The processing of pronouns and the production of referring expressions in L2 English

Carla Contemori and Paola E. Dussias
Pennsylvania State University, USA

1. Introduction

The ability to refer to entities in the surrounding world constitutes a prerequisite for successful language comprehension and production. For instance, a sentence such as (1) could not be possible if the speaker and the hearer could not refer to the entity *Anthony*.

(1) Anthony went on vacation

Psycholinguistic research investigating production and comprehension of referring expressions has shown that successful reference implies choosing/interpreting a form among a set of referring expressions (e.g., Arnold 2010 for a review). Native speakers are usually very fast and efficient in making these decisions in order to keep up with the flow of normal conversation, requiring them to comprehend and produce referring expressions at a fast pace. Research on second language (L2) acquisition has shown that, differently from native speakers, L2 learners may show residual indeterminacy in the L2 referential choice (e.g., Sorace & Filiaci, 2006), even at the highest levels of proficiency. Previous research on L2 comprehension and production of referring expressions has mainly focused on learners who acquire null subject languages (e.g., Spanish, Italian), and has primarily used offline experimental designs (see Wilson 2009, for a study on L2 German using eye-tracking during listening, and Roberts et al., 2008, for a study on L2 Dutch using eye-tracking during reading). The present study contributes to filling this gap by examining the process of comprehension and production in L2 speakers of a non null-subject language (English) whose first language is null-subject (Spanish), using both off-line and on-line methods. The paper is organized as follows. We will first present evidence from cross-linguistic research on anaphora resolution and referential choice showing the main differences between English and Spanish. Secondly, we will review evidence from the literature on L2 acquisition illustrating that L2 learners show a different pattern of anaphora resolution and referential choice than native speakers (L1). Then, we will present two main experimental studies that adopt off-line and online methodologies to address our research questions. We use eye-tracking measures during listening to examine the time course of pronoun resolution during language processing (Experiment 1) and behavioral measures of accuracy in a story telling task to test L2's production abilities (Experiment 2). To anticipate our findings, in Experiment 1 L2 speakers performed similarly to English native speakers on the eye-tracking task. In Experiment 2, we observed a difference in the type of referring expressions used by the two groups. When the native speakers prefer to use a full NP, L2 learners tend to frequently choose a pronoun, even in contexts where two-entities share the same biological gender, i.e., the use of a pronoun leads to ambiguity.

1.1 Anaphora resolution and choice of referring expressions in English and Spanish

Research focusing on the comprehension of referring expressions has provided evidence that the felicitous interpretation of anaphoric expressions is based on the interaction of a number of factors that determine their accessibility, including semantic, syntactic, lexical and discourse structural information (e.g., Kaiser, 2011; Kaiser & Trueswell, 2008). In English, for example, given a sentence such as (2), comprehenders are more likely to interpret the pronoun *he* as referring to *Anthony*, rather than *Simon*.

(2) Anthony went on vacation with Simon. He enjoyed the beach very much.

According to this bias (also called first-mention), English pronouns are expected to co-refer with the previous subject, even when there are no other clear cues pointing to such interpretation or when alternative interpretations are possible (Arnold, et al., 2000). However, in English the first mentioned-bias can be overridden when other cues to accessibility are present, or when other information about pronoun identity is available, such as semantic gender, as illustrated in (3).

(3) Tony went on vacation with Susan. She enjoyed the beach very much.

Semantic gender has been found to be a reliable cue during on-line pronoun resolution, and comprehenders can rapidly use it to determine the referent of a pronoun, even when other cues are pitted against it in order to decrease the accessibility of the intended referent (Arnold et al., 2000, Arnold et al., 2007). Languages may differ in the inventory and function of pronominal forms available to speakers. Cross-linguistic research on anaphora resolution suggests that depending on the set of referring expression and the grammatical properties of a language, native speakers have different antecedent biases. In null subject languages like Spanish, native speakers use null pronouns when the referent is in the addressee's focus of attention, while they prefer explicit pronouns when they want to refer to a non-topic antecedent. Consequently, for the sentence in (2), Spanish native speakers would likely use a null subject in the second clause to indicate that *Anthony* is the subject of the verb *disfrutó* (4). Conversely, they would use an explicit pronoun, to cue that *Simón* is the subject of the second clause, as illustrated in (5).

(4) Anthony$_i$ fue de vacaciones con Simón. *pro$_i$* Disfrutó mucho de la playa.
(5) Anthony fue de vacaciones con Simón$_i$. Él$_i$ disfrutó mucho de la playa.

Previous studies testing anaphora resolution in Spanish native speakers have observed the pattern of interpretation illustrated in (4) and (5) for null and explicit pronouns (e.g., Carminati 2002, for similar evidence on Italian; for Spanish: Keating et al., 2011; Filiaci et al., 2013[1]). Spanish also encodes semantic gender on pronouns, which can act as a cue to identify the referent of an overt pronominal form. However, differently from English, a gender informative pronoun referring to a topic antecedent would generally result in a pragmatically infelicitous utterance, as shown in (6).

(6) Maria$_i$ fue de vacaciones con Simón. Ella$_i$ disfrutó mucho de la playa.

Psycholinguistic research investigating production has shown cross-linguistic differences based on the sets of pronominal forms available to speakers. Native English speakers, for example, prefer to use pronouns, as shown in (7) when the referent is in the addressee's focus of attention (e.g., Arnold & Griffin, 2007). However, when their own focus of attention is distributed among more than one possible referent in the discourse, they prefer to use full NPs (8), as a more explicit form. In this case, there is evidence suggesting that both the visual presence (Fukumura et al., 2010) and/or the previous mention of another animate referent (e.g., Arnold & Griffin, 2007) can create competition for attentional resources in the speaker's representation of the discourse.

(7) Anthony went on vacation. He picked the location because he had been there many times before.
(8) Anthony and Simon went on vacation. Anthony picked the location because he had been there many times before.

---

1   Notice that recent results from self-paced reading and off-line accuracy by Filiaci et al. (2013) suggest that pronouns in the European variety of Spanish seem more sensitive to the linear distance of the antecedent than its actual syntactic position.

For Spanish the maximally reduced pronominal form is the null pronoun used to refer to the most discourse-prominent antecedent (i.e., the sentential subject), while the overt pronoun is preferably used to refer to a non-topic antecedent (4)-(5).

Psycholinguistic evidence on monolinguals indicates the presence of competing strategies across-languages that could pose a potential conflict for speakers of a null-subject language (Spanish) who learn a non null-subject language (English). The aim of our study is to investigate how L2 speakers of English whose first language is Spanish resolve this conflict during on-line processing and in sentence production. In the next section we review previous studies showing instability in the domain of anaphora resolution in L2 learners.

1.2 Anaphora resolution and choice of referring expressions in L2 learners
Studies on anaphora resolution and choice of referring expressions have mainly focused on learners of null subject languages. The findings suggest that L2 learners of Italian and Spanish may exhibit residual indeterminacy in the L2 referential choice, even at the highest levels of proficiency (e.g., Belletti, et al., 2007; Contemori et al., in press; Keating at al, 2011; Montrul & Rodríguez Louro, 2006; Rothman, 2008 , 2009; Sorace & Filiaci, 2006). In comprehension, studies have shown that L2 learners of a null subject language whose L1 is a non-null subject language tend to violate the pragmatic constraints on pronoun distribution in their null-subject language by extending the interpretive scope of overt pronouns. While L2 learners may have a native-like comprehension of the null subject pronoun in (4), repeated here as (9), they sometimes show a non-target like performance on the interpretation of the overt pronoun in (5) repeated here as (10), selecting the subject of the previous sentence as its referent.

(9) Anthony$_i$ fue de vacaciones con Simón. *pro$_i$* Disfrutó mucho de la playa.
(10) Anthony fue de vacaciones con Simón$_i$. Él$_i$ disfrutó mucho de la playa.

Similarly, in production it has been shown that the discourse distribution of overt and null subjects can be non-target-like, with L2 speakers overproducing overt pronouns in situations in which null subjects would have been pragmatically more appropriate. Similar results have been replicated both in comprehension and in production in a number of languages (i.e., Turkish, Spanish, Italian, Greek) with a variety of bilingual populations, including child bilinguals (Argyri & Sorace, 2007; Haznedar, 2010; Sorace, et al., 2009), L1 speakers in situation of attrition (Tsimpli, et al., 2004), and heritage speakers (Montrul, 2004, 2006; Keating et al., 2011).

Studies on anaphora resolution and choice of referring expressions in learners of non null-subject languages are much more limited (Wilson, 2009; Roberts et al., 2008), and have only tested the comprehension of referring expression in L2, but not production. In Roberts et al. (2008), for example, L2 Dutch learners whose L1 is a null subject language (Turkish) and learners whose L1 is a non null-subject language (German) participated in a reading eye-tracking study and an off-line sentence comprehension task. Roberts et al. found a difference between the two L2 groups in the off-line interpretation of Dutch subject pronouns, showing that L1Turkish L2-Dutch participants, but not the L1 German, optionally treated overt pronouns as signaling topic shift, as in their L1, experiencing L1 interference on the use of L2 pragmatics. In the on-line measures, the authors found that both L2 groups incurred into a processing cost when reading sentences in which they had to integrate potentially ambiguous syntactic information (i.e., find an antecedent for a gender ambiguous pronoun) and discourse pragmatic contextual information (e.g., *Hans and Peter are in the office. While Peter is working, he is eating a sandwich*). Roberts et al.'s results indicate two main effects during L2 pronoun resolution. The first is a general L2 processing effect occurring on-line and independent of the L1, suggesting that integrating information from multiple sources (e.g., pragmatic, syntactic) might be

more difficult for L2 learners than for native speakers. The second main result by Roberts et al. is a consequence of cross-linguistic interference that most likely occurred off-line (see Contemori et al. in press, for similar results on self-paced reading in L2 Italian), and that can again be attributed to the difficulty with coordinating information from different sources.

1.3 Aims and research questions
In the present research, we investigate the ability to interpret and produce pronouns in L2 speakers of English whose L1 is Spanish. In Experiment 1, we examine the participants' ability to comprehend ambiguous and unambiguous pronouns in English in contexts in which they can use (reliably and unreliably) the first mentioned-bias and semantic gender information on the pronoun. We will explore whether gender information and the first-mention bias have rapid on-line effects during pronoun interpretation by monitoring participants' eye movements. We expect native English speakers to rapidly integrate gender information and to flexibly override their initial preferences (i.e., the first mention bias) when gender and discourse cues are pitted against each other. For the L2 speakers, we might expect a processing penalty if participants have difficulty integrating different sources of information (i.e., semantic gender and contextual information), as suggested by Roberts et al. (2008). However, if L2 speakers have successfully acquired the first-mention bias and can use semantic gender in English, they might perform similarly to native English speakers. In Experiment 2, a subset of the L2 speakers recruited for Experiment 1 were administered a story telling task to investigate their ability to produce referring expressions in English; their responses were compared to a group of native English speakers. Given that the L2 speakers' first language has a different set of referring expressions than English (i.e., phonologically null forms and explicit pronouns), they cannot rely exclusively on the properties of their L1 when using referring expressions in the L2. There are two possible case scenarios for the L2 speakers. They might experience similar problems in the choice of referring expression as observed in learners of a null subject language. In this case, we would expect a difference between L1 and L2 speakers, with the L2 group showing cross-linguistic interference in the story-telling task. On the other hand, if L2 speakers show a similar pattern as L1 speakers and no difference is found between the groups, it would suggest that the problem with the production of referring expression is limited to the range of choices available in null subject languages.

Experiment 1: eye-tracking comprehension task

Participants
Twenty-nine native English monolingual speakers (mean age: 20; SD: 2) and 17 highly proficient learners of English (L1 Spanish) (mean age=26.5; SD: 7.5) participated in Experiment 1. Participants were students at Penn State University at the time of the testing and received compensation for their participation. L2 participants were born in a Spanish-speaking country (Central/South America) and moved to the US at different times in their lives. Table 1 shows information on the language background of the L2 learners (Language History Questionnaire, Marian et al. 2007) and their proficiency in English measured with an English proficiency test (MELICET, Michigan English Language Institute College English Test).

Table 1. Participant information: Mean (SD)

Self-reported measures		Spanish - L1	English - L2
	Age of exposure	0 (0)	10 (5)
	Age of arrival in a country where the language is spoken	0 (0)	19.9 (10)
	Length of residence in a country where the language is spoken		
	Speaking (1-10)	9.6 (0.5)	8.8 (1.2)
	Listening (1-10)	9.6 (0.5)	8.9 (1)
	Reading (1-10)	9.2 (0.3)	9.1 (0.3)
	Average daily exposure (%)	31.4 (20.5)	68.6 (28.2)
Language proficiency MELICET	Score (out of 50)	-	44.5 (4.5)

Material

In the eye-tracking task, three variables were manipulated: Gender on the pronoun (ambiguous or overt unambiguous), Antecedent of the pronoun (the subject vs. the object of the previous sentence), and Group (native vs. L2 English speakers). Participants heard a voice presenting a sentence in which two characters of different gender (Fig. 1-2) or same gender (Fig. 3-4) were introduced, followed by a sentence containing a pronoun referring to either the subject (Fig. 1-3) or the object (Fig. 2-4) of the previous sentence. Four counterbalanced lists containing eighty experimental sentences (each list containing five items per condition) were created in such manner that each participant saw just one version of the same item. In addition, twenty filler sentences were included in each list, half of them containing a third person singular and half of them containing third person plural pronoun.

Figure 1. Different Gender - First Mention

It was 9 o'clock at the fashion show; a model was talking to a tailor.
She apparently scared the tailor with her loud voice.

Figure 2. Different Gender - Second Mention

It was 9 o'clock at the fashion show; a model was talking to a tailor.
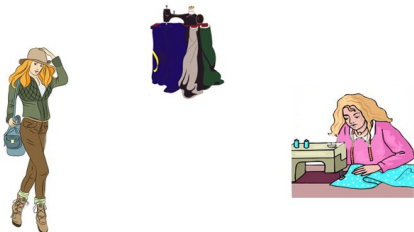He apparently scared the model with his loud voice.

Figure 3. Same Gender - First Mention

It was 9 o'clock at the fashion show; a model was talking to a tailor
She apparently scared the tailor with her loud voice.

Figure 4. Same Gender - Second Mention

It was 9 o'clock at the fashion show; a model was talking to a tailor
She apparently scared the model with her loud voice.

Stimuli were presented on a monitor using a desktop mounted Eyelink 1000 that records eye-movements at a 1000Hz sampling rate. The eye tracker was calibrated and validated for each

participant. A fixation cross was displayed on the screen prior to the start of each trial. After the fixation, participants saw three pictures on a screen (Target, Competitor and Distractor) and listened to a short story through audio speakers, while the two pictures remained on the screen. The short story was followed by the experimental sentence that contained an ambiguous/unambiguous pronoun. Eye-movements were time-locked to the onset of the pronoun (he/she), and were divided in 30 time-windows of 100 ms each (e.g., TW1: 0-100 ms; TW2=100-200 ms, etc.). For each time-window, proportions of looks to the two pictures inferior to 30% were discarded (1.5%). A female native speaker of English recorded the sentences presented auditorily. Sentences and pictures were normed to ensure the absence of any interpretation bias, besides the first-mentioned and the gender-congruency (e.g., one character was more likely to perform an action than the other). Pictures were also normed for naming and gender agreement. Target/Competitor pictures appeared either on the left or on the right side of the screen and the position was counterbalanced. The Distractor picture always appeared in the same position on the screen (center-back), as illustrated in Figures 1-4.

Results
Figures 5-8 show the eye-tracking results for the two groups in the four conditions.

Figure 5. Proportion of looks to Target, Competitor and Distractor in the different gender- first mention condition (L1 on the left; L2 on the right)
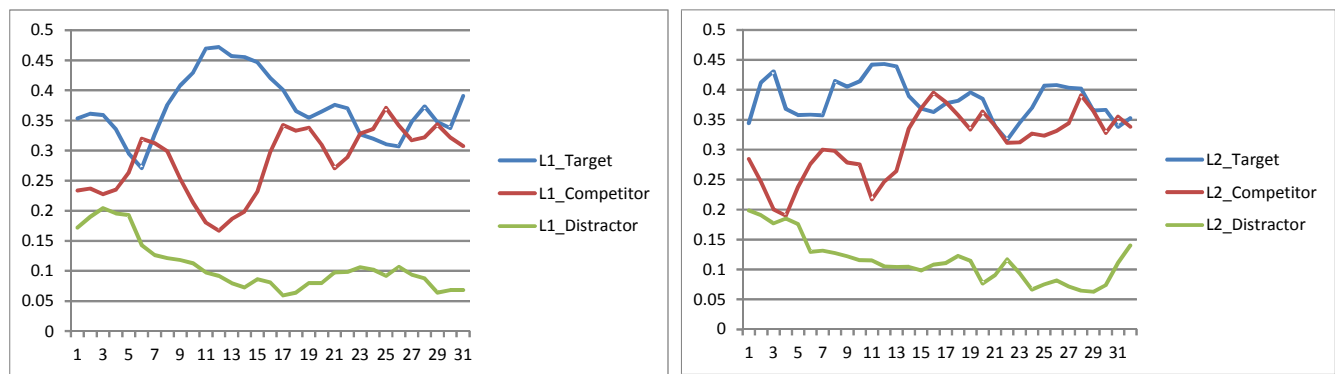

Figure 6. Proportion of looks to Target, Competitor and Distractor in the different gender- second mention condition (L1 on the left; L2 on the right)
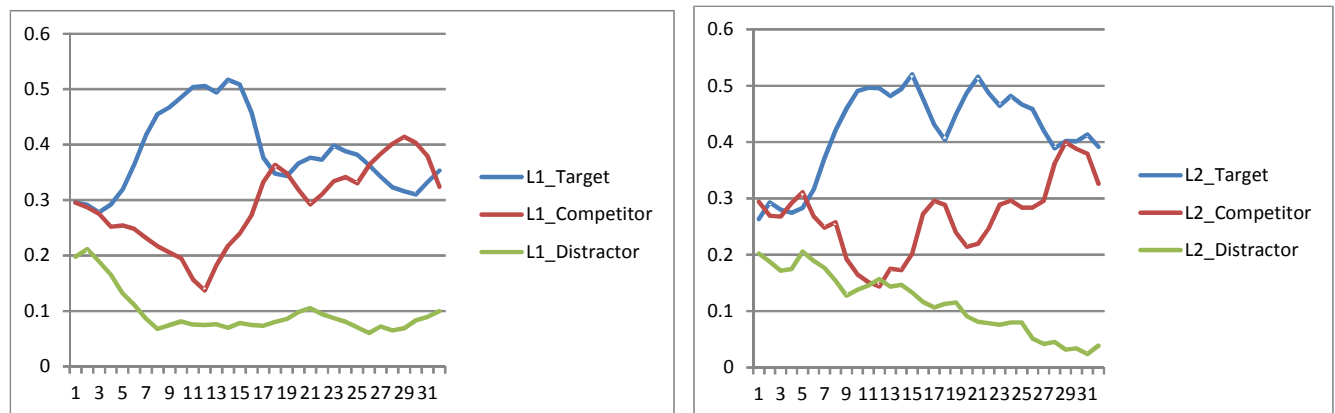
Figure 7. Proportion of looks to Target, Competitor and Distractor in the same gender- first mention condition (L1 on the left; L2 on the right)
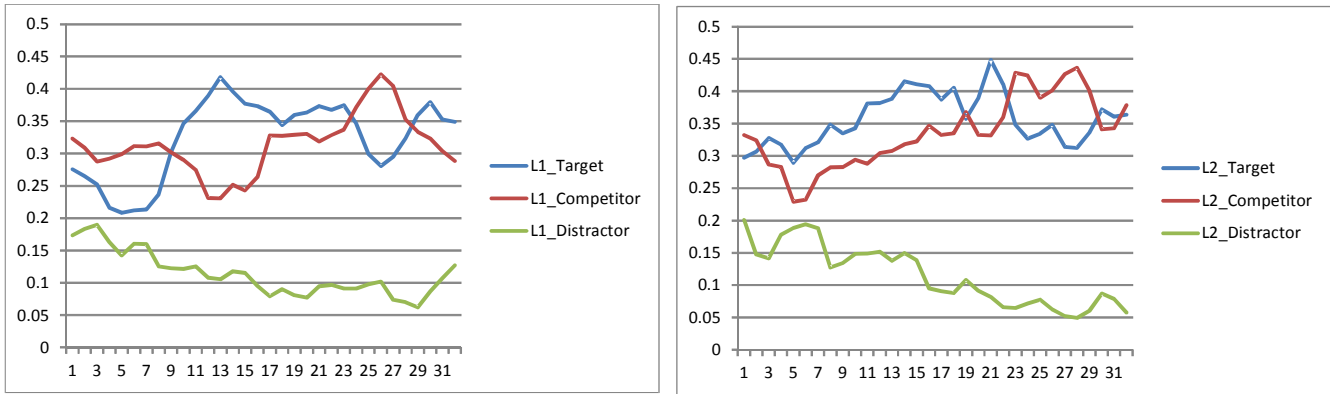
Figure 8. Proportion of looks to Target, Competitor and Distractor in the same gender- second-mention condition (L1 on the left; L2 on the right)
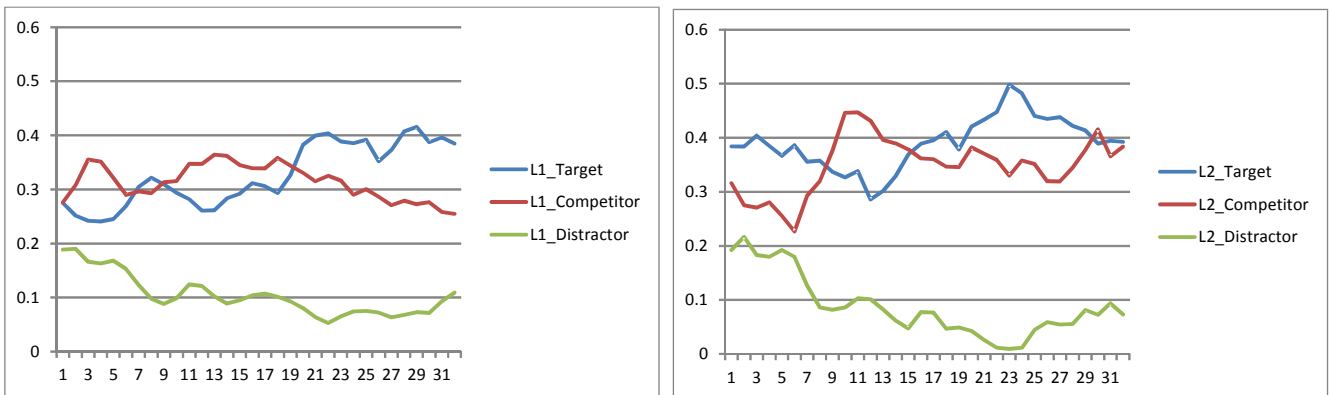
We analyzed the proportion of looks to the Target picture for each time window (TW). We used a repeated-measures ANOVA with Gender ambiguity (2-levels: ambiguous vs. unambiguous) and Referent (2-levels: first vs. second mentioned entity) as within-subjects factor and Group (2-levels: L1 vs. L2) as between-subjects factor, by subject (F1) and by item (F2). Interactions were followed using pairwise comparisons with Bonferroni correction. The ANOVA revealed three main effects. First of all, a main effect of Gender emerged in TW7 ($F1_{(1, 46)} = 5.528$; $p < 0.039$; $\eta p2 = 0.090$; $F2_{(1, 46)} = 7.715$; $p < 0.008$; $\eta p2 = 0.090$), and continued through TW15[2], showing that participants looked more at the target in different gender conditions, i.e., when the two characters were a male and a female, and gender was an informative cue on the pronoun. Secondly, the ANOVA showed a main effect of Referent in TW22 ($F1_{(1, 46)} = 4.464$; $p < 0.040$; $\eta p2 = 0.088$; $F2_{(1, 38)} = 4.261$; $p < 0.046$; $\eta p2 = 0.101$) through TW27[3], indicating that participants looked more at the Target picture in conditions in

---

[2] The main effect of Gender was found in TW8 ($F1_{(1, 46)} = 8.456$; $p < 0.006$; $\eta p2 = 0.155$; $F2_{(1, 47)} = 12.604$; $p < 0.001$; $\eta p2 = 0.249$), TW9 ($F1_{(1, 46)} = 10.119$; $p < 0.003$; $\eta p2 = 0.180$; $F2_{(1, 38)} = 12.975$; $p < 0.001$; $\eta p2 = 0.255$), TW10 ($F1_{(1, 46)} = 19.297$; $p < 0.0001$; $\eta p2 = 0.296$; $F2_{(1, 38)} = 18.621$; $p < 0.0001$; $\eta p2 = 0.329$), TW11 ($F1_{(1, 46)} = 22.773$; $p < 0.0001$; $\eta p2 = 0.331$; $F2_{(1, 38)} = 20.964$; $p < 0.0001$; $\eta p2 = 0.356$), TW12 ($F1_{(1, 46)} = 20.611$; $p < 0.0001$; $\eta p2 = 0.310$; $F2_{(1, 38)} = 24.997$; $p < 0.0001$; $\eta p2 = 0.397$), TW13 ($F1_{(1, 45)} = 14.440$; $p < 0.0001$; $\eta p2 = 0.243$; $F2_{(1, 38)} = 13.765$; $p < 0.001$; $\eta p2 = 0.266$), TW14 ($F1_{(1, 46)} = 10.264$; $p < 0.002$; $\eta p2 = 0.182$; $F2_{(1, 38)} = 8.692$; $p < 0.005$; $\eta p2 = 0.186$); TW15 ($F1_{(1, 46)} = 9.604$; $p < 0.003$; $\eta p2 = 0.173$; $F2_{(1, 38)} = 7.154$; $p < 0.011$; $\eta p2 = 0.158$).

[3] The main effect of Referent was found in TW23 ($F1_{(1, 46)} = 8.693$; $p < 0.005$; $\eta p2 = 0.159$; $F2_{(1, 38)} = 9.079$; $p < 0.005$; $\eta p2 = 0.193$), TW24 ($F1_{(1, 46)} = 10.693$; $p < 0.002$; $\eta p2 = 0.189$; $F2_{(1, 38)} = 11.849$; $p < 0.001$; $\eta p2 = 0.238$), TW25 ($F1_{(1, 46)} = 8.416$; $p < 0.006$; $\eta p2 = 0.155$, $F2_{(1, 38)} = 12.688$; $p < 0.001$; $\eta p2 = 0.250$), TW26 ($F1_{(1, 46)} = 6.131$; $p < 0.017$; $\eta p2 = 0.118$; $F2_{(1, 38)} = 7.883$; $p < 0.008$; $\eta p2 = 0.172$), and TW27 only by item ($F2_{(1, 46)} = 4.320$; $p < 0.044$; $\eta p2 = 0.102$)

which the referent was the object of the previous sentence compared to when it was the subject. We also found a main effect of Group in TW4 (F1 (1, 46) = 4.390; p < 0.042; ηp2= 0.087; F2 (1, 38) = 4.305; p < 0.045; ηp2= 0.102), and in TW 26 (F1 (1, 46) = 5.386; p < 0.025; ηp2= 0.105; F2 (1, 38) = 3.891; p < 0.056; ηp2= 0.093), indicating that L2 looked significantly more at the Target than native speakers (TW4: 0.27 vs 0.33; TW26: 0.32 vs 0.41).

The ANOVA revealed three interactions. A Group by Referent interaction emerged in TW28, which was significant by item and approached significance by subject (F1 (1, 46) = 3.603; p < 0.064; ηp2= 0.073; F2 (1, 38) = 3.906; p < 0.055; ηp2= 0.093). However, the post-hoc analysis for this interaction did not reveal any difference between the two groups in any of the conditions. Therefore, it is unclear why there is an interaction in this TW. Given that this difference emerges only in the item analysis and the effect size is small, we will not discuss this effect any further. An interaction between Gender and Group emerged in TW4 that was significant in the item analysis (F2 (1, 38) = 4.176; p < 0.048; ηp2= 0.099). The post-hoc comparisons showed that native speakers looked significantly more at the Target picture in the gender different-first mention condition compared to the two gender ambiguous conditions (gender different-first mention vs same gender-first: t1(30)=2.232, p<0.033, t2(19)=2.922, p<0.009; gender different-first mention vs same gender-second mention: t2(19)=3.270, p<0.004). No such difference was found in the L2 group. Finally, we found an interaction between Gender and Referent in TW1[4] (F1 (1, 46) = 4.237; p < 0.021; ηp2= 0.084; F1(1, 38) = 4.392; p < 0.043; ηp2= 0.104) that was significant through TW4[5]. We also found an interaction between Gender and Referent in TW12 (F1 (1, 46) = 7.383; p < 0.009; ηp2= 0.138; F2 (1, 45) = 7.665; p = 0.009; ηp2= 0.168) through TW16[6], and in TW28[7] (approaching significance in the by subject analysis: F1 (1, 46) = 3.603; p < 0.064; ηp2= 0.073; and significant in the by item analysis: F2 (1, 38) = 3.906; p < 0.055; ηp2= 0.093).

For the Gender by Referent interaction, we present the results of the pairwise comparisons per each TW. The pairwise comparisons revealed a significant difference between the gender different-first mention condition compared to the gender different-second mention in TW3 (F1 (1, 46) = 6.013; p < 0.018; ηp2= 0.113; F2 (1, 39) = 4.479; p < 0.041; ηp2= 0.103), TW14 (only in the by item analysis F2 (1, 39) = 4.878; p < 0.033; ηp2= 0.111) and TW15 (F1 (1, 47) = 3.845; p < 0.056; ηp2= 0.076; F2 (1, 39) = 7.485; p < 0.009; ηp2= 0.161). In TW3 participants looked significantly more at the Target picture in the gender different-first mention condition, while in TW14 and 15 the opposite was found, with participants looking more at the Target in the gender different-second mention.

We observed a significant difference between the gender different-first mention condition compared to the same gender-first in TW2 (F1 (1, 47) = 5.180; p < 0.027; ηp2= 0.099; F2 (1, 39) = 6.420; p < 0.015; ηp2= 0.141), TW3 (F1 (1, 46) = 6.149; p < 0.017; ηp2= 0.116; F2 (1, 39) = 7.494; p < 0.009; ηp2= 0.161), TW4 only subject analysis (F1 (1, 47) = 4.346; p < 0.043; ηp2= 0.085) and TW12 (F1 (1, 47) = 4.092; p < 0.049; ηp2= 0.080; F2 (1, 39) = 6.499; p < 0.015; ηp2= 0.143), indicating that participants looked significantly more at the Target picture in the gender different-first mention condition.

The pairwise comparisons revealed that participants looked significantly more at the Target picture in the gender different-first mention condition compared to the same gender-second mention in TW2

---

[4]   For TW1 the pairwise comparisons did not show any significant effect between the conditions. It is unclear why there is an interaction in this TW.

[5]   The interaction between Gender and referent emerged in TW2 (F1 (1, 46) = 4.336; p < 0.043; ηp2= 0.086; F2 (1, 38) = 5.737; p < 0.022; ηp2= 0.131), TW3 (F1 (1, 46) = 6.935; p < 0.011; ηp2= 0.131; F2 (1, 38) = 6.997; p < 0.012; ηp2= 0.156), TW4 (only in the subject analysis: F1 (1, 46) = 3.977; p < 0.052; ηp2= 0.080).

[6]   The interaction between Gender and referent was found in TW13 (F1 (1, 45) = 14.440; p < 0.0001; ηp2= 0.243; F2 (1, 38) = 8.038; p < 0.007; ηp2= 0.175), TW14 (F1 (1, 46) = 7.812; p < 0.008; ηp2= 0.145; F2 (1, 38) = 10.925; p < 0.002; ηp2= 0.223), TW15 (F1 (1, 46) = 6.693; p < 0.013; ηp2= 0.127; F2 (1, 38) = 10.132; p < 0.003; ηp2= 0.211), TW16 (only in the by item analysis: F2 (1, 46) = 5.542; p < 0.024; ηp2= 0.127).

[7]   For TW28 the pairwise comparisons did not show any significant effect between the conditions. Therefore, it is unclear why there is an interaction in this TWs. Given that this difference occurs only in the per item analysis and the effect size is small, we will not discuss this effect any further because it is not a reliable effect.

(only in the by subject analysis: $F1 (1, 47) = 3.847$; $p < 0.056$; $\eta p2 = 0.076$), TW3 (only in the by subject analysis: $F1 (1, 46) = 4.092$; $p < 0.049$; $\eta p2 = 0.080$), TW12 ($F1 (1, 47) = 26.422$; $p < 0.0001$; $\eta p2 = 0.360$; $F2 (1, 39) = 20.327$; $p < 0.0001$; $\eta p2 = 0.343$), TW13 ($F1 (1, 46) = 14.565$; $p < 0.0001$; $\eta p2 = 0.240$; $F2 (1, 39) = 13.765$; $p < 0.001$; $\eta p2 = 0.261$), TW14 ($F1 (1, 47) = 9.183$; $p < 0.004$; $\eta p2 = 0.163$; $F2 (1, 39) = 5.795$; $p < 0.021$; $\eta p2 = 0.129$), TW15 (only by subject $F1 (1, 47) = 5.034$; $p < 0.030$; $\eta p2 = 0.097$).

A significant difference emerged for the gender different-second mention condition in comparison to the same gender-first condition in TW12 ($F1 (1, 47) = 7.480$; $p < 0.009$; $\eta p2 = 0.137$; $F2 (1, 39) = 7.055$; $p < 0.011$; $\eta p2 = 0.153$), TW13 ($F1 (1, 46) = 4.857$; $p < 0.033$; $\eta p2 = 0.096$; $F2 (1, 39) = 4.418$; $p < 0.042$; $\eta p2 = 0.102$), TW14 (only by item analysis $F2 (1, 39) = 5.361$; $p < 0.026$; $\eta p2 = 0.121$), and TW15 ($F1 (1, 47) = 11.330$; $p < 0.002$; $\eta p2 = 0.197$; $F2 (1, 39) = 7.649$; $p < 0.009$; $\eta p2 = 0.164$), and in comparison to the same gender-second mention condition in TW12 ($F1 (1, 47) = 32.727$; $p < 0.0001$; $\eta p2 = 0.410$; $F2 (1, 39) = 22.031$; $p < 0.0001$; $\eta p2 = 0.361$), TW13 ($F1 (1, 46) = 18.220$; $p < 0.0001$; $\eta p2 = 0.284$; $F2 (1, 39) = 16.859$; $p < 0.0001$; $\eta p2 = 0.302$), TW14 ($F1 (1, 47) = 25.669$; $p < 0.0001$; $\eta p2 = 0.353$; $F2 (1, 39) = 14.755$; $p < 0.0001$; $\eta p2 = 0.274$), TW15 ($F1 (1, 47) = 24.907$; $p < 0.0001$; $\eta p2 = 0.346$; $F2 (1, 39) = 14.848$; $p < 0.0001$; $\eta p2 = 0.276$), and TW16 (only in the by item analysis: $F2 (1, 39) = 6.745$; $p < 0.013$; $\eta p2 = 0.147$), showing that participants looked more at the Target picture in the gender different-second mention condition compared to the two same gender conditions.

Finally, the pairwise comparisons revealed that participants looked significantly longer at the Target picture in the same gender-first condition compared to the same gender-second mention in TW12 ($F1 (1, 47) = 7.816$; $p < 0.007$; $\eta p2 = 0.143$; $F2 (1, 39) = 7.348$; $p < 0.010$; $\eta p2 = 0.159$), through TW14[8].

Interim discussion

The main result emerging from the eye-tracking task is that the native speakers and the L2 participants show a similar pattern in the interpretation of pronouns in English. The only difference that we found in the data analysis is a Gender by Group interaction in TW4 (300-400 ms after the onset of the pronoun). Notice however that the interaction does not point to a clear processing difference between the two groups, but rather indicates a wider gap between the looks to the gender different-first mention condition and the looks to the gender ambiguous conditions in the native speakers compared to the L2 group. Hence, our results seem to suggest that overall L2 learners can reach native-like performance in the processing of pronouns in English, when gender and the first-mention bias are manipulated. We will now discuss the qualitative differences between the four conditions that emerged from the data analysis.

First, the main effects of Gender and the Gender by Referent interaction in the early time-windows indicate that both groups process the gender information on the pronoun quickly and use it to disambiguate the referent of the pronoun (as shown in Figure 5 and 6, compared to Figure 7 and 9). The effect persisted up to 1600 ms after the onset of the pronoun, with participants looking longer at the target picture in the gender unambiguous conditions compared to gender ambiguous conditions. As demonstrated by the results, gender information had a different impact on participants' processing when the pronoun referred to the first or the second mentioned entity. Both native and L2 speakers showed an initial advantage (200-300 ms after the onset of the pronoun) when the gender unambiguous pronoun was the first mentioned entity in the preceding discourse (Figure 5) compared to when it was the second mentioned entity (Figure 6), as demonstrated by the pairwise comparisons in TW3. When the gender unambiguous pronoun referred to the second-mentioned character in the preceding discourse, participants had to override the first-mention bias, and it took them longer to activate the informative gender cue on the pronoun. For this condition, participants' disambiguation occurred later, and we observed an opposite pattern as the one observed in TW3 between 1300-1500 ms, with participants

---

8   Pairwise comparisons for TW13 ($F1 (1, 46) = 9.504$; $p < 0.003$; $\eta p2 = 0.171$; $F2 (1, 39) = 8.337$; $p < 0.006$; $\eta p2 = 0.176$) and TW14 ($F1 (1, 47) = 6.037$; $p < 0.018$; $\eta p2 = 0.114$; $F2 (1, 39) = 5.663$; $p < 0.022$; $\eta p2 = 0.127$).

looking more at the Target picture when the gender unambiguous pronoun referred to the second mentioned entity compared to when it referred to the first mentioned entity in the discourse.

From the results we also observed that in both groups, the gender unambiguous conditions showed an advantage over the gender ambiguous conditions at different points in time. Participants looked at the target picture more in the gender different-first mention condition (Figure 5) compared to the same gender-first mention (Figure 7) in the 100-400 and in the 1100-1200 ms time-window, and compared to the same gender-second mention (Figure 8) in the 100-300 and in the 1100-1500 ms time-window. This showed that the gender different-first mention condition had both an early and a late advantage over the gender ambiguous conditions.

For the gender different-second mention condition (Figure 6), in both groups the comparison showed significantly more looks compared to the same gender-first mention (Figure 7) in the 1100-1500 ms, and compared to the same gender-second mention condition in the 1100-1600 ms. These effects show no early advantage, but only a late advantage of the gender different-second mention condition over the gender ambiguous conditions. We hypothesize that the difference between early and late advantage found in the two gender unambiguous conditions is due to the effect of the first-mention bias discussed earlier. When the gender unambiguous pronoun referred to the second-mentioned character in the preceding discourse, participants had to override the first-mention bias in the earliest time-windows. Thus, the amount of looks to the target for gender different-second mention condition is not significantly higher than those for the ambiguous conditions until later (1100-1500, 1100-1600 ms).

Another interesting result emerging from our data is an effect of first-mentioned bias in the gender ambiguous conditions in both the L2 and the native speakers. This is shown by the higher number of looks to the target picture in the same gender-first mention condition (Figure 7) compared to the same gender-second mention (Figure 8) between 1100 and 1400 ms after the onset of the pronoun. When gender on the pronoun is not informative, participants do not wait for the disambiguating information to assign a referent, but rather assume that the first mentioned-entity is more likely to be the right antecedent.

Finally, we found a main effect of Referent starting later in the sentence, at around 2100 ms, indicating that L2 speakers and native speakers alike looked longer at the Target picture in conditions where the referent of the pronoun was the second-mentioned entity (Figure 6 and 9) compared to conditions where it was the first-mentioned entity (Figure 7 and 9). This effect is likely due to the late disambiguation provided by the experimental sentences in object-antecedent conditions, corresponding to the lexical NP presented after the main verb (e.g., A lawyer meets a cook near the door. She always irritates the *lawyer* with her attitude). Upon hearing the disambiguating NP, L2 and native speakers' looks to the target increase significantly in the object-antecedent conditions, lasting up to 2700 ms after the onset of the pronoun.

We will now present the story-telling task that we conducted with participant who took part in the eye-tracking experiment.

Experiment 2: story-telling task

Participants
Seventeen L2 speakers from Experiment 1 (mean age=26.5; SD: 7.5) and eighteen native English-speaking adults (mean age: 20; SD: 2) were recruited for the story telling task.

Material
The story telling task employed a subset of the materials used in Arnold & Griffin (2007). Participants were presented with two pictures that contained: (1) one character in the first panel and one in the second panel (Figure 9); (2) two characters in the first panel and two in the second panel (different

gender) (Figure 10); (3) two characters in the first panel and one in the second panel (different gender) (Figure 11); (4) two characters in the first panel and one in the second panel (same gender) (Figure 12).

Figure 9. 1-Referent

 Mickey went for a walk in the hills one day

Figure 10. 2-Referents (in both panels)

 Mickey went for a walk with Daisy in the hills one day

Figure 11. 2-referents (in one panel)

Mickey went for a walk with Daisy in the hills one day

Figure 12. 2-referents (gender ambiguous)

Mickey went for a walk with Donald in the hills one day

Two variables were manipulated: Condition (4-levels) as within-subjects factor and Group (native vs L2 speakers) as between-subjects factor. After listening and reading a description of the first panel, participants were asked to repeat the description and complete the story by describing the second panel. As shown by Arnold & Griffin (2007), English native speakers tend to use full NPs when more than one possible referent is present in the discourse. Based on their results, we expected L1 participants to complete the description with a pronoun in the 1-Referent condition, as shown in (11), and a full NP in all the other conditions, as shown in (12)-(14).

(11) 1-Referent condition: Mickey went for a walk in the hills one day; *he* was out of shape and got very tired
(12) 2-Referents (in both panels): Mickey went for a walk with Daisy in the hills one day; *Mickey* was out of shape and got very tired
(13) 2-referents (in one panel): Mickey went for a walk with Daisy in the hills one day; *Mickey* was out of shape and got very tired
(14) 2-referents (gender ambiguous): Mickey went for a walk with Donald in the hills one day; *Mickey* was out of shape and got very tired

Four counterbalanced lists containing sixty-four experimental sentences (each list containing four items per condition) were created in such manner that each participant saw just one version of the same item. In addition, sixteen filler sentences were included in each list, half of them containing a variable number of characters in the first and second panel.

Coding
Based on Arnold & Griffin (2007), we only included in the analysis productions in which a referring expression referred to the main character (the most prominent in the discourse and in the picture) as the grammatical subject. The main character also had to precede any mention of the second character. We excluded trials in which participants produced referring expressions that denoted more than one character (e.g., Donald and Mickey, they) and cases of ellipsis (e.g., ...and couldn't keep up...). We also excluded items with naming errors (e.g., Daisy for Minnie) or other errors. Based on these criteria, 69/516 (13%) experimental trials were discarded.

Results
Figure 13 illustrates the percentage of pronouns produced by the native speakers and the L2 participants out of the total number of pronouns and NPs.

Figure 13. Pronouns produced by L1 and L2 speakers in the four conditions
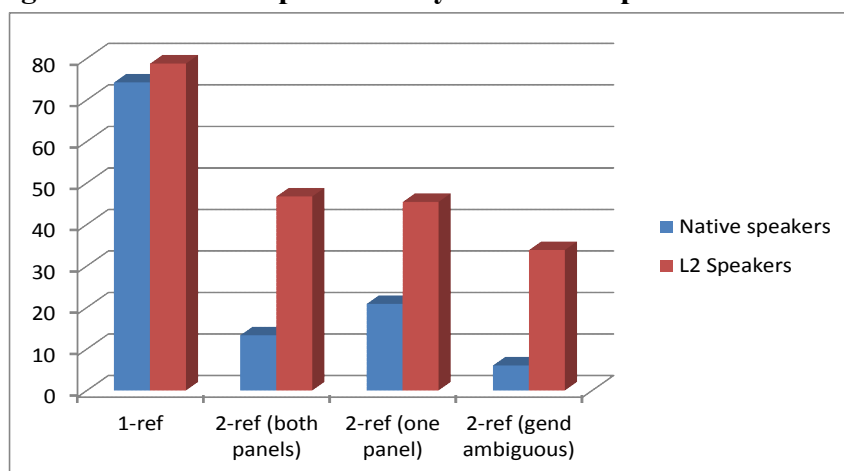
We analyzed the proportion of pronouns produced out of the total number of pronouns and NPs per group. Proportions were transformed using the arcsine transformation. We used a repeated-measured ANOVA with Condition (4-levels) as within-subjects factor and Group (2-levels) as between-subjects factor, by subject (F1) and by item (F2). Interactions were followed using pairwise comparisons with Bonferroni correction. The ANOVA revealed a main effect of Condition ($F1$ $(3, 99) = 30.362$; $p < 0.0001$; $\eta p2= 0.479$; $F2$ $(3, 90) = 27.004$, $p < 0.0001$; $\eta p2= 0.474$), a main effect of Group ($F1$ $(1, 33) = 7.116$; $p < 0.012$; $\eta p2= 0.177$; $F2$ $(3, 90) = 27.004$; $p < 0.0001$; $\eta p2= 0.474$) and an interaction between Group and Condition only by item ($F2$ $(3, 90) = 2.716$; $p < 0.049$; $\eta p2= 0.083$). In the post-hoc comparisons, we analyzed the proportion of pronouns produced by the two groups in each condition separately. We did not find a main effect for the 1-referent condition but we found a main effect of Group in all the 2-referents conditions (both panels: $F1$ $(1, 33) = 8.937$; $p < 0.005$; $F2$ $(1, 30) = 17.045$; $p < 0.0001$; one panel, only in the analysis by subject: $F1$ $(1, 33) = 3.843$; $p < 0.058$; gender ambiguous: $F1$ $(1, 33) = 5.796$; $p < 0.022$; $F2$ $(1, 20) = 17.621$; $p < 0.0001$), showing that L2 speakers produced a higher number of pronouns than native English speakers in conditions in which two referents were present in the discourse and or in the visual scene.

4. Discussion & Conclusion

The story telling task showed a main difference between native and L2 participants. For native speakers, even when a pronoun would not make the sentence ambiguous, the presence of another character in the discourse (either in the first panel or in both panels) decreased pronoun use to refer to the most prominent character in the discourse, confirming previous findings by Arnold & Griffin (2007). For the L2 group, results showed a higher production of pronouns than in native speakers when there were two characters in the preceding discourse with similar or different gender. L2 learners produced significantly more pronouns than native speakers also in the 2-referents (gender ambiguous) condition, in which the use of a pronoun leads to ambiguity. We hypothesize that L2 participants failed to suppress the assumption from their native language that overt pronouns are interpreted as referring to a non-topic referent. The learners produced more overt pronouns in two-referent contexts than native speakers of English likely because an overt pronoun is more explicit for them than a null form. While it may seem that they are being more explicit, in fact they are not yet explicit enough as an English native speaker. This result suggests that the production of referring expression is susceptible to L1 interference in highly proficient learners of English and is in line with previous studies on learners of

null-subject languages whose L1 is a non null-subject language (e.g., Belletti et al., 2007). Previous research has shown that the discourse distribution of overt and null-subjects can be non-target-like, with L2 speakers overproducing overt pronouns in their L2 in situations in which null subjects would have been pragmatically more appropriate, thus experiencing cross-linguistic interference from the L1. In our study, we looked at highly proficient learners of English (non null-subject language) whose L1 is Spanish (null-subject language), and showed that L2 participants overproduced pronouns in the L2 in a situation in which native speakers preferred to produce full NPs, proving evidence for a similar cross-linguistic interference. When the choice of referential expression is different between the L1 and the L2 (as in Spanish compared to English) the production of referring expression is susceptible to L1 interference even at a high level of proficiency.

In contrast with the production results, the processing data suggest that L2 learners process pronouns in English similarly to native speakers when gender and the first-mention bias are manipulated. L2 learners used gender information to quickly resolve a gender unambiguous pronoun. This result suggests that the L2 group has acquired the first-mention bias in English and can rapidly use it to guide pronoun interpretation when either gender on the pronoun is informative or not. We should notice that while gender is a cue that is also present in the learners' L1 (Spanish also marks semantic gender on third person pronouns), the first mentioned-bias has to be learned through language experience. In this respect, our results suggest that L2 learners are not just sensitive to the occurrence of third person pronouns as frequently co-referring with the subject of the previous sentence but rather can use this bias during processing in a native-like fashion. To our knowledge, this is the first study that shows the use of semantic gender and first mentioned bias on pronoun resolution in L2. We should point out that a study on the processing of pronouns in a non null-subject language (Dutch) has shown a processing penalty for L2 whose L1 was either a null subject (Turkish) or a non-null subject (German) language (Roberts et al., 2008). Our results do not contradict Roberts et al. (2008), but rather complement their findings. Roberts et al. used a reading task during eye-tracking to test local vs. disjoint interpretation of Dutch pronouns; in that study, the condition in which L2 learners experienced processing difficulties contained a potentially ambiguous pronoun and two entities in the discourse (e.g., *Peter and Hans are in the office. While Peter is working, he is eating a sandwich*) that native speakers resolved locally. Our sentence material is different than Roberts et al.'s and it is aimed at testing different aspects of the processing of pronouns in English (i.e., semantic gender and first-mentioned bias use) in sentences that are not globally ambiguous. We do not exclude that processing of pronouns is more costly in L2 compared to native speakers, but we have demonstrated here that under certain circumstances, pronoun resolution can be effortless and that highly proficient L2 speakers can make use of native-like strategies. Our production results, on the other hand, point to a residual difficulty emerging when L2 speakers have to choose a referential form between a range of options that is different from that of their native language (i.e., full pronouns and full NP in English; null and full pronouns in Spanish). Even though L2 participants have acquired the constraints on third person pronouns in English as shown by the comprehension results, they show residual indeterminacy in their pronoun choice as demonstrated with learners of a null-subject language (e.g., Sorace & Filiaci, 2006; Belletti, et al., 2007). To conclude, our study provides an investigation of referential expressions in L2; it explored both production and comprehension, showing an asymmetry between the two modalities and providing new data on the acquisition of a non-null subject language, English, which had not been investigated before.

References
Argyri, E., & Sorace, A. (2007). Crosslinguistic infl uence and language dominance in older bilingual children. *Bilingualism: Language and Cognition, 10,* 79 – 99.

Arnold, J. (2010). How Speakers Refer: The Role of Accessibility. Language and Linguistics Compass, 4, 187-203.

Arnold, J. & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language,* 56(4), 521–536.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eye tracking. *Cognition, 76,* B13–B26.

Belletti, A., Bennati, E., & Sorace, A. (2007). Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language and Linguistic Theory, 25,* 657 – 689.

Filiaci, F., Sorace, A. & Carreiras, M. (2013). Anaphoric biases of null and overt subjects in Italian and Spanish: a cross linguistic comparison. *Language and Cognitive Processes.* <DOI: 10.1080/016990965.2013.801502>

Fukumura, K., van Gompel, R. P. G., & Pickering, M. J. (2010). The use of visual context during the production of referring expressions. *Quarterly Journal of Experimental Psychology,* 63, 1700–15.

Haznedar, B. (2010). Transfer at the syntax-pragmatics interface: Pronominal subjects in bilingual Turkish. S*econd Language Research*, 26, 355 – 378.

Kaiser, E., & Trueswell, J. (2008). Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes*, 23(5), 709748.

Kaiser, E. (2011). Focusing on pronouns: Consequences of subjecthood, pronominalisation, and contrastive focus, *Language and Cognitive Processes,*26, 1625-1666.

Keating, G., VanPatten, B., & Jegerski, J. (2011). Who was walking on the beach: Anaphora resolution in Spanish heritage speakers and adult second language learners. *Studies in Second Language Acquisition*, 33, 193-221.

Marian, V., Blumenfeld, H.K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50, 940-967.

Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism: Language and Cognition, 7,* 125 – 142.

Liceras, J. M. (1988). Syntax and stylistics: More on the pro-drop parameter. In J. Pankhurst, M. Sharwood Smith, & P. Van Buren (Eds.), *Learnability and second languages: A book of readings* (pp. 71 – 93). Dordrecht: Foris.

Montrul, S. (2006). Incomplete acquisition as a feature of bilingual and L2 grammars. In R. Slabakova, S. A. Montrul, & P. Prévost (Eds.), *Inquiries in linguistic development: In honor of Lydia White* (pp. 335 – 359). Amsterdam : Benjamins.

Montrul, S., & Rodríguez Louro, C. (2006). Beyond the syntax of the null subject parameter: A look at the discourse-pragmatic distribution of null and overt subjects by L2 learners of Spanish. In V. Torrens & L. Escobar (Eds.), *The acquisition of syntax in romance languages* (pp. 401 – 418). Amsterdam : Benjamins.

Paradis, J., & Navarro, S. (2003). Subject realization and crosslinguistic interference in the bilingual acquisition of Spanish and English: What is the role of the input? *Journal of Child Language, 30,* 371 – 393.

Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition,* 30: 333-357.

Rothman, J. (2008). How pragmatically odd! Interface delays and pronominal subject distribution in L2 Spanish. *Studies in Hispanic and Lusophone Linguistics, 1,* 317 – 339.

Rothman, J. (2009). Pragmatic deficits with syntactic consequences? L2 pronominal subjects and the syntax-pragmatics interface. *Journal of Pragmatics, 41,* 951 – 973.

Silva-Corvalán, C. (1994). *Language contact and change: Spanish in Los Angeles.* Oxford: Oxford University Press.

Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, 22(3), 339-368.

Sorace, A., Serratrice, L., Filiaci, F., & Baldo, M. (2009). Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. *Lingua*, 119, 460 – 477.

Tsimpli, I., Sorace, A., Heycock, C., & Filiaci, F. (2004). First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism*, 8, 257 – 277.

Wilson, F. (2009). *Processing at the syntax-discourse interface in second language acquisition*. University of Edinburgh (Unpublished PhD dissertation), University of Edinburgh.