sup> For further information on *BELC* and *BAF*, see the *BilingBank Database Guide* at <http://talkbank.org/data/manuals/BilingBank.pdf>

<sup>5</sup> Information on this Research Group can be found at <http://www.ub.edu/GRAL/publications.php>

### 2.3. REAL, Research in English Applied Linguistics Research Group

As summarized by Cenoz (2003), this Research Group at the University of the Basque Country<sup>6</sup> has used various instruments to collect data on the use of the four skills (i.e. listening, speaking, reading and writing), by bilingual Basque/Spanish learners of English as a foreign language. Among the tasks, 135 primary and secondary school students, who began their instruction in the foreign language at different ages (4, 8, and 11), but had received the same number of instruction hours, were required to write a composition.

The texts in this learner corpus were graded with the holistic approach in Jacobs, Zinkgraf, Wormuth, Hartfield and Hughey (1981), thus considering issues related to content, organisation, vocabulary, language use and mechanics. By running t-tests, the significant differences between the older and the younger groups could be highlighted (Cenoz, 2003: 85-6).

In the same edited book, results on written data obtained by means of other elicitation devices are offered (cf. García Mayo, 2003). However, if the students' compositions are considered, the chapter by Lasagabaster and Doiz (2003) provides interesting results, since they study in depth these data by analysing the letters to an English family written by 62 students at different levels (6<sup>th</sup> grade of primary school, fourth grade of secondary education and second grade in high school). To do so, triangulation of the data is achieved by means of the holistic grading of the essays, a quantitative analysis of the measures of fluency, complexity and accuracy and a CEA (Lasagabaster and Doiz, 2003: 142-144).

### 2.4. SPAINWRITE

The first corpus by this research team was the *MADRID Corpus (MAD Corpus)* (Neff, Blanco, Dafouz, Díez and Prieto, 1992), which allows CAs and CIAs and ICMs, since it consists of three subcomponents: (i) argumentative texts written by more than 200 students of English as a foreign language in the first and fourth years of the degree in English Philology; (ii) argumentative texts by those students in their L1; and (iii) the essays written in English by third-year American students of Spanish Philology in the Middlebury Program in Madrid, as a control corpus.

The first studies based on the *MAD corpus* analysed the students' problems with coherence and cohesion (Díez Prados, 2001, 2003) and writer stance, information structure techniques, and the over-, under-, or misuse of metadiscourse connectors (Neff et al. 1992; cited in Neff, Ballesteros, Dafouz, Díez, Martínez, Prieto and Rica, 2006: 566).<sup>7</sup> The use of the *MAD Corpus*, together with a specialised corpus of editorial texts by professional writers in English and Spanish, fostered the study of various measures of lexical complexity, fluency, syntactic complexity, information-structure and the use of connectors, conjuncts and conjunctions in a number of CIAs, CAs and ICMs (cf. Neff and Prieto, 1994; Neff, Dafouz, Díez and Prieto, 1997; Dafouz, Neff, Díez and Prieto, 2001; Neff, Ballesteros, Dafouz, Martínez and Rica, 2004).

A related project was conducted by Neff, Dafouz, Díez, Prieto and Chaudron (2004). In that paper, the authors aimed at investigating the development of writers' abilities in the L1 and the FL in a cross-sectional way (1<sup>st</sup> and 4<sup>th</sup> year university students), and the role that the conventions characterizing good argumentative writing in both languages played as far as

---

<sup>6</sup> Information on this Research Group can be found at <http://www.vc.ehu.es/depfi/real/>

<sup>7</sup> See Neff et al. (2006) for a review of the publications by this research team.

transfer is concerned. With that objective in mind, the *English-Spanish Contrastive Corpus* (Marín and Neff, 2000) was also used.

Apart from the *MAD corpus*, this Research Group compiled the Spanish subcomponent of the *ICLE* (Granger, Dagneaux and Meunier, 2002), *SPICLE*, and joined the subsequent *ICLE* error tagging project.<sup>8</sup> The data in *SPICLE* has been analysed by means of various CIAs which have used the *LOCNESS* as a control corpus. Thus, the aspects studied by the members of this Research Group include the use of prepositions (cf. Martínez Osés and Neff, 2001), the use of certainty and doubt in adverbs (cf. Neff, Ballesteros, Dafouz, Díez, Herrera, Martínez, Rica and Sancho, 2002), the expression of modality and evidentiality in students from various L1s in the *ICLE*, namely Dutch, French, German, Italian and Spanish (Neff, Ballesteros, Dafouz, Díez, Martínez, Prieto, Rica and Sancho, 2004), etc. A CEA has also been undertaken by the research group to gain insight into the use of collocations by Spanish learners (Ballesteros, Rica, Neff and Díez Prados, 2006).

Finally, comparisons of the data in *SPICLE*, native data in the *LOCNESS* and the use of the *English-Spanish Contrastive Corpus* to conduct CIAs, CAs and ICMs has led to fruitful research on the formulation of writer stance by means of measures for fluency, syntactic complexity and information-structure (Neff, Dafouz, Díez, Martínez, Prieto and Rica, 2003).

## 2.5. Santiago University Learner of English Corpus (SULEC)

The *Santiago University Learner of English Corpus (SULEC)*, as described in Palacios Martínez (2005) and the website of the project,<sup>9</sup> aims at compiling at least 1,000,000 words from students at elementary, intermediate and advanced levels from various degrees, namely English Philology, Education (English), Law, Translation and Interpretation. However, the corpus currently contains 500,000 words by students at secondary and university levels,<sup>10</sup> whose proficiency level has been checked with the *Oxford Placement Test* (UCLES, 2001).

The design of the corpus considers two subcomponents, i.e. an oral and a written one. The latter consists of argumentative essays, compiled partly following the *ICLE* criteria, that is, written in class without any access to reference materials, although they could be directly typed in the computer room (thus allowing the use of a spellchecker).

Much research is being done by this research group, as can be seen in the list of publications, MA and PhD dissertations in the project web page. Among them, two learner corpora, *SULEC* and *ICLE*, and a control corpus, *LOCNESS*, were used to conduct CIAs to analyse the use of 'I think' by Spanish learners Fernández Granda (2005), 'there' constructions Palacios Martínez and Martínez Insua (2005) and the use of English negation (García Fuentes, 2008).

---

<sup>8</sup> As can be seen in the list of publications in the *Centre for English Corpus Linguistics*, available online at <http://cecl.fltr.ucl.ac.be/learner%20corpus%20bibliography.html>, many papers have been published with data from the *SPICLE*.

<sup>9</sup> Information on this corpus can be found at <http://www.usc.es/ia303/SULEC/SULeC.htm>

<sup>10</sup> Palacios Martínez (March 2009, personal communication).

## 2.6. UAM Corpus de Interlenguas Escritas

This learner corpus is composed of three subcomponents.<sup>11</sup> The first one is a set of 210 essays written by secondary school students in class (20,000 words approximately). Out of these essays, 174 texts correspond to a pre- and a post- task, before and after an innovative pedagogical intervention, by 87 students in the first, second and third years of *Bachillerato* and *COU*. The other 36 essays were written by students at the same levels and under the same conditions, but do not have a pre- or post- counterpart (Barrio Luis, 2005a: 64-65).

The second set of essays is composed of the production by 119 pre-university students from different high-schools who responded to three composition topics and a cloze test (cf. Martín Úriz and Whittaker, 2005).

Finally, the production by first-year students of English Philology at the *Universidad Autónoma de Madrid*, with the same topic as that in the first subcomponent of the learner corpus, was compiled (cf. Barrio Luis, 2004).

The data in the corpus, its compilation and error-tagging have led to various publications. The error tags used for the annotation of discourse, the toolbar in *Microsoft Office* which enables the insertion of those tags and examples of information retrieval are shown in Barrio Luis (2005).

CIAs and CEAs have been done to analyse the data. For example, Chaudron, Martín Úriz and Whittaker (2001) conducted a cross-sectional study with the written production by secondary school students from the two years of *Bachillerato LOGSE*, the third year of the late *BUP* system and *COU*. Barrio Luis and Martín Úriz (2001) analysed 14 essays by secondary school students by means of a CEA in order to analyse the interpersonal and textual metadiscourse in the written production by those students, and Barrio Luis (2005b) conducted a CIA with two groups of secondary school students (those whose essays had received the lowest and the highest score), university students and native speakers from a control corpus, the *ICE-GB* in this case.

The students' use of register in English was also analysed with a CIA which included data from argumentative texts by secondary schools students, a control corpus composed of 6 texts by two graduate students and 1 native speaker (Martín Úriz and Whittaker, 2005a). More recently, the subsection of the corpus which contains letters to friends by pre-university students was employed to study gender differences in the representation of experience (Martín Úriz, Hidalgo, Murcia, Ordoñez, Vidal and Whittaker, 2007). To do so, 81 letters were analysed in two ways. First, they were divided into the generic stages of recount – orientation, events and reorientation. Second, the Systemic Coder (O'Donnell, 2005) was used to divide the clauses in the text. Consequently, it was possible to analyse each clause in the generic stages of the recount considering the presence or absence of the writer and / or date, their position, the type of process which formed the pivot of the clause and the role of writer and/or date in that process.

The main results of the improvement of essay writing in secondary school students after the intervention programme were published in a book (Martín Úriz and Whittaker, 2005b), paying attention to aspects, such as the noun phrase (Martín Úriz, Blanco Paetsch, Hidalgo Downing and Whittaker, 2005), topic development in the students' essays (Martín Úriz, Hidalgo Downing and Whittaker, 2005) or metadiscourse resources by with a CEA (Martín Úriz, Barrio Luis, Hidalgo Downing and Whittaker, 2005).

---

<sup>11</sup> Further information on this learner corpus can be found at <http://www.uam.es/departamentos/filoyletras/filoinglesa/bin/docs/investigacion/UAM%20Corpus.pdf>

## 2.7. WOSLAC project

The main objective of the *WOSLAC (Word Order in Second Language Acquisition Corpora)* project<sup>12</sup> is to analyse the properties that affect word order at the lexico-syntax and syntax-discourse interfaces in the interlanguage of Spanish learners of English, and English learners of Spanish. With that objective in mind, the researchers in this project try to reveal if the unaccusative hypothesis plays a role in the word order in L2 learners' interlanguages, if the lexicon-syntax properties are acquired before the syntax-discourse ones and, finally, if interlanguages have structures which can only be explained by means of universal properties of languages. Therefore, formal and functional approaches to understand Second Language Acquisition (SLA) data are given a crucial role, and the hypothesis that understanding linguistic phenomena in non-native grammars helps understand native grammars is supported.

In order to test their hypotheses, CIAs are conducted with the data in the Spanish subcomponent of the *International Corpus of Learner English (ICLE)*, *SPICLE*. These CIAs involve: (i) the comparison of the production of inverted subjects by the learners in the Spanish subcomponent of the *ICLE* and that of students from other L1s (cf. Lozano and Mendikoetxea, 2008a, 2008b); and (ii) comparisons of those data with the ones obtained from a control corpus, the *Louvain Corpus of Native English Essays (LOCNESS)*<sup>13</sup> (Lozano and Mendikoetxea, 2007).

Another learner corpus compiled and used by this Research Group is the *Written Corpus of Learner English (WriCLE)*,<sup>14</sup> which is composed of approximately 750 essays (amounting to 750,000 words) written by Spanish students in the first and third years of English Philology at the *Universidad Autónoma de Madrid* (Chocano, Jiménez, Lozano, Mendikoetxea, Murcia, O'Donnell, Rollinson and Teomiro, 2007; O'Donnell, Rollinson, Teomiro and Mendikoetxea, 2009). Among the characteristics of this learner corpus, it is worth mentioning that each student contributing to the corpus took the *Oxford Quick Placement Test (UCLES, 2001)*,<sup>15</sup> and the availability of the corpus, as of March 2009.

The UAM Corpus Tool (O'Donnell, 2008),<sup>16</sup> which enables automatic segmentation and annotation into layers as specified by the user, has already been used to conduct a CIA with data in *WriCLE* (cf. O'Donnell, Rollinson, Teomiro and Mendikoetxea, 2009). In fact, by annotating 500,000 words written by first and third year students, it has been possible to study the relationship between the students' proficiency level and the frequency of use of voice (active clauses vs. passive clauses), finiteness (finite clauses vs. non-finite clauses), modality (modal vs. non-modal), and non-finite clause types (present participle, past participle, infinitive clause and imperative clauses).

---

<sup>12</sup> Information on this project can be found at <http://www.uam.es/proyectosinv/woslac/>

<sup>13</sup> Information on the *LOCNESS* corpus can be found at <http://www.fltr.ucl.ac.be/fltr/germ/etan/cecl/Cecl-Projects/Icle/locness1.htm>

<sup>14</sup> Information on this corpus can be found at <http://www.uam.es/proyectosinv/woslac/WriCLE/>

<sup>15</sup> As it was also the case with the *SULEC*

<sup>16</sup> This tool is freely available at <http://www.wagsoft.com/CorpusTool/index.html>

### 3. Conclusions

Many learner corpora have been compiled by Research Groups in Spain to analyse the written production by Spanish learners of English at different proficiency levels. In some cases, 'typical' learner corpora (Nesselhauf, 2004: 128) are the type of data gathered, while on other occasions 'peripheral' ones (Nesselhauf, 2004: 128) are preferred due to the research questions to be answered, or the students' command of the foreign language at the lowest levels. Nevertheless, both types of learner corpora can be used to elicit as much information from learners as possible, thus enabling triangulation of results (cf. *BAF project*, the corpus compiled by the *REAL Research Group*, etc.). Ranging from closed word classes to the use of register or interpersonal or textual discourse, the data in these learner corpora has encouraged the study of different aspects of the students' production in the foreign language either cross-sectionally or longitudinally.

The methodologies used to analyse the data in the learner corpora are CEAs, CIAs and ICMs. When CEAs are conducted, the error-tagging taxonomies differ, as they are determined by research interests. In fact, various descriptive categories and those having to do with error gravity are mainly used (cf. Dulay, Burt and Krashen (1982: 146-197), which leads to the impossibility to compare results. CIAs involve the use of a control corpus to compare the students' production with that by native speakers or the production in the foreign language by students of English from a different L1. In the former case, the *LOCNESS* is frequently used, while in the latter the subcomponents of the *ICLE* are frequently analysed. When a corpus of Spanish L1 writing is available, ICM studies can also be conducted. Finally, comparisons of the written production by Spanish learners of English may also involve the writings by novice native writers as well as expert ones.

In this paper, only the learner corpora compiled by seven Spanish Research Groups to improve the teaching materials offered to the students or to describe their interlanguage have been described, together with their main research interests. However, many more learner corpora have been compiled (and are being compiled) by individual researchers, which have allowed the publication of an important number of papers.

The wealth of information provided by researchers using learner corpora by Spanish students of English at various levels confirms the vitality of the field in Spain. Nevertheless, further research is encouraged so that a better understanding of the acquisition process of the foreign language is gained, interlanguage patterns are described, difficult aspects of the foreign language at different proficiency levels are highlighted, etc. With the scientific data obtained from rigorous learner corpus-based studies, teaching materials (i.e. textbooks, dictionaries, etc.) which meet Spanish students' real needs can be designed and used in our classes.



## References

- Ballesteros, F., Rica, J.P., Neff, J. and M. Díez (2006): "The ICLE error tagging project: analysis of Spanish EFL writers", in C. Mourón and T. Iciar (eds.), *Studies in contrastive linguistics. Proceedings of the 4<sup>th</sup> International Contrastive Linguistics Conference, September 2005, Santiago de Compostela*, Servizo de Publicacións e Intercambio Científico, 89-97.
- Barrio Luis, M. and A. M. Martín Úriz (2001): "Metadiscursio en textos en inglés de estudantes de secundaria españoles", in A. Moreno and V. Colwell (eds.), *Perspectivas recientes sobre el discurso*, León: Universidad de León.
- Barrio Luis, María (2004): "Lexical density and group elaboration as indicators of written style in native and EFL compositions", in J. M. Oro Cabanas, J. Anderson and J. Varela Zapata (eds.), *Linguistic perspectives from the classroom: language teaching in a multicultural Europe*, Santiago de Compostela, Servicio de Publicaciones de la Universidad de Santiago de Compostela, 203-216.
- Barrio Luis, María (2005a): "Diseño del corpus de interlenguas de textos escritos en inglés lengua extranjera", in A. Martín Úriz and R. Whittaker (eds.), *La composición como comunicación: una experiencia en las aulas de lengua inglesa en bachillerato*, Madrid, Ediciones de la Universidad Autónoma de Madrid, 61-75.
- Barrio Luis, Maria (2005b): "Lexical density and noun group elaboration as indicators of written style in native and EFL compositions", in J. Anderson, J. Varela Zapata and J. Manuel Oro Cabanas (eds.), *La enseñanza de las lenguas en una Europa multicultural*, Santiago de Compostela, Universidad de Santiago, 1005-1018.
- Celaya Villanueva, M. L., Pérez-Vidal, C., and M. R. Torras Cherta (2000/2001): "Matriz de criterios de medición para la determinación del perfil de competencia lingüística escrita en inglés (LE)", *RESLA* 14, 87-98.
- Celaya, M. L. and R. Torras (2001): "L1 influence and EFL vocabulary: do children rely more on L1 than adult learners?", in M. Falces Sierra, M. Díaz Dueñas and J. M. Pérez Fernández (eds.), *Actas del 25º Congreso AEDEAN (Asociación Española de Estudios AngloNorteamericanos), 13, 14 y 15 de Diciembre, Departamento de Filología Inglesa, Universidad de Granada*, Granada, Bernardi Producciones S.L.
- Celaya Villanueva, M. L. (2006): "Lexical transfer and second language proficiency: a longitudinal analysis of written production in English as a foreign language", in A. Alcaráz Sintes, C. Soto and M. C. Zunino-Garrido (eds.), *Proceedings of the 29<sup>th</sup> AEDEAN International Conference*, Jaén, Servicio de Publicaciones de la Universidad de Jaén, 293-297.
- Cenoz, J. (2003): "The influence of age on the acquisition of English: general proficiency, attitudes and code-mixing", in M. P. García Mayo and M. L. García Iecumberri (eds.), *Age and the acquisition of English as a Foreign Language*, Clevedon: Multilingual Matters, 77-93.
- Chaudron, C., Martín Úriz, A. and R. Whittaker (2001): "La composición como comunicación: influencia en el desarrollo general del inglés como segunda lengua en un contexto de instrucción", in I. de la Cruz Cabanillas, C. Santamaría García, C. Tejedor Martínez and C. Valero Garcés (eds.), *La Lingüística Aplicada a finales del siglo XX. Ensayos y propuestas. Tomo 1*, Alcalá de Henares, Servicio de Publicaciones de la Universidad de Alcalá de Henares, 53-62.
- Chocano, G., Jiménez, R., Lozano, C., Mendikoetxea, A., Murcia, S., O'Donnell, M., Rollinson, P. and I. Teomiro (2007): "An exploration into word order in learner corpora: The WOSLAC Project", in M. Davies, P. Rayson, S. Hunston and P.

- Danielsson (eds.), *Proceedings of the Corpus Linguistics Conference 2007, Conference of Birmingham*.  
 <[http://ucrel.lancs.ac.uk/publications/CL2007/paper/113\\_Paper.pdf](http://ucrel.lancs.ac.uk/publications/CL2007/paper/113_Paper.pdf)>
- Dafouz, E., Neff, J., Díez, M. and R. M. Prieto (2001): “Complejidad sintáctica del texto: adquisición de oraciones con verbo en forma no personal en lengua inglesa”, in I. de la Cruz Cabanillas, C. Santamaría García, C. Tejedor Martínez and C. Valero Garcés (eds.), *La Lingüística Aplicada a finales del siglo XX. Ensayos y propuestas, Tomo 1*, Alcalá de Henares, Servicio de Publicaciones de la Universidad de Alcalá de Henares, 45-51.
- Dagneaux, E., Dennes, S. and S. Granger (1998) : “Computer-aided error analysis”, *System*, 26, 163-174.
- Díez Prados, M. (2001): “Frecuencia de los mecanismos de cohesión en textos escritos en lengua inglesa”, in I. de la Cruz Cabanillas, C. Santamaría García, C. Tejedor Martínez and C. Valero Garcés (eds.), *La Lingüística Aplicada al final del s. XX. Ensayos y propuestas*, Alcalá de Henares, Universidad de Alcalá de Henares, 525-530.
- Díez Prados, M. (2003): *Coherencia y cohesión en textos escritos en inglés por alumnos de Filología Inglesa (estudio empírico)*, Alcalá, Servicio de Publicaciones de la Universidad de Alcalá.
- Dulay, H. C., Burt, M., K. and Krashen, S. (1982): *Language Two*, Oxford, Oxford University Press.
- Fernández Granda, A. B. (2005): “*I think* in argumentative student writing”, *GRETA*, 13, 1,2, 17-22.
- Flys, C., Valero, C., Mancho, G. and E. Cerdá (1999): “Error analysis in the teaching of academic writing in English”, in J. L. Chamosa González (ed.), *Proceedings of the 23<sup>rd</sup> International Conference of AEDEAN*, León, Servicio de Publicaciones de la Universidad de León.
- García Fuentes, A. (2008): “The use of English negation by Spanish students of English: A learner corpus-based study”, *Revista de lingüística y lenguas aplicadas*, 3, 48-58.
- Gilquin, G. (2000/2001): “The Integrated Contrastive Model. Spicing up your data”, *Languages in Contrast* 3, 1, 95-123.
- Gilquin, G., Papp, Sz. and M. B. Díez-Bedmar (eds.) (2008): *Linking up Contrastive and Learner Corpus Research*, Amsterdam and New York, Rodopi.
- Granger, S. (1993): “International corpus of learner English”, in J. Aarts, P. de Haan and N. Oostdijk (eds.), *English Language Corpora: Design, Analysis and Exploitation. Papers from the Thirteenth International Conference on English Language Research on Computerized Corpora, Nijmegen, 1992*, Amsterdam and Atlanta: Rodopi, 57-96.
- Granger, S. (1994): “The learner corpus: a revolution in applied linguistics”, *English Today* 39, 10, 3, 25-29.
- Granger, S. (1996): “From CA to CIA and back: an integrated approach to computerized bilingual and learner corpora”, in K. Aijmer, B. Altenberg and M. Johansson (eds.), *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies. Lund 4-5 March 2004*, Lund, Lund University Press, 37-51.
- Granger, S. (ed.) (1998): *Learner English on Computer*, London and New York, Addison Wesley Longman.
- Granger, S., Dagneaux, E. and F. Meunier (2002): *The International Corpus of Learner English. Handbook and CD-ROM*, Louvain-la-Neuve, Presses Universitaires de Louvain.

- Granger, S., Hung, J. and S. Petch-Tyson (eds.) (2002): *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, Atlanta and Amsterdam, John Benjamins.
- Hutchinson, J. (1996): *UCL Error Editor*, Louvain-la-Neuve: Centre for English Corpus Linguistics, Université Catholique de Louvain.
- Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfield, V. F. and J. B. Hughey (1981): *Testing ESL Composition*, Newbury, Rowley.
- Lasagabaster, D. and A. Doiz (2003): "Maturational constraints on foreign-language written production", in M. P. García Mayo and M. L. García Lecumberri (eds.), *Age and the Acquisition of English as a Foreign Language*, Clevedon, Multilingual Matters, 136-160.
- Lewandowska-Tomaszczyk, B. and P. J. Melia (eds.) (2000), PALC'99: Practical Applications in Language Corpora. Papers from the International Conference at the University of Łódź, 15-18 April 1999, Frankfurt am Main, Peter Lang.
- Lozano, C. and A. Mendikoetxea (2007): "Learner corpora and the acquisition of word order: A study of the production of Verb-Subject structures in L2 English", in M. Davies, P. Rayson, S. Hunston and P. Danielsson (eds.), *Proceedings of the Corpus Linguistics Conference 2007, University of Birmingham* <[http://ucrel.lancs.ac.uk/publications/CL2007/paper/112\\_Paper.pdf](http://ucrel.lancs.ac.uk/publications/CL2007/paper/112_Paper.pdf)>
- Lozano, C. and A. Mendikoetxea (2008a): "Postverbal subjects at the interfaces in Spanish and Italian learners of L2 English: a corpus study", in G. Gilquin, Sz. Papp and M.B. Díez-Bedmar (eds.), *Linking up Contrastive and Learner Corpus Research*, Amsterdam, Rodopi, 85-125.
- Lozano, C. and A. Mendikoetxea, A. (2008b): "Verb-subject order in L2 English: new evidence from the ICLE corpus", en R. Monroy and A. Sánchez (eds.), *25 años de Lingüística Aplicada en España: Hitos y retos*, Murcia, Editum, 97-113.
- Mancho, G., Valero, C., Flys, C. and E. Cerdá (2001): "ENWIL: un corpus computerizado de errores de estudiantes de EFL", in I. de la Cruz Cabanillas, C. Santamaría García, C. Tejedor Martínez and C. Valero Garcés (eds.), *La Lingüística aplicada a finales del siglo XX. Ensayos y propuestas*, Alcalá, Universidad de Alcalá, 419-425.
- Marín, J. and J. Neff (2000): *Spanish-English Contrastive Corpus*, UCM, Madrid.
- Marín Úriz, A., Barrio Luis, M., Hidalgo Downing, L. and R. Whittaker (2005): "El metadiscursio: el escritor como organizador y evaluador de su texto", in A. Martín Úriz and R. Whittaker (eds.), *La composición como comunicación: una experiencia en las aulas de lengua inglesa en bachillerato*, Madrid, Ediciones de la Universidad Autónoma de Madrid, 131-152.
- Martín Úriz, A., Blanco Paetsch, S. Hidalgo Downing, L. and R. Whittaker (2005): "Desarrollo y complejidad del sintagma nominal en composiciones en lengua inglesa de estudiantes de bachillerato" in A. Martín Úriz and R. Whittaker (eds.), *La composición como comunicación: una experiencia en las aulas de lengua inglesa en bachillerato*, Madrid, Ediciones de la Universidad Autónoma de Madrid, 99-114.
- Martín Úriz, A., Chaudron, C. and R. Whittaker (2005): "Caracterización lingüística de las composiciones de los estudiantes de bachillerato", in A. Martín Úriz and R. Whittaker (eds.), *La composición como comunicación: una experiencia en las aulas de lengua inglesa en bachillerato*, Madrid, Ediciones de la Universidad Autónoma de Madrid, 79-97.
- Martín Úriz, A., Hidalgo Downing, L. and R. Whittaker (2005): "El desarrollo del tema en la composición de estudiantes de secundaria: medidas de coherencia", in A. Martín Úriz and R. Whittaker (eds.), *La composición como comunicación: una experiencia en las*

- aulas de lengua inglesa en bachillerato*, Madrid, Ediciones de la Universidad Autónoma de Madrid, 115-129.
- Martín Úriz, A. and R. Whittaker (2005a): “Argumentative text in EFL in Spanish pre-university students”, in M. L. Carrió Pastor (ed.), *Perspectivas interdisciplinarias de la Lingüística Aplicada*, València, Universitat Politècnica de València, 203-124.
- Martín Úriz, A., and R. Whittaker (eds.) (2005b), *La composición como comunicación: una experiencia en las aulas de lengua inglesa en bachillerato*, Madrid, Ediciones de la Universidad Autónoma de Madrid.
- Martín Úriz, A. M., Hidalgo, A., Murcia, S., Ordoñez, L., Vidal, K. and R. Whittaker (2007): “Gender differences in recounts written in English by Spanish pre-university students”, in R. Monroy and A. Sánchez (eds.), *25 años de Lingüística Aplicada en España. Hitos y retos*, Murcia, Editum, 123-130.
- Martínez Osés, F. and J. Neff (2001): “Corpus analysis of prepositional patterns and non-native university writing”, in C. Muñoz, M. L. Celaya, M. Fernández-Villanueva, T. Navés, O. Strunk, E. Tragant (eds.), *Trabajos en Lingüística Aplicada*, Barcelona, Univerbook, 139-147.
- Mendikoetxea, A., Murcia, S. and P. Rollinson (2006): “Los corpus de aprendices como herramienta: descripción de una base de datos de errores para el desarrollo de material pedagógico para estudiantes universitarios”, in M. Amengual, M. Juan and J. Salazar (eds.), *Adquisición y enseñanza de lenguas en contextos plurilingües. Ensayos y propuestas aplicadas*, Illes Balears, Server de Publicacions i Intercanvi Científic, 334-354.
- Mendikoetxea, A., Murcia, S. and P. Rollinson (in press): “Focus on errors: learner corpora as pedagogical tools”, in M. C. Campoy, B. Bellés Fortuno and M. L. Gea-Valor (eds.), *Corpus-Based Approaches to English Language Teaching*, London, Continuum.
- Meunier, F. and S. Granger (eds.) (2008), *Phraseology in Foreign Language Learning and Teaching*, Amsterdam and Philadelphia, Benjamins.
- Miralpeix, I. (2006): “Age and vocabulary acquisition in English as a Foreign Language (EFL)”, in C. Muñoz (ed.), *Age and the Rate of Foreign Language Learning*, Clevedon, Multilingual Matters, 89-106.
- Muñoz, C. (2006a): “The effects of age on foreign language learning: the BAF project”, in C. Muñoz (ed.), *Age and the Rate of Foreign Language Learning*, Clevedon, Multilingual Matters, 1-40.
- Muñoz, C. (2006b): *Age and the Rate of Foreign Language Learning*, Clevedon, Multilingual Matters.
- Myles, F. (2005): “Interlanguage corpora and SLA research”, *Second Language Research* 21, 4, 373-391.
- Neff, J., Blanco, M., Dafouz, E., Díez, M. and Prieto, R. (1992): “*The MADRID Corpus (MAD)*”, Held in the Department of English Philology, Universidad Complutense de Madrid.
- Neff, J. and R. Prieto (1994): “First language influence on Spanish EFL University Writing Development”, in ERIC Clearing House on Language and Linguistics (ERIC Document Reproduction Service No. ED 385 144), Washington, D.C., Educational Resource Information Center, U.S. Department of Education.
- Neff, J., Dafouz, E., Díez, M. and R. M. Prieto (1997): “Information structuring in Spanish EFL University writing”, in R. J. Sola, L. A. Lázaro and J. A. Gurpegui (eds.), *XVIII Congreso de AEDEAN (Alcalá de Henares, 15-17 diciembre 1994)*, Alcalá de Henares, Servicio de Publicaciones de la Universidad, 789-797.

- Neff, J., Ballesteros, F., Dafouz, E., Díez, M., Herrera, H., Martínez, F., Rica, J. P. and C. Sancho (2002): "A contrastive study of certainty and doubt adverbs in native and non-native argumentative texts", in L. Iglesias Rábade (ed.), *Studies of Contrastive Linguistics*, Santiago de Compostela, Servicio de Publicaciones de la Universidad de Santiago de Compostela, 747-753.
- Neff, JoAnne, Francisco Ballesteros, Emma Dafouz, Mercedes Díez, Honesto Herrera, Francisco Martínez, Rosa Prieto, Juan Pedro Rica and Carmen Sancho (2003): "Contrasting learner corpora: the use of modal and reporting verbs in the expression of writer stance", in S. Granger and S. Petch-Tyson (eds.), *Extending the Scope of Corpus-based Research*, Amsterdam, Rodopi, 211-230.
- Neff, J., Ballesteros, F., Dafouz, E., Martínez, F. and J. P. Rica (2004): "Formulating writer stance: a contrastive study of EFL learner corpora", in U. Connor and T. A. Upton (eds.), *Applied Corpus Linguistics. A Multidimensional Perspective*, Amsterdam and New York, Rodopi, 73-89.
- Neff, J., Dafouz, E., Díez, M., Prieto, R. and C. Chaudron (2004): "Contrastive discourse analysis. Argumentative text in English and Spanish", in C. L. Moder and A. Martinoviz-Zic (eds.), *Discourse Across Languages and Cultures*, Amsterdam, John Benjamins, 267-284.
- Neff, J., Ballesteros, F., Dafouz, E., Díez, M., Martínez, F., Prieto, R. and J. P. Rica (2006): "A decade of English-Spanish contrastive research on written discourse: the MAD and SPICLE corpora", in M. Carretero, L. Hidalgo Downing, J. Lavid, E. Martínez Caro, J. Neff, S. Pérez de Ayala and E. Sánchez-Pardo (eds.), *A pleasure of life in words. A festschrift for Angela Downing, Volume II*, Madrid, CERSA, 199-218.
- Nesselhauf, N. (2004): "How learner corpus analysis can contribute to language teaching: A study of support verb constructions", in G. Aston, S. Bernardini and D. Stewart (eds.), *Corpora and Language Learners*, Amsterdam and Philadelphia, John Benjamins, 109-124.
- O'Donnell, M. (2005): *The Systemic Coder Version 4.68*, Available online at <http://www.wagsoft.com>.
- O'Donnell, M. (2008): "The UAM Corpus Tool: Software for corpus annotation and exploration", paper presented at the *XXVI Congreso de AESLA*, Almería, Spain, 3-5 April 2008.
- O'Donnell, M., Rollinson, P., Teomiro, I. and A. Mendikoetxea (2009): "Preparing a learner corpus for online distribution and search: the WRICLE corpus", paper presented at the *XXVII AESLA International Conference*, Universidad de Castilla la Mancha, 26-28 March 2009.
- Palacios Martínez, I. (2005): "Las nuevas tecnologías y la investigación en el campo de la adquisición de segundas lenguas", in M. Cal Varela, P. Núñez Pertejo and I. M. Palacios Martínez (eds.), *Nuevas Tecnologías en Lingüística, Traducción y Enseñanza de Lenguas*, Santiago, Servicio de Publicaciones de la Universidad, 205-225.
- Palacios-Martínez, I. M. and A. Martínez-Insua (2006): "Connecting linguistic description and language teaching: native and learner use of existential there", *International Journal of Applied Linguistics*, 16, 2, 213-231.
- Pravec, N. (2002): "Survey of learner corpora", *ICAME Journal* 26, 81-114.
- Schiftner, G. (2008): "Learner corpora of English and German: What is their status quo and where are they headed?", Vienna English Working PaperS 17,2, 47-78 (Available online at [http://www.univie.ac.at/Anglistik/views\\_0802.pdf](http://www.univie.ac.at/Anglistik/views_0802.pdf)).
- Tono, Y. (2003): "Learner corpora: design, development and applications", in D. Archer, P. Rayson, A. Wilson and T. McEnery (eds.), *Proceedings of the Corpus Linguistics*

- 2003 Conference (28-31 March), Lancaster: University Centre for Computer Corpus Research on Language, 800-809
- Torras, M. R., Navés, T., Celaya, M. L. and C. Pérez-Vidal (2006): "Age and IL development in writing", in C. Muñoz (ed.), *Age and the Rate of Foreign Language Learning*, Clevedon, Multilingual Matters, 156-182.
- UCLES (2001): *Quick Placement Test (Paper and pencil version)*, Oxford, Oxford University Press.
- Valero Garcés, C., Mancho Barés, G., Flys Junquera, C. and E. Cerdá Redondo (2000a): "Evolución de la interlengua y análisis de textos: ENWIL y el análisis de errores en la expresión escrita en EFL", in F. Alcántara Iglesias, A. Barreras Gómez, R. Jiménez Catalán and M. Á. Moreno Lara (eds.), *Panorama actual de la Lingüística Aplicada. Conocimiento, procesamiento y uso del lenguaje*, vol. 3, Logroño, Mugar Linotype, 1849-1860.
- Valero, C., Mancho, G., Flys, C. and E. Cerdá (2000b): "Análisis informatizado de errores: ENWIL", *Sintagma*, 12, 49-59.
- Valero, C., Mancho, G., Flys, C. and E. Cerdá (2003). *Learning to Write: Error Analysis Applied*, Universidad de Alcalá, Servicio de Publicaciones.