

INDEX

1. Introduction	4
2. Theoretical-methodological framework	5
3. Corpus linguistics and the learners' corpus	10
4. Error analysis in CEDEL2	15
4.1 Lexical and semantical errors	16
4.1.1 Lexical errors	16
4.1.1.1 Loanwords	17
4.1.1.2 Lexical creations	18
4.1.2 Semantical errors	19
4.1.2.1 Verb <i>to be</i>	19
4.1.2.2 False cognates	20
4.1.2.3 Collocations	21
4.2 Morphosyntactic errors	22
4.2.1 Determiners	23
4.2.2 Pronouns	25
4.2.3 Nouns	25
4.2.4 Adjectives	26
4.2.5 Adverbs	27

4.2.6	Verbs	28
4.2.7	Prepositions.....	29
4.2.8	Conjunctions	30
4.3	Orthographical errors.....	31
4.3.1	Diacritical marks	31
4.3.2	Capital letters	32
4.3.3	Other significant errors	32
5.	Conclusion.....	34
6.	Bibliography	36

ABSTRACT

In this dissertation we present a description about certain linguistic aspects those Anglophone speakers who are learning Spanish as a second language might have. In order to accomplish this, we have used the *Corpus Escrito del Español* (CEDEL2) as a starting point. After a very brief introduction, in which we highlight the importance of English as a universal language, we will be explaining our theoretical-methodological framework in which it is based the fulfilment of this project. Then, we will deal with the main part of our thesis: applied linguistics and the error analysis. Eventually, we will expose our conclusions and we will recapitulate the meaningful aspects of the main part of our dissertation.

1. INTRODUCTION

During the last few decades, we could appreciate how English has become one of the most spoken languages of all times. Nowadays, it is considered the universal language par excellence and it is the second language with more speakers in the world. English is used for many purposes: education, social media, computing, and, of course, leisure time. Because of this reason, it is quite difficult to question its usefulness and the necessity of learning it.

In the past few years, English learning has developed a huge significance and there are a lot of reasons why a person would want to learn it: for travelling, for having access to information on a specific field, for working etc. It is not unusual in these days that any company would entail their workers to know several languages, and English is included among them.

This increasing interest for English learning requires the use of different and varied specific components and provokes the appearance of dictionaries and other pedagogical materials. Due to this demand of educational and specific materials, corpus linguistics starts to gain relevance, considering, by means of the usage of corpora, researchers can find real examples on the use of the language in different contexts and in that way improve its investigation. The active use of this corpora allows for creating new teaching and learning methods, and thus improve the apprentice experience and ease the acquisition of the language they want to learn.

In current years, everything related to teaching and learning foreign languages gains importance and it is patently obvious in the relevance that the CEFL¹ gives to the plurilingüism. According to this institution, plurilingüism is the simultaneous presence

¹ *Common European Framework for Languages*

of two or more languages in the communicative competence of an individual and the interrelation and interaction that exists among them. Due to these facts, we thought that it was an interesting topic to cover in our dissertation and thus we could investigate about different methodologies used for the teaching-learning process. Also, we could analyse the difficulties of the students upon learning a foreign language. In the case of this dissertation, we will investigate and analyse the problems that the Anglophone students may have when they want to learn Spanish.

Our thesis begins with a description of the theoretical-methodological framework in which it is encompassed. In this part, we will contextualize our thesis, describing the linguistic branch in which it is framed, focusing on the approach we have chosen for our dissertation. In the next section, we will include a brief explanation about the corpus linguistics as introduction and we will continue with the exposition of the Spanish learners' corpora, contending why we have chosen those corpora. In this case, we have based our project in one corpus: *Corpus Escrito Del Español L2* (CEDEL2). After this chapter, we will continue with the heart of our thesis, the analysis of the problems and errors of the students upon learning a foreign language, in this case Spanish. Regarding the errors that we will analyse, we wanted to focus in those made by intermediate and advanced students of Spanish and carry out an exhaustive and well-structured analysis. In order to finish with this dissertation, we will add a brief conclusion in which we will recapitulate all the results we have obtained from our investigation.

2. THEORETICAL-METHODOLOGICAL FRAMEWORK

In this part of our dissertation, we will describe the theoretical-methodological framework we will use throughout the thesis. In order to do this, we will begin with the contextualization of our investigation. First, we will explain, briefly, the discipline in

which we can include our investigation and also the field in which it is applied. The linguistic discipline in which we can frame our thesis is applied linguistics and the field in which it is centred is the teaching of second languages. Straightaway, we will centre our attention in explaining our investigation through the methodology we have chosen, the error analysis, and in that way give solutions to the possible problems that Anglophone students may have when learning Spanish as second language.

2.1. Applied linguistics and the teaching of second languages

In order to contextualize this dissertation, we must define correctly the branch in which our study is inscribed, applied linguistics in this case. According to the *Longman Dictionary of Applied Linguistics*, applied linguistics has three different definitions:

It is used (1) to refer to the study of language teaching and learning; (2) to refer to the applications of language study in any area of practical concern; (3) to refer to the applications of the findings of theoretical linguistics. (1985: VI).

The definition that we will use in our thesis is the first one. Applied linguistics is divided in several areas, and one of these is the learning of second languages. This area is the one that will take up the central part of our project and this same dictionary defines it as “the process by which people develop proficiency in a second or foreign language” (1985: 252). In order to understand better this definition we must make a distinction between second language and foreign language. A second language is the one which is learnt in a country in which it is official or coofficial and coexist with other languages; meanwhile, a foreign language is the one which is learnt in a country in which this language does not have this status of officiality and coexistence.

Our analysis will deal with the errors that some Anglophone students make while they are learning Spanish and, because of this, we must clearly define the methodology employed for the teaching of a second language, something that will help

us describe the methodology we will use in our thesis. According to Pastor Cesteros (2004), this methodology can be defined dependant on three different aspects: the approach, the method and the technique.

The approach is a combination of theories that deal with the nature of the language and its learning process; the method is the manner of teaching a language based in certain systematic principles that imply the application of several stablished criteria about how to learn a language in a better way; the technique is the type of activity or activities that each method employs (Pastor Cesteros 2004: 133).

Due to this part of the project centres in the methodology, we are interested in explaining meticulously the aspect of the method. According to Sánchez Pérez, we can mention three different types of methods: traditional, direct and audio-oral (2008 [2004]: 679). The most used methods during the last decades are the direct and the audio-oral, despite this last one enjoys great proliferation since World War II. The audio-oral method structures the language in three levels: phonological, morphological and syntactical (2008 [2004]: 680). Additionally, it is centred in written language, and grammatical rules and semantics play a secondary role on it. Included in this method a distinction between Mother Tongue, or L1, and Second language, or L2, is made. This contrast is used to identify the differences and the similarities between L1 and L2, and also to detect and to try to correct the future errors caused by the influence of the L1 while an L2 is being learnt.

Another method used since 1970s was the “Contrastive Analysis Hypothesis”, whose purpose was to investigate about the possible interferences between the first language and the acquisition of the second language (Appel et al., 1996: 125-126). This hypothesis developed a series of difficulties that were directly related with the

interferences between L1 and L2. These difficulties were considered as errors and both words almost became synonyms.

However, as a result of the continuous use of this method, some exceptions in which the employment of this method was not correct appeared to offer a solution for the errors that were generated. One of these exceptions was the great similarity between the L1 and the L2, something that provoked that the learners of the L2 were frequently confused or, on the contrary, would acquire more complex structures in an easier way. Therefore, it remained demonstrated that these errors were not caused only for the existent interferences between L1 and L2. Due to this, another mode to focus these errors appeared, what is called nowadays *Error Analysis*.

2.2 The Error analysis as a new approach

This new approach began to be used after it was demonstrated that the “contrastive analysis” could not respond to several errors that were made by second language learners. According to the *Longman Dictionary of Applied Linguistics*, the *error analysis* can be defined as “the study and analysis of the errors made by second and foreign language learners” (1985: 96). The linguistic Beardsmore affirms that this method “is based on the examination of production by bilinguals of one kind or another with an attempt at explanation of the errors” (1991: 86), definition that is more centred in a bilingualistic vision.

The error analysis is focused on learners’ real productions and takes them as a starting point in order to create a series of steps to follow an adequate and exhaustive analysis. As S. P. Corder says, the steps we must follow are the next ones:

- a) Identification of the errors in their context;
- b) classification and description of them;

- c) explanation of their origin, searching its psycholinguistic mechanisms and strategies and the sources of each error: at this point enters the possible interference of the mother tongue, as another strategy;
- d) evaluation of the gravity of the error and the research of a possible treatment for it. (1967).

Since this conception of the approach, errors began to be treated as something normal and natural in the acquisition of a second language, and, according to the *Diccionario de Términos Clave de ELE*², even began to be considered as a learning strategy, because they constituted an unavoidable step for the appropriation of the new language that the learner experiences during the learning process.

Directly related to this last statement, appears the concept of *interlanguage*. The *interlanguage* is “the type of language produced by second –and foreign– language learners who are in the process of learning a language” (*Longman Dictionary of Applied Linguistics*, 1985: 145-146). This system has concrete characteristics that make it unique for each learner. It is an individual system regulated by its own coherent and variable rules, and which can be easily restructured and evolves following the different stages through which the learner goes on its learning. Due to the quick evolution and the several variations, the interlanguage can cause a phenomenon called *fossilization*, which is “a process which sometimes occurs in which incorrect linguistic features become a permanent part of the way a person speaks or writes a language” (*Longman Dictionary of Applied Linguistics*, 1985: 111). This process does not allow the learner to acquire the same competences as the native speakers.

As we have previously mentioned, the error analysis was born to substitute the contrastive analysis, which could not give solutions to the exceptions in the learning of second languages. With this new approach, errors are taken as something logical and inevitable for learners, which serve to generate certain situations and to carry out the

² From now on *CVC Dictionary*

subsequent tasks with the purpose of improving the learning of the learner as well as improving its interlanguage. Therefore, we believe that it is adequate to base our project in this method, in order to make a detailed analysis of the errors that lay out in the acquisition of a second language.

3. CORPUS LINGUISTICS AND THE LEARNERS CORPUS

3.1 Corpus linguistics

In this section of our dissertation, we will define what is a *corpus* and the corpus linguistics and its fields of application.

As what happens with other concepts related with the vast field of linguistics, corpus linguistics has created a great controversy referred to its definition. Nowadays, there are many descriptions of it, but none of them achieved a complete definition by itself. The corpus linguistics are defined as a branch of linguistics that base its investigations in data obtained from the corpus of a language, that is, real samples of the usage of that language (*CVC Dictionary*).

As we can observe, it is very complicated to correctly define the field of the corpus linguistics, but throughout several years, series of landmarks which served to develop and consolidate this branch of linguistics such as we understand it nowadays appeared. By the hand of this definition, we also have to explain what we know as *corpus*. According to the *Dictionary of Grammatical Terms in Linguistics*, a *corpus* is defined as “a body of linguistic data from a particular language, in the form of recorded utterances or written texts, which is available for analysis” (1993: 65). The first corpus, those that existed until XIX Century, were destined exclusively for the study of dead languages and they were centred in written texts. This manner of working was preserved until middle XX Century, when it appeared the American Structuralist School (Boas,

Sapir, Bloomfield etc.) that settled the bases of what we will call later corpus linguistics. Since this moment, the manner of working changed radically and the corpus were considered as the only appropriate and valid method for studying any language. Those structuralist corpus, unlike the first ones, were based in the study of oral language and they worked with phonetical-phonological analysis (Villayandre, 2008: 331). These types of corpus were broadly criticized, but the most relevant critiques were those established by Chomsky and Abercrombie. These critiques lead to the two aspects of corpus linguistics we know nowadays: the theoretical aspect and the practical aspect (Villayandre, 2008: 332). The theoretical aspect, accomplished by Chomsky, can be summarized in the following table:

First corpus linguistics	Chomsky's corpus linguistics
Centred on phonetics and phonology	Centred on sintaxis
Finite language	Non-finite language
Corpus as the unique explanation	Intuition as the unique explanation
Complete corpus	Partial corpus

Table 1. First corpus linguistics vs. Chomsky (Villayandre, 2008: 333)

For his part, Abercrombie, centred his research in the practical aspect, outlined by treating the corpus as “pseudo-techniques”, without distancing too much from Chomsky's vision.

As we could prove, the usage of the corpus for language learning purposes is, nowadays, indubitable above all since the appearance of the computer, fact that improved considerably all the aspects of corpus linguistics. However, corpus can be used also with other purposes such as the creation of dictionaries for learners of a second language (*COBUILD*, *CORDE* and *CREA*), the design of applications to facilitate translation, or even as different materials for the creation and development of the speech technologies.

3.2 Learner's corpus

After we commented on the definitions and applications of a corpus, in this part we will focus on defining what is a learner's corpus and we will describe briefly the most important ones, written in different languages: Spanish, Italian and English. We can define the learner's corpus as a "group of texts produced by people who are learning a concrete language with different mother tongues and several knowledge degrees of the language they want to learn" (*CVC Dictionary*). The first one we will mention is the *Japanese English as a Foreign Language Learner corpus (JEFL)*. This corpus is a collection of seven hundred-thousand words. It is formed by nearly ten thousand essays of different subjects made by Japanese-English speakers from beginners to intermediate learners. It is composed in its 100% of written texts and it is considered an open corpus because it is still in production. It was created by Dr. Yukio Tono and his students, from which he took the compositions after several 20-minute classes. It is destined for investigating written natural language in use and the spontaneous contributions of the students on its different learning steps (vid. <http://jefll.corpuscobo.net/#overview>).

Another corpus that we must mention is the *Varietà Lingua Italiana Corpus Online (VALICO)*. It was created in 2003 but it is still in production. It is also composed by written texts on its 100%. It was created by Manuel Barbera, Carla Marengo, Elisa Corino and a group of students. Its purposes are to show how students of different age and mother tongue write in Italian, to become a POS- and error-tagged corpus and to provide Italian Linguistics new insights into variation and acquisition (vid. http://www.valico.org/valico_en.html).

The last corpus that we will mention is the *Cambridge Learner Corpus* which is considered the biggest corpus in the world. It is composed by written texts made by English students from different countries all over the world. It was made by Cambridge University Press and Cambridge English Language Assessment and its purpose is to

compile universal and particular errors caused by the influence of the L1 and also to research the difficulties of the students learning English (vid. <http://www.cambridge.org>).

3.3 Learners corpus of Spanish as a second language

After accounting for some of the world's most important learner's corpus in several languages, we must focus on explaining the corpus used for learners of Spanish as a second language and also the one that we will use for the realization of this project.

The first one we will describe is the *Corpus de Aprendices de Español como Lengua Extranjera* (CAES), which is defined as a group of written texts produced by students of Spanish with different degree of linguistic knowledge (A1 to C1 of the CEFL applied in the CICP³) and that came from six L1: Arabic, Chinese, French, English, Portuguese and Russian. This corpus enjoys multiple applications, but the most important one is to provide the professionals dedicated to investigate on the Spanish as a Foreign Language field with objective and reliable data. Its data was collected between 2011 and 2013, and its development can be chargeable to some researchers of the University of Santiago with the funding of the Instituto Cervantes (vid. <http://galvan.usc.es/caes>).

The second corpus we will talk about is the *Spanish Learner Language Oral Corpora* (SPLLOC), which is divided into two parts: *SPLLOC 1* and *SPLLOC 2*. The main application of this corpus is to promote investigations in the acquisition of Spanish as a second language, but each one of its divisions is used for specific purposes. The *SPLLOC 1* was built between 2006 and 2008, and it is entirely composed by oral texts produced by Spanish as L2 learners in their different stages. Its main objective is to

³ Cervantes Institute Curricular Plan

investigate the acquisition of Spanish syntax and morphology, adding examples of Spanish native speakers to account for errors and interferences in an easier way. The *SPLLOC 2*, the second subproject in which this corpus is divided, is still in construction and its main applications will be centred on the investigation of the acquisition of time and aspect in Spanish verbs. The data is oral and it began to be recollected in 2008 thanks to a group of sixty participants, native Spanish students and Anglophone learners of Spanish as an L2.

The last corpus we will comment on is the *Corpus escrito del español L2* (*CEDEL2*), a project that is still in construction but it already has seven hundred and thirty thousand words, and as its own name indicates, it is 100% formed by written texts. (Lozano, 2009: 197-212). It was created by the researcher Cristóbal Lozano and other researchers of the UAM (Universidad Autónoma de Madrid) and the participants are more than two thousand students from all over the world, all of them learners of Spanish as a second language. These students are divided into native Spanish users, which represent the 28% of the participants, and learners of Spanish as a second language, the 72% of the participants. Its application is basically to provide data about the errors of the learners of Spanish as a second language and thus researchers could investigate about didactical methods for the teaching of second languages (vid. <https://www.uam.es/proyectosinv/woslac/collaborating.htm>).

Finally, and after the descriptions of the different corpus, we believe that we must contend the election of the corpus we will use for this project, the *CEDEL2*. The *Corpus escrito del español L2* is the more adequate election for the fulfilment of our investigation, because of its written form. It is composed by seven hundred and thirty thousand words, which will allow us to choose between a huge variety of errors. We also decided to base our project in this corpus because the L1 of all the participants is

English, despite there are data of native-Spanish speakers to compare. The diversity of the types of texts generated from the tasks that were proposed to the learners, allow us to obtain a wide vision of the interlanguage corresponding to the different levels, and a sufficiently complete vision of the error that were committed. This fact, will provoke an introduction of certain specific vocabulary related with the several themes that were suggested to the participants, something that will allow us to find a huge assortment of errors and choose the most suitable ones. For all these, we decided to focus in the usage of this corpus for our error investigation.

4 ERROR ANALYSIS IN CEDEL2

This section of our thesis will consist of showing the results of our investigation about the errors of the learners of Spanish as L2. We will divide the errors on its different classes and subclasses. We will be editing examples of the errors we could find in an organized way and following the order established by the types of errors we will be dealing with.

Previously in this dissertation, we have talked about CEDEL2, the corpus we will use to accomplish our investigation. The data of this corpus were written texts produced by Anglophone students from different ages which have different competences on their learning of Spanish as L2. In this thesis we will analyse the errors made by intermediate to advanced students of Spanish as L2. Thus, we will have more examples to introduce in our investigation.

Before we can begin with the error analysis, we must make a general classification of them. In this case, we have chosen the classification elaborated in the description of the term *error* in the *CVC Dictionary*, simply because it is very simple, but accounts for the fundamental types of errors made by the students at the same time. The *CVC*

Dictionary presents the most common classifications, in which it keeps in mind the following criteria:

- 1) Linguistic criteria: lexical and morphosyntactic errors, discursive errors etc.
- 2) Superficial strategies criteria: omission, addition, juxtaposition, wrong election and wrong positioning.
- 3) Pedagogical criteria: induced and transitory errors, fossilized errors etc.
- 4) Etiologic criteria (according to its cause or origin): interlingual and intralingual errors, simplification errors etc.
- 5) Communicative criteria: ambiguity errors and irritating errors among others.

In order to fulfil our investigation, we have decided to focus on the linguistic errors, because we consider it more appropriated for the nature of our dissertation. As we said before, the themes of the compositions will provoke the introduction of specific vocabulary and different syntactical structures, a fact that will favour the diversity of the errors and the ease of finding them. Due to this, our analysis will deal with the lexical, semantical and morphological errors which we will classify in subtypes as we make progresses in this project.

4.1 Lexical and semantical errors

In this section, as its own name implies, we will account for the lexical errors of the learners and the meaning of the words and expressions they use. This part will be divided in two subparts that include, in one hand, lexical errors and on the other hand, semantic errors.

4.1.1 Lexical errors

In this subpart, we will focus on explaining briefly the most common types of lexical errors and we will provide examples for each one of them. The two main types of lexical errors we will be dealing with are: loanwords and neologisms.

4.1.1.1 Loanwords

Before we could analyse and exemplify the *loanwords* we were able to find, we must clarify what is considered a loanword. According to the *Longman Dictionary of Applied Linguistics*, a loanword is “a word or a phrase which has been taken from one language and used in another language” (1985: 30). To this definition, we can add that this word or phrase could have been adapted from its original form (adapted loanword), or they could be simply an imitation of the original term (non-adapted loanword). The CEDEL2 participants have very different ages and levels and because of this it is easier to find lexical loanwords in it.

The non-adapted loanwords are those that have been imported from a language but did not suffer any substitution; that is, they are words that have been taken from a language and that have not been adapted to the recipient language. We were able to find examples of these borrowings in all the levels we have analysed, from lower intermediate to advanced level. One of these cases is the following one (we present the Spanish form in brackets):

(1) *Un “condo” [apartamento] (MAH, advanced, corpus CEDEL2).

This example caught our attention because we would not expect that an advanced learner would say it due to the differences with the Spanish word.

It is not the only example of non-adapted loanwords we found, though. We were also able to find the following ones:

- (2) **Era incredible [increíble]* (NER, intermediate, corpus CEDEL2).
- (3) **Ley de anti-tobaco [tabaco]* (gjj, advanced, corpus CEDEL2).
- (4) **Gente innocente [inocente]* (MV, intermediate, corpus CEDEL2).
- (5) **Recentalmente [recientemente]*, *La muestro de Oprah* (ARS, lower intermediate, corpus CEDEL2).

We also found examples of adapted loanwords, something normal that happens to learners in their different stages of learning. The first example we found is the following:

- (6) **Es la profesión [profesión] que quiero* (KJ, intermediate, corpus CEDEL2).

This example is very peculiar because, despite there not being in English a diacritical mark, the learner, in order to adapt the word, writes it following the Spanish accent mark rules.

We were also able to find the following examples of adapted loanwords:

- (7) **Estoy attendiendo [asistiendo] la Universidad de Georgia State* (KJ, intermediate, corpus CEDEL2).

In this example, we can see that, except for the spelling, we obtain a Spanish word, *atendiendo*, while in the next example, the situation changes:

- (8) **Una pictura [imagen]* (KSK, lower intermediate, corpus CEDEL2).

In this case, the word that concerns us, *pictura*, is not an Spanish word; it is an inexistent word in the Spanish vocabulary.

4.1.1.2 Lexical creations

In this section we will account for the words that do not correspond to any Spanish word or that they have a different meaning for the learners, and even for those words that are created for the learners when they try to imitate the Spanish word. These types of creations were very difficult to find and because of this, we will only mention two of them. The first one is the following:

(9) **Muy pecano* [*pequeño*] (NER, lower intermediate, corpus CEDEL2).

In this example, the learner tries to say “*el pueblo es pequeño*” using the word *pecano* instead of *pequeño*.

The second example we found is this one:

(10) **Un llamo* [*llamada*] (KSK, lower intermediate, corpus CEDEL2).

As we can see, the learner tries to explain that it received a call (*una llamada*), but it uses the word *llamo* instead of *llamada*, which leads us to think that a new word is created.

4.1.2 *Semantic errors*

In this part of our dissertation, we will account for the errors that deal with the meaning of words and expressions, false friends and the verb to be, for example.

4.1.2.1 Verb *to be*

We thought that it would be very interesting to treat these errors, because in English the verb *to be* is used for the Spanish verbs *ser* and *estar*, and this is a very frequent problem that Spanish learners have. The examples we found are the following ones:

(11) **El agua es* [*está*] *a la temperatura perfecta* (KSK, lower intermediate, corpus CEDEL2).

(12) * *Yo vivo in Atlanta, Georgia. Esta [es] muy bonito alli.* (CEP, intermediate, corpus CEDEL2)

(13) * *Pienso que él será [estará] muy orgulloso de Carlito* (MAH, Advanced, corpus CEDEL2).

(14) * *Era [estaba] alli con tres amigo* (NER, Intermediate, corpus CEDEL2).

(15) * *Sólo quieran ser [estar] casados para recibir los beneficios del gobierno* (EMO, lower intermediate, corpus CEDEL2).

(16) * *Necesitamos estar [ser] una buen influencia para ellos* (NW, lower intermediate, corpus CEDEL2).

4.1.2.2 False cognates

Within the semantic errors, we must highlight the importance of one type of errors called false cognates. A false cognate, also called false friend, is “a word which has the same or very similar form in two languages, but which has a different meaning in each”. (*Longman Dictionary of Applied Linguistics* 1985: 103). This might be a problem for those who are learning Spanish, because they tend to use the word wrongly. The examples we found about this errors are the following:

(17) * *Muchas veces los farmaceuticos [medicamentos] no funcionan o tienen malos efectos* (JAH, intermediate, corpus CEDEL2).

In this example, we were able to find that the learner made an association between the English word *farmaceuticals* and the Spanish word *medicamentos*. Instead

of using *medicamentos*, the learner decides to use *farmaceuticos* because of the formal similarity, despite *farmaceuticos* in Spanish has a different meaning.

(18) **Voy a attender [asistir] la escuela para mi masters en Educación* (KJ, intermediate, corpus CEDEL2).

(19) **Estoy attendiendo [asistiendo] la Universidad de Georgia State* (KJ, intermediate, corpus CEDEL2).

In the previous examples, we can see that the learner makes an association between the English verb *to attend* and the Spanish verb *atender*. Instead of using the verb *asistir* the learner uses the verb *attender* because it is very similar to the English verb.

(20) **Comanda [pide] respeto de todos que le conocen* (BEO, intermediate, Corpus CEDEL2).

(21) **Quiero terminar el colegio y el graduado con grados [notas] buenos* (SC, lower intermediate, corpus CEDEL2).

(22) **Quiero moverme [mudarme] a Florida* (SC, lower intermediate, corpus CEDEL2).

4.1.2.3 Collocations

In this section we will find examples of the use of collocations, but first we have to define the term: a collocation is “the way in which words are used together” (Longman Dictionary of Applied Linguistics 1985: 46). The first example we found is the following:

(23) **Ultimamente me ha interesado regresar a [volver a] leer* (LP, advanced, corpus CEDEL2).

In this case, we were able to find that the learner tried to use the English collocation *return to* to make a reference to the Spanish collocation *volver a*, because of the influence of their mother tongue.

We also found more examples of this type of error:

(24) **Voy a atender [asistir a] la escuela para mi masters en Educación* (KJ, intermediate, corpus CEDEL2).

(25) **Embarcamos en un autobus [subimos/montamos]* (JPG, intermediate, corpus CEDEL2).

(26) **Mi esposa y yo, nunca hacemos la comunicacion [nos comunicamos] escribiendo* (NER, intermediate, corpus CEDEL2).

(27) **Comemos en un bar y miramos [vimos] un partido de fútbol* (DMH, advanced, corpus CEDEL2).

4.2 Morphosyntactical errors

In this part of the thesis, we will be dealing with the most characteristic examples of morphosyntactical errors. In order to do this in an organized way, we will subdivide this section into eight different subgroups: determinants, pronouns, nouns, adverbs, verbs, prepositions and conjunctions. In each subgroup, we will account for the examples of this errors and we will classify them according to Vázquez's criteria (1998: 28). Before we can expose the cases we found, we must explain the criteria we will

follow to categorize them. According to Vázquez (1999: 30-32), we can divide the processes into:

- Addition: the process of adding redundant morphemes or words;
- Omission: the process of suppressing of necessary morphemes or words;
- False collocation: the process of changing the word order in a sentence;
- False selection: the process of choosing incorrect words or morphemes in a concrete context.

4.2.1 *Determiners*

Regarding the determiners, our research was very profitable and we were able to find plenty of examples of several types. First, we will begin with the omission errors. A huge number of learners tend to avoid adding the article between the preposition and the noun in cases in which is mandatory in Spanish:

(29) **Eso es [lo] que quiero de mi futuro* (MAH, advanced, corpus CEDEL2).

(30) **La compañía para [la] que trabajara es muy pequeña* (BEO, intermediate, corpus CEDEL2).

(31) **Hay muchas problemas con [el] terrorismo en el mundo* (FA, lower intermediate, corpus CEDEL2).

We also found two examples of addition errors, which are the following:

(32) **Estoy attendingo la Universidad de Georgia State y estoy estudiendo para ser maestra para los niños* (KJ, intermediate, corpus CEDEL2).

(33) **Aunque no podía tomar la cerveza (hay que recordar que en los EEUU tarda 21 años para beber), todavía me lo pasó bien* (DMH, advanced, Corpus CEDEL2).

The process in which we found more examples is the false selection process. All the cases we found are related with the gender concordance, probably because Anglophone learners do not have gender marks in their L1. They tend to relate the different gender with a different ending, that is, they usually associate the ending in *-a* with feminine nouns and the ending in *-o* with masculine nouns. However, this is the normal and most basic procedure, but in Spanish there are a few exceptions that can get the learners confused, something we can exemplify in the following cases:

(34) **No me gusta mucho hacer cosas con el césped ni con el nieve [la nieve]* (MAH, advanced, corpus CEDEL2).

(35) **Esta pequeña [este pequeño] problema* (MAH, advanced, corpus CEDEL2).

(36) **Siempre quiere ir con el bici [la bici] a las montañas* (NER, intermediate, corpus CEDEL2).

(37) **Podemos ir con los bicis [las bicis] por cuatro horas* (NER, intermediate, corpus CEDEL2).

(38) **No hay una idioma [un idioma] que hablan en casa* (NER, intermediate, corpus CEDEL2).

(39) **Gracias por el oportunidad [la oportunidad]* (NER, intermediate, corpus CEDEL2).

(40) * *Voy a estudiar para enseñar los niños con una programa [un programa] de internship (KJ, intermediate, corpus CEDEL2).*

4.2.2 Pronouns

In our investigation of the errors in pronouns, we were only able to find three examples, two about false selection and one of addition. As we can see, the two cases of false selection of the pronoun are both examples of false selection of relative pronouns:

(41) **Vamos una vez a Valencia, que tenemos una amiga quien [que] vive allí* (NER, intermediate, corpus CEDEL2).

(42) **No debo de hablar desde la perspectiva mía [mi perspectiva]* (TTA, advanced, corpus CEDEL2).

(43) **Como en el caso mío [mi caso]* (TTA, advanced, corpus CEDEL2).

The example we found of addition of a pronoun is probably related to the influence of learners' L1 and is the following:

(44) **Me encanta estar en España y vivir en un país extranjero, pero les echo de menos a mi familia y a mi perra* (MET, lower intermediate, corpus CEDEL2).

4.2.3 Nouns

It is very difficult to find examples of each one of the processes Vázquez described, and in the case of the nouns we were only able to find cases of false selection, all of them related to gender concordance and with number concordance:

(45) **Tomar la transportación público [el transporte público]* (MAH, advanced, corpus CEDEL2).

(46) **Digo muerto porque no hay mucha vida, pero hay genta [gente]* (NER, intermediate, corpus CEDEL2).

(47) **Era alli con tres amigo [amigos] americanos* (NER, intermediate, corpus CEDEL2).

(48) **Continuamos a Miami, para visitar a unos colegos [colegas] de mi hermano* (DMH, advanced, corpus CEDEL2).

4.2.4 Adjectives

Our research of errors in adjectives was not as productive as we wanted, but we were able to find two errors related to the false collocation of them:

(49) **Es el mas precioso pueblo [pueblo más precioso] de los pirineos* (NER, lower intermediate, corpus CEDEL2).

(50) **Y evitar peligrosas cosas [cosas peligrosas] como drogas* (CD, intermediate, CEDEL2).

(51) **Tengo acceso a cosas importantes culturales [importantes cosas culturales] de los estados unidos* (AB, advanced, corpus CEDEL2).

This type of error is very peculiar because the position of the adjective in English is different than in Spanish. In Spanish, the adjective tends to be after the noun. Meanwhile, in English, the adjective tends to precede the noun. This fact can cause some problems for learners because there are also exceptions to this statement. We found other errors that deal with the absence of gender and number concordance between adjective and noun:

(52) **Después de esta pequeña problema [pequeño problema]* (MAH, advanced, corpus CEDEL2).

(53) **Hay muchas festivas [muchos festivas] durante el año* (AB, advanced, corpus CEDEL2).

(54) **Yo fui al jardín de rosas y era muy bonita [muy bonito]* (RD, advanced, corpus CEDEL2).

4.2.5 Adverbs

Our analysis of errors in adverbs was very difficult to do, because there are not many examples of them. Despite this, we were able to find two examples of false selection and one of lack of concordance. The examples of false selection are the following:

(55) **Pero después de esta pequeña problema, cuando nadie necesita mi ayuda jamás [nunca más], quiero ser voluntaria* (MAH, advanced, corpus CEDEL2).

In the next case, we can observe the wrong utilization of the adverbial phrase *por si acaso*:

(56) **Por acaso [si acaso], so debe de haber ningún ley que regula la vida gay.* (TTA, advanced, corpus CEDEL2).

(57) **Atlanta tiene lugares ricos incluyendo Lenox Mall y Buckhead, pero tampoco [también] tiene lugares baratos.* (JDM, intermediate, corpus CEDEL2).

As we have said previously, we were only able to find one example of the lack of number concordance in adverbs and it is the following:

(58) **He visto bastante películas [bastantes películas] de el y me encantan*

(JEL, advanced, corpus CEDEL2).

4.2.6 Verbs

In this section, we will analyse examples of the incorrect use of verbs. Most of the cases we found are related to the false selection of the verb mode and person. As an example of the first type we found the following cases:

(59) **Cuando regreso [regrese] a los EE.UU., quiero encontrar un trabajo*

(MAH, advanced, corpus CEDEL2).

(60) **Echo de menos mi familia, y no pienso que puedo [pueda] vivir muy lejos de ellos por mucho tiempo* (MAH, advanced, corpus CEDEL2).

(61) **Creo que cuando estoy [esté] terminando con mi doctorado, vamos a Espana a vivir* (NER, intermediate, corpus CEDEL2).

(62) **Espero que un verano podemos [podamos] ir a Madrid un otro vez* (NER, intermediate, corpus CEDEL2).

In the next pair of examples, we observe how the learner wrongly chooses the verb person:

(63) **Me gusta [gustan] ciudades como Madrid y Nueva York* (MAH, advanced, corpus CEDEL2).

(64) **Conozco algunas personas que no les gustan [gusta] fumar* (LH, lower intermediate, corpus CEDEL2).

In these two cases, we find a false selection of the verb that affect directly to the concordance of subject and verb. The most probable reason why this errors are made

resides in the L1, because English only have person marks for the third person singular, while Spanish tend to have marks for each person.

4.2.7 Prepositions

In the case of prepositions, we can find a huge amount of examples of errors. It is very common, even for native Spanish speakers, to sometimes select the incorrect preposition and it is one of the challenges that Spanish learners have to surpass in order to learn the language correctly. The most numerous ones are those related to the false selection of the preposition:

(65) **Termine a [de] escribir una novela* (MAH, advanced, corpus CEDEL2).

(66) **Voy a trabajar por dos meses por [para] al gobierno* (NER, intermediate, corpus CEDEL2).

(67) **Gracias por el oportunidad a [de] escribir algo* (NER, intermediate, corpus CEDEL2).

(68) **Porque quiero saber si esto es por [de] verdad* (NER, intermediate, corpus CEDEL2).

(69) **Esta muy cerca a [de] los Pirineos* (NER, intermediate, corpus CEDEL2).

(70) **Fuimos para [a] comer afuera de casa la noche antes de la salida* (DMH, advanced, corpus CEDEL2).

We also found many examples of omission of the preposition, despite they are less numerous than the false selection ones:

(71) **[A] Él le gusta mucho leer* (MAH, advanced, corpus CEDEL2).

(72) **Echo de menos*[a] *mi familia* (MAH, advanced, corpus CEDEL2).

(73) **Cada ano, voy a Espana con mi esposa a ver* [a] *su familia* (NER, intermediate, corpus CEDEL2).

(74) **Pasamos tres días con el, jugando* [a] *juegos de mesa* (DMH, advanced, corpus CEDEL2).

(75) **Pasamos la noche* [en] *el coche* (DMH, advanced, corpus CEDEL2).

The group with less number of examples are the addition errors, in which we were able to find only two cases:

(76) **Mi mejor esperanza es que Minnesota decide* en *hacer un Metro* (MAH, advanced, corpus CEDEL2).

(77) **Mi hermano ayudó* con *instalar programas del ordenador* (DMH, advanced, corpus CEDEL2).

4.2.8 Conjunctions

This part of morphosyntactic errors is not very prolific, but we thought that it might be interesting to have a comment on these errors. In the examples we have found, we can observe that the conjunctions are taken directly from English. something that makes us think about *code-switching*. *Code-switching* is “a change by a speaker from one language or language variety to another one” (*Longman Dictionary of Applied Linguistics* 1985: 43). This change of code can provoke the apparition of the tag lines, which are like labels that, inside the sentence, are written in a different language (Appel & Muysken 1987: 175-176). The two errors we found deal with this tag lines and are the following:

(78) **Tarrega es como un ciudad como Lerida, or [o] Lleida in Catalan* (NER, intermediate, corpus CEDEL2).

(79) **Una hora en coche de Barcelona and [y] una hora de Tarragona* (NER, intermediate, corpus CEDEL2).

4.3 Orthographical errors

In this last section of our error analysis, we will be dealing with the orthographical errors; that is, we will account for errors of diacritical marks, capital letters and other significant mistakes.

4.3.1 Diacritical marks

In this part of our project, we found a lot of examples of omission and addition of written accents. In the case of omission, it is normal that these learners have problems with the accent mark, because in their L1 they do not exist. The examples we found of this type are the following:

(80) **Será un milagro si él gradua [gradúa] de la escuela secundaria* (MAH, advanced, corpus CEDEL2).

(81) **Pienso que seré autora y una tia [tía] muy amable para los niños* (MAH, advanced, corpus CEDEL2).

(82) **No se [sé] si sabes donde [dónde] esta Tarrega* (NER, intermediate, corpus CEDEL2).

(83) **Cada ano es menos porque tiene mas [más] y mas anos* (NER, intermediate, corpus CEDEL2).

In the case of the addition, we can say that these errors are probably caused because of the ignorance of Spanish diacritical mark rules, which are on occasion difficult even for Spanish natives. The examples are listed below:

(84) **Pero tengo fé [fe]* (MAH, advanced, corpus CEDEL2).

(85) **Su comedia [comedia] famosa* (JTSB, intermediate corpus CEDEL2).

(86) **Llegamos a Dublin [Dublín]* (SG, advanced, corpus CEDEL2).

(87) **Periodista de television [televisión]* (ML, lower intermediate, corpus CEDEL2).

4.3.2 Capital letters

There are many examples of this type of error, which is probably caused due to the ignorance of Spanish rules, which are difficult even for Spanish native speakers:

(88) **Solo quiero saber que es mi nivel a Espanol [español]* (NER, intermediate, corpus CEDEL2).

(89) **Mi mejor esperanza es que Minnesota decide en hacer un Metro [metro]* (MAH, advanced, corpus CEDEL2).

(90) **Si puedes ir a Tarrega por la fira, Ve Te [vete]!* (NER, intermediate, corpus CEDEL2).

(91) **masters [Máster] en Educación* (KJ, intermediate, corpus CEDEL2).

4.3.3 Other significant errors

This last subgroup of orthographical errors will be very brief and will deal with the errors that do not fit in any of the other parts of the analysis. Most of them are cases

of omission and confusion of letters, exclamation and interrogation marks or even the selection of *y* instead of *e*:

(92) **[j]* *Si puedas ir a Tarrega por la fira, Ve Te!* (NER, intermediate, corpus CEDEL2).

(93) **De colores extravagantes y [e] impresionantes* (DMH, advanced, corpus CEDEL2).

(94) **La Comunidad de Madrid tiene varias iniciativas para ayudar los inmigrantes buscar uno lugar para vivir y encontrar empleo etc.... [etc.]*
(PCLDO, intermediate, corpus CEDEL2).

5. CONCLUSION

The aim of this dissertation was to account for the errors committed by intermediate and advanced learners of Spanish as L2. It is important to remember that all these participants have English as their mother tongue, something that allowed us to comprehend the dissimilar competences they have dependant on each level. Most of the errors, are caused by the influence of their L1 and even for the ignorance of Spanish grammar rules, which are difficult even for native-Spanish speakers. We accounted for the diverse assortment of errors in each level we have analysed, and we can say that we achieved the main aim of this thesis.

While we were carrying out our investigation, we realized about the importance of the corpus and their main aim. The existence of the corpus benefits teachers and learners: the continuation of this corpora allows teachers to create dictionaries, didactic units, and other pedagogical materials to improve and ease learners knowledge of Spanish; for the learners, the corpora facilitate their experience in the learning process and also provides them with a huge variety of errors that will allow them to be autocritics of their own work, in order to polish their use of Spanish language.

The error analysis has to be considered as something normal and natural, processes that all learners have to surpass to aquire the proficiency level in a second language. In the case of our thesis, we wanted to recapitulate and remember the most difficult aspects for Anglophone speakers who are learning Spanish as L2. The most difficult forms for a Spanish as L2 learner are the prepositions, the Spanish diacritical mark rules and the correct use of verb modes. These difficulties can be derived from the incorrect knowledge of Spanish grammar and orthographic rules, or even from the ignorance of them, because in English there are not an institution (as the Spanish *Real*

Academia Española) that elaborates these guidelines arbitrarily. . However, we have to say that most of these problems are also difficult for a Spanish native, because Spanish orthography is regulated by concrete rules and, in some cases, there are a lot of exceptions that escape the rules, something that complicates even more the process of learning the Spanish language.

The corpus linguistics field and, in particular the error analysis specialty is quite extensive and difficult, and it is crucial that we will keep investigating on it in order to develop new teaching methods and even reinforce the possible weaknesses in teaching and learning processes.

6. BIBLIOGRAPHY

Articles, dictionaries and books

APPEL, R. and MUYSKEN, P. (1987): *Language Contact and Bilingualism*. London: Edward Arnold.

BAETENS, H. (1986): *Bilingualism: Basic Principles*. England: Multilingual Matters LTD

COUNCIL FOR CULTURAL COOPERATION (2002): *Common European Framework for Languages: Learning, Teaching, Assessment*, 4-5. Strasbourg: ANAYA.

CORDER, S. (1981 [1967]): "The significance of learners' error". In Corder, S. Pit (ed), *Error Analysis and Interlanguage*, 161-170. Oxford: Oxford University Press.

INSTITUTO CERVANTES. CENTRO VIRTUAL CERVANTES (2007-2015): *Diccionario de términos claves de ELE*, http://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/indice.htm [Consulted: june 2016].

LADO, R. (1957): *Linguistics Across Cultures: applied Linguistics for Language Teachers*. Michigan: The University of Michigan Press.

LOZANO, C. (2009): 'CEDEL2: Corpus escrito del español L2'. In Bretones Callejas, C. et al. (eds), *Applied Linguistics Now: Understanding Language and Mind / La Lingüística Aplicada Hoy: Comprendiendo el Lenguaje y la Mente*, 197-212. Almería: Universidad de Almería.

- PASTOR, S. (2004). *Aprendizaje de segundas lenguas. Lingüística aplicada a la enseñanza de idiomas*. Alicante: Universidad de Alicante.
- REAL ACADEMIA ESPAÑOLA (2012): *Diccionario de la lengua española* (22ª edic.), <http://lema.rae.es/drae/> [Consulted: june 2016].
- RICHARDS, J. et al. (1985): *Longman Dictionary of Applied Linguistics*. Essex: Longman
- SÁNCHEZ, J. and SANTOS, I. (dirs.) (2008). *Vademécum para la formación de profesores: enseñar español como segunda lengua (L2)/lengua extranjera (LE)*. Madrid: SGEL.
- TRASK, R. L. (1993): *A Dictionary of Grammatical Terms in Linguistics*. London: Routledge.
- VÁZQUEZ, G. (1999): *¿Errores? ¡Sin falta!* Madrid: Edelsa.

Corpus

- Cambridge Learner Corpus*, < <http://www.cambridge.org> > [Consulted: june 2016].
- Corpus de aprendices de español*, <http://galvan.usc.es/caes> [Consulted: june 2016].
- Corpus Escrito del Español L2*, <https://www.uam.es/proyectosinv/woslac/cedel2.htm> [Consulted: june 2016].
- Spanish Learner Language Oral Corpora*, www.splloc.soton.ac.uk [Consulted: june 2016].
- Japanese English as a Foreign Language Learner Corpus*, <http://jefll.corpuscobo.net/index.htm#overview> [Consulted: june 2016].

Varietà Apprendimento Lingua Italiana Corpus Online, <http://www.valico.org/>

[Consulted: june 2016].