Spanish pre-university students' use of English: CEA results from the University Entrance Examination

MARÍA BELÉN DÍEZ-BEDMAR
Universidad de Jaén

ABSTRACT

In this paper an updated overview of the main errors that Spanish students make when writing the English exam in the University Entrance Examination is provided. To do so, a Computer-aided Error Analysis (CEA) (Dagneaux, Denness & Granger, 1998) was conducted on a representative sample of the students who took the exam in June 2008 in Jaén, and wrote a composition on the same topic. The use of the most widely-used error taxonomy, the *Error Tagging Manual* version 1.1. (Dagneaux, Denness, Granger & Meunier, 1996), and the analysis of the results by means of descriptive statistics foster the possibility to replicate this study and move forward in the description of the students' written command in the foreign language at this stage. The comparison of the findings obtained in this study and those from previous (C)EAs on the English exam reveals that some common tendencies may be shown.

KEYWORDS: English exam, University Entrance Examination, Computer-aided Error Analysis

RESUMEN

Este artículo aporta una visión general actualizada de los errores más importantes que los alumnos españoles cometen cuando escriben el examen de inglés de la Prueba de Acceso a la Universidad. Con este fin, se realizó un Análisis de Errores Informatizado (Dagneaux, Denness & Granger, 1998) en una muestra representativa de los estudiantes que realizaron el examen en Junio de 2008 en la Universidad de Jaén, y que escribieron sobre un mismo tema. El uso de la taxonomía de errores más usada, el *Error Tagging Manual* version 1.1. (Dagneaux, Denness, Granger & Meunier, 1996), y de la estadística descriptiva para analizar los resultados, favorecen la réplica de este estudio y la posibilidad de mejorar la descripción de la destreza escrita en la lengua extranjera. La comparación de los resultados en este artículo y otros Análisis de Errores (Informatizados) previos muestran que se pueden exponer algunas tendencias comunes.

PALABRAS CLAVE: Examen de Inglés, Pruebas de Acceso a la Universidad, Análisis de Errores Informatizado

---

*Address for correspondence: Universidad de Jaén. Paraje las Lagunillas. Edificio D-2. 23071 Jaén. T. +34 953213555. Email: belendb@ujaen.es

1. INTRODUCTION

The University Entrance Examination (UEE, henceforth) in Spain is a high-stakes exam which may determine the degrees which students may access, so it is crucial that the reliability and validity of the individual subject exams which compose it are analysed and improved, if necessary. Being conscious of the importance of these two "measurement qualities" (Bachman & Palmer, 1996: 19, 23), studies have been conducted on individual subject exams of the UEE at different universities, regarding rater-related issues (the rater's gender, his/her working place, or the board of raters), exam-related issues (the exam formats, the type of evaluation done, i.e. holistic, focused-holistic or analytic, the individual subject exam rated, i.e, biology, foreign language, etc.), and candidate-related issues (for instance, the test-taker's attitude towards the exam)[1].

Among the individual subject exams analysed, the one for the English language has also been the focus of research. According to García Laborda (2006), the studies which have considered any aspect related to the English exam in the Spanish UEE have been divided into three categories: first, the studies in which the validity the exam design is tackled; second, the ones which offer results regarding the issues of construct validity, inter- and intra-rater reliability, the raters' scoring process, the raters' gender or working place and their effect on the rating process, etc.; and, finally, those which highlight the need for the improvement of this individual subject exam and suggest future lines of research.

No doubt, these publications have proved crucial for the progressive improvement of the English exam and its evaluation along the years. However, there is a fourth block of studies which have paid attention to the students' written command of the foreign language (FL) at this stage of their FL learning process or interlanguage (see Díez-Bedmar, 2011), as reflected in their pieces of writing in the English exam in the UEE. This block is composed of the studies by Crespo García (1999), Doval Suárez (1999), González Álvarez (1999), Iglesias Rábade (1999b), Woodward Smith (1999), Wood Wood (2002) and Rodríguez Aguado (2004), which have provided their results by means of a (Computer)-aided Error Analysis[2] (CEA) (Dagneaux, Denness & Granger, 1998) or an Interlanguage Analysis (IA) (Selinker, 1972, 1992)[3].

However, four main limitations may be found in the studies included in the fourth block, which call for an updated description of the students' main problems when writing in the FL. First, the most recent one dates from 2004 (Rodriguez Aguado, 2004); second, the type of quantitative data provided, if any, is percentages; third, each study focuses on a limited number of linguistic aspects in the FL. For instance, Crespo García (1999) dealt with morpho-syntactic errors, Doval Suárez (1999) focused on spelling errors, González Álvarez (1999) on lexical errors, Iglesias Rábade (1999b) analysed textual organization, Woodward Smith (1999) considered the use of closed word classes, and Wood Wood (2002) article use. It was Rodriguez Aguado's (2004) PhD dissertation the one which offered a broader view of

the students' written command of the FL, since he provided results concerning morphology, syntax, orthography and vocabulary use. Consequently, the findings in these studies offer an incomplete description of the students' command of the FL. Fourth, the methodology used in these studies is different, since various error taxonomies are employed, different writing tasks (e.g. short answers vs. composition writing) are analysed, etc. Therefore, the results in these analyses are not fully comparable either between them, or with other CEAs conducted, even though they offer a valuable insight into the description of the Spanish students' command of the FL at this stage.

The current need to describe and define the students' (written) level of English when entering Higher Education, specially in the European Higher Education Area (EHEA), by means of a methodology which may allow the comparison of data regarding the type of error taxonomy used and the type of results (i.e. descriptive statistics, rather than percentages) fostered the research presented in this paper. Thus, this study pursued two main objectives. First, to offer an updated overview of the students' main problems when using the FL to write a composition for the English exam in the UEE. Second, to do so by employing a methodology which would provide comparable results, thus overcoming this limitation in previous studies on the written language used by students when taking this individual subject exam.

Therefore, this paper presents the results obtained from a CEA conducted on a representative sample of the compositions written on the same topic for the English exam in the UEE in Jaén in June 2008. To ensure the comparability of findings, the *UCL Error Editor* (Hutchinson, 1996), and the widely-used *Error Tagging Manual, version 1.1.* (Dagneaux, Denness, Granger & Meunier, 1996) were employed. The use of this error-taxonomy by a wide range of scholars from all over the world, as can be seen in many edited books (Granger, 1998; Granger, Hung & Petch-Tyson, 2002; Gilquin, Papp & Díez-Bedmar, 2008, etc.), the publications listed in the learner corpus bibliography compiled by the *CECL*, or the development of the *ICLE Error Tagging Project* (Ballesteros, Rica, Neff & Díez Prados, 2006), allows the needed replica studies (see, for instance, Polio & Gass, 1997).

This paper is divided into five sections. After this introduction, a summary of the studies which have provided an overview of the students' main errors when writing in the FL in the English exam in the UEE is offered in section 2. The methodology employed for this study is described in section 3, and the results obtained are offered in section 4 by means of two sub-sections. The first one, (section 4.1.), presents the findings obtained as a consequence of the CEA conducted, which allows an updated description of the students' profile at this stage of their FL acquisition. In section 4.2. a comparison is made between the findings in this study and those presented in the above-mentioned research on the students' written production in the English exam in the UEE, so that it is possible to point to interesting tendencies regarding the common errors made by secondary-school leavers. Although the limitations posed by the different methodologies prevent a full comparison of results,

tendencies in the students' problems when writing in the FL at this stage will be shown. Finally, the main data in this study are summarized and the main limitations and lines for future research are described in section 5.

2. AN OVERVIEW OF THE STUDENTS' MAIN ERRORS WHEN WRITING THE ENGLISH EXAM IN THE UNIVERSITY ENTRANCE EXAMINATION

The seven studies devoted to the analysis of the students' written production in the English exam in the UEE can be found in an edited book by Iglesias Rábade (1999a), and in two PhD dissertations, namely the one by Wood Wood (2002) and that by Rodriguez Aguado (2004)[4]. The five papers in Iglesias Rábade (1999a) correspond to the analysis of the English exam in the UEE in June 1995 in Galicia (Crespo García, 1999; Doval Suárez, 1999; González Álvarez, 1999; Iglesias Rábade, 1999b; Woodward Smith, 1999). Doval Suárez (1999) focused on spelling errors, González Álvarez (1999) on lexical ones, Iglesias Rábade (1999b) analysed textual organization, and Woodward Smith (1999) the use of closed word classes. Finally, the chapter by Crespo García (1999) offered a description of the students' main morpho-syntactic errors when writing the exam. As far as the two PhD dissertations are concerned, the purpose of Wood Wood (2002) was to conduct an IA on the students' article use, whereas the PhD by Rodríguez Aguado (2004) showed the findings of a CEA conducted on different aspects of the FL, as divided in different error categories, namely vocabulary, orthography, morphology and syntax

In this section, a summary of the main findings in the studies which analyse the students' main errors in various error-categories when writing in the FL (i.e. Crespo García, 1999; and Rodríguez Aguado, 2004) will be offered to provide a general overview of the students' written command of the FL at this stage. Although the wealth of data in the other studies is valuable to gain a deep insight into the students' use of specific aspects of the FL, they will not be considered here.

2.1. Morpho-syntactic errors

Morpho-syntactic errors were the focus of research in the publication by Crespo García (1999). In the learner corpus used, 2,117 errors were identified in 500 exams by means of a (C)EA which used a linguistic category classification and a surface structure taxonomy (Dulay, Burt & Krashen, 1982: 146-197). As a result, most errors proved to be morphological (80%), then followed by morpho-syntactic (17%) and syntactic ones (3%).

The analysis of the errors made by students in each word class revealed that the word class which posed more problems was that of verbs (58%), followed by adjectives (11%) and nouns (7%). In fact, open word classes were the ones with more errors (76.1%), followed by

the closed word classes (14.4%). If the errors in the word class verbs are considered, main verbs posed more problems than modal auxiliary verbs[5]. Within the former, finite forms showed the highest percentage (81%), especially due to the problems that students had with the inflectional morpheme for the third person singular, the verb forms in main verb phrases and the problems related to their regular or irregular pattern. Derivational morphology caused a high percentage of errors in the production of items in open word classes. In fact, 39% of the errors in adjectives corresponded to derivational morphemes, 33.3% of the errors in the word class noun, and 66.6% in adverbs. The second typical error was related to inflectional morphology, involving the use of a morpheme to erroneously mark the plural number of adjectives in English (21.7%), or problems of agreement in nouns (33.3%).

As far as closed word classes are concerned, relative pronouns are the items which showed the highest percentage of errors (37.5%). Regarding determiners in the learner corpus, the use of the zero article was the most problematic aspect (77% of errors), as compared to the indefinite (14.6%) and the definite (8.4%) ones. In the case of demonstratives, the grammatical category number was the main cause of error, since the demonstrative was used in the singular when the head of the noun phrase was in the plural.

2.2. Vocabulary, morpho-syntax, spelling and discourse

The learner corpus compiled by Rodríguez Aguado in Valladolid in 1996 consists of 123 compositions (12,204 words) which were analysed by means of a (C)EA in which four major error categories were considered, namely vocabulary, morpho-syntax, spelling and discourse. Thus, the 1,325 errors found in the learner corpus were distributed into the four categories as follows, in decreasing percentage of errors, morpho-syntax (47%), vocabulary (27.8%), spelling (12.7%) and discourse (12.5%).

To begin with morphosyntactic errors, the word class verb was the one which presented the highest percentage of errors (38%), closely followed by the errors found in determiners (37%). If the word class verbs is analysed, the data reveals that the errors which stem from the omission of the necessary inflectional morphemes for person and number were the most frequent ones (31%), followed by the form in non-finite verb phrases (25%). Among determiners, articles were the ones posing more problems to students at this level (71% of the errors), followed by the use of other determiners (14%), possessives (9%) and demonstratives (6%). The analysis of the errors in the definite article reveals that the overuse of *the* in contexts where the zero article would have been preferred became the most frequent error by far (88%). However, in the case of demonstratives, errors in their form were the ones with the highest percentage of errors (77%), and the confusion between the pronoun and the adjective was the most important one in the use of possessives (55%).

The morphosyntactic errors in adjectives were mainly related to the use of adjectives in the plural form (56%), although the confusion in the comparative forms (17%) and their erroneous position in the sentence (15%) were also present in the learner corpus. The position of the adverbs was the main problem when using them, specially in the cases of *only*, *also* and *too* (50%), and frequency adverbs (32%). The use of the singular instead of the plural forms in nouns, or the opposite case, the use of the plural instead of the singular ones (69% and 21%, respectively), were also errors found in the learner corpus. Finally, the errors posed by wh-pronouns represented almost half of the errors in the category of pronouns (44%).

The analysis of lexical errors revealed that verbs were the word class which posed more problems to students (39.12% of errors), followed by nouns (21.5%), prepositions (20.92%), adjectives (10%), adverbs (5.16%) and conjunctions (3.3%). As far as spelling errors are concerned, the most frequent ones were found in double consonants (29%), problems with *y* and *i* (29%), and plurals in *-es* (23%).

Reference errors were the most frequent ones when textual cohesion was analysed. In fact, these errors were equally made when using demonstrative and personal reference (49% of errors each, that is, 98% of the errors in the reference subcategory), and comparative reference only entailed the other 2% of the errors in this subcategory. The problems with substitutors were posed by the use of *some*, *all* and *other(s)* (39% of the errors with substitutors), closely followed by the errors when trying to substitute sentences (32%). Ellipses in comparative clauses were the most difficult ones to do by students (46% of the errors in ellipsis), and the omission of conjunctions constituted their most representative problem, specially in the case of additive ones (42% of all the problems related to conjunctions). Finally, the students' limited knowledge of cohesion mechanisms was the most frequent cause behind their misuse in the learner corpus.

3. METHODOLOGY

The learner corpus employed to analyse the students' written production in the English exam in the UEE in June 2008 in the University of Jaén (Spain) is composed of a representative sample of the 1,406 compositions[6] which were written on the same topic 'Where, outside Spain, would you go on a short pleasure trip?'

In order to obtain a representative sample of these 1,406 compositions, simple random sampling was used (Cochran, 1977) (CI= 95%, *p*= *q* = .50) with the program *Stats 1.1*. The findings revealed that 302 compositions were needed to obtain a representative sample. After randomly selecting 302 compositions and faithfully transcribing them (only removing the title of the composition when it had been copied from the exam paper), 34,403 words were analysed.

As stated in the introduction, one of the main objectives in this study was to offer comparable results to those published by other researchers around the world and with students at different proficiency levels. Therefore, the learner corpus was error-tagged by means of the *UCL Error Editor* (Hutchinson, 1996), and the error taxonomy in the *Error Tagging Manual, version 1.1.* (Dagneaux, Denness, Granger & Meunier, 1996) developed by the *Centre for English Corpus Linguistics* at the *Université Catholique de Louvain*, Belgium[7]. This error taxonomy provides a general overview of the students' errors, as determined by those errors which have been proved to be frequent in students with a Romance language background (Granger, 1993: 60; Dagneaux, Denness & Granger, 1998: 168). As described in the *Error Tagging Manual 1.1.* (Dagneaux, Denness, Granger & Meunier, 1996), the forty-three tags considered are divided into seven error categories, namely Form (F), Grammar (G), Lexico-Grammar (X), Lexis (L), Word Redundant, Word Missing and Word Order (W), Register (R) and Style (S). Figure 1 shows that the level of description of each error category varies from one error-tag, for the error category Register (R), to fifteen tags for Grammar (G).[8]

Apart from these error categories, the *UCL Error Editor* (Hutchinson, 1996) also includes another error category, (P), which is further subdivided into three error-tags, namely (PX), (PM) and (PR). Although these are not described in the *Error Tagging Manual* (Dagneaux, Denness, Granger and Meunier, 1996), they were used in this paper to identify and error-tag those errors dealing with wrong punctuation, missing punctuation and redundant punctuation, respectively.
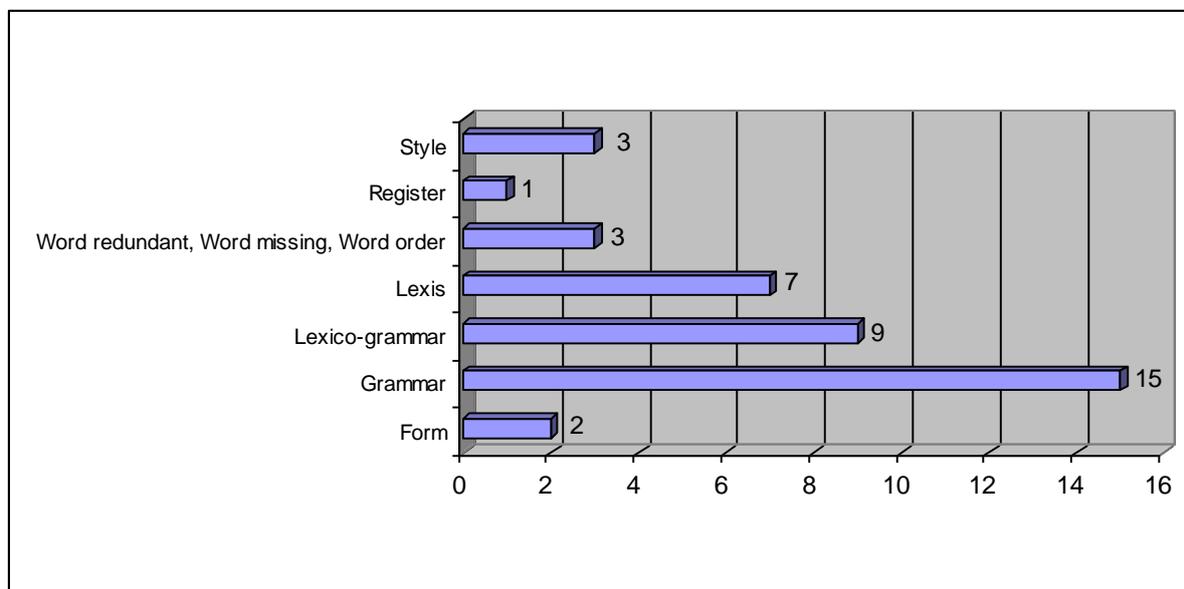

Figure 1. Breakdown of error categories and error-tags in the *Error Tagging Manual* v 1.1. (Dagneaux, Denness, Granger & Meunier, 1996).

The programme Regex Tokenizer 2.0 (Alcaraz Calero, 2009) was used to retrieve the types and frequencies of errors from the learner corpus and translate it into an SPSS document.

4. RESULTS AND DISCUSSION

4.1. Results in this study

All in all, 5,811 errors were identified in 34,403 words, that is, 16.89 errors per 100 words, or a mean of 19.24 errors (*SD*= 10.63) per composition.

To have a general overview, the first analysis made considers the eight error categories, i.e. Form, Grammar, Lexis, Punctuation, Register, Style, Word and Lexico-Grammar. As seen in Figure 2, the error category with the highest mean of errors per composition is Grammar (*M*= 7.22; *SD*= 4.86), then followed by Lexis (*M*= 4.22; *SD*= 3.28) and Form (*M*= 3.65; *SD*= 2.78). Contrarily, the two error categories which show the lowest means are Style (*M*= .28; *SD*= .62) and Register (*M*= .00; *SD*= .06).
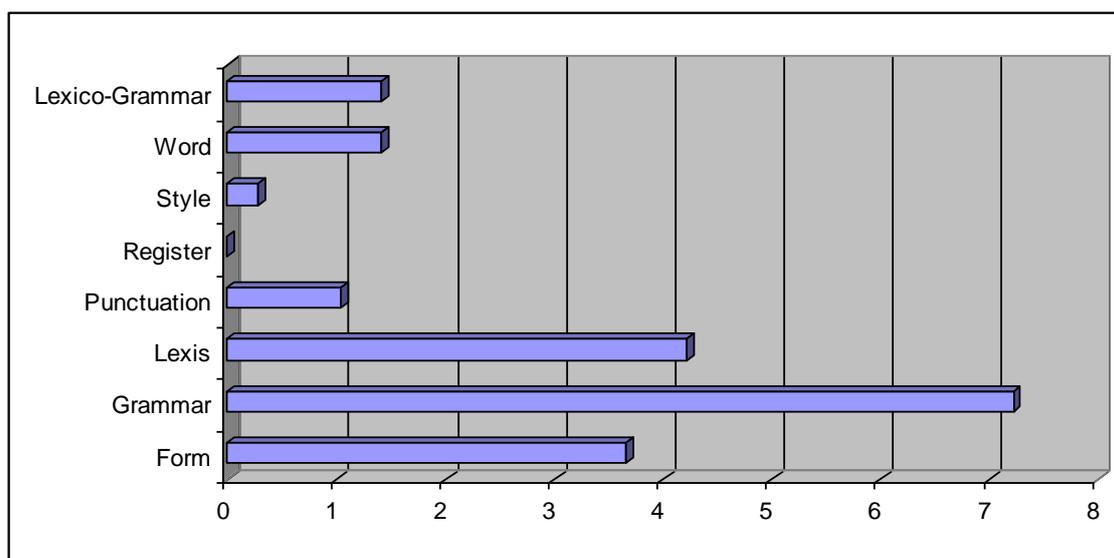

Figure 2. Means of types of errors per composition: overview of the 302 compositions.

A more comprehensive analysis of the students' main errors is shown in Figure 3, which summarizes the means of errors per composition obtained in the error-tags in the taxonomy.
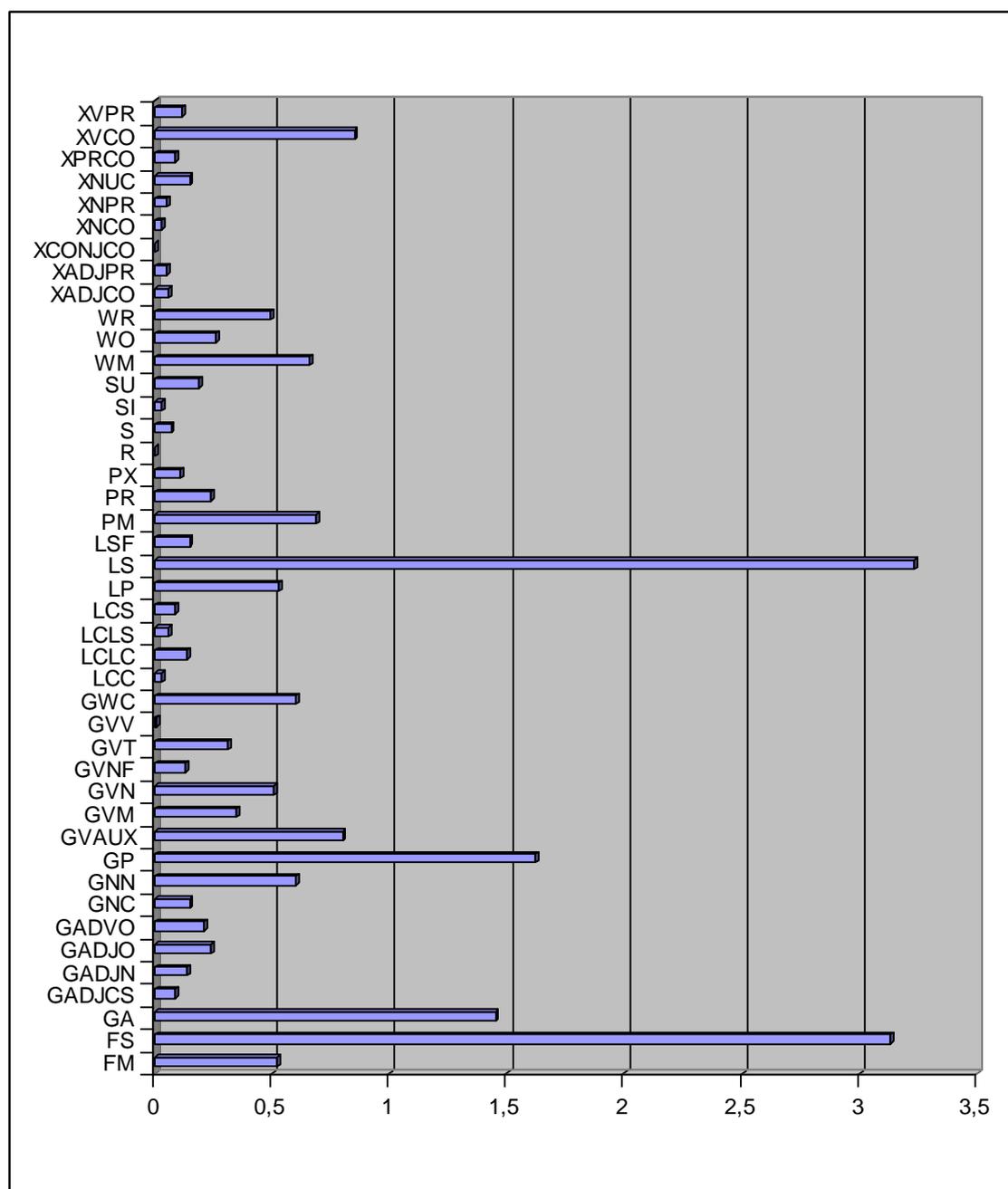
Figure 3. An overview of the means of the errors per composition made in the *Production* section.

As can be observed, the error-tags which present the highest means (*M*> 1, in all cases) are the ones related to the appropriate selection of vocabulary, (LS), (*M*= 3.23; *SD*= 2.75), and the one for spelling errors, (FS), (*M*= 3.13; *SD*= 2.53), then followed by the inappropriate use of pronouns, (GP), (*M*= 1.62; *SD*= 1.70), and the incorrect use of articles (GA) (*M*= 1.45; *SD*= 1.56). Thus, errors such as the ones in examples 1 to 4 were frequent in the learner corpus:

      (1)   'I will swim in the **(LS)** beach sea […]' (et_6988)

      (2)   '**(FS)** Althought $Although$ I have never been to […]' (et_6388)

(3)   ' […] tourists choose Spain for **(GP)** they \$their\$ holidays.' (et_6684)

(4)   ' […] and people see **(GA)** the \$0\$ life in another way […]' (et_6504)

On the opposite end, that is (*M*< .10), the error-tags with the lowest means of error per composition were those related to the problems found in the use of comparative and superlative grades in adjectives (GADJCS), the use of subordinating conjunctions (LCS), the complementation of prepositions (XPRCO), style (S), the use of conjuncts or disjuncts (LCLS), the complementation of adjectives (XADJCO) and (XADJPR), the complementation of nouns (XNPR) and (XNCO), the use of coordinating conjunctions (LCC) and, finally the complementation of conjunctions (XCONJCO).

Once the error-tags which show the highest and the lowest means have been highlighted, the analysis of the error-tags which present the highest means of errors per composition in each of the error categories will offer a more detailed scenario.

In the case of formal errors, spelling errors (FS) are the ones presenting the highest mean (*M*= 3.13, *SD*= 2.53), with examples such as

(5)   'Europe is a **(FS)** beatiful \$beautiful\$ place […]' (et_7721)

The grammatical errors which showed the highest means were the incorrect use of pronouns (GP) and articles (GA) (*M*= 1.62, *SD*= 1.70; *M*= 1.45, *SD*= 1.56, respectively), then followed by the incorrect use of auxiliary verbs (GVAUX) (*M*= .80, *SD*= 1.21), the problems in the selection of incorrect word classes (GWC), and the appropriate selection of number in nouns (GNN) (*M*= .60, *SD*= .85; and *M*= .60, *SD*= .92, respectively), as exemplified in 6 to 10 below.

(6)   '[…] because it is famous for **(GP)** her \$its\$ carnival […]' (et_6834)

(7)   '[…] sunbathing in **(GA)** the \$a\$ Caribbean beach […]' (et_6630)

(8)   'If I could go on a short pleasure trip, I **(GVAUX)** could \$would\$ go to Paris' (et_6881)

(9)   'The **(GWC)** beautiful \$beauty\$ of this city […]' (et_7447)10

(10)   '[…] there are a lot of interesting **(GNN)** museum \$museums\$ […]' (et_6739)

The appropriate selection of words, (LS), was the most outstanding lexical error (*M*= 3.23, *SD*= 2.75), with instances such as

(11)   ' […] it is a very **(LS)** famous \$popular\$ country.' (et_5913)

Within punctuation errors and errors related to the error category Word, the omission of punctuation marks, (PM), and the omission of words, (WM), were the errors which showed the highest means of errors per composition in each error category (*M*= .69, *SD*= 107; and *M*=

.66, *SD*= .95, respectively). Thus, errors such as the ones in examples 12 and 13 were frequent in the learner corpus

> (12)   'However **(PM)** $,$ last year I went to Holland […]' (et_6481)
> (13)   '[…] I would **(WM)** $like$ to return to this special city.' (et_6610)

Examples like the one in 14 below, in which there is a communication breakdown, (SU), proved the type of error which showed the highest mean of errors per composition as far as the error category style is concerned (*M*= .19, *SD*= 51).

> (14)   '(GA) The 0 (FM) persons $people$ would like to go (WM) on a short trip **(SU)** an the price is cheap but they are two-day $and the price is cheap but they only last two days?/ at a cheaper price than they are today?$.' (et_6946)

Finally, the error-tag which included the highest mean of lexico-grammatical errors per composition was the one related to the verbal complementation of verbs (XVCO) (*M*= .85, *SD*= 1.27). See, for instance,

> (15)   'I am **(XVCO)** thinking to travel $thinking about travelling$ to […]' (et_7458)

To finish this section, the type of analysis conducted to obtain these results is worth mentioning so that the findings are interpreted correctly. As stated in section 3, a CEA was employed, i.e. only the errors which students made were considered. In other words, no reference or comparison is made to the correct uses of the linguistic aspects in the FL, that is, an IA (Selinker, 1972; 1992) was not conducted. Consequently, the low mean (*M*< .10) of some errors in the learner corpus may be due to three reasons. The first one is avoidance (Schachter, 1974). As previously reported (Schachter, 1974; Kleinmann, 1977, 1978; Dagut & Laufer 1985; Hulstijn & Marchena, 1989), students may decide not to use those aspects of the FL which they do not master, or are not sure of. Consequently, they avoid using them not to make mistakes, specially in a high-stakes examination such as the English exam in the UEE. The findings obtained in the students' mean of errors per composition when using lexico-grammatical patterns may illustrate this idea. All in all, the means of errors per composition when using those patterns (*M*< 1) seem to indicate that students avoid using the ones which require the use of prepositions or non-finite verb forms. This is specially so when the patterns are employed to complement nouns or adjectives (*M*< .09), as compared to those used with verbs (*M*= .12 and *M*= .85).

The second aspect to be considered is that students may reduce their variety of resources when writing. They may use "lexical teddy bears" (Hasselgren, 1994: 237), i.e. the items of vocabulary with which students feel comfortable and overuse when writing in the

FL, to the detriment of a wider array of paradigmatic relationships. This may also be the case of the students' use of a limited number of lexico-grammatical patterns, and coordinating and subordinating conjunctions.

The type of language elicited by the topic of the essay is the third important factor to take into account. The topic on which the students wrote, "Where, outside Spain, would you like to go on a short pleasure trip?" (see section 3), may not elicit the use of the passive voice, or the use of different varieties of the language. As a result, students may not use those linguistic resources, not having opportunities to make errors when using them.

Apart from these three causes, the results in this section describe the students' command of the FL at this level, as reflected in the means of errors below 1 of those aspects of the FL which cannot be avoided, which cannot be varied or which are equally elicited by any topic. That is, linguistic aspects related to communicative breakdowns (SI), (SU), the correct expression of the grammatical category number in nouns or verbs (GNN) and (GVN), as well as the incorrect expression of the grammatical category number in adjectives (GADJN).

4.2. Comparison of the results obtained with those by Crespo García (1999) and Rodríguez Aguado (2004)

The comparison of the results obtained in this paper and the findings revealed in Crespo García (1999) and Rodríguez Aguado (2004) point to interesting tendencies regarding the common errors made by students when writing their English exam in the UEE. However, a note of caution is necessary here, since the data obtained are not fully comparable due to a number of reasons. First, the error taxonomies employed are not the same, so the errors which are considered within an error-tag or error category in a study may be included in other error-tags or another error category. Second, percentages are the only quantitative data provided in those studies. Third, some of the previous analyses on the English exam do not only refer to the *Production* section of the exam, and may also include closed answer questions. Fourth, the groups of students who took the exam do not only differ in their location in Spain (Galicia, and Valladolid), but also in the year when they took the exam, which may indicate a difference in the methodology in the English teaching classes in their Secondary Education. Finally, the topic(s) on which students wrote were different. For this reason, the comparisons drawn should be considered tentative, and only informative of the tendencies observed.

Despite these differences in the methodology, sample and population used to obtain results on the students' written production in the English exam, the coincidence of some findings in those studies and in this paper may point to the existence of some tendencies in the students' written production.

The first one may indicate that the error category which includes most of the students' errors is the one related to grammatical aspects, then followed by the lexical error category

and the formal one, as indicated in the data by Rodríguez Aguado (2004) and this study. If word classes are considered, the use of the open word class verb also points to the students' tendency to make more errors when using main verbs, rather than modal auxiliary verbs, as indicated in Crespo García (1999) and the results in this paper. Within the errors made in main verbs, another tendency may highlight that the most problematic aspect is related to the correct use of the third person singular, as observed in the data by Crespo García (1999), Rodríguez Aguado (2004) and the data in this paper. Finally, tendencies in the use of adjectives and nouns may also be revealed. As far as the use of adjectives is concerned, the incorrect addition of the plural morpheme is reported to be frequent in Crespo García (1999), Rodríguez Aguado (2004) and this study. When considering the word class noun, the correct expression of number proves to be the linguistic aspect which triggers most problems to students, as indicated by Crespo García (1999), Rodríguez Aguado (2004) and this study.

The comparison of the results in this paper and the studies by other researchers who focused on specific aspects of the FL, rather than in morpho-syntactic or morphological categories (e.g. the studies by Doval Suárez, 1999; González Álvarez, 1999; Iglesias Rábade 1999b; Woodward Smith, 1999; and Wood Wood, 2002) cannot be drawn without a detailed qualitative analysis of the instances of errors in some error-tags in the learner corpus in this study. For instance, to analyse personal reference, the error-tag for the use of pronouns (GP) would need to be analysed, as is the case of the error-tags for the use of pronouns (GP), lexical selection (LS), use of articles (GA) and the grade in adjectives (GADJCS) for demonstrative reference. Similarly, the analysis of article use would entail a further qualitative analysis of the (mis)use of the articles in the learner corpus, as has already been done with other learner corpora (Díez-Bedmar & Papp, 2008; Díez-Bedmar, 2010).

5. CONCLUSION

Two objectives were pursued in this paper. First, to provide an updated overview of the students' main problems when using the FL to write a composition in the English exam in the UEE. Second, to offer results which would be comparable to the ones obtained by other researchers around the world who also use the widely-known error taxonomy used in this paper.

Regarding the first objective, the findings show that grammatical aspects are the ones which pose more difficulties to Spanish students of English at this stage of their language learning process, then followed by the lexical and the formal ones. If specific types of errors, rather than error categories, are considered, the aspect of the FL which triggers the highest mean of errors per composition is the appropriate selection of vocabulary, then followed by the problems in spelling, the correct use of pronouns and the use of the article system. Although the findings obtained in this paper and those which were offered by previous

research are not fully comparable due to the differences in methodology, task type, etc., tendencies may be shown regarding the prevalence of grammatical errors and the problems that students have when using some aspects of the FL in some word classes. Nevertheless, these tendencies are only informative and a wide range of studies which use the same methodology would be needed to ensure comparability of results and allow the replication of studies.

The second objective was met by using the most widely-known error taxonomy for Romance languages nowadays, which was fostered by the need to obtain a comparable description of the students' written production in the FL at this stage. The European Higher Education Area and the importance of the students' use of the (written) FL during their degree, and as a requirement to finish it, demand that FL teachers design courses which may help the students improve their (written) competence in the FL. Using this error taxonomy to describe the students' most important errors when writing in the FL motivates the use of comparable methodologies and presentation of data, so that researchers may establish comparisons between this representative sample and any other in Spain, or in any other country by means of a Contrastive Interlanguage Analysis (CIA) (Granger, 1996). Thus, it will be possible to cater for the students' specific needs and help them move forward in their FL acquisition process.

Finally, the main limitation of this study is that the results obtained can only be generalized to the students who took the English exam in the UEE in June 2008 in Jaén, and wrote on the topic 'Where, outside Spain, would you go on a short pleasure trip'. Therefore, it would be desirable to replicate this study with students writing a short piece of writing in the English exam in the UEE in different universities, with different text types and, for comparative purposes, in other countries with a similar exam. An interesting line of research is the consideration of the Common European Framework of Reference for Languages (Council of Europe, 2001) (CEFR) levels which are awarded to the compositions in the exam (Díez-Bedmar, 2012), so that it may be possible to characterise the type of errors which students at different CEFR levels show. Finally, teaching materials could be designed with the results in this study. Since the students' main problems when writing in the FL at this level are now known, the design of appropriate 'tailor-made' materials could help them improve their command of the written language in the FL and help them face the European Higher Education Area better.

their help in the compilation of the data. Last, but not least, I would like to thank Dr. Antonio Bueno González and Dr. Pascual Pérez Paredes for their help and support.

NOTES

1. In this respect, the *Centro de Investigación y Documentación Educativa* (*CIDE*) has funded research on various aspects related to the UEE, which has resulted in a wide number of publications.
2. Brackets are used with the word 'computer' because on some occasions the learner corpora analysed are not in electronic version, but in paper format.
3. The analysis of the students' command of the L1 was also investigated in the frame of the Corpus'92 (see, for instance, Battaner, 1995; Pujol, 1995-1996, 1997; and Atienza-Cerezo & López Ferrero, 1997).
4. García Laborda and Bakieva (2010) have recently published another paper on the students' written performance regarding spelling and morphosyntactic errors when using a computer-based pilot version of the exam. However, this publication has not been included in the literature review because the spelling or morphosyntactic errors considered in the analysis are not specified, and the data on the learner corpus are not provided either.
5. This finding is expected, since EFL students at this stage are likely to use more main verbs than modal auxiliary verbs. However, an IA should be conducted to analyse the number of correct and incorrect uses of main verbs and modal auxiliary verbs.
6. The nature of the UEE, and the students' anonymity while the exams are sat and marked, limited the amount of information on the students' variables. Consequently, the data concerning the students' mother tongue, the use of other languages at home, age, gender, type of high-school they had attended (i.e. monolingual, bilingual, etc.), or whether they had had any additional instruction in the English language (e.g. Official Language School), or taken any University of Cambridge ESOL examination were not available.
7. Two native speakers of English helped in the identification of errors in the learner corpus, so that the identification of errors was as accurate as possible.
8. Appendix 1 offers an overview of the error-tags in this error taxonomy.

An overview of the error-tags in the *Error Tagging Manual*, version 1.1. (Dagneaux, Denness, Granger & Meunier, 1996)

Form	FM	form, morphology
	FS	form, spelling
Grammar	GA	grammar, articles
	GADJCS	grammar, adjectives, comparative / superlative
	GADJN	grammar, adjectives, number
	GADJO	grammar, adjectives, order
	GADVO	grammar, adverbs, order
	GNC	grammar, nouns, case
	GNN	grammar, nouns, number
	GP	grammar, pronouns
	GVAUX	grammar, verbs, auxiliaries
	GVM	grammar, verbs, morphology
	GVN	grammar, verbs, number
	GVNF	grammar, verbs, non-Finite / Finite
	GVT	grammar, verbs, tense
	GVV	grammar, verbs, voice
	GWC	grammar, word class
Lexis	LCC	lexis, conjunctions, coordinating
	LCLC	lexis, connectors, logical, complex

	LCLS	lexis, connectors, logical, single
	LCS	lexis, conjunctions, subordinating
	LP	lexical phrase
	LS	lexical single
	LSF	lexical single, false friends
Register	R	register
Style	S	style
	SI	style, incomplete
	SU	style, unclear
Word	WM	word missing
	WO	word order
	WR	word redundant
Lexico-grammar	XADJCO	lexico-grammar, adjectives, complementation
	XADJPR	lexico-grammar, adjectives, dependent preposition
	XCONJCO	lexico-grammar, conjunctions, complementation
	XNCO	lexico-grammar, nouns, complementation
	XNPR	lexico-grammar, nouns, dependent preposition
	XNUC	lexico-grammar, nouns, uncountable / countable
	XPRCO	lexico-grammar, prepositions, complementation
	XVCO	lexico-grammar, verbs, complementation
	XVPR	lexico-grammar, verbs, dependent preposition
Punctuation	PM	punctuation missing
	PR	punctuation redundant
	PX	wrong punctuation

REFERENCES

Atienza Cerezo, E., & López Ferrero, C. (1997). Caracterización sintáctica del discurso escrito por aspirantes a estudios universitarios. In J.L. Otal, I. Fortanet & Victòria Codina (Eds.), *Estudios de Lingüística Aplicada* (pp. 581-585). Castelló de la Plana: Publicacions de la Universitat Jaume I.

Bachman, L.F., & Palmer, A.S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Ballesteros, F., Rica, J.P., Neff, J. & 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 4th International Contrastive Linguistics Conference, September 2005, Santiago de Compostela* (pp. 89-97). Servizo de Publicacións e Intercambio Cientifico.

Battaner, M. P. (1995). Escritura, redacción y composición en las P.A.U. In J.M. Ruiz, P. Sheerin & E. González-Cascos (Eds.), *XI Congreso Nacional de Lingüística Aplicada* (pp. 125-130). Valladolid: Servicio de Publicaciones de la Universidad de Valladolid.

Cochran, W.G. (1977). *Técnicas de muestreo [Sampling techniques]*. México. Trillas.

Council of Europe (2001). *Common Europeak Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Crespo García, B. (1999). Errores morfosintácticos. In L. Iglesias Rábade (Ed.), *Análisis de los errores del examen de inglés en las pruebas de acceso a la universidad en el*

distrito universitario de Galicia (pp. 167-206). Santiago de Compostela: Servicio de Publicacións e intercambio Cientifífico, Universidade de Santiago de Compostela.

Dagneaux, E., Denness, S., Granger, S., & Meunier, F. (1996). *Error tagging manual version 1.1*. Louvain-la-Neuve: Centre for English Corpus Linguistics, Université Catholique de Louvain.

Dagneaux E., Denness S., & Granger, S. (1998). Computer-aided error analysis. *System, 26,* 163-174.

Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs: A case for contrastive analysis. *Studies in Second Language Acquisition, 7,* 73-79.

Díez-Bedmar, M.B. (2012). The use of the Common European Framework of Reference for Languages to Evaluate Compositions in the English Exam Section of the University Admission Examination. *Revista de Educación, 357,* 55-79.

Díez-Bedmar, M.B. (2011). The English exam in the University Entrance Examination: An overview of studies. *Revista Canaria de Estudios Ingleses, 63,* 101-112.

Díez-Bedmar, M.B. (2010). From secondary school to university: The use of the English article system by Spanish learners. In B. Bellés-Fortuño, M.C. Campoy-Cubillo & M.L. Gea-Valor (Eds.), *Exploring corpus-based research in English language teaching* (pp. 45-55). Castelló de la Plana: Publicacions de la Universitat Jaume I.

Díez-Bedmar, M.B., & Papp, Sz. (2008). The use of the article system by Chinese and Spanish learners. In G. Gilquin, Sz. Papp & M. B. Díez-Bedmar (Eds.), *Linking up contrastive and learner corpus research* (pp. 147-175). Amsterdam and New York: Rodopi.

Doval Suárez, S.M. (1999). Análisis de los errores ortográficos. In L. Iglesias Rábade (Ed.), *Análisis de los errores del examen de inglés en las pruebas de acceso a la universidad en el distrito universitario de Galicia* (pp. 115-165). Santiago de Compostela: Servicio de Publicacións e intercambio Cientifico, Universidade de Santiago de Compostela.

Dulay, H.C., Burt, M.K., & Krashen, S. (1982). *Language two*. Oxford: Oxford University Press.

García Laborda, J. (2006). Analizando críticamente la Selectividad de inglés. ¿Todos los estudiantes españoles tienen las mismas posibilidades? *TESOL Spain, 30,* 9-12.

García Laborda, J., & Bakieva, M. (2010). Análisis primario de la expresión escrita de estudiantes de 2º de Bachillerato en una prueba de acceso a la universidad informatizada. In J.L. Bueno Alonso, D. González Álvarez, U. Kirsten Torrado, A.E. Martínez Insua, J. Pérez Guerra, E. Rama Martínez & R. Rodríguez Vázquez (Eds.), *2010: Analizar datos > Describir variación/ Analysing Data > Describing variation* (pp. 756-764). Vigo: Universidade de Vigo, Servizo de Publicacións.

Gilquin, G., Papp, Sz., & Díez-Bedmar, M.B. (Eds.). (2008). *Linking up contrastive and learner corpus research*. Amsterdam and New York: Rodopi.

González Álvarez, E. (1999). Errores léxicos. In L. Iglesias Rábade (Ed.), *Análisis de los errores del examen de inglés en las pruebas de acceso a la universidad en el distrito universitario de Galicia* (pp. 207-270). Santiago de Compostela: Servicio de Publicacións e intercambio Cientifífico, Universidade de Santiago de Compostela.

Granger, S. (1993). The International Corpus of Learner English. In J. Aarts, P. de Haan & N. Oostdijk (Eds.), *English Language Corpora: Design, analysis and exploitation (*pp. 57-69)*. Amsterdam: Rodopi.

Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg & M. Johansson (Eds.), *Languages in Contrast. Text-based cross-linguistic studies* (Lund Studies in English 88, pp. 37-51). Lund: Lund University Press.

Granger, S. (Ed.). (1998). *Learner English on computer*. London & New York: Addison Wesley Longman.

Granger, S., Hung J., & Petch-Tyson, S. (Eds.). (2002) *Computer Learner Corpora, Second Language acquisition and foreign language teaching*. Atlanta and Amsterdam: Benjamins.

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics, 4,* 237-260.

Hulstijn, J.H., & Marchena, E. (1989). Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquisition, 11,* 241-255.

Hutchinson, J. (1996). *UCL Error Editor*. Louvain-la-Neuve: Centre for English Corpus Linguistics, Université Catholique de Louvain.

Iglesias Rábade, L. (Ed.). (1999a). *Análisis de los errores del examen de inglés en las pruebas de acceso a la universidad en el distrito universitario de Galicia*. Santiago de Compostela: Servicio de Publicacións e intercambio Cientifífico, Universidade de Santiago de Compostela.

Iglesias Rábade, L. (1999b). Análisis de los errores de organización textual. In L. Iglesias Rábade (Ed.), *Análisis de los errores del examen de inglés en las pruebas de acceso a la universidad en el distrito universitario de Galicia* (pp. 357-397). Santiago de Compostela: Servicio de Publicacións e intercambio Cientifífico, Universidade de Santiago de Compostela.

Kleinmann, H.H. (1977). Avoidance behavior in adult second language acquisition. *Language Learning, 27,* 93-107.

Kleinmann, H.H. (1978). The strategy of avoidance in adult second language acquisition. In W.C. Ritchie (Ed.), *Second language acquisition research: Issues and implications* (pp. 157-174). New York: Academic Press.

Polio, C., & Gass, S. (1997). Replication and reporting. A commentary. *Studies in Second Language Acquisition, 19,* 499-508.

Pujol, M. (1995-1996). El uso incorrecto de la coma como señal de textos defectuosos. *Tabanque Revista Pedagógica*, *10-11,* 131-160.

Pujol, M. (1997). Algunas cuestiones sobre la evaluación de la puntuación. In A. Mendoza Fillola, F.J. Cantero Serena & M.C. Romea Castro (Eds.), *Didáctica de la Lengua y la Literatura para una sociedad plurilingüe del Siglo XXI: Actas del IV Congreso Internacional de la Sociedad Española de Didáctica de la Lengua y la Literatura* (pp. 1209-1212). Barcelona: Universitat de Barcelona.

Rodríguez Aguado, J.I. (2004). *Análisis de errores en el ejercicio de redacción en las pruebas de acceso a la universidad. Asignatura, lengua extranjera, inglés*. Unpublished doctoral dissertation. Universidad de Valladolid.

Schachter, J. (1974). An error in error analysis. *Language Learning, 24,* 205-214.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics 10,* 209-231.

Selinker, L. (1992). *Rediscovering Interlanguage*. New York: Longman.

Woodward Smith, E. (1999). Análisis de los errores en las palabras funcionales (preposiciones, conjunciones y partículas). In L. Iglesias Rábade (Ed.), *Análisis de los errores del examen de inglés en las pruebas de acceso a la universidad en el distrito universitario de Galicia* (pp. 271-356). Santiago de Compostela: Servicio de Publicacións e intercambio Científico, Universidade de Santiago de Compostela.

Wood Wood, M. (2002). La transferencia de la L1 a la L2 (Español-inglés): El artículo en la lengua escrita de los alumnos en las pruebas de acceso a la universidad. Unpublished doctoral dissertation, Universidad de Las Palmas de Gran Canaria.