

## A Corpus-Based Analysis of Errors in Adult EFL Writings

Ana M Pérez Sánchez

Universidad Autónoma.

a.perez.sanchez@uam.es

Pérez Sánchez, A. M. (2013). A Corpus-Based Analysis of Errors in Adult EFL Writings. *Revista Nebrija de Lingüística Aplicada* (2013) 13.

### RESUMEN

En España, las Escuelas Oficiales de Idiomas aplican el Marco Común Europeo de Referencia en la enseñanza y certificación de los distintos idiomas. Algunos profesores han constatado el hecho de que la Expresión Escrita es una de las destrezas más problemáticas para los alumnos, que a menudo no alcanzan los resultados esperados en los exámenes de Certificación, y sobre todo en el nivel avanzado. El objetivo principal del estudio que aquí se presenta, es el análisis sistemático de un corpus computerizado de textos escritos por alumnos con el fin de identificar los errores más comunes en el nivel intermedio. Además en vista de los resultados obtenidos, se sugieren medidas específicas de actuación para mejorar los resultados de los alumnos en futuros exámenes de expresión escrita. Otros estudios relacionados con el análisis de errores en textos de alumnos incluyen : Muehleisen, 2006; Meunier, 2007; O'Donnell et al. 2009 or Granger, 2009.

Palabras clave: Aprendizaje de inglés, expresión escrita, análisis de errores, estudios de corpus

### ABSTRACT

*In the context of adult education in Spain, Official Schools of Languages set the standard of teaching and certification of the different levels described in the Common European Framework of Reference. It has been noticed by a number of teachers that writing has become one of the most problematic skills for the students, who often fail to achieve the results they expect in the different Certificate Exams, in particular at the advanced level. The main objective of this study was to analyse a computerized learner corpus of written texts in order to identify the most common errors made by students at the intermediate level. In addition, in light of the results of the study specific measures are suggested to improve students' results in future writing tests. Recent research concerning errors and learner corpora can be found in: Muehleisen, 2006; Meunier, 2007; O'Donnell et al. 2009 or Granger, 2009.*

*Keywords: writing, EFL, learner corpus, error analysis*

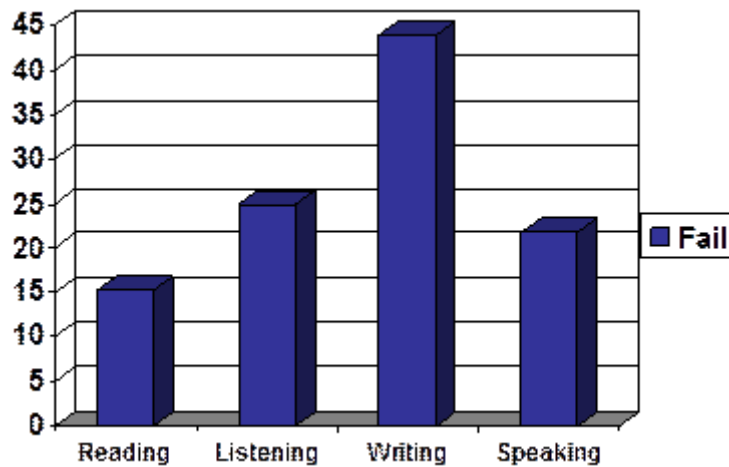
### INTRODUCTION

The teaching of writing in English to adult students involves facing a number of challenges. Some of these challenges are common to any aspect of foreign language teaching, namely: a large number of students in class, time restrictions, learner differences in terms of aptitude and motivation, and in general, a lack of experience and (or) practice of language use in the L1. In addition, there are less obvious challenges, which make the teaching of writing different from other skills. On the one hand, the necessary investment of time to develop good writing habits is sometimes considered too high, since adult learners often need to use the language immediately. Moreover, the teacher faces the need to make writing meaningful, so that learners can perceive it as a means of accomplishing their language goals. These different needs and challenges of adult language learning are important reasons why research that focuses on this large group of learners is necessary.

In the Spanish education system, specialized language teaching is organized and implemented by State-run Language Schools (Escuelas Oficiales de Idiomas, henceforth EOI) following the levels recommended by the Council of Europe within the *Common European Framework of Reference (CEFR)*. According to Spanish education law (BOE, 2006), the teaching of foreign languages should prepare students (outside the regular grades of the educational system) to use the foreign language adequately. Moreover, the law establishes three different levels: basic, intermediate and advanced, which correspond to the A2, B1 and B2 levels described by the CEFR.

It has been noted by a number of teachers (in personal communication), that the transition from the intermediate level to the advanced level is a difficult one for these language learners, who tend to be adults and have entered the workforce after completing their studies, some at the secondary level and others at university. Although a good number of students pass the certificate exam to obtain an Intermediate level diploma (B1) only a few are able to pass the advanced level exam after two more years of classes (240 teaching hours). All exams leading to certification from the EOI assess students' skills in reading, writing, listening and speaking.

One of the most problematic areas seem to be the writing component. At times, students get good results in other parts of the test but fail the writing part of the test. To illustrate the problem, table 1 shows the percentages of failure in the different skills among the students who took the Intermediate Level Certificate exam in June of 2011 at the EOI of Carabanchel.



**Table 1. Intermediate Certificate results. EOI Carabanchel, June 2012**

Contrary to what many of the students believe, it is not the listening or speaking part of the test that are the biggest challenge. The percentage of students who fail the writing part of the test is much higher than any of the other skills. This trend is even more noticeable when we look at the results of the Advanced Certificate (table 2). It should be noted that the percentages in these two tables include only those students who took the test and not the students who were registered in the class but who did not take part in the exam.

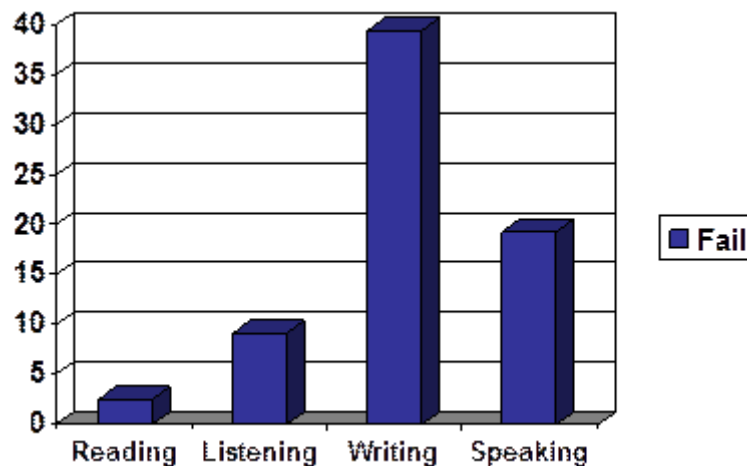


Table 2. Advanced Certificate results. EOI Carabanchel, June 2012

What makes writing different from other language skills is that it is school-centered and school-driven. The number of writing tasks that the average citizen might need to undertake is highly dependent on his or her profession. In most cases writing is reduced significantly after an individual finishes his or her school years. Since the EOI caters to adults, most students are in their thirties and no longer studying. Their experience of writing and their practice outside the classroom can vary greatly. For these reasons, it is important not only that the teaching of writing becomes a significant part of the teaching scheme of the two school years leading to the Advanced level certificate, but also that teachers know what the most problematic areas for the learners are.

## 1. OBJECTIVES

The idea behind the project was to analyse a learner corpus of written texts in order to identify the most common errors at the Intermediate level. The main purpose was to identify patterns of errors that could help teachers target the problem areas and find ways to improve learner's writing skills so that they can be successful in future writing tests. The Intermediate level was chosen for two reasons. Firstly, it is at this level that the learners are developing their writing beyond the sentence, which allows the researcher to access more varied data. Secondly, there is still time to intervene in order to help learners improve their writing.

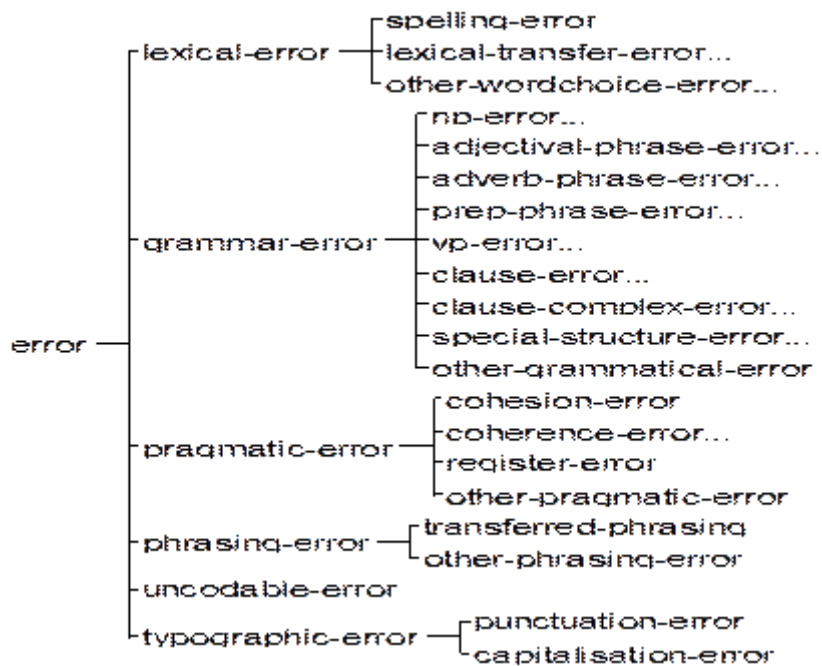
## 2. METHODOLOGY

The study focused on the writings of Spanish students of English as a foreign language enrolled at the Escuela Oficial de Idiomas (EOI, "Official School of Languages") in Puente de Vallecas, an extension of the EOI in Moratalaz, Madrid. For the purpose of the study, 18 students were selected. Those 18 students were all enrolled in the same Intermediate level 2 class and received 4 ½ hours of English instruction per week from the same teacher during the academic year. The majority of the students had been enrolled in the school the three previous academic years and only one person was new to the school. The age of the participants ranged from 18 to 41 years. There were 13 women and 5 men. All of the participants were native speakers of Spanish, 8 of them had completed university studies while 3 were enrolled at university at the time the data was collected. All of the participants had completed their compulsory secondary education as it is a requirement to enter the school. Only one of the participants had stopped her education after completion of compulsory education. The class was very representative of the type of student that enrolls at the EOIs

The data collected for this project included two tasks written for their final Certificate Exam, which took place in June 2011. The texts collected were very suitable for this project since they represented good examples of student writing: they were written according to a specific set of instructions, with no external help or other resources and with a time limit. The exam gave the students the chance to try their best writing and the teacher the chance to explore their strengths and weaknesses.

Once the exam was graded and the course was completed, each of the written texts from the exam had to be typed and saved as a plain text document in order to compile a small learner corpus. For the analysis of the data, UAM Corpus Tool was used. UAM Corpus Tool is a software programme designed and developed by linguist and computer expert Mick O'Donnell in 2008. This software allows the researcher to manually select and code errors following a pre-designed error taxonomy or creating a new taxonomy. Corpus Tool is currently being used by researchers in the TREACLE project (O'Donnell, et al. 2009)

The learner corpus compiled for this project contained a total of 36 texts and 5,908 words. For the error coding process, the text files were incorporated into the corpus and each error was manually selected and assigned a category from the available features of the taxonomy. The TREACLE taxonomy is easy to use, since it moves from very general error types to very specific ones. The following is an illustration of the taxonomy showing the first two subdivisions of the different categories. The ellipses indicate further subdivisions:



### 3. RESULTS

The first step in analyzing the data was to perform a general count of error types across all texts in the corpus, including the total number of errors and the errors per main category: lexical, grammatical, pragmatic, phrasing or typographic. Table 3 shows these percentages. Grammar errors are the largest sub-group of errors, making up 39.2% of all errors in the corpus, followed by lexical errors (33.4%) and the other types with lower amounts.

**Total number of errors = 730**

Percentage

Grammar 39.3%

Lexical 33.4%

Pragmatic 11.0%

Phrasing 7.9%

Typographic 7.9%

Uncodable error 0.4%

Table 3. Percentages of the different type of errors.

The most common grammatical errors were NP errors and PP errors. Within NP, errors in the use or choice of determiner accounted for 21.6% of the total, being by far the most common NP error. On the other hand within PP errors, the choice of preposition was the most common problem.

**NP errors**Percentage

Determiner-error 21.6%

Head-error 10.5%

Premodifier-error 2.8%

**PP errors**Percentage

Prep. choice error 17.8%

Missing preposition 4.9%

Complement error 2.8%

Table 4. Percentages of the different type of NP and PP errors.

Within the different types of lexical errors, we find over half of them are spelling errors (57%), followed by errors choosing specific vocabulary (36.1%), with the choice of nouns (12.7%) and verbs (14.3%) being the most problematic.

A comparison was done between the errors found among the students who passed the test and those who did not. Although the results could not be conclusive due to the small number of participants, two tendencies were clear. First, the number of errors per 1000 words was found to be higher among the students who did not pass, which was expected. Secondly, students who did pass had a higher percentage of verb-vocabulary errors, and prepositional-phrase errors (lexical and grammar errors). Students at a lower level seemed to be making use of structures found in Spanish and translating those forms into English, as they had a higher number of transfer errors. Those students seemed also less aware of punctuation rules, and had a high percentage of punctuation errors overall. Nevertheless, one conclusion that could be drawn from the data was that the areas that learners are having most problems with are lexis and grammar, and that this fact becomes more evident as the student progresses.

---

## 5. SUMMARY OF FINDINGS

---

The most conclusive finding was that grammatical errors were the most common for all groups, and within grammar error there were clearly two problematic areas: determiner error and errors in the use of prepositions. From a teacher's point of view both of these areas would need to be reinforced and they should be addressed. In fact, 21.7% of all grammar errors were related to students misusing determiners. To be more precise, 13.6% of grammar errors involved including a determiner where none is required or not using one when required. On the other hand, 17.8% of grammar errors involved the use of an incorrect preposition in a prepositional phrase. Finally, 57% of all lexical errors were spelling errors.

---

## 6. RECOMMENDATIONS

---

In view of the findings summarized above, the following recommendations would be useful for those who teach Advanced level English at the EOI:

- Provide more exposure to the use of determiners in English, either through explicit teaching of the rules, or through some non-explicit means such as drilling.

- Try to focus student's attention on prepositions used in the right context.
- Give enough opportunity to the students to practice spelling.
- Increase noun and verb vocabulary acquisition.

---

## 7. CONCLUSION

---

Finding the areas that students at the Intermediate level need to improve can help focus teaching time on those specific issues, in order to maximise the limited class time that we can dedicate to the teaching and practice of writing.

Despite the unquestionable validity of the communicative method, the fact is that more and more students and teachers are realizing that successful communication at advanced levels requires a level of accuracy that students find difficult to achieve. The empirical evidence presented in this project demonstrates that the students' grammatical accuracy needs improvement. For that reason, including more grammar teaching within the communicative method, in the line suggested by Long (1990), Larsen-Freeman and Long (1991) or Meunier (2007, 2010) would seem a better approach for the future.

## REFERENCIAS BIBLIOGRÁFICAS

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Cumming, A. (2001). "Learning to write in a second language: two decades of research." *International Journal of English Studies*, 1.2: 1-23.

<http://revistas.um.es/ijes/article/view/48331>.

Granger S. (2009). "The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation", In Aijmer, K. ed(s). *Corpora and Language Teaching*, 13-32. Amsterdam & Philadelphia, Benjamins.

Larsen-Freeman, D. and Long, M. (1991). *An Introduction to Second Language Acquisition Research*. London: Longman.

[Ley Orgánica 2/2006, de 3 de mayo, de Educación](#). BOE nº 106 de 4 de mayo de 2006.

Long, M. H. (1990). "Maturational constraints on language development". *Studies in Second Language Acquisition* 12, 3, 251-85.

Meunier, F. (2007). "The pedagogical value of native and learner corpora in EFL grammar teaching". In :Teubert W. and R. Krishnamurthy, *Corpus Linguistics: Critical Concepts in Linguistics*, London & New York : Routledge, 119-141.

Meunier, F. (2010). "Learner Corpora and English Language Teaching: Checkup Time". *Anglistik: International Journal of English Studies*, v. 21, n. 1, 209-220.

Muehleisen V. (2006). "Introducing the SILS Learners' Corpus: A Tool for Writing Curriculum Development". *Waseda Global Forum*, No. 3. 119-125.

O'Donnell, M., S. Murcia, R. García, C. Molina, P. Rollinson, P. MacDonald, K. Stuart, and M. Boquera. (2009). *Exploring the proficiency of English learners: The TREACLE project*. Proceedings of the Fifth Corpus Linguistics, Liverpool.