

# EXPLORING THE ERRORS MADE BY EFL STUDENTS AT THE UNIVERSITY LEVEL<sup>1</sup>

M.<sup>a</sup> Esther Mediero Durán, Ainhoa Robles Baena

Universidad Autónoma de Madrid

This paper reports on some preliminary results of the TREACLE project, which analyzes a learner corpus of texts written in English by Spanish university students in order to explore the structures that students use at different levels of linguistic competence and the errors they make in producing them. We will focus, in particular, on the error annotation of the corpus. First we will explain the methodology used, briefly present the taxonomy of errors which was designed by the research team for annotating the corpus. We will also mention the problems found by the coders in the project during the corpus annotation and the ways in which these problems have been minimized. Finally, we will present the results of the broadest error categories and their distribution across levels, providing detail about the most relevant results within the category of grammar errors found in the corpus.

Key words: Learner corpora, error analysis, EFL.

## 1. Introduction

*I am indeed amazed when I consider how weak my mind is and how prone to error.*

Rene Descartes

*Error is discipline through which we advance.*

William Ellery Channing

In the light of these two initial quotes, we would like to predispose ourselves and the reader to adopt a positive attitude towards errors, which represent motivation to progress and new opportunities to learn.

Error is present in all aspects of our lives, especially when it comes to acquiring a new language. Various approaches to the treatment of errors have been propounded over the years with the aim of providing correction and achieving accuracy. In his pioneering study, Corder (1982: 12) compared the learning process of one's mother tongue to the acquisition of foreign languages. Corder considered the learner's native language facilitative and that errors were not to be considered as a sign of inhibition but as evidence of the student's learning strategies. In the same way infants formulate utterances that do not follow adults' speech rules, learners of English will attempt to test if their mother tongue rules are transferable to the new language. This approach accounts for the large number of errors which are related to the interference of their mother tongue. Selinker (1972) called this class of idiosyncratic dialects *interlanguage* and defined it as a dialect whose rules share characteristics of two social dialects of languages, whether these languages themselves share rules or not. These rules will refer to the areas of grammar that are problematic for our students. We share Larsen-Freeman's view on the potentiality of grammar when she states that "... grammar affords speakers of a particular language a great deal of flexibility in the ways they can express propositional, or notional, meaning and how they present themselves in the world" (Larsen-Freeman 2002: 104).

---

<sup>1</sup> The TREACLE project is funded by the Spanish Ministerio de Economía y Competitividad (MINECO) under project grant FFI2009-14436/FILO. *TREACLE stands for Teaching Resource Extraction from an Annotated Corpus of Learner English.*

The purpose of this paper is to explore errors as a natural phenomenon in the learning process of English as a foreign language at the university level. In doing so, we want to offer an opportunity for both students to reassess their own work and for teachers to re-schedule the syllabus or prioritise certain areas if necessary to meet students' needs and improve efficiency based on empirical data obtained from TREACLE.

Our study forms part of the TREACLE project. TREACLE stands for Teaching Resource Extraction from an Annotated Corpus of Learner English, and it seeks to create a methodology for producing grammatical profiles of Spanish university learners of English. The project links the relevant areas with the CEFR (Common European Framework of Reference) levels to improve the curriculum. The CEFR uses some abbreviations for the traditional levels: beginners and pre-intermediate students correspond to A1 and A2, the intermediate levels is designated by B1 and B2, the advanced level is identified as C1 and C2. We will refer to these levels when reporting the results.

The empirical evidence around which our study revolves is university students' written production. This thorough analysis allows us to explore the areas that are more difficult for our students. In highlighting errors in their compositions, we learn which parts of the language need to be reinforced at every level. Thanks to this type of work we can observe the general trend and propose alternatives to improve the development of the linguistic competence in the syllabus design. We must point out that our notion of error follows Corder's classification of errors. In our work we will focus exclusively on the errors of competence as opposed to the errors of performance such as slips of the tongue which are unsystematic.

As English teachers we are aware that the acquisition of foreign languages is demanding. In addition to the challenge of learning foreign languages, the task of writing poses some difficulty for students, as Widdowson (1983: 34) acknowledges: “[...] getting the better of words in writing is commonly a very hard struggle. And I am thinking of words which are in one's native language. The struggle is all the greater when they are not”.

Although this project is mainly concerned with grammar-related errors, the broad error taxonomy will give the reader a better understanding that the notion of grammar that we refer to goes beyond formal accuracy. As we have mentioned before, we believe grammar is a powerful tool that enables students to use the language system not only to convey meaning, but also to match intentions to particular contexts, emphasizing the actual impact grammar has on communication as a social phenomenon with multiple aspects, linguistic and sociolinguistic (politeness principles and appropriate register to the communicative context).

In this paper, we will discuss some preliminary conclusions that we reached from a manually annotated corpus. This project aims at analysing students' written production to explore the more problematic areas for students at different levels of linguistic competence. For this purpose we will explain some specific features of the project: the corpus, the methodology and error taxonomy. Then we will present some results of the texts we have coded, the conclusions we have reached so far, TREACLE contributions to the field of learner corpus and some future areas of research.

## **2. The corpus**

The project was born with the compilation of a corpus that includes university students' compositions both from Universidad Autónoma de Madrid and Universidad Politécnica de Valencia. Most writings deal with current topics such as immigration.

The corpus that we use is the result of combining the MiLC Corpus from the Universidad Politécnica de Valencia (UPV) and the WriCLE Corpus from Universidad Autónoma de Madrid (UAM). MiLC has 780 texts totalling up to 150,000 words whereas WriCLE includes fewer texts but the total amount of words is higher due to the genre the

students were covering. Most compositions in MiLC discuss topics related to immigration and those in WriCLE comprise a wider variety of current affairs such as immigration, homosexuality, marriage, and traffic problems. Research has proved that writing in a foreign language combines L1 and L2 resources, and that the mother tongue resources can be particularly useful at the stage of pre-planning the contents (Manchón 1999: 439-78). In both cases the level of students contributing to MiLC and to WriCLE was assessed with the Oxford Quick Study Placement Test (UCLES 2001) and they range from A1 to C1 at MiLC and from A2 to C2 in WriCLE as per CEFR.

Texts coming from MiLC are generally shorter than those from WriCLE and most of them are opinion texts. The latter are longer and sometimes adopt the form of essays, and therefore include opinions.

The most important characteristics of the corpus can be summarized as follows:

	<b>Amount of Texts</b>	<b>Length of texts</b>	<b>Genre / Topic</b>	<b>Students</b>	<b>Level</b>
<b>MiLC Corpus – UPV (Andreu et al 2010)</b>	950 compositions ◊ 180,000 words	220-250 words per text	Opinion texts mainly devoted to Immigration.	Spanish University Students of all Levels from UPV	A1- C1
<b>WriCLE Corpus – UAM (Rollinson &amp; Mendikoetxea 2010)</b>	750 essays ◊ 500,000 words (Selection of 521 texts)	1,000 avg. words per text	Essays / Variety of topics such as immigration, homosexual marriages, traffic problems.	Students of English Philology	A2 - C2

Figure 1.

### 3. Methodology

A relevant and essential part of our work could not have been achieved without the help of ICT. This section will first give some information on the tool, then about the error coding process, the steps we follow manually, the coding criteria and finally about the reliability, that is, how we standardised the coding criteria across the corpus.

Regarding the tool, the UAM Corpus Tool was designed and developed by Michael O'Donnell in 2008. It is free software and can be downloaded from <http://www.wagsoft.com/CorpusTool/>. It allows you to manually code the errors of each text following the error taxonomy that was created for this project. The error taxonomy has 113 errors at the narrowest level in the hierarchy and the software allows the scheme to be changed to cope with novel cases. The tool also includes glosses (coding criteria) associated with each feature, helping coders choose the correct category and greatly increasing inter-coder agreement.

#### 3.1. Error Taxonomy

One of the main differences between the TREACLE corpus and other existing learner corpora is the error taxonomy. If we compare TREACLE with other relevant corpora (Pravec 2002) such as *The International Corpus of Learner English* (ICLE) directed by Grangers, we observe a slight difference in the approach. The ICLE scheme deals with grammar errors, but it does so in terms of word classes, with no notion of phrase or clause. Our taxonomy, in

trying to relate errors to the university grammar curriculum, is focused on placing errors in relation to the grammar category in which they are taught, as we teach clause and phrase more than word classes. Despite the greater resemblance of TREACLE with the error taxonomy of NICT JLE Corpus (Izumi et alii: 2005) that focuses on *intelligibility* and *naturalness*, TREACLE goes much further in the total amount of errors.

Due to space constraints we cannot offer a full description of each error category. However, the major categories are presented below.

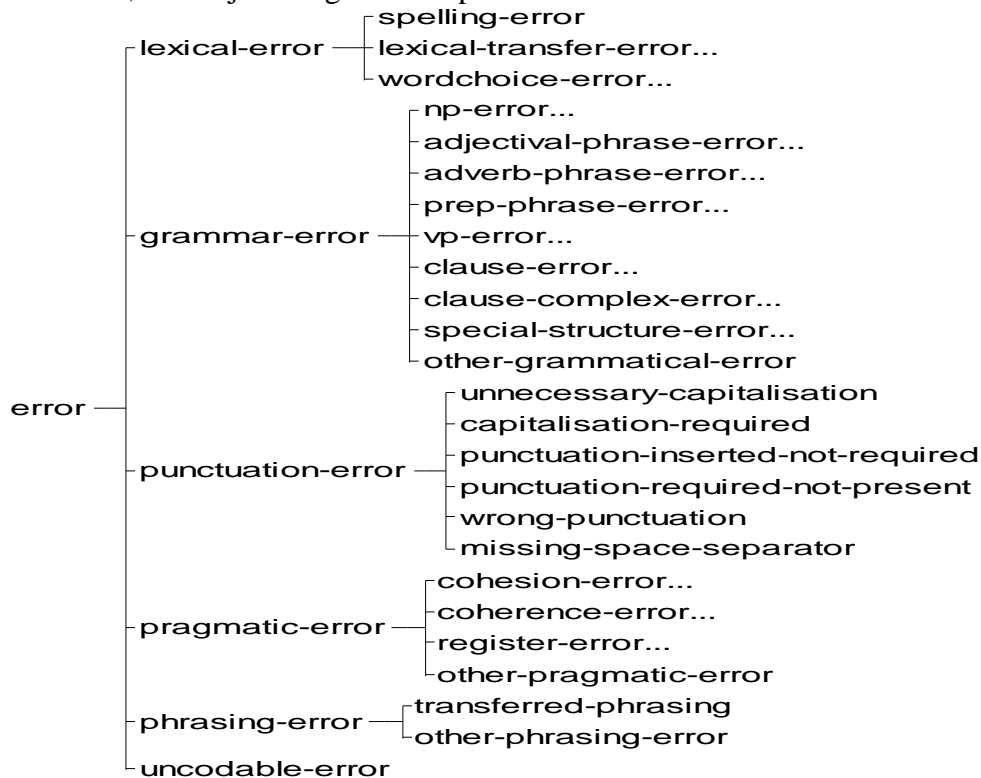


Figure 2.

\*The dots mean that there are further subcategories.

To better cater for the specific requirements of the project, we redefined the error scheme on the basis of two inter-coder reliability studies (ICRS). In the latest version, the tool also allows a visualisation of the text with corrections and description of the error to be saved which can be sent to the student. Additionally, the error coding manual has been extended with more description of error categories, and examples drawn from real texts to help coders select the appropriate error type.

### 3.1.1 Major Error Categories

The error taxonomy includes five main error types: lexical, grammatical, punctuation, pragmatic and phrasing. We also have an extra category for uncodable errors which is used when the coder cannot sufficiently understand the student's text to decide what the error is. At the current stage of development, the error scheme accounts for almost all errors students make.

One of the main design principles has been to ensure that the error scheme should map cleanly onto the organization of grammar topics which are taught within EFL courses. This design has a pedagogical orientation as the coding of errors moves from the more generic grammatical classes to more precise ones, thus it allows us to see which units of grammar present in normal grammatical courses are challenging for students. As for grammar references, we generally follow Quirk and Greenbaum (2008).

### 3.2. Error Coding Process

The error coding process is rather straightforward and usually follows three steps:

1. We first select the text that is incorrect. To provide better agreement between coders, coders are advised to select only the text that needs to be corrected, rather than whole grammatical units.
2. We assign error codes to the segment, starting with the broadest category appropriate for the error and then successively choosing more specific categories in the taxonomy. Given the large number of error codes in the taxonomy, this process allows the coder to quickly zero in on the appropriate category without searching through hundreds of categories.
3. We propose a correction.

To illustrate the steps, we include a shot of the screen coders work on.

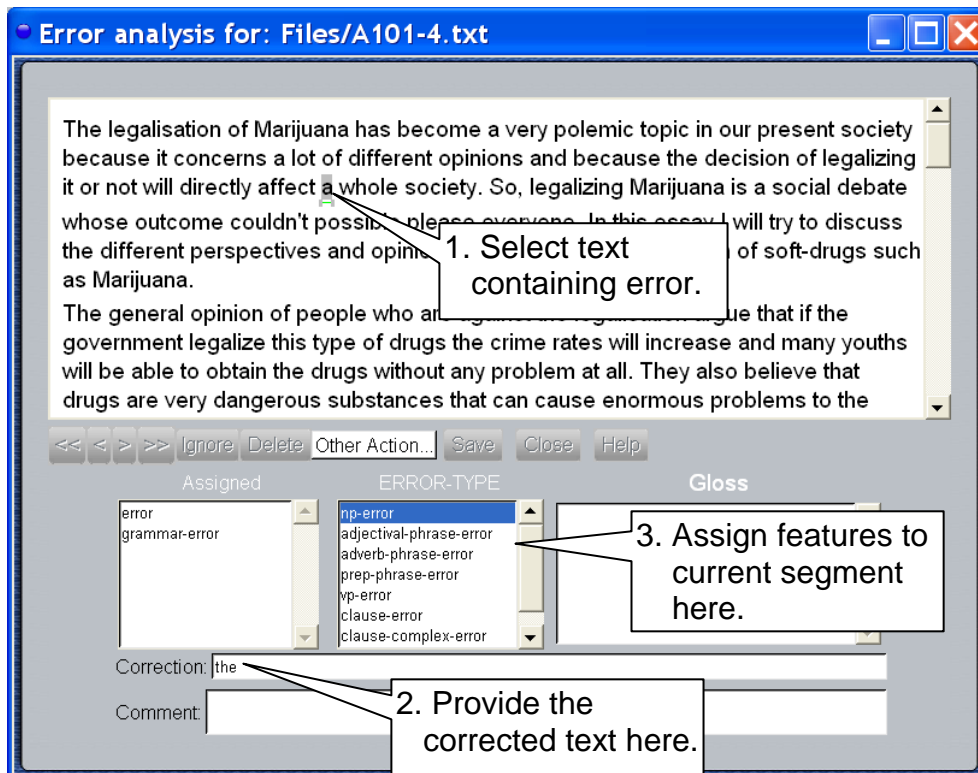


Figure 3.

At the top, we can see the text written by the student. Immediately below, we have three successive boxes. The first two correspond to the error taxonomy, the last one *glosses* shows coding criterion for the currently selected error category, sometimes with examples. This additional application has proved to be very useful for coders to familiarize themselves with the error taxonomy, especially at early stages. At the bottom there is space to write the correction and any comments that the researcher may find relevant for the team to discuss.

These notes by the team of coders were extremely relevant while performing the inter-coder reliability studies. The tool allowed the TREACLE members to compare the cases where all coders agreed and those where a consensus needed to be reached.

#### 4. Problems and Solutions, Reliability

The more delicate aspect of the manual tagging is the mastery of the error categories and the standardisation of error classification across the corpus. For this purpose, two inter-coder reliability studies (ICRS) were carried out.

As an initial step, the leaders of the research team created a Coding Criteria Manual, specifying the criteria for the applicability of error categories. The manual seeks to maximize the degree to which all members follow the same criteria. Although most rules were established in the first version of the manual, after working on some texts we held meetings to compare the cases that were classified differently by some coders and discussed problematic areas. As a result, some new error categories were created or redistributed in the hierarchy and a new version of the manual was created. Moreover, we calculated *Consensus Rates* whereby reliability scores were obtained by comparing each individual's coding to the consensus model.

One major problem we encountered related to the extent of text selected in identifying an error. Some coders selected the entire phrase (e.g., a noun phrase), while others selected only the word or words within the phrase which were erroneous. To resolve this error, we proposed as the basic segment identification criteria that only the words that need to be corrected should be selected.

The final principle is straightforward but points at the fuzzy barrier between coding errors and corrected versions that are quite often used interchangeably. We must code what the learner has written and not what should have been written. Following this principle, the error “the car of John” would not be classified as a determiner problem but as a postmodifier error.

The procedure just described is pedagogically oriented as we focus on the phrase where the error occurs and anticipate which areas of the syllabus need to be reinforced. Let us suppose that a student wrote “very browner”. Although the error might be related to the inclusion of an adverbial premodifier, for didactic purposes, we will cover this type of error when teaching the adjectival phrase whether the error is in the adjective itself or in the adverbial premodifier.

Let us now focus on some specific cases that were problematic while coding and then see how we solved them. We will see how we dealt with problems of segmentation and coding.

In this example we can see how coders mostly agreed on the type of error but segmented differently. We observed that some members used to select the phrase where the error was inserted. We decided against this practice for the reasons we have explained above.

*The education in Spain is a subject that given a lot of play because **for one people***

versus

*The education in Spain is a subject that given a lot of play because **for one people.***

The following image shows another example.

players who have an important paper. In this aspect I have to men

	Consensus:	lexical-error: wordchoice-error: other-wordchoice-error: noun-vocab-error	role
✓	R4		✓
✓	R1		✓
✗	R7	phrasing-error: transferred-phrasing	✓
✗	R6	lexical-error: wordchoice-error: transferred-word: borrowing	✓
✗	R3	lexical-error: wordchoice-error: false-friend	✓
✗	R2	lexical-error: wordchoice-error	✓

Figure 4.

Here coders agreed regarding segmentation but disagreed on how to code the error type. After the meetings for the inter-coder reliability studies, a consensus was reached.

The changes that we implemented involved creating new error categories such as “missing-space-separator” or “incoherent-connector-for-cotext” under “pragmatic-error”. We also added error types for the three types of non-finite clause: “infinitive-clause-formation-error”, “present-participle-formation-error” and “past-participle-formation-error”. In other cases we modified the hierarchy such as when we decided to include “vp-missing” under “clause-error”. The tool was also refined and some additional functions were included such as the option to visualize the corrections provided to students.

## 5. Preliminary Results

So far, we have coded 233 of the student texts (88,600 words) in the corpus, and identified a total of 12,000 errors. The large number of errors coded allows us to draw fairly clear statements, at least regarding general trends in the kinds of errors our students make. Clearly, a larger corpus would allow us to make clearer statements as to more specific types of errors.

Figure 5 shows the number of errors per 1000 words. As you can see there is a clear inverse trend in the errors students make as they become more proficient in the language.

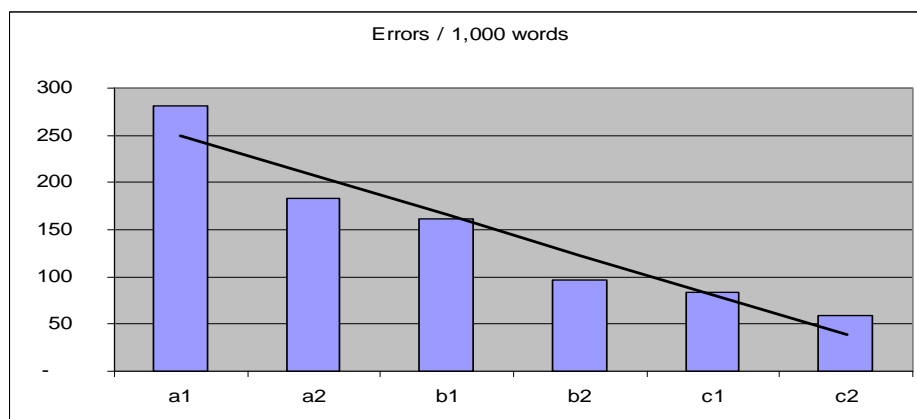


Figure 5.

The sharpest drop takes place from A1 (starters) to A2. Out of 1000 written words 300 are erroneous on average. Another significant decrease is observed in between B1 and B2.

### 5.1 Results per Level of CEFR

As students improve their linguistic skills, the type of error evolves. At the A1 level, grammar errors account for 46% of errors, whereas at the B1 level, grammatical errors increase to account for almost 50% of all errors. At higher levels such as C2, punctuation and pragmatic errors now surprisingly become the two most significant ones. This could be due partially to the fact that at higher levels the message becomes more elaborate and students tend to use longer sentences. We also must bear in mind that coders may pay more attention to this type of error as the number of grammatical and lexical errors decreases.

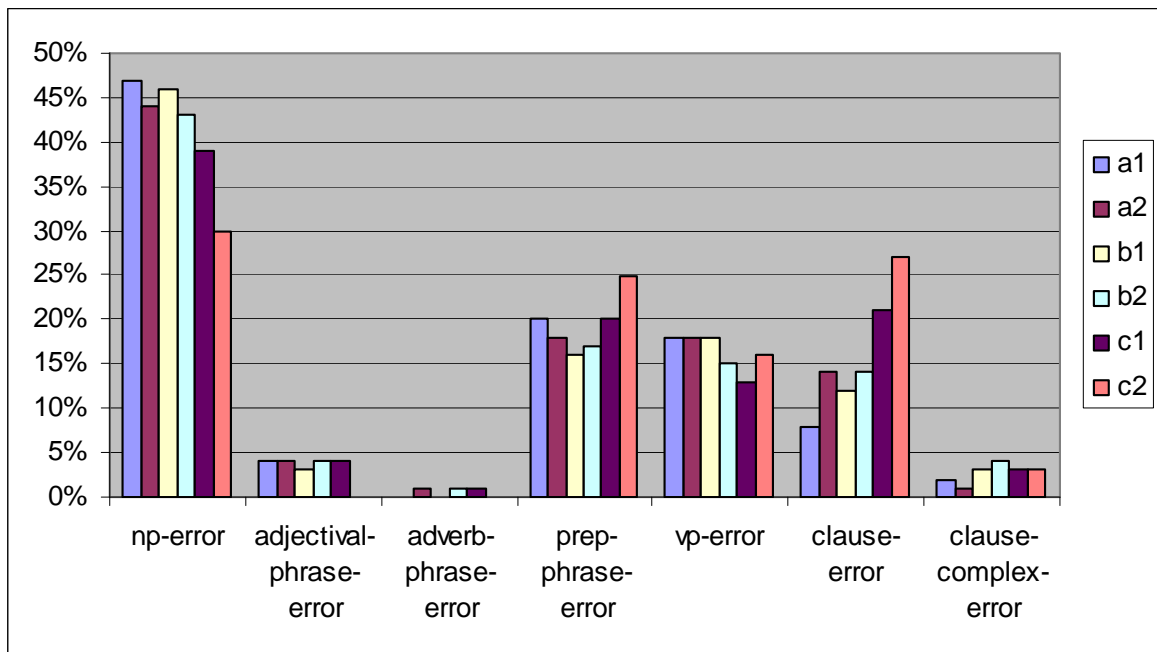


Figure 6.

Grammar errors are the most common type of error across all levels. What seems surprising is the rise of grammar errors as we move from beginners (A1 and A2) to the intermediate level (B1). A possible answer is that students start to experiment more with the language, trying to form new structures. As we progress with the coding we will be able to confirm or dismiss this hypothesis. Regarding lexical errors, we observe very similar percentages across different levels. However they decrease at higher levels. We also observe an increase in punctuation and pragmatic errors from B1 level up to C2.

### 5.2 Main error categories

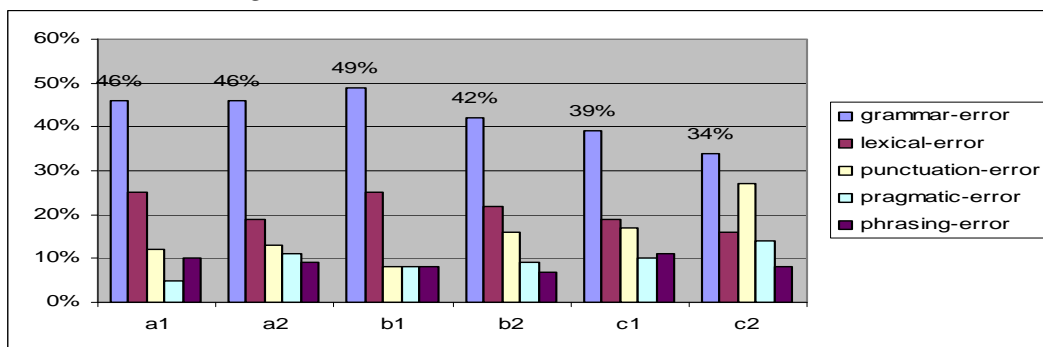


Figure 7.



As a general trend, most errors fall into the grammatical error type, and within this category most of them occur in the noun phrase, with the determiner being the most problematic area for Spanish students. Besides the noun phrase, prepositional phrases are the next biggest problem, in particular concerning the selection of the preposition. Noun phrase errors are definitely the most widely spread across all levels. Prepositional errors also seem to be quite consistent at lower and intermediate levels; they even rise in higher ones. Clause errors reach a peak at C1 and C2.

## **6. Conclusions**

As we have just seen, the empirical study based on TREACLE Learners Corpus suggests that the error-coding system we are using is viable and provides codes for most errors encountered in error-coding a text.

The data compiled so far suggests that the number of errors decreases as the proficiency level increases. With regards to the type of errors, we would like to highlight two areas which require more emphasis in the teaching: the noun phrase and prepositions.

Although time-consuming, projects like TREACLE are extremely useful to provide data on learners' interlanguage and contribute to a potential improvement in the curriculum design. This type of studies is essential for a more fine-grained grammar syllabus at the university level.

## **7. Contributions**

We believe the use of learner corpora can promote learners' autonomy which is one of the more important goals in current language teaching, especially in higher education as stated by Littlemore (2001).

In this regard, the UAM Corpus Tool has proved to be very versatile as it includes the necessary functions for researchers to analyse students' errors in a meaningful context (task based approach) and it can also be used to give feedback to students and to promote learners' autonomy. Indeed, providing students with feedback is one of the five components that Holec (1996: 89-93) includes in his definition of "self-directed process of learning".

In addition to learners' autonomy, language awareness plays a key role in the assimilation of accurate grammar use. Many authors such as Mendikoetxea et alii (2010: 180-94) agree on the benefits of error exploration in the EFL context as a tool to raise language awareness.

We hope that this work contributes to spreading the notion of grammar as not only referring to rules governing form but also to the grammatical knowledge that matches language uses to students' intentions in particular contexts (Larsen-Freeman 2002).

If we adopt a broader perspective, we can appreciate the relevance of this work as it may contribute to further postulate hypotheses on the acquisition of foreign languages. As Storjohann states, the use of corpora to study language use in an empirical way has revealed new research possibilities in linguistics and cognitive linguistics.

Through the use of corpora, for example, we gain a different notion of language as it emerges from language use. The central function of language as a means of natural communication and its role in social interaction are no longer ignored. (...) Both cognitive linguists and corpus linguists share an interest in contextualised, dynamically constructed meaning and in the grounding of language use in cognitive and social-interactional process. (Storjohann 2010: 8)

We believe TREACLE is a pioneering work in learner corpora at higher education in the Spanish context. The detailed annotation and granularity of the corpus certainly offer a wide variety of options for further research (Meunier 1998: 20). We will mention some of them in our final section.

## 8. Areas of Further Research

Although we have not devoted time to exploring transfer from L1 to L2, it could be interesting to analyse how students' interlanguage evolves at different levels of proficiency. At present we do not have statistics in this area. However, during the coding of texts from various levels, the degree of lexis and structures transferred from L1 to L2 was salient. At lower levels, students tend to rely more on L1 linguistic resources, but as they evolve, this dependency on L1 resources tends to disappear progressively (Larsen-Freeman 1978). A factor that can discontinue this progression could be related in some cases to the writing process's own cognitive dynamics. In general, students with stronger writing skills in their mother tongue may find it easier to overcome the transfer from L1 to L2. Nonetheless, we have observed that some of these errors persist in lower and more advanced levels.

Another possible area of research would be finding a rationale for the increase of grammar errors when we move from the basic to the intermediate level. Although we need to progress with the coding to allow more significant results, it may be worth exploring further samples by *cautious learners* and other more adventurous students. Moreover, it would be interesting to observe the general trend in errors and compare it with the sequence of elements in the curriculum and with the *processability model* advocated by Pienemann (1998).

With regards to vocabulary acquisition, a contrastive analysis of the lexis used in essays per topic with other compositions by native speakers on the same topics could offer a more accurate idea of the lexical gap that students need to cover. In the long term, it would be desirable to create specific materials catering for these special needs.

### Works Cited

- Andreu, M<sup>a</sup> Ángeles, Aurora Astor, María Boquera, Penny Macdonald, Begoña Montero and Carmen Pérez 2010. Analysing EFL learner output in the MiLC project: An error it's\*, but which tag? Eds. Begoña Belles-Fortuno, M<sup>a</sup> Carmen Campoy and M<sup>a</sup> Lluisa Gea-Valor. *Corpus-Based Approaches to English Language Teaching*. London: Continuum. 167-179.
- Cook, Vivian 2008. *Second Language Learning and Language Teaching*. London: Hodder Education.
- Corder, S. Pit 1982. *Error Analysis and Interlanguage*. Oxford: Oxford University Press.
- Díaz-Negrillo, Ana and Jesús Fernández Domínguez 2006. *Error Tagging Systems for Learner Corpora*. RESLA 19: 83-102.
- Greenbaum, Sidney and Randolph Quirk 2008. *A Comprehensive Grammar of the English Language*. Harlow: Longman.
- Holec, Henri 1996. Self-directed learning: an alternative form of training. *Language Teaching* 29: 89-93. <<http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=2768924>>. 24 September 2011.
- Izumi, Emi, Kiyotaka Uchimoto and Hitoshi Isahara 2005. *Error Annotation for Corpus of Japanese Learner English*. Proceedings of the Sixth International Workshop on Linguistically Interpreted Corpora (LINC-2005) <<http://acl.ldc.upenn.edu/I/I05/I05-6009.pdf>>. 20 July 2011.
- Larsen-Freeman, Diane 1978. Evidence of the need for a second language acquisition index of development. Ed. William C. Ritchie. *Second Language Acquisition Index of Development*. New York: Academic Press. 127-36.
- Larsen-Freeman, Diane 2002. The Grammar of Choice. Ed. Eli Hinkel. *New Perspectives on Grammar Teaching in Second Language Classrooms*. Mahwah: L. Erlbaum Associates. 105-20.

- Littlemore, Jeannette 2001. Learner autonomy, self-instruction and new technologies in language learning: Current theory and practice in higher education in Europe. Eds. Angela Chambers and Graham Davies. *ICT and Language Learning: A European Perspective*. Lisse: Swets & Zeitlinger Publishers. 39-52.
- Manchón, Rosa M. 1999 La investigación sobre la escritura como proceso. Algunas implicaciones para la enseñanza de la composición en una lengua extranjera. Ed. Sagrario Salaberri. *Lingüística aplicada a la enseñanza de lenguas extranjeras*. Almería: Universidad de Almería. 439-78.
- Mendikoetxea, Amaya, Michael O'Donnell and Paul Rollinson 2009. WriCLE: A Learner Corpus for Second Language Acquisition Research. *Proceedings of the Corpus Linguistics Conference. CL2009. 20-23 July 2009*. <[http://ucrel.lancs.ac.uk/publications/cl2009/351\\_FullPaper.doc](http://ucrel.lancs.ac.uk/publications/cl2009/351_FullPaper.doc)>. 02 July 2011.
- Mendikoetxea, Amaya, Susana Murcia and Paul Rollinson, 2010. Focus on errors: learner corpora as pedagogical tools. Eds. Begoña Belles-Fortunato, M<sup>a</sup> Carmen Campoy and M<sup>a</sup> Lluisa Gea-Valor. *Corpus-Based Approaches to English Language Teaching*. London: Continuum. 180-94.
- Meunier, Fanny. The pedagogical value of native and learner corpora in EFL grammar teaching 2002. Eds. Sylviane Granger, Joseph Hung and Stephanie Petch-Tyson. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins. 119-142.
- Pravec, Norma A. 2002. Survey of learner corpora. *ICAME Journal 26*: 81-114. <<http://nora.hd.uib.no/icame/ij26/pravec.pdf>>. 27 September 2011.
- Quirk, Randolph 1998. *A University Grammar of English*. Harlow: Longman.
- Richards, Jack C. 1970. *A Non-Contrastive Approach to Error Analysis*. San Francisco: TESOL Convention.
- Ritchie, William 1978. *Second Language Acquisition Research. Issues and Implications*. New York: Academic Press.
- Rollinson, Paul Martin and Amaya Mendikoetxea 2008. *The WriCLE corpus*. <<http://web.uam.es/woslac/WriCLE/>>. 10 September 2011.
- Storjohann, Peter 2010. *Lexical Semantic Relations*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Widdowson, H. G. 1983. *Learning Purpose and Language Use*. Oxford: Oxford University Press.