



# The Use of Pronouns in Oral Productions of Spanish as a Second Language: a pilot study

Kim Collewaert<sup>1</sup> and An Vande Castele<sup>1</sup>

<sup>1</sup>Vrije Universiteit Brussel, Brussels, Belgium

Kim.Collewaert@vub.ac.be, An.Vande.Castele@vub.ac.be

## Abstract

In this paper, some preliminary results on the use of pronouns in oral discourse of language learners of Spanish will be discussed. The article mainly focuses on the use of different kinds of personal pronouns and the pro-drop phenomenon, namely the existence of a null subject, typical of the Spanish language. The absence of an explicit subject due to a rich verbal conjugation opposes Spanish to other languages, such as French, English and Dutch, where an explicit subject pronoun is obligatory.

As to investigate the use of the pronouns by language learners of Spanish, we compiled a corpus of oral productions of second language learners of Spanish who are all native speakers of Dutch and also learned French and English, which means that for them the pro-drop phenomenon is new. We will investigate which kinds of pronouns are used in which syntactic contexts and indicate in what contexts the use of a pronoun is not required. Next to this, we observe in our learners' corpus an unnecessary repetition of proper names and an over-use of personal pronouns as subjects. This can be related to the concept of "over-explicitation" or "overspecification", whereby learners of a second language tend to use more explicit forms than necessary.

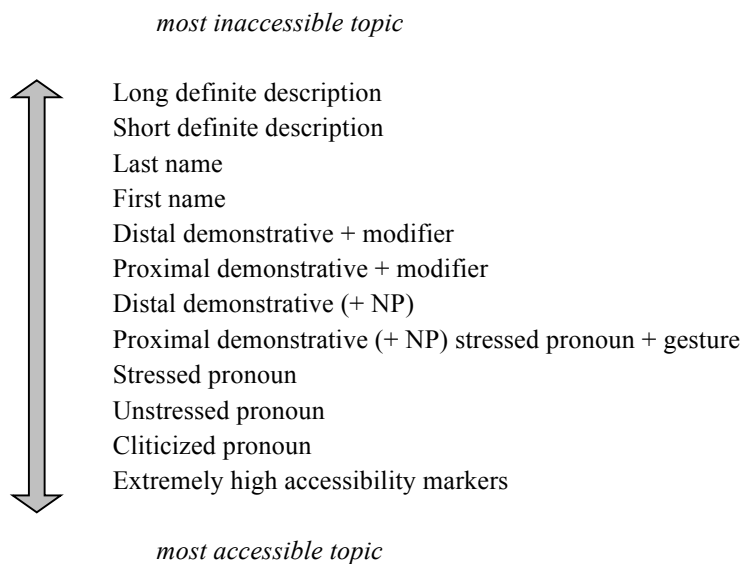
## 1 Introduction

The aim of the present study is to investigate the use of pronouns in oral discourse by Dutch-speaking language learners of Spanish. The analysis of pronouns for this specific language combination is particularly relevant, because both languages have a different pronoun system, the most important difference being the fact that Spanish is a pro-drop language, which has a rich verbal system and consequently favours the use of zero anaphora, while Dutch requires other referential mechanisms, such as pronouns or full noun phrases. The use of the subject pronoun, for example, is obligatory in Dutch, while it is reserved for some specific cases in Spanish. In this paper, we will highlight two phenomena that are directly linked to the opposition between the two studied languages. In addition to the subject pronoun, we will also focus on the (in)direct object pronoun and, more

precisely, analyse erroneous uses in the participants' interlanguage. Finally, we will reflect on the use of the demonstrative pronoun in our corpus. Our database consists of 33 transcriptions of an oral test we designed, based upon the Spanish telenovela *Yo Soy Bea*. The participants are second-year students of Spanish Linguistics and Literature at our university (Free University of Brussels, VUB). We would like to note that this pilot study is embedded in a broader PhD-research project on referential mechanisms in language learners of Spanish' discourse. The project is based on discourse-pragmatic research on referential units and their accessibility status, as investigated by Ariel (1990), Figueras (2002) and Cuenca (2010).

## 2 Theoretical framework

As stated before, this exploratory study on the use of pronouns by language learners of Spanish is embedded in a larger research project on referential mechanisms. Consequently, we refer to the accessibility theory of Mira Ariel (1990) – and its adaptation to Spanish (Figueras 2002) – for the theoretical framework of the present study. Ariel's accessibility marking scale presents a hierarchy of referents which are more or less accessible in the interlocutors' mind. A representation of the scale can be seen below:

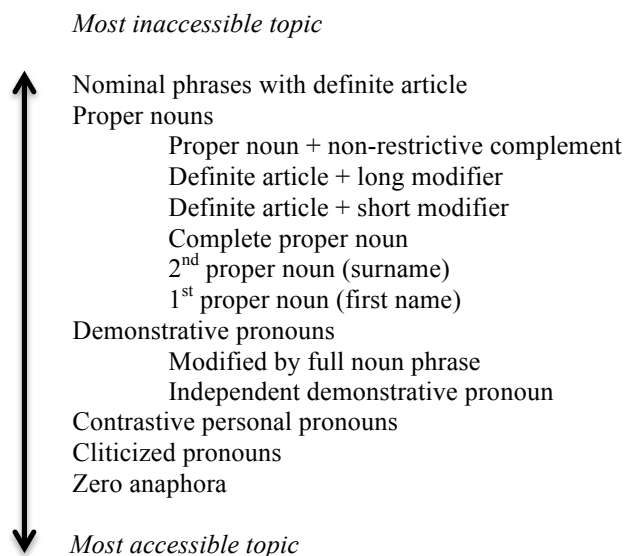


**Figure 1:** Accessibility marking scale (Ariel: 1990)

The scale clearly indicates that longer expressions (such as definite descriptions) represent less accessible entities in the interlocutors' mind. As such, they are used, for instance, to introduce new entities in a discourse. Shorter expressions, on the other hand, – such as the pronouns treated in this paper – represent more accessible entities, which are already present in the interlocutors' mind, since these entities were already introduced in the discourse. The scale thus indicates the degree of accessibility of a referring expression, which does also correlate with the difficulty of the recipient to capture information. Actually, the choice of a referential expression depends on its cognitive status, in other words, on the assumptions made by a speaker about the mental representations of an entity by the recipient. As such, the entities that are less accessible to the recipient require the most descriptive

content, while more accessible entities require less descriptive content. Therefore, low accessibility correlates with full lexical forms, while shorter forms are used to refer to more accessible entities.

Ariel's scale is a common model to represent the degree of accessibility of referential expressions. Nevertheless, as we focus on Spanish, the adaptations of Figueras (2002) and Cuenca (2010) – who added the possessive determiner to the medium accessible markers – are of particular interest for our study. In Figueras' scale, represented below, we note that the highest accessible marker in Spanish, a pro-drop language, is the zero anaphora.



**Figure 2:** Accessibility marking scale (Figueras: 2002)

As Montrul (2004: 175) states, the omission of lexical and pronominal subjects in Spanish is possible, because “*the information is recoverable from the rich person and number morphology on the verb*”. However, in Dutch, the mother tongue of our participants, and in French and English, languages that our participants acquired in primary and secondary education, the use of an explicit subject pronoun is obligatory. So this could impact on the referential choices our participants make in Spanish.

### 3 Research design and methodology

In order to investigate the use of referring expressions – and more specifically, the use of pronouns – in oral discourse, a corpus of oral narratives was composed. All participants, second-year students of the Spanish Linguistics and Literature BA-program at the *Vrije Universiteit Brussel* (VUB), were assigned an oral story-telling task, based on the Spanish telenovela *Yo soy Bea*. A compilation of seven fragments in which the telenovela's protagonists appear has been created for the task. The main characters in this compilation are Álvaro, the new handsome manager of a fashion journal, and Beatriz, his smart, but ugly secretary. All other characters appearing in the fragments are linked to Álvaro: his best friend Gonzalo, his brother-in-law and enemy Diego, his pretty girlfriend Cayetana and his parents Francisco and María. The consecutive references to each of these seven characters were examined. Moreover, as the video contains seven fragments, the narratives have also been divided into seven sequences, which allowed us to analyse the consecutive references within these

sequences. We assume thus that each part responds to the structure of the marking scale, as a new sequence should imply the start of a new referring chain. As the preliminary results of our study will show, some of the obtained results regarding the pronouns can be explained by the transition from one sequence to another. As an illustration, an example of a sequence in a narrative of one of our students (Note that errors have not been corrected):

(1) *aquí estamos en el baño*  
*eso no ento no he entendido muy bien*  
*euhm creo que Beatriz está en la en el asiento*  
*y euhm Alvaro habla con Gonzalo que es su compañero de trabajo no sé*  
*y bueno es lo que lo que he entendido es que euhm es una una acto es un actor porque no está*  
*realmente enamorada de Beatriz*  
*y tiene que que hacer como sí está enamorado por qué no sé*  
*y euhm Beatriz euhm escucha todo*  
*y euhm después*  
*pero es muy es mucho lo que dicen pero*  
*euhm pf Beatriz está triste y después vas a va a desmayarse*  
*euhm es escucha todo demasiado que no quiero escu quiere es escuchar*  
*entender es major (NN1516204)*

The oral task was designed as follows. First, the participants watched the video with the sound, while they had the opportunity to write down some important elements of the story, which could help them achieving the second part of the task, namely the retelling of the story while watching the video for a second time, but this time without sound. This procedure allows us to analyse recordings with a similar duration and avoids that participants keep to a mere summary of the history. Finally, the participants were not informed about the research's purpose. The activity was presented to them as an ordinary oral task within a language acquisition course.

## 4 Preliminary results

Before we will discuss in a more detailed way our findings about the use of pronouns in the oral discourse of language learners of Spanish, we will present some general results on their use of referring expressions, of which the pronouns form part. The most frequently used referential mechanism found in the corpus is the proper noun, a lowly accessible referential expression, which, together with the noun phrase, another lowly accessible marker, constitutes almost half of the items (47,5% of the expressions used in the entire corpus). On the other hand, highly accessible markers, as the zero anaphora and the (different types of) personal pronouns, represent almost the other half (42,5% of the expressions found in the corpus). Other, less frequent, referring expressions include the possessive determiner, the appositional construction, the proper noun with complement and the demonstrative pronoun. Now, the target markers of this study, i.e. the pronouns, represent a 14,5% of the corpus.

The table below offers a complete overview of all referring expressions used by the language learners, each kind illustrated with an example from the corpus.

Spanish learners	language		
Referential expression	Occurrences	Percentage with respect to total number of referents	Examples
Proper noun	1171	35,08%	“ah <i>Beatriz</i> está aquí también pero <i>Álvaro</i> no ve eh a <i>Beatriz</i> y euhm” (NN1516202)
Zero anaphora	962	28,82%	“Bea pues está allí despierta, pero pero no <i>dice</i> nada, <i>intenta</i> de decir algo cuando una mujer había salido pero entra de nuevo y <i>se calle</i> euhm” (NN131415)
<b>Personal pronoun</b>	<b>457</b>	<b>13,69%</b>	“y cuando <i>Beatriz</i> euhm lleva a es muy euhm confusa euhm pero <i>Álvaro</i> se euh preocupe con euh <i>ella</i> ” (NN141512)
Noun phrase	409	12,25%	“ <i>El hombre</i> es muy simpático todos son alegres” (NN141514)
Possessive determiner	219	6,56%	“Francisco euhm cuenta <i>su</i> historia de la de <i>su</i> vida y euhm de la revista [...]” (NN141512)
Appositional construction	54	1,62%	“Euhm creo que <i>Beatriz la secretaria fea</i> se presenta para un puesto en la en la revista de moda” (NN141501)
Proper noun + complement	38	1,14%	“Ahora euhm se ven fotos y <i>Beatriz, que escribe un diálogo o algo</i> ” (NN131405)
<b>Demonstrative pronoun</b>	<b>28</b>	<b>0,84%</b>	“ <i>eso</i> es Diego y pienso que es muy euhm “jaloers” euhm euhm celoso de <i>Álvaro</i> ” (NN141502)
<b>TOTAL</b>	<b>3338</b>	<b>100%</b>	

**Table 1:** General overview of referring expressions used by language learners of Spanish

As stated before, the corpus contains both personal and demonstrative pronouns. Within the personal pronouns, three types can be distinguished: the subject pronoun, the (in)direct object pronoun and the prepositional pronoun.

Table (2) below presents the number of occurrences and an example for each category of personal pronouns.

Personal pronouns				
Type	Occurrences	Percentage (within pronoun types)	Percentage (total)	Examples

<b>Subject pronoun</b>	174	38,07%	5,21%	“y comienzo comienza con una mujer que mira las mujeres perfectas y se realiza que <i>ella</i> no era una mujer perfecta porque tiene lunetes” (NN131409)
<b>(In)direct object pronoun</b>	192	42,01%	5,75%	“Y Álvaro aquí está simpático con <i>ella</i> pregunta que hay quiere ayudarla y Beatriz está tímide” (NN131407)
<b>Prepositional pronoun</b>	91	19,91%	2,73%	“Y Álvaro aquí está simpático con <i>ella</i> pregunta que hay quiere ayudarla y Beatriz está tímide” (NN131407)

**Table 2:** Use of personal pronouns by language learners of Spanish

As table (2) clearly shows, the (in)direct object pronouns are most frequently used by the language learners to refer to the different characters. Nevertheless, the learners usually commit errors when they use this type of pronouns, as can be seen in the following examples. These errors include the use of a direct object pronoun instead of an indirect one, the lack of congruency between the direct pronoun and the gender of the character referred to and finally, various errors with respect to the indirect pronoun in combination with the verb *gustar*. As our participants had a Spanish level between A2 and B1 at the moment of testing, it could be expected that these types of errors would form part of their interlanguage.

- (2) “Álvaro entra y **lo** dicen que **lo** [= *Bea*] encontraron en el baño” (NN131402)
- (3) “euhm pero eh **la lo** dije que **lo** dijo que la imagen es todo porque *ella* es no es un chica una chica com como nosotros” (NN1516204)
- (4) “euhm aquí euhm Diego va a golpear a Álvaro pero euhm gol pero euhm pega a *Bea* es un accidente piense pero no estoy seguro por qué **lo** [= *Bea*] golpe” (NN141509)
- (5) “Y **se** gustan mucho las modelas y mujeres guapas” (NN131403)
- (6) “En este momento hay algunas chicas que vienen con bikini para hablar con euhm Álvaro y él **a él les gustan mucho**” (NN131406)
- (7) “Aquí hay Beatriz, es la secretaria fea euhm y no se gusto **no se gusto mucho Álvaro**” (NN131410)
- (8) “Ehm pero es es enamorado de Álvaro y ehm le gusta pero **Álvaro no la gusta**, pero **Álvaro gusta las mujeres guapas y los modelos**” (NN141511)

Second, we encountered a surprisingly high percentage of subject pronouns in the learners' corpus. Previous research on the L2 acquisition of the Spanish null and overt subject pronoun investigates the subject from various perspectives (cf. for example, Licerias & Díaz 1999, Lubbers Quesada & Blackwell 2009, Lubbers Quesada 2014, Pérez Leroux & Glass 1999, Lozano 2002). Traditional grammars indicate that the use of the subject pronoun in Spanish, a pro-drop language, is reserved to cases that express contrast/opposition or emphasis, as illustrated by the following

examples.

(9) *Cuando Juan conduce, no bebe.*

*Cuando Juan conduce, **él** no va en coche. (él = Paco, otra persona)*

(10) *Carmen es española y habla inglés. (=*

*Carmen es española y **ella** habla inglés (ella = Clara, otra persona)*

In the corpus, several examples can be found wherein the subject pronouns are correctly used, i.e. where they express contrast. In example (11), Álvaro wants to talk with Beatriz, but she refuses, while in example (12), the contrast between the ugly Beatriz, wearing very decent clothes, and the handsome models in sexy bikinis, is expressed.

(11) *“... y euhm después Álvaro entra de nuevo y euhm propone hablar con ella pero **ella** todavía no quiera eh no quería” (NN1516201)*

(12) *“Creo que todas las chicas vienen a solicitar y la Beatriz también está ahí pero **ella** sí está llevando ropas” (NN1516207)*

Nevertheless, the opposition between the null and overt pronoun can also be studied from another angle. Montrul (2004: 176), for example, argues that the (non-) expression of both null subjects and overt pronouns “*is regulated by discourse-pragmatic principles*”. As such, the use of an overt pronoun seems pragmatically odd with continuous references to the same referent, while the null subject results odd when a new referent is introduced. Other discourse-pragmatic studies on the null versus overt pronoun use include, among others, Lubbers Quesada & Blackwell (2009), Matos Amaral & Schwenter (2005), Blackwell (2003; 1998) and Davidson (1996). Montrul’s statement can also be easily correlated with the accessibility theories of Ariel (1990) and Figueras (2002), as continuous references tend to be introduced by highly accessible markers (i.e. null subject for Spanish), while the overt pronoun can be used to (re)introduce a referent (or to mark contrast or emphasis (*cf. supra*)). Nevertheless, in our corpus several examples can be found in which the subject pronoun is used in continuous references to a character, while the zero anaphora thus would have been a more logical option, according to the referential scales. In other words, participants used a lower accessible referential expression than necessary for the retrieval of the referent in question, as can be seen in the following examples.

(13) *“y ahora **la** secretaria Beatriz es en el baño y **ella** oye una conversación de Álvaro y de Gonzalo” (NN131409)*

(14) *“Álvaro es muy contento de ver sus modelas modelos y mujeres guapas y **él** juega un poco con ellas y sonrisa a ellas” (NN131409)*

(15) *“Ellos se van y Beatriz se queda toda sola, totalmente sola y **ella** se va a mayes-desmayecer y sus dos amigas del trabajo van a ayudarle para euh para salir del servicio y ponerla en un lugar euh euh en el sofá donde él ella puede recuperar” (NN141504)*

(16) *“Y ahora **el** padre va a agradecer su mujer y pedir perdón porque **él** no ha logrado tiempo juntos porque para trabajar por su empresa” (NN141504)*

(17) *Creo **los** hombres van a hacer esa cosa porque euhm **ellos** no quieren que Diego euhm es el nuevo jefe de la revista” (NN141513)*

This phenomenon, namely the use of a referential expression that “contains more information than what is necessary for unique identification of the referent” (Arts *et al.* 2011: 362) is called “overspecification” (*cf.* Leclercq & Lenart 2013; Koolen 2013). Besides the overuse of the subject pronoun, another typical example of overspecification that appears frequently in our corpus is the use of a proper name instead of a zero anaphora, but this aspect will not be further discussed in the present article.

Next, we also encountered the opposite phenomenon in our corpus, namely cases in which the subject pronoun offers insufficient information for the identification of the referent. By analogy, we could classify these examples as cases of “underspecification”. In example (18), the first sentence is marking the end of a sequence and the second one the beginning of the next sequence. Generally speaking, the references after episode boundaries are prone to present overspecification, because a speaker tends to reintroduce a referent by using a lower accessible marker. Nevertheless, in the present example, the opposite occurs: the sequence ends with a reference to Álvaro’s girlfriend, Cayetana, and the next sequence starts with a reference to *ella*. Consequently, the interlocutor will assume that it is again Cayetana who is hearing the conversation between her boyfriend Álvaro and Gonzalo, but actually it is Beatriz, the secretary, who is. In order to account for this character switch, the speaker should thus have used a lowly accessible referential expression, (re)introducing Beatriz.

In example (19), the participant describes a sequence in which alternatively appear Álvaro and Gonzalo, on the one hand, and Beatriz, on the other. After the first reference to the two friends, the participant refers to Beatriz with a proper noun and with two possessive determiners, which are medium accessible referential expressions. Next, after having switched briefly to the discussion of the two friends, the subject pronoun *ella* is used as reference to Beatriz. Although the use of a highly accessible marker may be in accordance with the aforementioned accessibility scales, as previous references were realized by lowly and medium accessible referents, we argue that in this case another factor, namely distance (*cf.* Vázquez Rosas 2004), interferes. The last reference to Beatriz and the use of the subject pronoun are separated by three clauses, in which other characters are mentioned. Therefore, it would be useful to reintroduce Beatriz with a lower accessible referential expression. Nevertheless, as Beatriz is the only woman who appears in this sequence, the reference with *ella* is not fully ambiguous, but it could take the interlocutor a little more time to relate the pronoun and the character referred to.

(18) “*la mujer de Álvaro es enfadado con todos*”//  
 “*Y aquí es en la lavadora ella oye una conversación entre Álvaro y Gonzalo*” (NN131403)

(19) “*ahora hay una conversación entre Álvaro y Gonzalo en los aseos y lo que no sepan es que Beatriz está en un aseo su herida ahora vemos su cara **los dos chicos** dicen algo que es una secreto no es una cosa bien pero no sé lo que es el secreto creo que **van** a engañar de alguno o abusar de alguien”  
 “*ella no se sienta bien pero no puede salir del aseo porque tiene miedo pienso*”  
 (NN131411)*

Finally, the corpus also contains 28 instances with demonstrative pronouns, as could be seen in table 2. The use of this type of pronouns could be explained by the experiment’s setting. At the beginning of the pilot study, participants watched the video fragment together with the researcher, before they had to retell the story to the interviewer, who was next to them, while watching the video a second time. As such, participants could easily point to the video and use a demonstrative pronoun at the same time, while retelling the story. Later, however, we slightly adapted the study’s design and



so, the participants were seated in front of the researcher and watched the video fragment alone before the retelling. In the transcriptions from this second phase, we note that the demonstrative pronoun is not used. Thus, as the aim is to further investigate the choice of referential expressions by Spanish language learners in both settings, we will be able to observe if the experimental setting really proves to have an influence on the use of demonstrative pronouns.

(20) *Esto es Álvaro que es el jefe de este euhm revista de mo euhm de moda (NN131416)*

(21) *Eso es Diego que euhm dice cosas a Cayatana euhm cosas muy mal (NN141507)*

## 5 Conclusion

As a conclusion, this exploratory study on the use of pronouns by language learners of Spanish is part of a larger discourse-pragmatic investigation project about referential mechanisms and is based upon the accessibility theory of Mira Ariel (1990). Our general analysis of the referential mechanisms used by the participants showed that the different types of pronouns (both personal as demonstrative) constitute almost 15% of the corpus. The zero anaphora, a typical feature for Spanish, also accounts for 28,82% of the corpus. As such, the markers that we discuss in this study represent almost half of the referring expressions in our database (43,5%). Next to an error analysis with respect to the (in)direct object pronoun, the study especially focused on the opposition between the null and overt subject pronoun, typical of Spanish as a pro-drop language. We encountered some interesting features regarding the subject pronoun that need to be further investigated, such as the over- and underspecification phenomena, which are a direct result of the (non-)use of the subject pronoun. Finally, we also mentioned a particular use of demonstrative pronouns, where the influence of the experiment's setting proved to be an interesting factor that has to be monitored in future research, although the demonstrative pronouns only form a small part of the total corpus.

## References

- Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.
- Arts, A., Maes, A., Noordman, L. & Jansen, C. (2011). Overspecification facilitates object identification. *Journal of Pragmatics*, 43, 361-374.
- Blackwell, S.E. (1998) Constraints on Spanish NP anaphora: The syntactic versus the pragmatic domain. *Hispania*, 81, 606-618.
- Blackwell, S.E. (2003) *Implicatures in discourse: the case of Spanish NP anaphora*. Amsterdam: John Benjamins.
- Cuenca, M. J. (2010). *Gramática del texto*. Madrid: Arco Libros.
- Davidson, B. (1996). Pragmatic 'weight' and Spanish subject pronouns: The pragmatic and discourse uses of 'tú' and 'yo' in spoken Madrid Spanish. *Journal of Pragmatics*, 26, 543-565.
- Figueras, C. (2002). La jerarquía de accesibilidad de las expresiones referenciales en español. *Revista española de lingüística*, 32, 1, 53-96.
- Koolen, R. (2013). *Need I say more? On overspecification in definite reference*. PhD Dissertation. Tilburg University.
- Leclercq, P. & Lenart, E. (2013). Discourse Cohesion and Accessibility of Referents in Oral Narratives: A Comparison of L1 and L2 Acquisition of French and English. *Discours*, 12.

Liceras, J.M. & Díaz, L. (1999). Topic-drop versus pro-drop: null subjects and pronominal subjects in the Spanish L2 of Chinese, English, French, German, Japanese and Korean speakers. *Second Language Research*, 15.1, 1-40.

Lozano, C. (2002). The interpretation of overt and null pronouns in non-native Spanish. *Durham Working Papers in Linguistics*, 8, 53-66

Lubbers Quesada, M. (2014). Subject Pronouns in Second Language Spanish. In Geeslin, K. (ed.) *The Handbook of Spanish Second Language Acquisition* (pp. 253-269). Oxford: Wiley-Blackwell.

Lubbers Quesada, M. & Blackwell, S.E. (2009). The L2 Acquisition of Null and Overt Spanish Subject Pronouns: A Pragmatic Approach. In Collentine, J. et al. (eds.) *Selected Proceedings of the 11th Hispanic Linguistics Symposium* (pp. 117-130). Somerville, MA: Cascadilla Proceedings Project.

Matos Amaral, P. & Schwenter, S.A. (2005). Contrast and the (non-) occurrence of subject pronouns. In D. Eddington (ed.), *Selected Proceedings of the 7<sup>th</sup> Hispanic Linguistics Symposium* (pp. 116-127). Somerville, MA: Cascadilla Proceedings Project.

Montrul, S. (2004). The acquisition of Spanish: morphosyntactic development in monolingual and bilingual L1 acquisition and adult L2 acquisition. Amsterdam: John Benjamins.

Pérez-Leroux, A.T. & Glass, W.R. (1999). Null anaphora in Spanish second language acquisition: probabilistic versus generative approaches. *Second Language Research*, 15.2, 220-249.

Vázquez Rosas, V. (2004) Algunas reflexiones sobre el cálculo de la distancia referencial. *D.E.L.T.A.*, 20(1), 24-47.