

## Error coding in the TREACLE project<sup>184</sup>

PENNY MACDONALD (*Universitat Politècnica de València*), SUSANA MURCIA (*Universidad Autónoma de Madrid*), MARÍA BOQUERA (*Universitat Politècnica de València*), ANA BOTELLA (*Universitat Politècnica de València*), LAURA CARDONA (*Universitat Politècnica de València*), REBECA GARCÍA (*Universidad Autónoma de Madrid*), ESTHER MEDIERO (*Universidad Autónoma de Madrid*), MICHAEL O'DONNELL (*Universidad Autónoma de Madrid*), AINHOA ROBLES (*Universidad Autónoma de Madrid*), KEITH STUART (*Universitat Politècnica de València*)

### *Abstract*

*This paper presents the approach to error analysis within the TREACLE project, the aim of which is to profile learner proficiency to help inform teaching curriculum design. We will introduce the error annotation methodology used on a corpus of texts written by Spanish learners of English at University level. After a short introduction concerning Computer Learner Corpora and error analysis, we will discuss the underlying principles of the error coding scheme and then provide more details about the coding scheme itself. To ensure coders are annotating the texts in the same way, two steps were followed. Firstly, we developed a comprehensive coding criteria description giving full details as to how to code particular instances. Secondly, we performed two inter-coder reliability studies to help us identify areas where coders were differing, so that we could address these areas. We will present the preliminary results of the error analysis and discuss the repercussions of these results for grammar teaching.*

*Keywords: learner corpora, error analysis, English as a Foreign Language*

### *Resumen*

*Este artículo presenta el enfoque del análisis de errores del proyecto TREACLE, cuyo objetivo es proporcionar perfiles de competencia de aprendices con el objeto de poder informar el diseño curricular. Se presente la metodología utilizada en la anotación de errores de un corpus de aprendices escrito por estudiantes españoles de inglés a nivel universitario. Después de una breve introducción a los corpus computerizados de aprendices y al análisis de errores, se relacionan los principios subyacentes al esquema de codificación de errores y se proporcionan más detalles sobre el esquema de codificación mismo. Para asegurar que los codificadores anotan los textos de la misma manera, se siguieron dos pasos. En primer lugar, se realizó una amplia descripción de los criterios de codificación que contiene detalles sobre cómo codificar ciertos casos. En segundo lugar, realizamos dos estudios de fiabilidad inter-codificadores que nos permitieran identificar aquellas áreas en las que los codificadores discrepaban, de modo que pudiéramos aclararlas. Se presentan los resultados preliminares del análisis de errores y se discuten las posibles repercusiones de estos resultados para la enseñanza de la gramática.*

*Palabras clave: corpus de aprendices, análisis de errores, Inglés como Lengua Extranjera*

<sup>184</sup> The TREACLE project is funded by the Spanish Ministry of Science and Innovation (research grant: FFI2009-14436/FILO)

## 1. INTRODUCTION

The development of Computer Learner Corpora (CLC) in the early 1980s prompted the creation of a new and lasting relationship which has served to bridge the gap between the field of Corpus Linguistics and foreign language learning research and pedagogy. One of the first CLC projects was the Danish PIF - Project in Foreign Language Pedagogy (Faerch et al., 1984). In the early 1990s, The International Corpus of Learner English (ICLE), founded and coordinated by Sylvianne Granger at the Université Catholique de Louvain in Belgium (Granger, 1993; 1998) was developed and has become the most cited in the literature. However, since then, the development of new learner corpora has increased tenfold as can be observed on the 'Learner Corpora around the World' web page created by the Centre for Corpus Linguistics at Louvain.

Learner corpora, which tend to be much smaller than native corpora, have been used in the literature to provide evidence concerning what language learners have acquired and not acquired, sometimes comparing the output with native speakers (NSs), and other times comparing it with other mother tongue groups. Granger et al. (2002) identify two main ways in which corpora have been studied from a linguistic-based methodological perspective: Contrastive Interlanguage Analysis and Computer-aided Error Analysis. The first method deals with those studies that compare the output of either two or more non-native speaker (NNS) groups or two or more NS and NNS groups. Computer-aided Error Analysis involves the study of learner output (i) in order to detect the difficulties that learners have, which in turn contributes to finding out more about the processes involved in learning, (ii) for identifying the instances of crosslinguistic influence, (iii) for the development of new materials, etc.

Despite the affirmation of one of the most prestigious researchers in the field of Error Analysis, Carl James, that Contrastive and Error Analysis (EA) "are still going strong" (1994:179), there has not been much evidence of this in the literature in recent years. Indeed, since the heyday of Error Analysis in the 60s and 70s, there have been fewer and fewer studies dedicated to the analysis of errors in learner output in foreign language learning environments, which, as Leech (1998: xvii) points out, may be due to the fact that "...the negative attitudes to Error Analysis inherited from that period have coloured many people's thinking ever since".

Yet on a day to day basis, teachers are doing it continually. And in the research on second language acquisition and teaching methodologies, studying learner output, both correct and incorrect forms, is still central to the agenda although the shift has been more towards finding ways of being successful with corrective feedback in order to improve the learners' linguistic and communicative competence. This may involve teacher-centred feedback; peer feedback, or computer-mediated feedback in the form of grammar checkers, online tutorials, and so on.

From a different point of view, although, with pedagogical goals as the main focus, there are a number of more recent studies which are analysing errors in order to create specific profiles of learner competence (Capel, 2010; Granger and Thewissen, 2005a; 2005b). In the case of the English Profile Project (Hawkins & Buttery, 2009) the aim is to develop

Reference Level Descriptions for English linked to the Common European Framework of Reference for Languages (CEFR).

The TREACLE project (O’Donnell et al., 2009),<sup>185</sup> is also concerned with the creation of a methodology for producing grammatical profiles of Spanish university students’ written English language, linking them to the CEFR levels, with the aim of redesigning the English Language grammar curriculum to improve its efficiency and projection.

The project is described in detail in O’Donnell in the present volume, and therefore the purpose of this paper is to present only those aspects of the project related to error-annotation, discussing its methodology and some of the preliminary results, as well as their possible repercussions on the teaching and learning of English in the Spanish higher education context.

## **2. THE TREACLE PROJECT AND THE CORPUS FOR ERROR ANNOTATION**

The TREACLE<sup>186</sup> project (O’Donnell et al., 2009) involves the development of an annotated corpus of learner English for pedagogical application. One of the aims of the project is to carry out a computer-aided error analysis on the corpus to find out what students tend to get wrong at each level of proficiency.

This project uses corpora from two Spanish universities: Universidad Autónoma de Madrid (UAM) and Universidad Politécnica de Valencia (UPV). Since these universities have different degree courses on offer, the written output of these students can be said to represent the kind of English as a Foreign Language which learners in both Humanities and Technically-oriented universities produce in Spain.

The two corpora we are using for our study –the WriCLE Corpus (UAM) and the UPV Learner Corpus have been developed following strict design criteria.

The UPV Learner Corpus is part of the MiLC corpus<sup>187</sup> (Andreu et al., 2010). The UPV Learner Corpus consists of 950 written compositions (180,000 words) from Spanish students of all levels, mostly centring on the topic of Immigration.

The WriCLE Corpus (Rollinson and Mendikoetxea, 2010), consists of 750 essays written by Spanish learners of English. For the TREACLE project, we use 521 of these essays (about 500,000 words) written by Spanish students of all levels of proficiency. The essays deal with the topics of immigration, homosexual marriages and traffic problems.

In both cases, the metadata is carefully recorded (i.e. details of sex, age, year of study, mother tongue and other languages spoken/learnt, etc.) and all students gave their permission for the texts to be used for research purposes and did the Oxford Quick Placement Test (UCLES 2001) close to the time of writing in order to identify their levels according to the CEFR levels.

---

<sup>185</sup> More information on this project can be found at <http://www.uam.es/treacle/index.html>

<sup>186</sup> *TREACLE stands for Teaching Resource Extraction from an Annotated Corpus of Learner English.*

<sup>187</sup> MiLC is a multilingual learner corpus involving the written work (formal and informal letters, summaries, essays, reports, translations, simulations, computer-mediated communication, etc.) of students learning English, Spanish, French, and German as a foreign language, and also Catalan, as a first, second or foreign language.

The length of the texts varies – the UPV compositions tend to be much shorter (between 200 and 250 words) as the levels, in general, are lower. In the case of the WriCLE essays, these are mostly 1,000 word essays produced by students of English Philology. In this way we have a collection of essays that represent the different levels from A1 through to C2.

### 3. METHODOLOGY

The error annotation of the corpus is carried out manually with the *UAM CorpusTool* (O'Donnell, 2008), which uses an error coding scheme devised by the researchers for this particular purpose and which can be modified as needed.

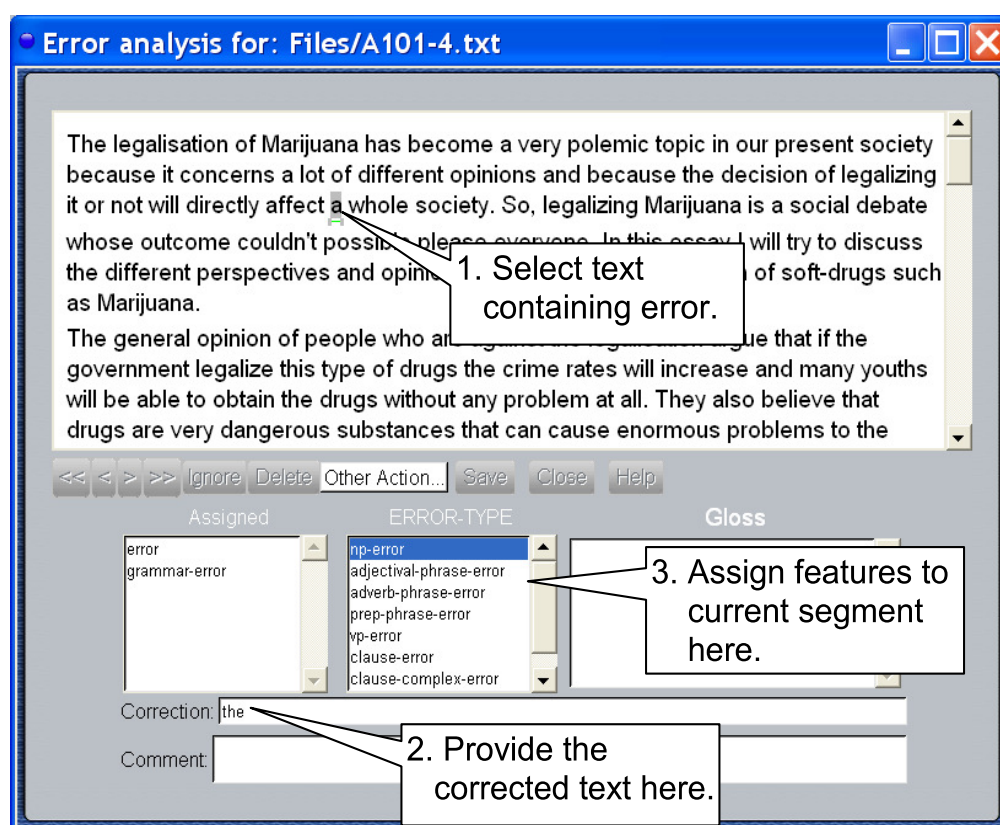


Figure 1. Error coding a text

As shown on the error annotation window in Figure 1, the UAM CorpusTool allows the coder to select the text of the error (step 1 in the diagram), provide the correction for that error (step 2), and assign error codes to it (step 3). The system is provided with a hierarchically organised set of error codes (the coding scheme) which the user walks through to assign a code, e.g., selecting first “grammar-error”, then “np-error”, then “determiner-error” and then “determiner-choice-error”. This process of gradual refinement of error codes facilitates the coding process, because often the coder does not know what leaf of the error tree they should code, but can make a series of decisions on more general grounds (e.g., Is it a grammatical or lexical error?).

To facilitate the coder's job, the error scheme incorporates coding criteria ('glosses') with each feature in the scheme, which are displayed in the coding window. These glosses are also included in a lengthy Coding Criteria Manual (20 pages long) which provides clear guidelines for determining which of the categories is appropriate for a given error.

### 3.1. The Error Scheme

The TREACLE error scheme has been designed from the start to integrate into a University level language teaching programme. So far, the coding scheme contains 113 different errors at the most delicate level in the hierarchy. Figure 2 shows the more general categories, as the whole scheme does not fit here.<sup>188</sup>

The main design principle has been to ensure that the error scheme should map cleanly onto the organization of grammar topics which are taught within EFL courses. The main reason for this is that our goals are pedagogical, and we later want to be able to recover the errors relevant to each grammatical structure, so as to inform our teaching of that topic.<sup>189</sup>

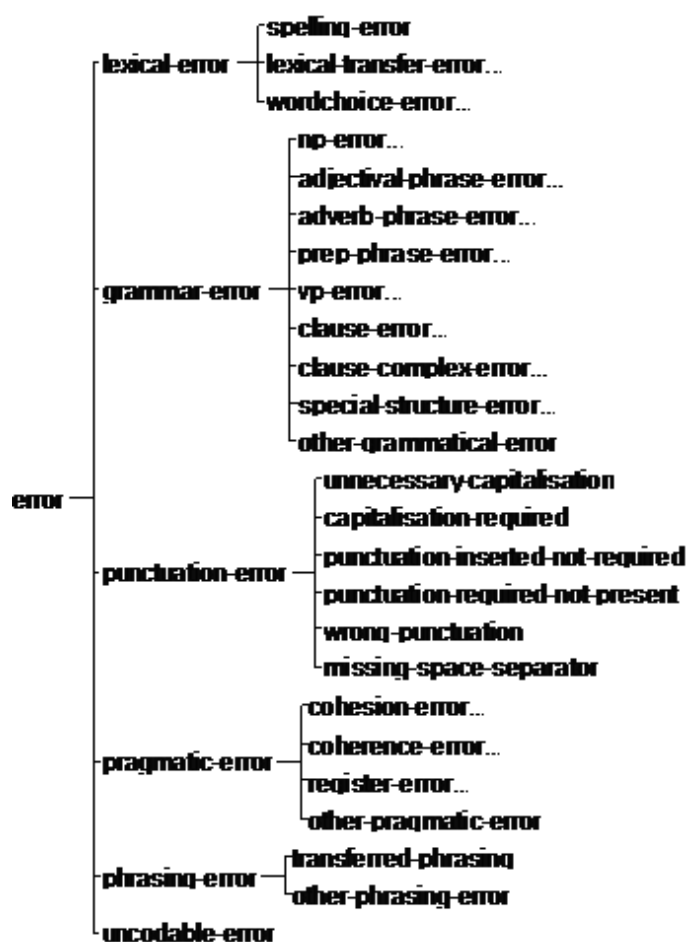


Figure 2. The coding scheme

<sup>188</sup> The three dots after an error category indicate that sub-categorization has been omitted.

<sup>189</sup> Our scheme, however, includes errors other than grammatical ones, as shown in Figure 2.

To demonstrate our approach, we will focus on one particular part of the error scheme. Let us assume the student has written “\*this results are...”. At the root of the error hierarchy, we distinguish between various types of error, including lexical errors, grammatical errors, pragmatic errors, etc. (see Figure 2).

Assuming the error is grammatical, one next chooses the type of grammatical error. While some error coding systems are more oriented to coding errors in terms of the part of speech of the word concerned, our approach is more focused on the grammatical phrase in which that word occurs. Thus, while teaching about adjectival phrases, we can find errors within adjectival phrases, whether they involve the adjective itself, or any adverbial premodifier of the adjective, e.g. “very browner”.

Continuing with our example, assume the error is within the selection of the determiner. We thus select *np-error*. This leads to the next level of delicacy, as shown in Figure 3. Our division of error codes within the NP reflects the fact that, in many courses, the teaching of the Noun Phrase is divided into topics: determiners, pre-modifiers, the head, and post-modifiers.

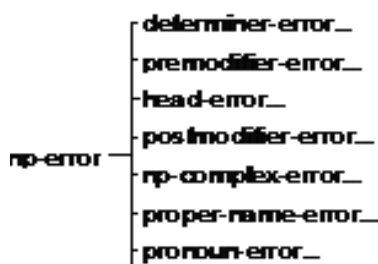


Figure 3. NP Error sub-classes

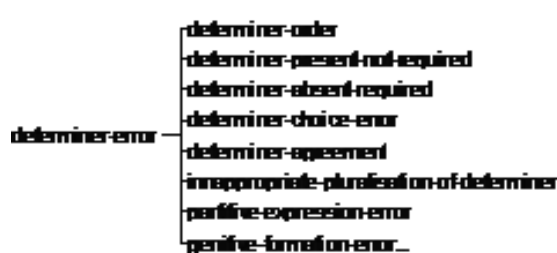


Figure 4. Subtypes of determiner errors

Selecting *determiner-error*, we are presented with the next set of choices, as shown in Figure 4. Note that the error codes shown here are not exhaustive; the coder can add new error codes as examples are encountered in the learner texts. The particular example we started with, “*this* results are...”, would be coded as *determiner-agreement*.

Our error coding is still at an early stage. So far, we have coded 128 texts containing 57,000 words, with 7,428 errors. However, this trial period has established that the error scheme and annotation process is viable, and that it caters for almost all of the errors encountered. We intend to double the number of errors coded by the end of 2011.

To ensure that the coding scheme and the coding criteria manual were valid instruments for our coding purposes and that they can render reliable coding among the researchers taking part in the project, we carried out two inter-coder reliability studies which will be explained in the following section.

#### 4. INTERCODER RELIABILITY STUDY

According to Polio (1997: 102), error analysis studies ‘have rarely reported intra- and interrater reliabilities, which can call into question’ the conclusions reached in the research results, and certainly ‘make replication of a study’ in a different context, somewhat problematic.

We performed two inter-coder reliability studies with the aim of:

1. Refining the error scheme and coding criteria document.
2. Ensuring all coders were complying with the coding criteria.

One of the first stages in the error analysis was for the coders to become accustomed to the programme and the different levels and types of error codes. Several texts were coded on an individual basis and meetings were then held to discuss either problematic phrases or structures, or doubts concerning how to actually code the errors that had been detected. The second stage in the inter-coder reliability study (ICRS) involved the coding by all researchers of several texts, followed by consensus meetings in order to detect any individual ‘doubts’ that needed to be clarified, to add on new codes, or to make any necessary changes to the error hierarchy.

The coders had to familiarize themselves with the guidelines concerning segmentation with the UAM Corpus Tool. The rule is that we use minimal segmentation, i.e. only select the amount of text that would be necessary in order to make the correction of the erroneous form. So, in the case of a frequent error such as: *\*in the other hand* we would only select the preposition ‘*in*’. The reason for this is that the automatic syntactic analysis<sup>190</sup>

will locate this word within the phrase on another level of analysis, and also, as the rest of the connector is correct, the learner has got most of it right and therefore it is not necessary to highlight the whole phrase. However, to begin with, it was found that several researchers tended to select more words than was necessary, since the error could be seen to be affecting the whole phrase.

We also code in regards to what the learner has written rather than what they should have written, which is also common practice in the ICLE learner corpus especially prioritising the identification of the ‘grammatical’ errors which is, after all, the *raison d’être* of the project, always keeping in mind the use of the different grammatical forms made by Spanish university students.

For the inter-coder study, rather than comparing each coder to each other coder (requiring 120 distinct comparisons), we derived a ‘consensus model’ (or ‘golden standard’), this being the subset of all the error codings on which at least 51% of the coders agreed. We could then compare each individual coder to this consensus project to see how the individual coder differed from the consensus. A report was automatically generated (see figures 5 and 6 below), showing each identified error, how many participants coded it, which participants provided alternative codings of it, cases where the text correction differed, and so on. This document provided the basis of our discussion of the various differences in coding practice, and helped us move towards a consensus.

Reliability scores were calculated by comparing each individual’s coding to the consensus model.

As regards consensus rates, some codes were easier to pinpoint by all coders, such as

---

190 Our scheme, however, includes errors other than grammatical ones, as shown in Figure 2.

those involving spelling, punctuation, article errors, verb tense errors, etc. On the other hand, at least in the initial stages, certain errors involving larger chunks of text, for instance, phrasing errors, proved to be more difficult to assign codes to in a unanimous way. The following examples (extracted from an essay on education in Spain) illustrate some of the issues that have interfered with achieving a higher score in the ICRS.

### Example 1

*'The education in Spain is a subject that given a lot of play because for one people this system of education is great, but for other people is awful.'*

In Figure 5 below, it can be seen that R6 did not highlight the minimum segment (**one**) although the code and the correction coincide with the rest of the coders. Technically speaking, consensus was reached, although this is not reflected in the results.

ven a lot of play because for **one** people this system of educati

	Consensus:	grammar-error: np-error: determiner-error: determiner-choice-error	some
✓	R7		✓
✓	R1		✓
✓	R2		✓
✓	R5		✓
✗	R4	grammar-error: np-error: premodifier-error: incorrect-premodifier-category	✓
✗	R3	lexical-error: wordchoice-error: other-wordchoice-error	✓

ven a lot of play because for **one people** this system of education is g

✗	R6	grammar-error: np-error: determiner-error: determiner-choice-error	some people
---	----	--	-------------

Figure 5. Example 1 of ICRS

### Example 2

There are other cases where the coders did not agree at first with the codes chosen for a certain error. However, after the consensus meetings, acceptable solutions were found, such as in the example below – the error was due to L1 transfer, and should be coded as *Lexical error – lexical transfer error – false friend*.

*'there are a lot of players who have an important **paper**'.*



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	✓

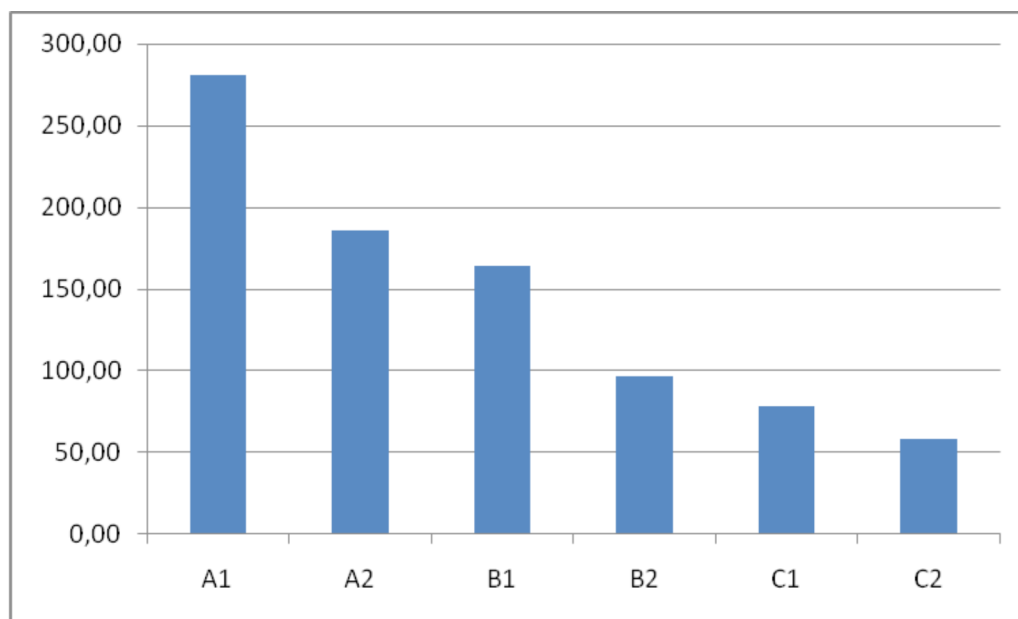
**Figure 6. Example 2 of ICRS**

False friends groups together those words that have similar ancestors, such as is the case with cognates, but whose meanings (or some of their meanings) have diverged over time e.g. Spanish *éxito* and English *exit* or Spanish *remover* and English *remove*. At times there may only be partial semantic identity, as Odlin (1989:79) explains – as in Spanish *suceder*, and English *succeed*, which is the case of the word *paper* as the writer took for granted s/he could transfer the metaphorical meaning of Spanish *papel* (*role*) for the literal translation into English *paper*.

The Error Criteria Manual has undergone several changes as a result of the ICRS, and indeed, the coding tool is designed so as to expand on any of the branches already established and in this way, include a wider variety of choices which can pinpoint certain errors with greater accuracy.

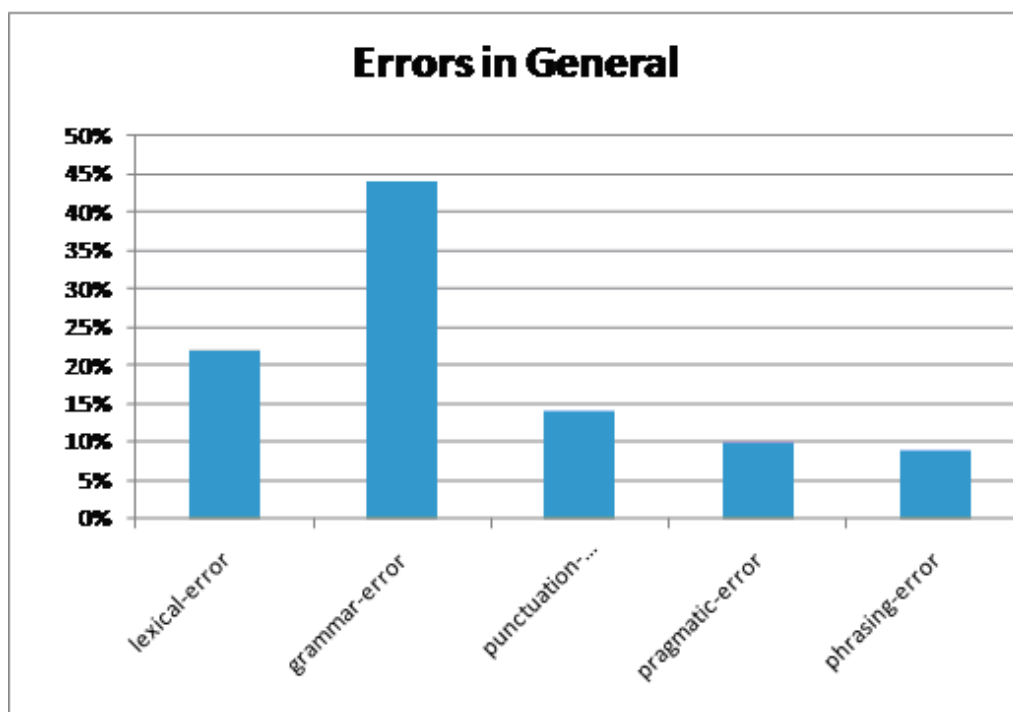
## 5. RESULTS

As we mentioned above, so far, a total of 128 texts (57,000 words) have been coded, and 7,428 errors have been identified. The coding done so far reveals some interesting results.



**Figure 7. Number of errors per 1,000 words**

Figure 7 shows that the number of errors decreases as the proficiency level increases. At the same time, these results provide evidence that the Oxford Quick Placement Test is a valid instrument for identifying students' proficiency level



**Figure 8. Results for general error categories**

Figure 8 shows the percentage of errors in relation to all the errors made by the learners. Of the five main categories of errors, we can observe that grammar errors account for 44% of all errors.

The object of this study is to make profiles of what learners know and do not know (or do not use) according to the different proficiency levels and with particular reference to grammar. Figure 9 indicates how the degree of error occurrence varies according to the CEFR level. Thus, although grammar errors are still the highest, there are some notable differences within this group depending on the level. For example, it is the B1 Group which makes most grammar errors – nearly 50%, and this is 15% more than the C2s. This group, with the A1 group, also makes more errors as regards lexis. On the other hand, the C2 group makes the most errors as regards punctuation and what we have called ‘pragmatic’ errors. In the case of punctuation errors, the high rate among the C2 proficiency level may be due to the fact that the texts tend to be written with longer sentences, involving more subordination and linking devices, which in turn, require more punctuation. Also, we understand that when analysing the errors of the lower level groups, the coders may not actually be able to concentrate on finding the punctuation errors since first they have to make sense of, and code the other types of errors (grammar, lexical) present in the texts.

As regards these preliminary results, there are some levels of errors which do not seem to follow the expected pattern, for example, the unexplained drop in A2 as regards lexical errors, or the sharp rise in punctuation errors within the C2 level. This is most likely to be a sampling error, and will probably smooth out as more data is obtained.

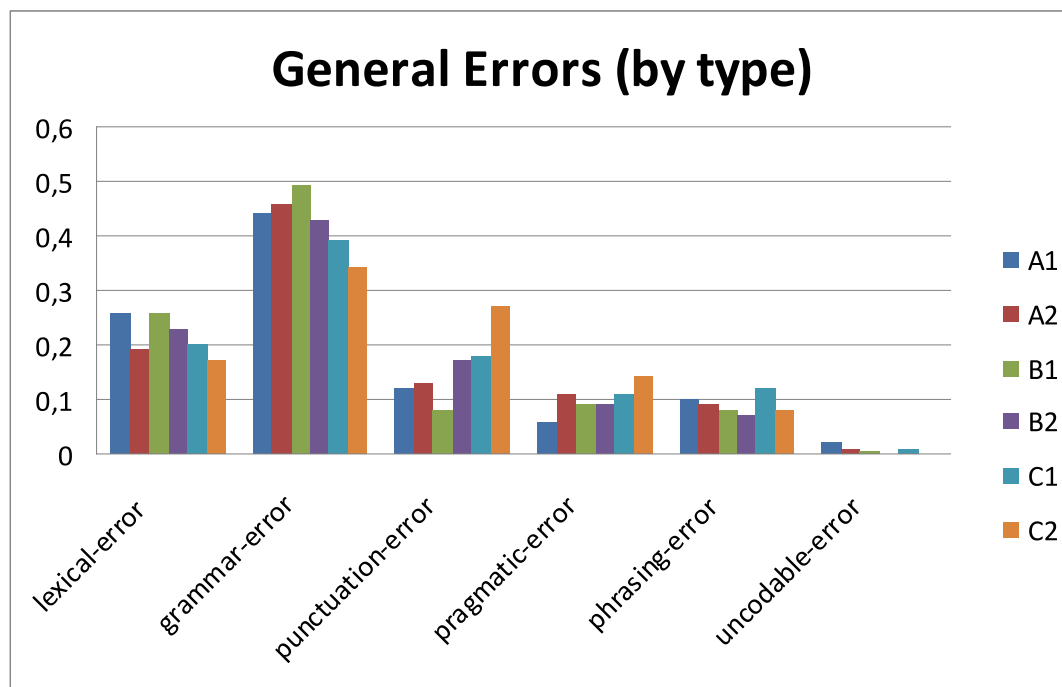


Figure 9. General errors produced at each CEFR level

Figure 10 looks at the distribution of errors within Grammar Error category. Around 40% of grammar errors are within the NP, which suggests that more emphasis in the teaching of NP syntax is needed. The fact that this percent falls in general as proficiency rises suggests that more attention to NPs is needed for the lower proficiency learners. Errors in prepositional phrases are also significant (around 20% of grammar errors), with most of these consisting of wrong selection of the preposition itself. This is another area where increased teaching emphasis is indicated.

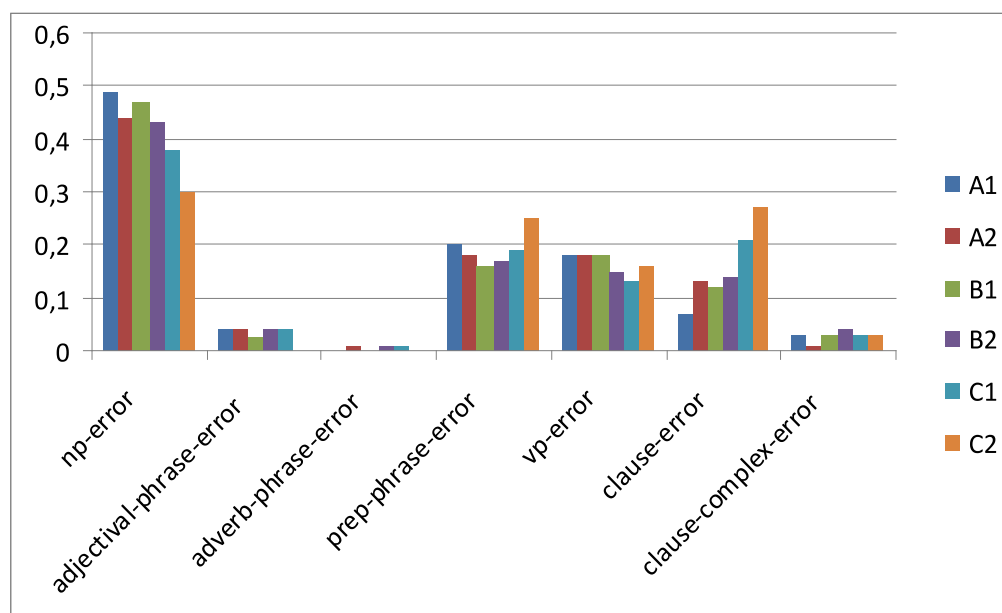


Figure 10. Grammar errors produced at each CEFR level

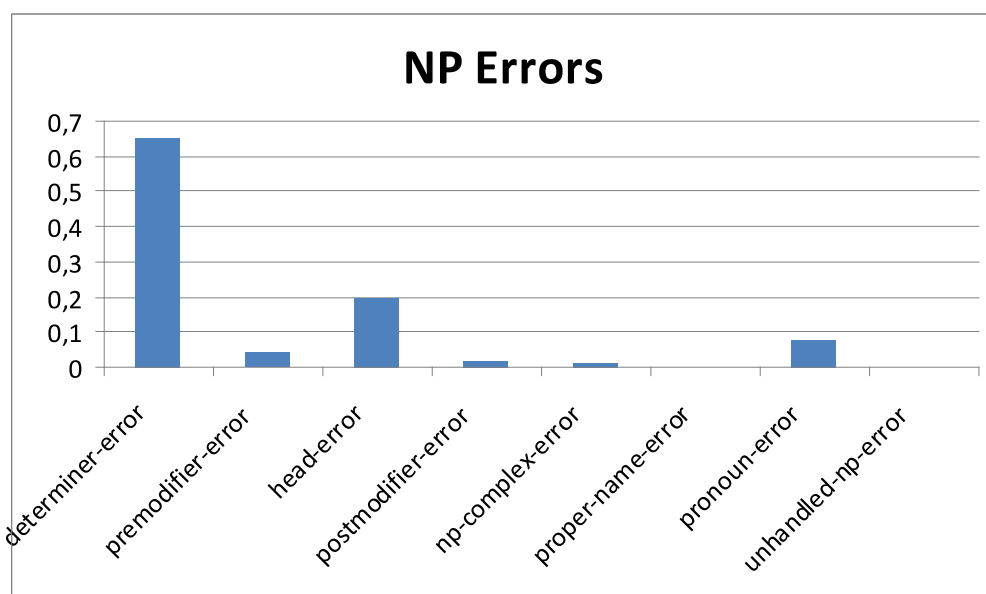


Figure 11. Results for NP error subclasses

In particular, within the noun phrase category (Figure 11), the most frequent errors by far were those related to determiner use (30% of grammar-related errors), and 75% of these involved the wrongful presence or absence of the determiner. The students' mother tongue syntactic structures may have a certain influence here. As Swan & Smith (1987: 83) point out, in Spanish, as regards one determiner, the definite article, this goes with mass nouns and plural count nouns when used with a general meaning, whereas in English this is not the case. Likewise, there are certain contexts in English (i.e. with single count nouns) where articles are needed, and are not required in Spanish e.g. *\*Do you have bicycle* (¿Tienes bicicleta?) or in the following case: *\*My sister is teacher* (Mi hermana es profesora).

In addition to helping with the process of error coding, the *UAM Corpus Tool* allows one to search for particular structures and phrases containing specific examples of errors in context.

Enter Search Query Below:

clauses + containing immediately noun-phrase + containing immediately determiner-present-not-required +

Only Partially Coded Segment with Comment 29 matches

P	C	Pretext	Contained	
		o contaminating movements, increase in	a 30 %	the rail bus, develop even more
		a 30 % the rail bus, develop even more	the public transport	
		Secondly,	the education at school	must be considered as another
		ces, they have reached their position on	the society	, through a lot of demonstration
		having the same rights like	the other people	
		to adopt	some children	
		that	the children	go to be adopted by the couple
		e twenty of december, Holand legalized	the marriage	and the adoption.
		ber, Holand legalized the marriage and	the adoption	.
		and let too	the marriage and adoption	.
		In Canada in	courious, because the civil marri	but the church marriage is com
		that the amount of cars within	the population	is increasing

Figure 12. Corpus search of article errors

As figure 12 shows, the researcher wishes to find all NPs which contain an error of type *determiner-present-not-required*. Such searches can be used to recover examples which can be then be used for the teaching of particular topics.

## 6. CONCLUSIONS

From the study carried out so far in the TREACLE project, we can draw conclusions as regards the methodology used, the results obtained from the error coding, and the potential applications of studies such as this.

As regards methodology, the error-coding system we are using is viable and seems to provide codes for almost all of the error types we have encountered, since the main design principle underlying the project has been to ensure that the error scheme should relate

directly to the organisation of grammar topics which are taught within EFL courses at university level.

Our coding scheme differs from previous work in two aspects:

- It is oriented towards the pedagogical organisation of grammar, rather than being based on part of speech (e.g. the ICLE taxonomy of errors)
- As a result, we have a large number of codes, necessary to align the error-analysis with a fine-grained grammar syllabus.

The results from the error coding show that the number of errors decreases as students gain proficiency, although some of the results show trends that must be taken in a tentative way until we have more texts coded.

So far, the numbers and types of errors point towards potential improvement, or re-designing, of the teaching curriculum, and considering other teaching methods:

- The types of errors produced by students change at different levels of proficiency. The teaching curriculum, therefore, needs to adapt to the students' needs at each level.
- In Spanish university English language programmes, whether dedicated English Studies degrees or other more technically-oriented degree programmes, more emphasis could be given to certain grammatical structures, such as, for example, the use of the English determiner system (which does not receive much attention in the Spanish university EFL study programmes), or the noun phrase in general, since the results show that from A1 to B2 levels, there are twice as many noun phrase errors as verb phrase errors. Also, the amount of teaching time devoted to prepositions is minimal, but the results from the coding show that students struggle with prepositions at all levels.

These preliminary results, although tentative, lead us to think about how we can help our students. We propose to reinforce those grammar areas where problems are made. We need, however, to distinguish between explicit teaching of concepts (which is often time-constrained and classroom-centred) versus out of class online drilling and self-study. This latter is another area of interest within TREACLE.

Combining the data obtained from the error analysis (on which this paper has focused) with the automatic syntactic analysis (which is another part of the TREACLE project) will help to create learner profiles which will determine which grammatical features need to be taught, in what order, and with what degree of emphasis and attention in the different Spanish university English language study programmes. Although error analysis is a highly time-consuming activity, it does provide valuable information concerning the learners' interlanguage, partially indicating which direction the latter stages of the project will be taking in the near future. Error analysis, for example, cannot be considered in isolation. On the contrary, it needs to be seen in the context of what the students are attempting. Later work, therefore, will explore learner types based on the experimental versus cautious learners.

## 7. BIBLIOGRAPHY

- ANDREU, M., ASTOR, A., BOQUERA, M., MACDONALD, P., MONTERO, B. AND PÉREZ, C. (2010) Analysing EFL learner output in the MiLC project: An error it's\*, but which tag?. In Campoy, M.C., B. Belles-Fortuno and M.L. Gea-Valor (Eds.), *Corpus-Based Approaches to English Language Teaching*. London: Continuum.
- CAPEL, A (2010) Insights and issues arising from the English Profile Wordlists project, *Research Notes*, 41, 2-7.
- DAGNEAUX, E., DENNESS, S., GRANGER, S., & MEUNIER, F. (1996) *Error Tagging Manual Version 1.1*. Centre for English Corpus Linguistics, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
- FAERCH, C., HAASTRUP, K., PHILLIPSON, R. (1984) *Learner language and language learning*. Clevedon: Multilingual Matters.
- GRANGER, S. (1993) 'International Corpus of Learner English.' in J. Arts, P. de Haan & N. Oostdijk (Eds.), *English Language Corpora: Design, Analysis and Exploitation* (pp.57-71). Amsterdam: Rodopi.
- GRANGER, S. (Ed.) (1998) *Learner English on Computer*. London: Longman.
- GRANGER, S., HUNG, J. & PETCH-TYSON, S. (2002) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.
- GRANGER, S. AND J. THEWISSEN (2005a) 'Towards a reconciliation of a "Can Do" and "Can't Do" approach to language assessment'. Paper presented at the Second Annual Conference of EALTA (European Association of Language Testing and Assessment), Voss, Norway, 2-5 June 2005.
- GRANGER, S. AND THEWISSEN, J. (2005b) "The contribution of error-tagged learner corpora to the assessment of language proficiency. Evidence from the International Corpus of Learner English". Paper presented at the 27th Language Testing Research Colloquium, Ottawa (Canada), 18-22 July 2005.
- JAMES, C. (1994) Don't shoot my dodo: On the resilience of contrastive and error analysis. *International Review of Applied Linguistics*, 32 (3), 179-200.
- LEECH, G. (1998) 'Preface' in S. Granger (Ed.) *Learner English on Computer* (pp. xiv-xx). London: Longman.
- HAWKINS, J. A., & BUTTERY, P. (2009) Using learner language from corpora to profile levels of proficiency: Insights from the English Profile Programme. In Proceedings of the 3rd ALTE Conference 2008. Cambridge University Press.
- O'DONNELL, M., MURCIA, S., GARCÍA, R., MOLINA, C., ROLLINSON, P., MACDONALD, P., STUART, K. AND BOQUERA, M. 2009. Exploring the proficiency of English learners: The TREACLE project. *Proceedings of the Fifth Corpus Linguistics Conference, Liverpool*.
- ODLIN, T. (1989) *Language Transfer*. Cambridge: Cambridge University Press.

- O'DONNELL, M. (2008) Demonstration of the UAM CorpusTool for text and image annotation. *Proceedings of the ACL-08:HLT Demo Session (Companion Volume)*, Columbus, Ohio, June 2008 (pp. 13-16). Association for Computational Linguistics.
- POLIO, C. (1997) Measures of linguistic accuracy in second language writing research. *Language Learning*, 47, 101-143.
- ROLLINSON, P. AND MENDIKOETXEA, A. (2010) Learner corpora and second language acquisition: Introducing WriCLE In: J. L. Bueno Alonso, D. Gonzáliz Álvarez, U. Kirsten Torrado, A. E. Martínez Insua, J. Pérez-Guerra, E. Rama Martínez & R. Rodríguez Vázquez (Eds.), *Analizar datos>Describir variación/Analysing data>Describing variation* (pp. 1-12).Vigo: Universidade de Vigo (Servizo de Publicacións).