



**Universitat de Lleida**

Document downloaded from:

<http://hdl.handle.net/10459.1/67968>

The final publication is available at:

<https://doi.org/10.1558/jrds.39063>

Copyright

©Equinox Publishing, 2019

# **Analysis of the production of pronominal constructions in Spanish in a learner corpus**

Ana M. Fernández-Montraveta (Universitat Autònoma de Barcelona)

Glòria Vázquez (Universitat de Lleida)

## **Abstract**

It is a well-established fact that pronominal constructions are one of the most difficult areas for learners of Spanish as a second language. This study aims to contribute to the area of language learning and material design by drawing a general picture of the production of pronominal constructions applying corpus linguistics methodology. The analysis has been carried out taken into consideration three aspects: the students' L1, the semantics of each construction, and the type of errors found. In order to conduct our study, a corpus of 2,532 real pronominal sentences (written by learners of Spanish from levels B1 and C1) was compiled. Findings show that students with a Romance L1 do not perform better than their peers. Also, the semantics of the different types of pronominal constructions seem to have an influence on the students' performance. Finally, errors of overgeneralization are more common than errors of omission. In addition, the latter decrease in higher levels whereas the former seem to increase. These results may be used to reconsider some aspects of the teaching of Spanish as a foreign language.

**Keywords:** pronominal constructions, learner corpus, Spanish as second language, Spanish as a foreign language, error analysis.

## **1. Introduction**

Pronominal constructions are undoubtedly one of the most difficult areas

students have to face when learning Spanish as an L2 (Escutia 2016, Gómez Soler 2015, Tremblay 2006, Toth 2000). Moreover, pronominalization is an essential intransitivization mechanism in this language used to detopicalize the logical argument of transitive structures and to formally mark coindexation of arguments (Fernández-Montraveta and Vázquez 2017). Thus, mastering pronominal construction<sup>5</sup>s contributes to perceiving students as near-natives.

It is precisely for all these reasons that pronominal constructions have raised the scholars' interest in the last decades. However, most of the research on this topic focuses on just one or two specific pronominal constructions and, therefore, they do not provide a general description. Gómez Soler's work (2015) deals with the construction known as anticausative with *se*, i.e. an intransitive pronominal construction whose grammatical subject is not semantically marked as an agent and that alternates with a transitive causative structure. The author studies errors of omission (absence of the clitic) and overgeneralization (use of a pronoun when it is not required) at 4 different levels of proficiency. The methodological approach is based on scalar grammaticality judgment tests and results from the experimental group are compared to a control group. In line with Toth 1999 (cited by Gómez Soler 2015), she shows that low-proficiency learners still show difficulty with the management of the clitic whereas highly proficient learners exhibit native-like command of this construction.

Escutia (2010) studies the use of the clitic *se* by learners of Spanish within the class of unaccusative predicates, that is, monadic predicates with an internal argument that is realized as external. This class of predicates includes verbs that express an internal change of state (*desaparecer* 'disappear') or a change of position (*llegar* 'arrive') and also presentational or existential predicates (*ocurrir* 'happen'). Since the most common function of the pronoun *se* is the suppression of the agent, as is the case in passive and anticausative constructions, and given the fact that unaccusative predicates share some properties with these two constructions (postverbal subject marked as theme that can be syntactically anticipated by an expletive pronoun), learners tend to overgeneralize the pronoun when using these specific predicates. Escutia (2010) discusses the relationship between the process of acquisition and the role Universal Grammar, and students' L1 and L2 play in it, whereas Escutia (2016) defends that learners of Spanish, with English as their L1, interpret the pronoun *se* as equivalent to the expletive form *it*, especially with unaccusative predicates. Others scholars that also study intransitive verbs and pronominalization include unergative predicates to compare

them to unaccusative verbs (Montrul 2000, Toth 2000 and Toth and Guijarro-Fuentes 2013).

Other pronominal constructions, such as the passive and the impersonal constructions, have also been the focus of much interest. Hence, Tremblay (2006) investigates the acquisition of these constructions using grammaticality judgment tests. The participants in her study are learners of Spanish whose L1 is either French and English. This author proposes that there is no complete acquisition of said constructions, not even by advanced students. She also states that no significant differences that could be attributed to the learners' L1 can be found.

Finally, we would like to briefly review Zyzik's work (2006). She focuses on transitive alternations; namely, transitive verbs that can also be used in intransitive pronominal constructions, that is passive, impersonal and anticausative constructions. This author is also interested in the overgeneralization of pronouns, and she analyzes oral productions of students of Spanish in the USA. Participants in the study range from beginners to advanced. Based on the results of her study, the author suggests that the acquisition of these pronominal constructions depends not only on the student's level but also on individual factors.

In contrast to the studies just reviewed, in our work we take into account all the pronominal constructions, thus, adopting a more inclusive approach (Fernández and Vázquez 2017 which represents a contribution). From this approach, pronominal constructions are conceived as a whole family that share a formal feature: a formal mark (a clitic); however, several meanings can be distinguished. One of the most relevant meanings, because of its high productivity, is the change of the discursive focus. This communicative feature is shared by the anticausative, the passive and the impersonal constructions. Other constructions, such as the reflexive or reciprocal constructions, have in common that they are used when there is correferentiality of the participants in the event. In addition, other less frequent pronominal constructions in Spanish are used for emphatic purposes. Finally, there are verbs that include the pronoun in their lexical form and this clitic does not contribute to the semantic interpretation of the sentence.

As we have seen above, the most common methodological devices used in this field are grammaticality and/or acceptability judgment tests (Gómez Soler 2015, Tremblay 2006). Nevertheless, some scholars have questioned the validity and reliability of this kind of instruments (Zyzick 2006, Guijarro-Fuentes and Clibbens 2002); for example, Zyzick (2006, p. 460) considers that "grammaticality judgments

represent only one kind of performance, specifically metalinguistic performance. Thus, it would be beneficial to complement the existing grammaticality judgment data with production data". Following this idea, this piece of research intends to contribute to the study of these constructions from a different methodological approach, based on corpus linguistics. Thus, a corpus of written productions of learners has been created in order to achieve a wider perspective of the students' productions at 2 different stages of the learning process: intermediate (B1) and advanced (C1). The corpus compiled for this work can be considered extensive since it contains an unusual number of texts for this field (see Section 2.2).

Lastly, there are a few studies that research into the role the students' L1 play in the acquisition of pronominal constructions comparing groups of learners with different L1. Mention should be made of Tremblay's (2006) work, which presents this kind of contrastive analysis but, as mentioned above, only passive and impersonal constructions are considered in French and English learners. In our work we follow the line initiated by Tremblay (2006) but comparing a more comprehensive group of pronominal constructions produced by learners from a very wide range of different L1s (see Section 2.1).

Thus, the aim of this study is to make a contribution from the field of corpus linguistics in order to present quantitative data that can be used to advance in the field of learning and teaching of Spanish as a foreign language. In this study we will consider three aspects. First, to what extent the influence of the students' L1 is significant in the acquisition of pronominal sentences. Thus, our first hypothesis (H1) predicts that those students with a Romance L1 will perform better when producing pronominal constructions. Second, we believe pronominal constructions are not acquired uniformly and that the semantics of the constructions plays a determinant role (H2). Finally, following Zyzik's (2006), our third hypothesis (H3) looks into the type of errors and postulates that, even though overgeneralization errors are more frequent than omission errors, they are expected to be overcome in more advanced levels.

## **2. Methodology**

The methodological approach followed in this study is based on corpus linguistics. The data has been gathered from authentic material, as defined in Granger (2002). Therefore, the material used in this research, more specifically the written essays, were the outcome

of real instructional activities: essay writing.

We examined several learner corpora in Spanish and decided to use the CORANE corpus (Cestero Mancera and Penadés Martínez 2009) for this study (see Section 2.1 for a full description). Our choice was based on the necessities established in this work. First, the corpus had to be in digital format and freely available. Second, it should contain texts from learners of Spanish, from different L1s, including Romance and non-Romance languages, at different levels of competence. Only the CEDEL corpus (Lozano 2008) also presented these features. Nevertheless, this resource was dismissed because it only includes 231 texts versus almost 1,000 texts made available by CORANE.

From CORANE we have created a subcorpus (Section 2.2) with the sentences of interest for this study, that is, pronominal constructions produced by the students together with other sentences that have been considered candidates to be pronominal even though students did not use a pronominal structure (incorrect pronominal constructions). As far as this latter group is concerned, the most frequent errors, including errors of omission and of overgeneralizations were annotated. The subcorpus has also been annotated as for the semantics expressed by each construction following the procedure described in Section 2.3.

### ***2.1 The CORANE corpus***

CORANE was compiled as a pedagogical resource in the year 2000 at the Universidad de Alcalá to be used as a dataset for the study of the acquisition of Spanish as a foreign language. The aim, according to its developers, was to obtain the greatest number of essays written by students with different levels of proficiency throughout the year so that it was also possible to see the progression of a student, if desired.

The total amount of texts we have accessed is 926. From this total, 43 texts were written by 10 A2 level students, 436 texts by 132 B1 students, 126 texts by 53 B2 students and 321 texts by 95 C1 students.

The texts compiled in the corpus belong to two different types of writing tasks, referred to as Type 1 and Type 2 in this corpus. Type 1 texts were written as a take-home assignment and collected weekly. Students could use any complementary material they thought necessary when writing them; the sentences and grammatical structures produced were free but the vocabulary limited by the topic proposed. Once a month, students wrote an essay in class without the help of any extra materials; these texts are

labeled as Type 2. They only represent approximately 10% of the total. In this sense, the corpus is specially unbalanced.

Personal information was gathered for each participant and has been made available online together with the texts. According to the authors, the total number of participants is 290 (206 women and 84 men) with 20 different mother languages belonging to very different language families, such as Romance, Germanic, Slave, Semitic, Bantu, Turkic, Sudanic, Altaic, Uralian, Slavic.

## ***2.2 Constitution of the subcorpus***

For the present study, and taking into consideration the huge amount of data offered in the resource, we decided to only consider the essays written by students at just two different stages of their learning process (B1 and C1). The reasons for this choice were both the availability and representativity of texts. That is, essays written by students in level A were discarded since not all pronominal constructions are taught at this early stage. In addition, the number of texts available in level B2 was considerably lower than in B1 or C1. Also, the gap between the two levels chosen for this study, B1 and C1, can help us to observe more clearly the progression of the students in the learning process.

Thus, the total number of essays we have been working with is 436 in the intermediate level and 321 in the advanced; so, the final corpus to be annotated contained a total of 757 texts produced by 227 students from 20 different L1s (695 texts Type 1 and 62 texts Type 2). We believe the number of texts to be sufficient to shed a valid insight into our research questions.

Finally, the total number of sentences that constitutes our corpus is 2,532. The procedure for the annotation process is as follows. Initially, all the texts produced by the students were selected and read. All the sentences in which the students had used a pronominal construction as well as all the sentences that should be pronominal but were not were selected to constitute our corpus. All these sentences are linked to the participant who produced them and, therefore, they are connected to their personal information and language level. Some of this personal information is relevant for the analysis carried out in this study, specifically the L1 language.

## ***2.3 Annotation of the subcorpus***

Once the subcorpus was created, sentences were manually labeled to create the annotated version. The pronominal sentences produced by the students were either annotated as ‘correct’ or as ‘incorrect’, when written as a pronominal structure that was not truly a pronominal construction. In this last case, another label is added, ‘overgeneralization’, to further specify the kind of error. On the other hand, those sentences that should be pronominal but were not were also considered ‘incorrect’ but tagged with the label ‘omission’ to specify the type of error.

Incorrect sentences that include an overuse of the clitic (overgeneralization) were further annotated with respect to the type of formal error they exemplify. We have used the following labels:

- Error 1: the pronoun is required but agreement with the grammatical person is not correct (1).

(1) Mi madre *me rió*. (...*se rió*)<sup>i</sup>  
‘My mother laughed.’

- Error 2: the pronominal form is never possible with the verb (2).

(2) Pero el dinero siempre *se desaparece* muy rápido. (...*se desaparece*...)  
‘But money always disappears fast.’

- Error 3: a verb that can sometimes be pronominal but that in the particular sentence it should not be; for example, the verb *enviar* (‘send’) can be found in pronominal constructions with a passive meaning, but this is not the case in (3).

(3) Y he usado ordenador, *se he enviado* e-mail a mis amigos. (...*se he enviado*...)  
‘And I have used the computer and I have sent an e-mail message to my friends.’

- Error 4: a verb that can be used pronominally but with a difference in meaning. The verb form –*se me ocurrió*– means ‘I had an idea’ but this is not the meaning intended in sentence (4), where, according to our interpretation, the writer meant ‘that somebody was involved in an accident’.

(4) .... en relación con el accidente que *se me ocurrió* en el día 10 de abril de 2000.  
(... *se me ocurrió*...)  
‘... in relation to the accident that in which I was involved on April, 10th, 2000.’

- Error 5: the whole sentence is incoherent so it is impossible to decide whether it should or should not be a pronominal construction (5).

(5) Sabría todos que *se ocurren* en el mundo.

- Error 6: the student uses a pronominal verb that exists but they should be using a completely different lexical item. For example, the correct lexical item to use in the sentence below (6) would be *quedar* ‘meet’; the verb used, *se concerta* ‘is agreed’, can be used in Spanish as a passive form but it does not make sense in this context.

(6) A las 8 en la tarde *se concerta* con un amigo para beber un café con leche.  
(...queda...)  
‘At 8 in the evening, she makes an appointment with a friend to have a coffee.’

In addition, as stated above, whenever a sentence required a pronominal verb form (whether or not the student used it), it was further annotated regarding its semantics. As can be seen below, 8 labels have been considered:

- 1 - Reflexive construction: a construction in which there is coindexation of arguments; that is, the subject and the object of the sentence refer to the same entity. In (7) the verb *ducharse* (‘take a shower’) is used reflexively because the action is applied to the subject.

(7) *Me he duchado*.  
‘I have taken a shower.’

- 2 - Reciprocal construction: it expresses a complex event in which the subject and the object are crossed; that is, there are at least two subevents and the logical subject of event 1 is the logical object of event 2 while the logical object of event 1 is the logical subject of event 2. Even though the verb *besar* has a non-reciprocal interpretation, in (8) it refers to a reciprocal action (‘kiss each other’) and should be used with a pronoun.

(8) He visto los novios (...) *besando* en la calle. (...besándose...)  
‘I have seen the couple (...) kissing each other in the street.’

- 3 - Pronominal passive or impersonal construction: these two construction differ syntactically in that the latter does not have a syntactic subject but are semantically equivalent in that in both of them the object is the discursive focus and the agent is omitted (9).

(9) Hay zonas a dónde ya no *se puede ir* por las noches.  
'There are areas where you can no longer go at night.'

- 4 - Anticausative construction: the verbs that participate in this construction denote a change of state and are mostly dyadic. In the pronominal structure, the logical object of the transitive variant becomes the syntactic subject. There is no change of meaning, just a different focalization of the arguments. In (10), which is incorrect, the clitic is missing, so formally it looks like the transitive structure with an elided subject (*algo preocupa a alguien* 'something worries somebody' vs. *alguien se preocupa* 'somebody worries').

(10) Y *ha preocupado* mucho porque estoy en España sola. (...se ha preocupado...)  
'And he/she has worried a lot because I am all alone in Spain.'

- 5 - Construction with a pure pronominal verb: in fact, some of these verbs could also be found in a non-pronominal construction but with a complete different meaning; e.g., in (11) the verb used by the student is *producirse* ('happen') but *producir* ('produce') also exists in Spanish (*Esta empresa produce chocolate* 'This company produces chocolate').

(11) El mes pasado *se produce* dos acontecimientos. (...se produjeron...)  
'Last month two events happened.'

- 6 - Construction with verbs that are lexically pronominal and convey a reciprocal meaning<sup>ii</sup>: some of these verbs can also be used without a pronoun but, if it is the case, they are no longer reciprocal and, would, therefore, be considered different verbs. For example, in (12) the verb *reunirse* describes the act of holding a meeting, whereas *reunir*, without the pronoun, describes the act of collecting things or people.

(12) es como si todos estuvieran esperando para esta semana para *reunirse* con viejos compañeros de bar.  
'is as if they were all waiting for this week to meet with old bar friends.'

- 7 - Construction with lexically pronominal verbs expressing autonomous movement. These verbs can also be used without a pronoun but with a different, yet related, meaning. In (13), a pronominal form is needed because it refers to the action of *sentarse -uno mismo-* 'take a sit/sit'. The same verb, without the pronoun, is used to refer to the non-autonomous motion (two arguments, the subject and the object: *sentar a otra persona* 'sit someone else').

(13) Pero un hombre *sentó* a lado de mí pronto. (...se *sentó*...)  
 ‘But a man sat next to me soon.’

- 8 - Construction with an alternating pronominal verb<sup>iii</sup>: the pronominal form of the verb is optional and there is no change of meaning. For example, in Spanish the verb *acabarse* in (14) can also be used without clitic with the same meaning.

(14) En Hollanda ya *se acabó* el juego y una de los personas *ganó*.  
 ‘In the Netherlands the game was already over and one of the two people won.’

### 3. Results and discussion

This section presents first general information extracted after the process of annotation. Thus, the total number of correct and incorrect sentences is presented first and then, in order to present results in a more structured way, the data used to refute or validate each hypothesis are presented separately in subsections.

Below, Table 1 displays the exact number for correct and incorrect sentences in our corpus. As can be observed, the overall number of correct sentences is really high (81.20%). Obviously, if we examine the data by levels (Table 2), as could be expected, the number of mistakes is higher in B1 than in C1, but still the figures (25,30% and 15,61%) are quite low.

	Number of sentences	%
<b>Incorrect pronominal sentences</b>	476	18.8%
<b>Correct pronominal sentences</b>	2056	81.20%
<b>Total of pronominal sentences</b>	2532	

Table 1: General distribution of correct and incorrect sentences

	Number of incorrect sentences	%	Total of pronominal sentences
<b>B1</b>	211	25,30	834
<b>C1</b>	265	15,61	1698

Table 2: Distribution of correct and incorrect sentences by levels

### 3.1 Hypothesis 1

H1 stated that those students whose L1 belongs to the family of Romance languages were expected to show a higher competence on the production of Spanish pronominal constructions, and correspondingly faster improvement was expected. Table 3 below presents a graphic overview of the general information about the participants' profiles including their L1.

	Men	Women	L1 Romance (Group 1)	L1 Other (Group 2)
<b>Intermediate</b>	90	42	38	94
<b>%</b>	68.18	31.82	28.79	71.21
<b>Advanced</b>	66	29	12	83
<b>%</b>	69.47	30.53	12.63	87.37

Table 3: General information about the participants

As can be seen, the number of participants in both groups is a bit unbalanced: in Group 1 there are 50 students whose L1 belongs to the Romance family of languages and, in Group 2, we find 177 students whose L1 does not belong to the Romance family. For this reason, we decide that in order to test the validity of H1 we would use the Mann-Whitney-U-test considering the number of correct sentences an indicator of proper acquisition. Results are presented in Table 4.

Variable	N	Means (SD)	P-value	U
L1 Romance (B1 & C1)	50	4.82 (5.7)	$p < 0.05$	4424.5
L1 Non-Romance (B1 & C1)	177	10.25 (14.08)		
L1 Romance (B1)	38	3.97 (4,84)	$p > 0.05$	1786
L1 Non-Romance (B1)	94	4,97 (6.5)		
L1 Romance (C1)	12	7.5 (7.53)	$p > 0.05$	252

L1 Non-Romance (C1)	83	16.18 (17.5)		
---------------------	----	--------------	--	--

Table 4: Statistics results taking into consideration students' L1

First, we tested the two groups regardless their level of Spanish, that is, grouping B1 and C1 students together. Second, we compared each level independently. As can be seen, if we consider all the p-values in Table 4, Hypothesis 1 cannot be either confirmed or refuted since p-values are not conclusive ( $p < 0.05$  and  $p > 0.05$ ) and, therefore, the statistical significance of students' L1 is controversial. Moreover, according to our data, it seems that having a Romance L1 works contrary to our expectation since the means of correct sentences show that students with a non-Romance L1 perform better, especially in the advanced group (7.5 (7.53) versus 16.18 (17.5)).

In order to further investigate into the relation between L1, levels and improvement, another test was carried out to compare the improvement from level B1 to C1 within each group.

Variable	N	Means (SD)	P-value	U
L1 Romance (B1)	38	3.97 (4,84)	$p > 0,05$	114
L1 Romance (C1)	12	7.5 (7.53)		
L1 Non-Romance (B1)	94	4,97 (6,5)	$p < 0,05$	3901
L1 Non-Romance (C1)	83	16.18 (17.5)		

Table 5: Statistic significance in the improvement from B1 to C1

As can be observed in the data presented in Table 5, the correct use of pronominal constructions highly improves from B1 to C1 in Group 2 (Non-Romance L1) and, in addition, this improvement is only statistically significant in this group. This could be one of the factors affecting the results presented above. Another factor that could explain the lack of significant results could be the difference in the number of participants, which has to be acknowledged as a clear limitation of the present study.

These results are partially in line with other authors who also suggest that students' L1 does not seem to play a determinant role in the production of pronominal constructions. So, Tremblay (2006) did not observe any relevant differences between French and English learners of Spanish. Escutia (2012, 2016) proposed as an

explanation for this fact that other linguistic constructions, like the expletive in English, could be consider equivalent to the clitic in Spanish and this might help the acquisition of constructions that require this kind of pronoun.

### 3.2 Hypothesis 2

In order to corroborate our second hypothesis, which stated that the semantics of the construction also plays a role in the acquisition of pronominal constructions, correct and incorrect sentences (only omissions) have been further annotated to see if a trend can be observed (for a full description of the labels used see Section 2.3).

Table 6 below presents the results of this annotation process. The figures for some constructions are presented together; namely, the pronominal passive and impersonal construction (number 3 in Section 2.3) and the anticausative construction (number 4 in Section 2.3) have been grouped since they all share a common semantic trait: they are used when the discursive focus wants to be shifted from the actor to the patient. The other pronominal constructions taken into account in the annotation of semantics are the reflexive construction (number 1 in Section 2.3) and constructions with pronominal verbs (numbers 5, 6 and 7 in Section 2.3). As for the reciprocal construction (number 2 in Section 2.3), it has not been considered in this analysis because its use was extremely marginal in our corpus. Finally, construction 8, where the clitic is optional, has not been considered either since the sentences that exemplify it were always classified as correct (see Section 2.3).

	<b>Passive, impersonal &amp; anticausative constructions</b>	<b>%</b>	<b>Reflexive construction</b>	<b>%</b>	<b>Lexical pronominal verbs</b>	<b>%</b>
<b>B1</b>						
<b>Incorrect omission</b>	11	5.16%	11	14.86 %	22	6.29%
<b>Correct</b>	202	94.84%	63	85.14 %	328	93.71%
<b>Total</b>	213		74		350	
<b>C1</b>						
<b>Incorrect omission</b>	17	3.33%	1	2%	25	3.17%
<b>Correct</b>	494	96.67%	49	98%	764	96.83%

<b>Total</b>	511		50		789	
--------------	-----	--	----	--	-----	--

Table 6: Classification of pronominal constructions depending on their semantics

Some remarks can also be made from the data in Table 6. First, it can be noted that the number of errors for reflexive constructions decreases considerably in C1 with respect to B1 (12.86 points), which means a real improvement in the performance from B1 to C1. As for the other two types of constructions, the improvement in passive, impersonal and anticausative constructions is lower than 2, and approximately 3 for lexical pronominal verbs. So, the progress is especially noticeable for reflexive constructions. Nevertheless, we need to mention that its frequency of use is rather lower when compared to the other two types.

Second, if we analyze data without comparing levels, we can see that our hypothesis is also confirmed in level B1, where the percentage of correct reflexive constructions (85.14%) is lower than in the other two cases, which are very similar (95% for passive, impersonal and anticausative constructions, and 94% for lexical pronominal verbs, approximately).

When students reach level C1 all these constructions seem to have been practically fully acquired. In this sense, our results seem to be different from Gómez Soler (2015), who states that the anticausative construction is acquired only by students with a very high level. Taking into account the data presented in this paper, the acquisition of this construction, as the rest of pronominal constructions mentioned in Table 6, is almost complete in level C1.

### **3.3 Hypothesis 3**

As explained above, whenever a student produced a pronominal sentence that was not a real pronominal it was considered an overgeneralization error. On the other hand, when students did not produce a pronominal construction when they should have, it was considered an omission error. Generally speaking, observing the differences of frequency of these two types of errors also might shed light on the acquisition of pronominal constructions and help to draw a general description. Hypothesis 3, based on Zyzick's (2006) proposal, predicted that overgeneralization errors are, in general, more frequent than omission errors and that they are easier to correct as students

progress in the process of language learning. Table 7 presents the data gathered in order to analyze the distribution of these types of errors.

		Number of sentences		%	
Incorrect pronominal sentences	Omission	96	476	20.17%	18.8%
	Overgeneralization	380		79.83%	
Correct pronominal sentences			2056		81.20%

Table 7: Distribution of correct and incorrect pronominal sentences and type of errors

These results seem to be in line with Zyzik's (2006) work in that overgeneralization mistakes are more frequent than omission errors: the former represent 79.83% of the total number of errors committed by the learners of Spanish as an L2, whereas the latter represent 20.17%, in the two levels examined in this work. According to Zyzik, this tendency can be explained because students perceive the use of pronominal constructions as a very common linguistic mechanism in Spanish and, therefore, try to use them even when it is not required to sound more native-like.

In addition, and following the same author, it was also hypothesized that overgeneralization mistakes would tend to decrease in higher levels. As can be seen from the data in Table 8, this does not seem to be the case, since overgeneralization errors are more common in C1 level (79.65 %) than in level B1 (75.82 %), whereas omission errors are the ones that do decrease (from 24.17 % in B1 to 20.45 % in C1). These findings are in line with White's (1991), cited by Zyzik (2006), who argued that overgeneralization is difficult to correct because of the lack of negative direct evidence. Also, they seem to somehow support Cabrera's (2010) study, that argues that advanced students accept pronominal constructions more than beginners and intermediate students, who almost always rejected the acceptability of said constructions in her study.

Incorrect pronominal sentences					Number of pronominal sentences
Overgeneralization	Omission	Total	%		

<b>B1</b>	160 75.82%	51 24.17%	211	25,30	834
<b>C1</b>	220 79.65%	45 20,45%	265	15,61	1698

Table 8: Distribution of incorrect pronominal sentences considering the students' level and the type of errors

Taking into consideration Cabrera's findings, it would be expected that lower level students should present more errors of omission than of overgeneralization. However, this is not supported by the results presented in this work, as can be seen in Table 8, since in B1 omission errors represent 24.17% of the total number of 211. This might be due to the fact that this author considers both beginners and intermediate together whereas we have only considered data from B1 students.

In order to further refine our analysis, we have also classified the type of errors according to the construction (Table 9). Thus, we distinguish between verbs that can never be used in passive, impersonal or anticausative constructions (because they are either lexically pronominal or never pronominal; Group A) and verbs that can participate in pronominal constructions in order to change the discursive focus and suppress the actor (Group B). Thus, in the former group overgeneralization happens when a clitic is used with these verbs (*se desaparece* in (15)) whereas in the latter this type of error happens when the pronoun is used in a transitive structure (*se escuchó* in (16)). Omission errors in Group A imply that students don't use the pronoun when the verb is lexically pronominal (*darse cuenta* in (17)) whereas in Group B the construction requires the pronoun and the student does not write it (the verb *romper* in (18) participates in an anticausative construction and the pronoun is compulsory: *se han roto*).

(15) Pues ahora casi *se desaparece*. (...~~se~~ desaparece)  
'Then now it almost disappears'.

(16) Era una calleja muy tranquila y *se escuchó* el grito todo el mundo.  
(...~~se~~ escuchó...)  
'It was a small quiet Street and everybody could hear the scream'.

(17) ¿No *has dado cuenta* de lo que sé tu secreto desde hace tres meses?  
(...~~te~~ has dado cuenta...)  
'Haven't you realized that I have known about your secret for three months?'

(18) Las cosas, por ejemplo, grifos, cristales, la cerradura de la puerta, etc. *han roto*.  
(...~~se~~ han roto)  
'Things, for example, taps, glass, the door lock, etc. broke.'

		<b>Group A</b>	<b>%</b>	<b>Group B</b>	<b>%</b>	<b>Total</b>
<b>B1</b>	<b>Overgeneralization</b>	19	47.50	35	72.92	54
	<b>Omission</b>	21	52.50	13	27.08	34
	<i>Total</i>	<b>40</b>		<b>48</b>		<b>88</b>
<b>C1</b>	<b>Overgeneralization</b>	35	59.32	55	76.39	90
	<b>Omission</b>	24	40.68	17	23.61	41
	<i>Total</i>	<b>59</b>		<b>72</b>		<b>131</b>
<b>Total</b>		<b>99</b>		<b>120</b>		<b>219</b>

Table 9: Semantic distribution of a subset of incorrect pronominal sentences considering the type of errors

From the results presented in Table 9, conclusions may be further refined. First, the assertion that overgeneralization errors are more common is generally true but cannot be kept for all the constructions at all levels. More specifically, as can be observed, in level B1 Group A, the number of omission errors is slightly higher, even though the difference is really low. Second, overgeneralization errors rise in level C1 for all the constructions presented in Table 9, but the increase is higher in the verbs in Group A (11.82 points) than in the verbs in Group B (3.47 points).

#### 4. Conclusions

Applying Corpus Linguistics methodology, this paper represents a new approach to the analysis of the production of pronominal sentences. An important amount of written sentences produced by learners of Spanish as an L2 was collected and analyzed. This analysis has proven that, generally speaking, the percentage of corrected pronominal constructions produced by the students is reasonably high. Furthermore, it increases from intermediate level to advanced level, as was to be expected, but it does not homogeneously.

Our findings are only partially in line with those provided by other researchers, which in some cases might be due to the different methodological approach. In the first place, regarding Hypothesis 1, the influence of the students' L1 on the correct or

incorrect production of pronominal seem to be in line with Escutia's findings (2010, 2012) who does not find students' L1 is a relevant factor.

Our second hypothesis was aimed to prove the relevance of the semantics of pronominal construction in the correct production of pronominal constructions. Our findings show that semantics play a role at some stages of the learning process. More specifically, students at the intermediate level (B1) show higher competence with pronominal verbs and with the passive, impersonal and anticausative constructions than with the reflexive construction. Nevertheless, when students reach level C1 this construction is totally acquired and the percentage of the correct use is almost the same for all subtypes of pronominal constructions. In this regard, our findings contrast with Gómez Soler (2015), who states that the anticausative construction is only fully acquired by students of Spanish with a very high level.

Finally, in general, errors due to overgeneralization seem to be more frequent than omission errors, especially in verbs that can participate in the passive, impersonal and anticausative construction. In addition, in our analysis we have seen that overgeneralization errors increase from lower levels to higher ones, which contrasts the idea supported in other studies (Zyzik's, 2006).

From the results presented in this study we believe that better pedagogical materials could be created and more accurate syllabi could be designed to better cover the problems associated to the mastering of pronominal constructions in Spanish. For example, at level B1 more emphasis should be placed on reflexive constructions and, in general, more attention should be paid to the problem of overgeneralization of the clitic.

Regarding future improvements, oral data or more levels could be added to the study and compared to see if results are in line with the ones presented in this paper. For example, cases of near-native students could also be analyzed, following Escutia's (2016) line of research. The number of participants in each group (Romance vs. non Romance L1) should also be balanced to further investigate the relevance of this variable in the learning process or further study the production of the non-Romance group discriminating between the different L1s. Finally, the relation between verb classes and the cause for some errors could also be approached.

## **References**

- Cabrera, M. (2010) “Intransitive/inchoative structures in L2 Spanish”. In C. Borgonovo, M. Español-Echevarría and P. Prévost (Ed.), *Selected Proceedings of the 12th Hispanic Linguistics Symposium*, (p.160-170). Somerville, MA: Casacadilla Proceedings Project.
- Cestero Mancera, A. M. and I. Penadés Martínez (2009). *Corpus de textos escritos para el análisis de errores de aprendices de E/LE (CORANE)*. Alcalá de Henares: Universidad de Alcalá.
- Escutia, M. (2010). “El uso de *se* con verbos inacusativos por estudiantes avanzados de español como lengua extranjera: Transferencia y reestructuración”. *Revista Española de Lingüística Aplicada*, 23, 129-151.
- Escutia, M. (2012). “Expletives and unaccusative predicates in L2A”. *Higher Education of Social Science*, 2:3, 1-14.
- Escutia, M. (2016). “L2 Spanish preverbal ‘se’ in analysis and production data”. *Revista Española de Lingüística Aplicada*, 29:1, 30-63.
- Fernández, A. and G. Vázquez (2017). *Las construcciones con se*. Madrid: Arco Libros.
- Gómez Soler, Inmaculada (2015). “Acquisitional patterns of Spanish anticausative *se*: The end of the road”. *Revista de Lingüística Aplicada*, 28:2, 349-381.
- Granger, S. (2002). “A bird’s eye view of learner corpus research”. In S. Granger, J. Hung and S. Petch-Tyson (Ed.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam and Philadelphia: Benjamins, 3-33.
- Guijarro-Fuentes, P. and J. Clibbens (2002). “Las pruebas de gramaticalidad ¿instrumentos fiables en la recogida de datos en el español como L2?”. *Revista Española de Lingüística Aplicada*, 15, 53-73.
- Lozano, C. (2008). “Cedel2: Corpus escrito del español como L2”. In Carmen M. Bretones et al. (Ed.), *Applied Linguistics now: Understanding language and mind*. Almería: Universidad de Almería.
- Montrul, S. (2000). “Transitivity alternations in L2 acquisition. Toward a modular view of transfer”. *Spanish Second Language Acquisition*, 22, 229-273.

- Toth, P. D. (1999). “Preemption in instructed learners of Spanish as a foreign language: Acquiring a rule for se”. *Spanish Applied Linguistics*, 3:2, 195-246.
- Toth, P. D. (2000). “The interaction of instruction and learner–internal factors in the acquisition of L2 morphosyntax”. *Studies in Second Language Acquisition*, 22, 169–208.
- Toth, P. D. and P. Guijarro-Fuentes (2013). “The impact of instruction on L2 implicit knowledge: Evidence against encapsulation”. *Applied Psycholinguistics*, 34, 1163–1193. doi:10.1017/S0142716412000197
- Tremblay, A. (2006). “On the second language acquisition of Spanish reflexive passives and reflexive impersonals by French- and English-speaking adults”. *Second Language Research*, 22:1, 30-63.
- White, L. (1991). “Argument structure in second language acquisition”. *French Language Studies*, 1:2, 189-207.
- Zyzik, E. (2006). “Transitivity alternations and sequence learning: Insights from L2 Spanish production data”. *Studies in Second Language Acquisition*, 28, 449-485.

---

<sup>i</sup> In the examples, the correct form (always depending on our interpretation of the sentence) is presented between parentheses. This interpretation is the one translated. Only errors related to pronominal constructions are corrected.

<sup>ii</sup> Lexical reciprocal verbs express reciprocity lexically, unlike verbs that can occur in construction 2 in which reciprocity is expressed syntactically.

<sup>iii</sup> This label is only used for correct sentences.