Null and Overt Subject Biases in Spanish and Italian: A Cross-linguistic Comparison

Francesca Filiaci
University of Edinburgh

1. Introduction

Over the last twenty years a great deal of linguistic research has investigated how anaphoric expressions retrieve their antecedents in the discourse showing that a variety of pragmatic factors together with grammatical and cognitive constraints contribute in determining the distribution of different types of expressions. A particularly interesting case for the study of such phenomena is that of Null Subject Languages, and of the principles that in such languages underlie the alternation of null and overt subjects (NSs and OSs).

Many languages have the possibility of leaving the subject of a tensed verb phonetically empty and this generalisation is captured by the Null Subject Parameter (for a formulation of the parameter see Rizzi, 1986a, 1997). As Rizzi points out, in order for a subject to be dropped it has to be possible to recover its content through the context that is overtly realised. The mechanisms that allow for such recovery vary cross-linguistically, nevertheless it has been observed that NS languages often possess a rich verbal morphology, expressing overtly tense and, more crucially, person features on the verbal head. This observation is valid for Spanish and Italian as we can see from the Indicative Present morphology shown below.

Italian	Spanish
parl*o*	habl*o*
parl*i*	habl*as*
parl*a*	habl*a*
parl*iamo*	habl*amos*
parl*ate*	habl*ais*
parl*ano*	habl*an*

Given the morphological similarities and the fact that the two languages are closely typologically related, it seemed plausible to assume that we are dealing with an equivalent phenomenon and much cross-linguistic research has assumed, more or less explicitly, that the pragmatic principles that determine the distribution and interpretation of null and overt pronominal subjects should be roughly equivalent in these two languages (see for example Sorace et al., 2009; Serratrice et al., 2009; Alonso-Ovalle et al., 2002).

As a matter of fact linguistic phenomena at the interface between syntax and pragmatics have increasingly attracted the attention of researchers, since it is emerging that interfaces are the areas of language that are vulnerable to cross-linguistic influence in situations of contact, like bilingual first language acquisition, child and adult second language acquisition, language attrition etc. (among others Hulk & Müller, 2000; Müller & Hulk, 2001; Paradis & Navarro, 2003; Serratrice et al., 2004; Sorace et al., 2009). It is therefore crucial to understand the principles governing interface phenomena in different languages before considerations about cross-linguistic influence can be drawn.

The aim of this study is to compare directly data from Italian and Spanish and find the extent to which the assumptions above are true. More precisely, this study tests the validity in Spanish of

the Position of Antecedent Strategy (PAS), a parsing strategy proposed by Carminati (2002) for intra-sentential anaphora resolution in Italian.

In the next sections we will introduce the PAS and review the research already conducted on Italian and Spanish; on the basis of the evidence available we will formulate a more precise hypothesis about the possible cross-linguistic differences between the two languages. We will then describe the methodology used and finally present and discuss the results.

2. The Position of Antecedent Strategy

The PAS is a pragmatically motivated parsing strategy proposed by Carminati (2002) for intra-sentential anaphora, suggesting that Italian NSs and OSs are biased towards looking for their antecedents in different structural positions within the preceding syntactic structure. The PAS is formulated as follows (Carminati, 2002:33):

The Position of Antecedent Strategy:

The null pronoun prefers an antecedent which is in the SpecIP position (or in the AgrS position under Pollocks split INFL hypothesis), while the overt pronoun prefers an antecedent which is not in the SpecIP position.

Carminati tested the validity of the PAS through a series of self-paced reading experiments and sentence completion tasks. Her experiments manipulated different variables, like for example the structural position of possible antecedents, their number and the use of gender, number, and person features as cues for the resolution of the anaphora. The basic evidence in support of the PAS comes from a self-paced reading experiment manipulating the structural position of the antecedent. The experiment consisted of semantically disambiguated sentences formed by a subordinate clause, introducing two human referents of the same gender, followed by a main clause starting with either a NS or an OS that could ambiguously co-refer with either the preceding subject or object. The four conditions tested in the experiment are shown in example (1) below:

(1)   a.   Quando Maria$_i$ è andata a trovare Vanessa$_j$ in ospedale, lei$_{i(/j)}$ le ha portato un mazzo di fiori.

         b.   Quando Maria$_i$ è andata a trovare Vanessa$_j$ in ospedale, $\emptyset_{i(/j)}$ le ha portato un mazzo di fiori.
            'When Maria went to visit Vanessa at the hospital, she brought her a bunch of flowers.'

         c.   Quando Maria$_i$ è andata a trovare Vanessa$_j$ in ospedale, lei$_{(i/)j}$ era già fuori pericolo.

         d.   Quando Maria$_i$ è andata a trovare Vanessa$_j$ in ospedale, $\emptyset_{(i/)j}$ era già fuori pericolo.
            'When Maria went to visit Vanessa at the hospital, she was already out of danger.'

Sentences (1a) and (1b) are semantically plausible if the anaphoric subject of the main clause co-refers with the preceding subject (Maria), whereas in (1c) and (1d) the anaphora has to co-refer with the preceding object (Vanessa) for the sentence to be plausible. Carminati measured the reading times for the second clause of each sentence and found a significant processing penalty (indicated by longer reading times) when a NS was forced, against its processing bias, to retrieve an antecedent in the object position (as in condition (1d)) and similarly when an OS was forced to retrieve an antecedent in the preverbal subject position (that is in condition (1a)). While these results support the validity of both biases formulated in the PAS, other experiments in the same study showed that overriding the OS bias seems less costly than overriding the NS bias, and that a violation of the NS bias yields smaller penalties in extra-sentential contexts compared to intra-sentential ones.

Notice that, according to Carminati, the biases encoded by the PAS are not motivated by general cognitive limitations, as it is the case for other parsing strategies, nor do they belong to the core grammar, since their violation does *not* produce outright ungrammaticality: they simply predict preference of interpretation. Based on these facts, Carminati concludes that the PAS is motivated by universal pragmatic principles, and she refers in particulare to Accessibility Theory (AT) principles (Ariel, 1990).

The basic idea of AT is that all types of referring expressions available in a given language can be ranked according to three criteria: informativity (the amount of information contained in a referring

expression), rigidity (how uniquely an expression refers to an antecedent) and attenuation (related to the phonological size of the expression). A speaker will chose which expression to use to refer to a certain entity depending on how 'accessible' s/he thinks the mental representation of the entity is to the addressee. Highly accessible entities, that is entities that are (deemed to be) readily available in the mind of the hearer, will be referred to using expressions that are relatively less informative, less rigid and more attenuated, or 'high accessibility markers' (for example pronouns); entities that are deemed to be less accessible in the mind of the addressee will be referred to using 'low accessibility markers', that is to say expressions that are highly informative, rigid and less attenuated (like definite NPs or proper names). In order to chose a given expression to retrieve an antecedent it is crucial to know what influences the antecedent degree of accessibility. Ariel (1990, 1991) mentions: discourse cohesion, recency of mention, number of competitors; but more in general any factor that influences the prominence of an antecedent will have an impact on its accessibility. Topicality (Almor, 1999) or lexical frequency of the antecedent (van Gompel & Majid, 2004) have been shown to influence the salience of an entity in the discourse; on the other hand, the PAS states clearly that the prominence of an antecedent, at least in the type of discourses examined, is determined by its syntactic position: the higher the structural position, the more prominent the expression. In the sentences tested by Carminati, the highest position is occupied by the preverbal subject, and which is assumed to be placed in SpecIP.

3. Cross-linguistic considerations on the PAS

There are two aspects to consider in order to make predictions about the cross-linguistic validity of the PAS: the first one concerns the cross-linguistic validity of the pragmatic principles that motivate it; the second one the syntactic structure of the languages analysed.

3.1. Pragmatic considerations

From the point of view of the pragmatic principles, Ariel (1990, 1991) argues that the ranking of referring expressions from low to high accessibility markers is universal. An example of such scale is reported below (from Ariel, 2006):

> Full name + Modifier > Full name > Long definite description > Short definite description > Last name > First name > Distal demonstrative (+ Modifier) > Proximal demonstrative (+ Modifier) > Stressed pronouns + Gesture > Stressed pronouns > Unstressed pronouns > Cliticizd pronoun > Verbal person agreement markers > ∅[1]

Notice that Ariel claims that only the relative ranking is universal, as matter of fact, we should expect cross-linguistic variation at least along two dimensions. Firstly, the inventory of expressions can change from language to language, for example NS languages will have ∅ in their inventory, non-NS languages will not. Secondly, it is plausible to expect that the relative distance between expressions along the scale may vary form language to language. This means that, even in languages with the same inventories, expressions that could be considered morphologically or semantically equivalent (like personal or demonstrative pronouns) may still be used to retrieve antecedents with different degrees of accessibility within each language.

In other words, given two languages with similar inventories, the least informative, most attenuated and least rigid expression available should always be placed at the bottom of the scale and used as the highest accessibility marker, which means that we would *never* expect to find a language where ∅ prefers to retrieve a *less prominent* antecedent *than* an overt pronoun. As for overt (third person) personal subject pronouns, they should be situated, both in Italian and in Spanish, just above ∅s, given that they express overtly person, number, and gender features, but this doesn't necessarily mean that their distance from ∅ is the same in both languages or that they should prefer antecedents with the *same* level of accessibility across the two languages. As a matter of fact, we predict to find languages in which

---

[1] Strictly speaking, Italian and Spanish would not be considered by Ariel as including ∅ in their inventories, since person and number features are obligatorily expressed by the verbal agreement markers. The two languages are equivalent in this respect and the cross-linguistic considerations discussed here remain valid even if we assume that their accessibility markers scale ends at 'Verbal person agreement markers' rather than ∅.

personal pronouns are used specifically to retrieve antecedents that are inaccessible to ∅s (like Italian, according to Carminati's data), but languages in which overt pronouns and ∅s are relatively close to each other, are not ruled out (see for example Bresnan (1997, 2001) discussion of Chicheŵa *unmarked* pronoun). What *is* ruled out by Ariel's scale is a language where an overt pronoun can only retrieve antecedents that are more accessible than the antecedents retrieved by ∅.

3.2. Syntactic considerations

As for the antecedents structural position, that Carminati assumes to be the crucial factor to determine prominence in the type of discourses analysed, we are interested here in the preverbal subject position and in the object position. There is no agreement in the literature regarding the structural position occupied by preverbal subjects in NS languages: Cardinaletti (1997), presenting data from Italian and Italian dialects, claims that in they occupy a position internal to the sentence, the specifier of the functional projection AgrSP. Other researchers (Alexiadou & Anagnostopoulou, 1998; Ordóñez & Treviño, 1999), making reference in particular to (Greek and) Spanish data, argue that Spanish preverbal subjects occupy a higher, left dislocated A' position.

As for the position of postverbal object antecedents, this could be another locus of cross-linguistic variation between the two languages analysed, since in Spanish direct objects introduced by the dative preposition *a* are argued to raise to a particular structural position (the specifier of the functional projection v, taking VP as its object), whereas this movement is not required in Italian direct or indirect objects (Torrego, 1998).

What is crucial for our cross-linguistic comparison at the moment is not so much whether the position of preverbal subjects and postverbal objects is exactly the same across the two languages, but rather the fact that in each language such subjects occupy a relatively higher, and therefore relatively more prominent, position than lower (direct and indirect) objects.

To summarise, from a syntactic point of view, we assume here that the preverbal subject occupies the most prominent position in the sentence and should therefore be relatively more accessible than its competitors. This means that, given the pragmatic considerations discussed in the previous section, we expect that in both Italian and Spanish ∅ should be better at retrieving a subject antecedent than an object antecedent; we don't expect to find a language in which an OS is better than a ∅ at retrieving a subject and worse than ∅ at retrieving an object antecedent; but we can't necessarily make a prediction with regard to the specific preferences of the OS: it may be a lot faster than ∅ at retrieving the non-prominent antecedent, if it is specialised to retrieve antecedents that are relatively less accessible, but it may show no definite preference for non-prominent vs. prominent antecedents, if it is relatively close to ∅ along the accessibility markers scale.

4. Empirical Data on Spanish

Some empirical data on the use and interpretation of OS pronouns in Spanish may help us make more precise predictions about their behaviour compared to the Italian ones. Although Carminati's experiments have not been replicated in Spanish up to now, there has been at least one explicit attempt to test the validity of the PAS in this language using psycholinguistic methods by Alonso-Ovalle et al. (2002). In addition, a number of variationist studies has investigated the distribution of NS and OS in different varieties of Spanish as well as possible changes in their use in situations of language contact.

Starting with the latter type of studies, Cameron (1992) conducted a large corpus study on Puerto Rican Spanish spoken in San Juan. He found a significant correlation between the expression of overt pronominal subjects and switch in subject reference as he found that personal subject pronouns tend to be used significantly more often to refer to an antecedent that is not the subject of the previous tensed verb. Conversely, NSs were found to be used significantly more often to maintain the reference to the subject of the previous tensed verb. These results are confirmed by many other studies on varieties of Spanish spoken in the US, Mexico and the Caribbean (for a review see Flores-Ferrán, 2004), as well as for Iberian Spanish (Enríquez, 1984).

If on the one hand these results suggest that the alternation of NS and OS in Spanish must be governed by the sort of biases encoded by the PAS, on the other hand they are also compatible with a situation where only one type of subject (that would have to be the NS) is subject to pragmatic

restrictions, but not the other type. That is to say, if we are counting the percentages of two expressions in complementary distribution, only one needs to be governed by a bias, whereas the other may be *unmarked* (in Bresnan's 2001 sense) and we would still obtain the same distribution. So this evidence cannot help us directly to refine our hypotheses concerning the OS behaviour in Spanish.

As mentioned in the introduction, areas of language where the choice between syntactic options is affected by discourse pragmatics have been shown to be vulnerable to cross-linguistic influence in situations of language contact (Müller & Hulk, 2001; Hulk & Müller, 2000; Serratrice et al., 2004; Serratrice, 2007; Tsimpli et al., 2004; Sorace & Filiaci, 2006). This means that the association between OS pronouns and switch reference should become underspecified under the influence of English, in which OS use is not restricted by pragmatic principles.

This hypothesis seems to be supported by empirical data in Italian (Tsimpli et al., 2004; Sorace & Filiaci, 2006; Sorace et al., 2009), whereas the evidence is not considered to be very robust for Spanish: Flores-Ferrán (2004), discards the hypothesis that English contact may have affected the use of subject pronouns in a group of Puerto Rican Spanish speakers living in New York, in the sense of obliterating the difference between switch and non-switch reference. Silva-Corvalán (1994), studied a group of Mexican speakers resident in Los Angeles and found evidence of English influence in a number of constructions, but no clear evidence of increased use of OSs as an influence of English. Montrul (2004), found changes in both NS and OS use in a group of heritage speakers of Mexican Spanish in the US, but her results, are difficult to interpret in support of the hypothesis of change induced by contact with English since they show that also the pragmatic restrictions on the use of NS seems to be affected in these speakers, but English does not allow NS, so it shouldn't in principle affect their use.

Finally, Alonso-Ovalle et al. (2002) run a psycholinguistic study that was not directly meant to replicate Carmianti (2002), nevertheless the authors claim that their data confirm the validity of the PAS for Spanish. Alonso-Ovalle et al. asked participants to identify the antecedent of the anaphoric subject in ambiguous sentences like the following:

(2)   a.   Juan$_i$ pegó a Pedro$_j$. Él$_{i/j}$ está enfadado.

          b.   Juan$_i$ pegó a Pedro$_j$. $\emptyset_{i/j}$ está enfadado.
             'Juan hit Pedro. He is angry.'

The results show that with a NS (like in (2b)) 73% of the participants identify the previous subject as the antecedent but this preference drops to 50.2% with an OS yielding a highly significant difference. The authors infer that the PAS makes valid predictions in Spanish, but if we consider the data more carefully, we notice that the antecedent preferences for the OS are at chance level, or in other words, that a sentence containing an OS pronoun remains permanently ambiguous, which would confirm the impressionistic observations by one (of two) Spanish native speaker, reported in Carminati's study, who found 'the overt pronoun more ambiguous than the null.' (Carminati, 2002: 196)[2].

4.1. Hypotheses for the present study

The data reviewed above suggest that there may be slight differences regarding the use of OSs in Spanish and Italian, more precisely, we argue that the second clause of the PAS, the one concerning the OS bias, may not apply to Spanish leaving Spanish OSs free(er) to retrieve both prominent and non-prominent antecedents without incurring a significant processing penalty. This type of cross-linguistic variation would be compatible with AT restrictions and with the findings of corpus based variationist studies, as pointed out above. A lack of a direct association between OS use and switch in subject reference may also shed light on why it has been so difficult to find evidence for the loss of pragmatic restrictions on the OS use due contact with English.

---

[2]Alonso-Ovalle et al. argue that the lack of cross over effect may be due to the use of extra sentential anaphora, and claim that also in Italian the set of biases observed in extra-sentential anaphora are milder than the ones observed in intra-sentential anaphora cases. These claims are not completely true. What Carminati found is that a violation of the (null) subject bias yields smaller penalties in extra-sentential contexts compared to intra sentential ones, but this cannot explain the fact that in Spanish we do find a bias guiding the null subject resolution but no bias at all for the overt subject.

In order to test this hypothesis, we used the same experimental paradigm as Cariminati (2002) and run the same experiment in Spanish and Italian. The procedure is described in the next section; we expect to find longer reading times when an anaphora is forced to co-refer with an antecedent against its processing bias and take this difference as an indication of increased processing cost.

5. Method
5.1. Participants

The participants in the study were 32 adult monolingual speakers of Italian (mean age 26.6, sd = 5.6) and 32 adult monolingual speakers of Spanish (mean age 26.47, sd = 5.5). They were recruited in Edinburgh among international students at Edinburgh University, and among people enrolled in summer English Language courses in different schools around the town. Participants had been living in Edinburgh (or in an English speaking country) for a maximum of a few months at the time of the experiment (for the Italian group the mean number of months spent abroad was 2.4, sd = 4.3; for the Spanish group the mean was 3, sd = 4.9) Spanish speakers were asked about their origin and only speakers from Spain were included in the study in order to control for some of the variation found in different varieties of the language.

5.2. Materials and design

The experimental materials and design were the same as those used by Carminati (2002) for Experiment 1, and are illustrated in (1a) to (1d). The experimental sentences were adapted and translated into Spanish in order to obtain two equivalent sets of items. Each version of the materials (Italian and Spanish) was checked separately by a monolingual native speaker. The sentences were formed by two clauses, the first clause was always a subordinate clause (temporal or reason clause) that introduced two antecedents of the same gender, one in the subject and one in the object position[3]. The second clause was the main clause and it began with either a null or an overt pronominal subject which could ambiguously refer to either of the two antecedents. The meaning of the sentence would disambiguate the antecedent of the second clause subject as it is shown in examples (3) for Italian and (4) for Spanish:

(3)   a.   Dopo che Giovanni$_i$ ha criticato Franco$_j$ così ingiustamente, lui$_{i(/j)}$ si è scusato ripetutamente.

   b.   Dopo che Giovanni$_i$ ha criticato Franco$_j$ così ingiustamente, $\emptyset_{i(/j)}$ si è scusato ripetutamente.
      'After that John$_i$ has criticised Franco$_j$ so unjustly, he$_{i(/j)}$ apologized repeatedly.'

   c.   Dopo che Giovanni$_i$ ha criticato Franco$_j$ così ingiustamente, lui$_{(i/)j}$ si è sentito offeso.

   d.   Dopo che Giovanni$_i$ ha criticato Franco$_j$ così ingiustamente, $\emptyset_{(i/)j}$ si è sentito offeso.
      'After that John$_i$ has criticised Franco$_j$ so unjustly, he$_{(i/)j}$ felt offended.'

(4)   a.   Cuando Ana$_i$ visitó a María$_j$ en el hospital, ella$_{i(/j)}$ le llevó un ramo de rosas.

   b.   Cuando Ana$_i$ visitó a María$_j$ en el hospital, $\emptyset_{i(/j)}$ le llevó un ramo de rosas.
      'When Ana$_i$ visited Mary$_j$ in the hospital, she$_{i(/j)}$ brought her a bunch of roses.'

   c.   Cuando Ana$_i$ visitó a María$_j$ en el hospital, ella$_{(i/)j}$ ya estaba fuera de peligro.

   d.   Cuando Ana$_i$ visitó a María$_j$ en el hospital, $\emptyset_{(i/)j}$ ya estaba fuera de peligro.
      'When Ana$_i$ visited Mary$_j$ in the hospital, she$_{(i/)j}$ was already out of danger.'

Two variables were manipulated: the anaphoric subject in the main clause, which could be either null or overt; and the structural position of its antecedent, which could be either in the preverbal subject position

---

[3]The object could be either a direct or an indirect object, and in Spanish the direct object was always introduced by the dative preposition *a*; as pointed out in section 3.2, these objects occupy a particular structural position, different from Italian direct or indirect objects Torrego (1998). Leaving the details aside, here we are simply assuming that the preverbal subject both in Italian and in Spanish occupies a higher (and therefore relatively more prominent) structural position than any of the verb complements and should therefore be the most accessible antecedent in the type of sentences used, in both languages.

(like in sentence (3a-b) and (4a-b)), or in the (postverbal) object position (sentences (3c-d) and (4c-d)). The experimental materials consisted of 16 experimental sentences and 86 filler sentences[4].

5.3. Procedure

The experimental hypotheses were tested using a moving window self-paced reading task. The items were presented clause by clause and the reading times for each clause were recorded. The experiment was run on an iBook laptop with a 12" screen, the software used was Psyscope X. The responses were provided through the laptop keyboard. The instructions were presented in written form, at the beginning of the experiment, in the same language as the experiment. The participants were instructed to read normally and to press the space bar to see the following part of the sentence. Half of the items were followed by comprehension questions asking to identify the antecedent of the main clause (anaphoric) subject. The two possible answers to the question appeared at the bottom of the screen, to the left and to the right, and the participants were instructed to chose the correct one using the 'F' key (under their left index finger) for the left answer or the 'J' key (under their right index finger) for the right answer. The order of presentation of the sentences was randomised at every run.

6. Results

We analysed the mean reading times for the second clause of the experimental items; this is the clause that contained the null or overt subject anaphora to be resolved. Longer reading times are taken to indicate a processing penalty in the resolution of the anaphora. The mean reading times for Italian and Spanish participants are shown in Table 1 and 2.

		N	Mean	Std Deviation	Minimum	Maximum
NULL	Object Antecedent	32	2569.53	797.95	1221.75	4303.25
SUBJECT	Subject Antecedent	32	1940.92	635.93	1022.00	4031.00
OVERT	Object Antecedent	32	2266.43	698.29	960.75	4364.00
SUBJECT	Subject Antecedent	32	2749.77	776.67	1435.50	4264.75

Table 1: Descriptive statistics for Italian: mean reading times (msec) for the second clause of pragmatically disambiguated sentences.

		N	Mean	Std Deviation	Minimum	Maximum
NULL	Object Antecedent	32	2319.09	915.98	1131.25	4145.25
SUBJECT	Subject Antecedent	32	1998.32	587.26	1043.75	3748.50
OVERT	Object Antecedent	32	2389.32	784.52	1302.25	4354.33
SUBJECT	Subject Antecedent	32	2507.60	880.53	1351.75	4214.50

Table 2: Descriptive statistics for Spanish: mean reading times (msec) for the second clause of pragmatically disambiguated sentences.

As for the Italian data it replicates the results obtained by Carminati (2002): in the NS condition there is a significant penalty when the $\emptyset$ anaphora is semantically forced to corefer against its bias with the previous object (2569.53 msec. vs.1940.92 msec.; $F (1, 31) = 18.831; p < .001$); in the OS condition the opposite effect obtains and we find a significant penalty when the OS is forced to corefer with a subject antecedent (2749.77 msec. vs. 2266.43 msec.; $F (1, 31) = 18.124; p < .001$).

As for the Spanish data, similarly to Italian, we find a significant penalty when the NS is forced to corefer with an object rather than a subject antecedent (2319.09 msec. vs. 1998.32 msec.; $F (1, 31) =$

---

[4]The filler sentences tested other types of anaphora in different experimental conditions and were taken from Carminati (2002) Experiment 6, 8 and 10.

5.0327; p = .032), but crucially we find no significant difference in the OS condition for the pronoun retrieving its antecedent in any structural position (2389.32 msec. vs. 2507.6 msec.).

The data from the comprehension questions confirm the pattern found in the reading times for the sentences. The percentage of correct answers is shown in table 3 and the average times taken for reading and answering the questions are shown in table 4.

Anaphora	Antecedent	Italian	Spanish
Null	Object	73%	67%
Null	Subject	**95%**	**84%**
Overt	Object	**89%**	73%
Overt	Subject	73%	**88%**

Table 3: Percentage of correct answers to the comprehension questions in each condition.

Anaphora	Antecedent	Italian	Spanish
Null	Object	4344.82	3190.65
Null	Subject	**2883.41**	**2687.30**
Overt	Object	**3126.23**	3183.38
Overt	Subject	4100.58	2840.02

Table 4: Mean RTs (combination of reading time and answering time in milliseconds) for the comprehension questions in Italian and Spanish.

From a first look at the data, in Italian we see a higher accuracy and speed when a NS retrieves a subject antecedent and an OS retrieves an object antecedent as we would predict from the PAS. In Spanish, there seems to be higher speed and accuracy when the antecedent is in the subject position, no matter if the anaphora is null or overt. An ANOVA was run on the RTs. In the Italian data there is a significant effect for Antecedent with questions read and answered significantly faster for Subject than Object antecedents (6984 msec. vs. 7471 msec.; *F(1, 30) = 7.699; p = .009*), and a significant interaction between Antecedent and Anaphora (*F(1, 31) = 10.677; p = .002*). The Spanish data showed a significant effect for Antecedent, with faster answers to sentences with Subject rather than Object antecedents (5527 msec. for Subjects and 6374 for Objects, *F(1, 30) = 5.6641; p = .024*). A separate analysis of the NS and OS conditions though, revealed that the statistical significance for this effect must come from the NS condition alone (*F(1, 31) = 4.395; p = .044*), since no significant difference is found when the OS condition is analysed on its own.

Finally, we analysed the Italian and the Spanish data together, using Language as a between subjects variable. The raw data was adjusted to take into account the systematic differences in the length of the stimuli between the two languages. This was done computing the deviations from regression[5]. The data obtained is plotted in Figure 1.

The ANOVA performed on the transformed data showed a significant main effect for the Anaphora *F (1, 62) = 6.3136; p = .0145* with NS anaphora read overall significantly faster than OS anaphora (-240 msec vs. 307.99 msec); a significant interaction of Anaphora by Antecedent *F (1, 62) = 37.8143; p < .000* showing that NSs are read faster when they retrieve a subject antecedent and that OSs are read faster when they retrieve an object antecedent. Most interestingly, there is a significant three-way interaction between Language, Anaphora and Antecedent (*F (1, 62) = 6.6452; p = .0123*). In order to understand the three-way interaction, we further analysed the OS and the NS conditions separately, looking at the effects of the other variables. For the OS condition, this analysis reveals a highly significant effect for

---

[5]This is the standard way to adjust reading times taking into account the length of the stimuli. For each participant the expected reading time for each stimulus is calculated in relation to its length. The expected reading times are then subtracted from the observed times; so, a negative value indicates a reading time faster than expected and a positive value slower than expected.

Spanish and Italian (adjusted RTs)
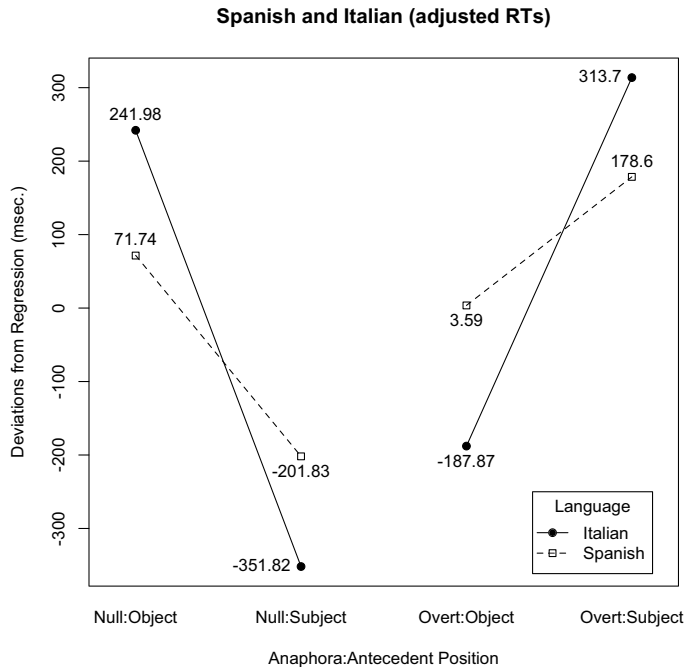
Figure 1: Adjusted RTs for the second clause of pragmatically disambiguated sentences in Spanish and Italian.

Antecedent (*F(1, 62) = 16.123; p < .000*), and a marginally significant interaction between Language and Antecedent (*F (1, 62) = 3.7574; p = .057)*. In the NS condition there is a significant effect for Antecedent (*F(1, 62) = 16.123; p < .000*), no main effect for Language and no significant interactions.

A last analysis on the OS condition for the two languages combined revealed that for a sentence that contains an OS retrieving an object antecedent, reading times are significantly faster in Italian than in Spanish (-187.87 vs. 3.59; *F (1, 62) = 4.4235; p = .0395*), whereas no significant difference is found between the two languages when the OS retrieves a subject antecedent.

7. Discussion and Conclusions

The results from this study seem to agree with the cross-linguistic considerations about the PAS discussed in section 3 and with the hypothesis about the Spanish OSs outlined in section 4.1, since on the one hand they confirm the cross-linguistic validity of the NS bias, on the other they suggest the presence of cross-linguistic differences on the pragmatic biases that guide the resolution of personal subject pronouns.

The cross-linguistic validity of the NS bias is confirmed by the significant penalty incurred in both Italian and Spanish when a NS is resolved against its bias, and by the lack of a significant interaction between the Language and Antecedent variables when the two languages are analysed together. As it was discussed in section 3, following AT considerations we expected this bias to hold universally.

As for the OSs they seem to be resolved differently in each language as it is suggested by the fact that in Italian the experiment yielded a significant processing penalty for sentences where an overt pronoun is forced to co-refer with the most prominent antecedent against its bias (replicating the results of Carminati (2002)), but the same construction in Spanish does not seem to produce any significant extra processing cost neither in terms of reading times nor in terms of ease and accuracy of the answers to the comprehension questions. As for this last piece of data, it actually seems that questions are always answered more easily in Spanish when they follow a sentence where the antecedent of the anaphoric subject is the previous subject, but this effect turns out to be fully significant only in the NS condition.

When the two languages are analysed together the presence of a cross-linguistic difference is confirmed by an interaction between the Language and Antecedent variables in the OS condition, although this difference only came out marginally significant (p = .057); this could be due to a lack of power of the test and should be investigated further.

To summarise, while Italian OSs seem to bear a strong 'Switch Reference' connotation, indicating to the addressee a switch of reference from the preceding subject, Spanish OSs seem not to carry such a strong connotation and be more compatible with a 'non-Switch Reference' reading. So, while in Italian both NS and OS are specialised in retrieving different antecedents, in Spanish only NSs seem to obey a bias and we may argue that this bias alone could be responsible for the patterns of distributions found in corpus studies. Remember that the PAS does not predict ungrammaticality, but simply preference of interpretation and the type of variation documented by the present study is compatible with AT predictions about possible languages.

But what could be the reason(s) for this cross-linguistic difference? The question can be approached from two angles: cross-linguistic differences in the position of the antecedents and cross-linguistic differences in the nature of the personal pronouns.

As for the first possibility, we have hinted in section 3.2 that there are cross-linguistic differences regarding the syntactic position of both subjects and objects in Spanish and Italian: Spanish objects may occupy a relatively higher position than Italian ones, and Spanish subjects may occupy a left dislocated position higher than SpecIP. Notice that the idea adopted by Carminati, that syntactic position may determine the prominence of an antecedent, is not uncontroversial, for example information structure is often assumed to play a crucial role, but in the type of sentences tested by Carminati (where the preverbal subject is always also a topic) these two variables cannot be teased apart. Nevertheless, if we follow Carminati's hypothesis and argue that Spanish objects may be 'closer' to subjects because of their higher structural position, and therefore sharing with subjects a similar level of accessibility, then we would expect both NSs and OSs to be affected by the structural differences. In other words, if subjects and objects are structurally closer to each other, then also the NS should retrieve them both with the same ease. The chart in figure 1 seem to provide some support for this idea, in the sense that the bias for the NS in Spanish seems to be somewhat weaker than the bias in Italian, as it can be seen from the slightly less steep slope; on the other hand, the statistical analyses did not yield any significant effect for Language in the NS anaphora condition, so we have to assume that the NS bias holds equally for Italian and Spanish. Therefore the hypothesis that the cross-linguistic differences between Italian and Spanish may be due to the different syntactic positions occupied by the antecedents does not seem to receive much support, at least from the data available at the moment.

As for the nature of the overt pronoun, we said that, according to Ariel, the distance between expressions along the accessibility markers scale can vary from language to language, but what does this mean in practice? Cardinaletti & Starke (1999), while producing a cross-linguitic typology of pronouns, provided evidence that the Italian pronominal system possesses two series of overt pronouns, a strong one, including *lui* and *lei* and a weak one, including *egli/ella*. One of the systematic semantic differences between weak and strong pronouns, the authors argue, is that weak pronouns, due to their structural deficiency, can co-refer with prominent discourse antecedents whereas strong pronouns can't; as a matter of fact Cardinaletti & Starke predict that in Italian *egli* and *pro* (which is also analysed as a weak pronoun) should pattern together and be in complementary distribution with *lui*, that, being strong, is only allowed in contexts where weak pronouns are independently ruled out. The anaphoric behaviour of the Spanish pronouns *él* and *ella* could then be taken as an indication of structural deficiency, in the sense of Cardinaletti & Starke.

Bresnan (1997, 2001) provides another cross-linguistic analysis of pronouns making predictions that are partly different from Cardinaletti and Starke, but compatible with the analysis of the Spanish OS anaphoric preferences presented here. Within the Optimality Theory framework, Bresnan argues that, although, cross-linguistically, reduced pronouns (∅, bound pronominals, clitics and weak pronouns) are used to refer anaphorically to topic antecedents *and* strong pronouns for focus or contrast (and are therefore excluded from topic-anaphoricty), only the reduced forms are marked by the grammar with a topic-anaphoricity (TOP) feature, whereas the strong forms are actually unmarked. The evidence comes from the fact that strong pronouns are incompatible with topic-anaphoricity only as long as they can alternate with a reduced form, but where a reduced form is unavailable (typically because of a gap in

the paradigm), then the strong form can be used instead without implying any meaning of contrast. This shows, according to Bresnan, that the strong pronoun is not inherently marked for non-topic-anaphoricity, it inherits this interpretation when it is used instead of a weaker form. As Bresnan (1997) points out, this sense of morphosyntactic *unmarkedness* as neutralisation of oppositions also used by Jakobson (1984) in his analysis of the Russian verb system. Following Bresnan's analysis then, the different interpretation of OS personal pronouns in Italian and Spanish should be related to a gap in the paradigm of reduced pronominal forms affecting Spanish, but not Italian, so what does the paradigm of Spanish reduced forms consist of exactly? Both Italian and Spanish allow for the use of the most reduced pronominal form, the NS, the availability of which has been traditionally related to the presence of a rich verbal morphology; moreover several analyses of the NS phenomenon have suggested that the verbal morphology may actually have a (pro)nominal content or should be analysed as a clitic pronoun, although there is no agreement in the literature regarding the details of the proposal (Rizzi, 1986b; Alexiadou & Anagnostopoulou, 1998; Ordóñez & Treviño, 1999; Platzack, 2004). In particular Ordóñez & Treviño (1999) present strong evidence supporting the analysis of verbal morphology as a clitic based on Spanish data, whereas Cardinaletti (1997) rejects it on the basis of Italian data and Alexiadou & Anagnostopoulou (1998) acknowledge Cardinaletti's objections and suggests that the Italian data may present some difficulties for their analysis.

These are only a very sketchy ideas, but it seems to me that the most promising way to approach the cross-lingustic differences between Italian and Spanish regarding the data presented in this paper is by investigating further the differences regarding the properties of their pronominal systems, the nature of the verbal morphology in the two languages and the interactions between the two.

References

Alexiadou, Artemis & Elena Anagnostopoulou (1998). Parametrizing Agr: Word order, V-movement and EPP-checking. *Natural Language and Linguistic Theory* 16, 491–539.

Almor, Amit (1999). Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review* 106:4, 748–765.

Alonso-Ovalle, Luis, Susana Fernández-Solera, Lyn Frazier & Charles Clifton (2002). Null vs. overt pronouns and the topic-focus articulation in Spanish. *Rivista di Linguistica* 14:2.

Ariel, Mira (1990). *Accessing Noun Phrase Antecedents*. Routlege, London.

Ariel, Mira (1991). The function of accessibility in a theory of grammar. *Journal of Pragmatics* 16, 443–463.

Ariel, Mira (2006). Accessibility theory. Brown, Keith (ed.), *Encyclopedia of language and linguistics*, Elsevier, Oxford, 15–18.

Bresnan, Joan (1997). The emergence of the unmarked pronoun: Chicheŵa pronominals in optimality theory. Bailey, Ashlee C, Kevin E Moore & Jeri L Moxley (eds.), *Berkeley Linguistics Society: Proceedings of the Annual Meeting*, Berkeley Linguistics Society, vol. 23spe, 26–46.

Bresnan, Joan (2001). *The emergence of the unmarked pronoun*, vol. Optimality-theoretic Syntax. The MIT Press, Cambridge, MA.

Cameron, Richard (1992). *Pronominal and null subject variation in Spanish: Constraints, dialects, and functional compensation*. Ph.D. thesis, University of Pennsylvania.

Cardinaletti, Anna (1997). Subjects and clause structure. Haegeman, Liliane (ed.), *The New comparative Syntax*, Longman, New York, 33–63.

Cardinaletti, Anna & Michal Starke (1999). The typology of structural deficiency: A case study of the three classes of pronouns. van Riemsdijk, Henk (ed.), *Clitics in the languages of Europe*, Mouton de Gruyter, The Hague, 145–233.

Carminati, Maria Nella (2002). *The Processing of Italian Subject Pronouns*. Ph.D. thesis, University of Massachusetts Amherst.

Enríquez, Emilia-V. (1984). *El pronombre personal sujeto en la lengua española hablada en Madrid*, vol. 4. Consejo Superior de Investigaciones Cientfica, Instituto Miguel de Cervantes, Madrid.

Flores-Ferrán, Nydia (2004). Spanish subject personal pronoun use in New York City Puerto Ricans: Can we rest the case of English contact? *Language Variation and Change* 16, 49–73.

van Gompel, Roger P. G. & Asifa Majid (2004). Antecedent frequency effects during the processing of pronouns. *Cognition* 90, 255–264.

Hulk, Aafke & Natascha Müller (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition* 3:3, 227–244.

Jakobson, Roman (1984). Structure of the russian verb. Waugh, Linda R. & Morris Halle (eds.), *Roman Jakobson. Russian and Slavic Grammar. Studies 1931-1981*, Mouton, Berlin, 1–14.

Montrul, Silvina (2004). Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism: Language and Cognition* 7:2, 125–142.

Müller, Natasha & Aafke Hulk (2001). Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition* 4:1, 1–21.

Ordóñez, Francisco & Esthela Treviño (1999). Left dislocated subjects and the pro-drop parameter: A case study of Spanish. *Lingua* 107, 38–68.

Paradis, Johanne & Samuel Navarro (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.

Platzack, Christer (2004). Agreement and null subjects. Dahl, Anne, Kristine Bentzen & Peter Svenonius (eds.), *19th Scandinavian Conference of Linguistics*, vol. 31.2.

Rizzi, Luigi (1986a). Null objects in Italian and the theory of *pro*. *Linguistic Inquiry* 17:3, 501–557.

Rizzi, Luigi (1986b). On the status of subject clitics in romance. Jaeggli, Osvaldo & Carmen Silva-Corvalán (eds.), *Studies iin Romance linguistics*, Foris, Dordrecht, 391–419.

Rizzi, Luigi (1997). A parametric approach to comparative syntax: properties of the pronominal system. Haegeman, Liliane (ed.), *The New Comparative Syntax*, Longman, New York, 268–285.

Serratrice, Ludovica (2007). Cross-linguistic influence in the interpretation of anaphoric and cataphoric pronouns in English-Italian bilingual children. *Bilingualism: Language and Cognition* 10:3, 225–238.

Serratrice, Ludovica, Antonella Sorace & Sandra Paoli (2004). Cross linguistic influence at the syntax-pragmatics interface: subjects and objects in Italian-English bilingual and monolingual acquisition. *Bilingualism: Language and Cognition* 7, 183–205.

Serratrice, Ludovica, Antonella Sorace, Francesca Filiaci & Michela Baldo (2009). Bilingual children's sensitivity to specificity and genericity: Evidence from metalinguistic awareness. *Bilingualism: Language and Cognition* 12:2.

Silva-Corvalán, Carmen (1994). *Spanish in Los Angeles*. Oxford University Press, New York.

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

Sorace, Antonella, Ludovica Serratrice, Francesca Filiaci & Michela Baldo (2009). Discourse conditions on subject pronoun realization: testing the linguistic intuitions of older bilingual children. *Lingua* .

Torrego, Esther (1998). *The Dependencies of Objects*, vol. 34. MIT Press, Cambridge, Massachusetts.

Tsimpli, Ianthi Maria, Antonella Sorace, Caroline Heycock & Francesca Filiaci (2004). First language attrition and syntactic subject: The role of discourse pragmatics in the acquisition of subjects. *Applied Psycholinguistics* 8:3, 257–277.

Selected Proceedings of the 12th Hispanic Linguistics Symposium

edited by Claudia Borgonovo,
Manuel Español-Echevarría,
and Philippe Prévost

Cascadilla Proceedings Project   Somerville, MA   2010

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Filiaci, Francesca. 2010. Null and Overt Subject Biases in Spanish and Italian: A Cross-linguistic Comparison. In *Selected Proceedings of the 12th Hispanic Linguistics Symposium*, ed. Claudia Borgonovo, Manuel Español-Echevarría, and Philippe Prévost, 171-182. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2415.