

Introduction

One area in which syntactic aspects of language interact with discourse and pragmatic considerations is anaphora resolution, the process of resolving a pronoun in accordance with an earlier or a later item in discourse, the antecedent. Anaphoric dependencies are one of the most important linguistic mechanisms employed to maintain discourse coherence. The use of pronouns to refer to previously mentioned referents or to anticipate arguments mentioned later is a common linguistic device that serves to avoid redundancy or repetition. In so-called 'null-subject languages', the existence of two separate forms to express pronouns (both overt and null) has led researchers to believe that different types of pronouns have different preferences in terms of determining their antecedents. The expression or omission of the subject, albeit grammatically optional, relies heavily on discourse-pragmatic considerations, such as whether a new referent is introduced in the discourse or not.

Since the 1990s, it has been argued that linguistic phenomena at the syntax-discourse interface level are somewhat more complex and costly (in terms of processing) than purely syntactic structures, as multiple types of information must be integrated (for a discussion, see Sorace 2011) and hence are principal loci of developmental delays. Although the syntactic knowledge necessary for dependencies between reflexives and their antecedents seems to be attained by children without any difficulty at an early age, the same children exhibit delays with respect to the interpretation of discourse dependencies between non-reflexive pronouns and their antecedents. This phenomenon, known as the Delay of Principle B Effect, has been extensively discussed in the literature on the acquisition of binding conditions (i.e. English monolingual children allow coreference between pronouns and local antecedents, as in **Peter_i likes him_i*; Avrutin & Wexler 1992, Chien & Wexler 1990).

The claim that the syntax-discourse interface is an area that is particularly vulnerable to instability has become increasingly prominent in contemporary research, especially since the formulation of the Interface Hypothesis (Sorace & Filiaci 2006). Evidence for the validity of the 'syntax before discourse' hypothesis has emerged from a substantial body of research on anaphora resolution in various bilingual developing grammars in both children and adults acquiring a null-subject language together with a non-null-subject language (see, among others, Serratrice, Sorace & Paoli 2004, Tsimpli, Sorace, Heycock & Filiaci 2004). Early studies on bilingual child production and (near-native and attrited) adult comprehension pointed towards a unidirectional non-target-like pattern consisting of an overuse of overt pronouns due to the overextension of their scope to topic-continuity contexts.

Interestingly, these bilingual populations showed native-like behaviour with respect to null pronouns. Thus, it was concluded that the overt pronoun becomes the default option for learners when they fail to integrate syntactic knowledge with peripheral systems such as discourse-pragmatics. However, the pattern established in earlier studies has been challenged by more recent child comprehension data from several Romance languages: monolingual and simultaneous bilingual (English-Spanish, Spanish-Italian and Spanish-Catalan) children of 6-8 years of age have been reported to display a bidirectional non-adult-like extension of the scope of both null and overt pronouns (Barquin & Costa 2011 and Shin & Cairns 2012 for Spanish; Sorace, Serratrice, Filiaci & Baldo 2009 for Italian).

In recent decades, the field of bilingual acquisition in childhood has experienced a significant shift in the focus of study from the simultaneous acquisition of two languages from birth to child second language development. The successive acquisition of a second language beyond age 3 has become an emerging field of study, and the number of studies on children whose exposure to the second language occurs between the ages of 4 and 7 has steadily grown. At this point, the question of whether successive early L2 learners will follow the pattern of monolingual learners or instead that of adult L2 learners has already been clarified in some grammatical domains, such as inflectional morphology, where child L2 learners seem to behave like adult L2 learners (e.g. Meisel 2008, Schlyter 2011). However, in anaphoric dependencies, a linguistic phenomenon situated at the interface between internal and external modules (i.e. the syntax-discourse interface), there has been practically no investigation concerning the performance of child L2 learners, as research has mainly focused on the mastery of simultaneous bilinguals in comparison to monolingual children with regard to the syntactic and discourse properties of subject pronouns. However, preliminary results from English child L2 learners of Spanish have revealed residual optionality in the discourse constraints in line with the response patterns of adult advanced learners (Pladevall 2010).

The study presented here is intended to contribute to the ongoing debate on anaphora resolution in null-subject languages by presenting comprehension data on intrasentential anaphoric dependencies in Basque, a language that lacks true third-person pronouns. The aim of this work is threefold. Firstly, it investigates whether null pronouns and overt referential devices – namely, the demonstrative *hura* ‘that’ and the quasipronoun *bera* ‘(s)he (him/herself)’ (de Rijk 2008: 114) – differ in their antecedent choices in Basque in a manner consistent with the crosslinguistic patterns. Secondly, it examines whether the linguistic descriptions established in Basque traditional grammars of *hura* ‘that’ as a *obviative/neutral* pronoun vs. *bera* ‘(s)he (him/herself)’ as a *proximate/intensive* pronoun (Ortiz de Urbina 1989: 147) and their equivalence to third-person pronouns in other languages (de Rijk 2008: 115, 209) can be supported by experimental evidence. Finally, it com-

compares the interpretation of null and overt pronouns by three groups of Basque-Spanish bilinguals: native adults (L1 adults), native children (the cL1 group) and early successive bilingual children (the cL2 group). This last comparison allows us to investigate the effect of extralinguistic factors (such as biological age, age of onset of acquisition and amount of exposure) on the interpretation of null and overt pronouns in Basque. Data from child L2 learners will provide more insights on the behaviour of early sequential bilinguals in the acquisition of syntactic and discourse subject properties.

To this end, the results of six off-line tasks are presented. Using two different methodologies, the interpretations of Basque null and overt pronouns in subject position are analysed from various perspectives. In a Picture Selection Task, the referential properties of *hura* and *bera* in comparison to null pronouns were tested in forward and backward anaphora conditions with different clause orders (main-subordinate and subordinate-main) in one-referent (Experiments 1 and 2) and two-referent sentences (Experiments 3 and 4). Another methodology was used for Experiments 5 and 6, namely an Acceptability Judgement Task in which participants had to choose between null and overt pronouns in topic-continuity and topic-shift contexts. The acceptability status of null and overt pronouns in this task, based on speakers' preferences, allows examination of whether Basque learners have any default form when confronted with a choice between null and overt pronouns.

The structure of the book is as follows. The first sections of Chapter 1 concentrate on defining pronominal anaphora and discussing the different factors that may affect antecedent assignment, such as: i) clause order, ii) anaphora type and iii) the overt realisation of the subject in null-subject languages. The troublesome crosslinguistic category of the third-person personal pronoun is also discussed; this category may not be universal, since the so-called 'two-person' languages have no true third-person pronouns (Bhat 2004). Instead, demonstratives fulfill this role, blurring the dividing line between the two anaphoric forms. The subsequent sections focus on reviewing the three major theories of referring expressions that attempt to capture the restrictions governing anaphoric dependencies from different angles: Accessibility Theory (Ariel 1990), Binding Theory (Chomsky 1981) and the Position of Antecedent Strategy (Carminati 2002).

Chapter 2 starts by pointing out a distinction between different types of early bilingual language acquisition, such as simultaneous bilingual acquisition (2L1) and child L2 acquisition (cL2). The following sections give a thorough and up-to-date discussion of the background literature on the instability at the syntax-discourse interface, providing the results from child studies in the generative (Binding Theory) model (Principle B). Results obtained from child studies on Principle C, which directly sets syntactic constraints on the interpretation of backward anaphora structures, are also reported. In addition, empirical research on children's interpretations of null and overt

subject pronouns is discussed in relation to the Interface Hypothesis (Sorace & Filiaci 2006) and the possible explanations proposed in the literature to account for the delays observed at the syntax-discourse interface.

In Chapter 3, comprehension data from the two tasks focusing on the interpretation of null and overt subject pronouns are reported. First, a description of the criteria for using third-person pronouns and a discussion of their frequency in prior non-experimental empirical work are provided. The linguistic profiles of the participants are then described, including information about differences in the age of onset of acquisition between the cL1 and the cL2 groups. Subsequently, the materials and the methods used are introduced and the results from the six experiments are presented.

Chapter 4 offers an analysis of the data presented in Chapter 3 in light of the prior research presented in Chapter 2. The discussion is centred on the research questions formulated for the present study: a) whether a division of labour is found between null and overt pronouns in Basque in line with the pattern observed crosslinguistically, b) whether the experimental evidence supports the observations of usage found in the descriptive grammars of Basque and c) to what extent children's performance is similar to that of adults and homogeneous between linguistic profiles (cL1 and cL2 groups). Finally, Chapter 5 presents conclusions, emphasising the findings of the study and providing suggestions for future research.

1 Pronominal anaphora

On issues related to pronominal elements and the establishment of coreference, one of the central questions is how a pronoun is interpreted – that is, how the pronoun selects its referent. The process of finding an antecedent for a pronoun is relevant because a pronoun itself does not carry referential content (apart from morphological feature specifications such as gender, number or case). However, this does not mean that a pronoun can pick any antecedent available in the mental representation or in the immediate context of a given discourse.

Ever since the publication of *Lectures on Government and Binding* (Chomsky 1981), anaphoric dependencies have been one of the most common topics of research in the generative framework. The beginning of the research paradigm of principles and parameters was based on (among other theories) Binding Theory, which addresses anaphoric relations and the syntactic restrictions found between coreferents. However, it was soon discovered that anaphoric dependencies could not be addressed exclusively from a syntactic perspective, since in certain coreference relations, syntactic considerations seem to be irrelevant (see Section 1.3). Thus, more recently, the problem of resolving a pronoun in accordance with an earlier or later linguistic item in the discourse has been examined from a discourse-based and processing point of view. Nevertheless, the underlying process involved in resolving what a pronoun or a noun phrase refers to is still unclear. Hence, it is still a challenging and active area of research.

The structure of this introductory chapter is as follows. Section 1.1 starts off with the possible relationships between an anaphoric noun phrase and its antecedent and continues with the potential effects of clause order (main-subordinate vs. subordinate-main) and anaphora directionality (forward vs. backward anaphora) on pronominal-antecedent dependencies. A discussion of each of these components present in anaphoric relations is followed by an overview of the findings reported in the literature, especially with regard to the interpretation of pronouns. Section 1.1.3 is devoted to the description of third-person (personal and demonstrative) pronouns and their different antecedent biases. In addition, in Section 1.1.4, special attention is focused on the different interpretative behaviour exhibited by null and overt pronouns. The remainder of this introductory chapter reviews different theoretical streams in the analysis of anaphoric relations that will be relevant for the interpretation of null and overt subject pronouns in the context of intrasentential anaphora: Accessibility Theory (Ariel 1990, 2001), Binding Theory (Chomsky 1981) and the Position of Antecedent Strategy (Carminati 2002). The first account argues for the selection of antecedents in relation to the activation status of the mental representations (Ariel 1990), while the second

approach views anaphoric dependencies as constrained by the syntactic configuration (Chomsky 1981). The third approach is a processing account that predicts different antecedent biases for null and overt pronouns based on syntactic prominence (Carminati 2002). These three accounts enable us to incorporate both syntactic and extra-syntactic (discourse) information in order to develop a better understanding of the establishment of dependencies linked to pronominal interpretation.

1.1 Anaphora

A number of definitions of the term ‘anaphora’ exist in the literature, but the same concept underlies all of them: reference to something mentioned or implied in previous discourse (Green 1989). This characteristic element of tracking is related to the word’s etymology. The Ancient Greek word *anaphora* (*αναφορα*), made up of the separate words *ana* (‘back, upstream, back in an upward direction’) and *phora* (‘the act of carrying’), denotes the act of carrying (something) back upstream (Peral & Ferrández 2003). More precisely, Huang (2000: 1, referring to Huang 1994, Lust 1986, Wasow 1986) states that anaphora is a relation between two linguistic forms “wherein the interpretation of one (called an anaphor) is in some way determined by the interpretation of the other (called an antecedent)”. Here, Huang alludes to the referential dependency of the anaphor (usually the pro-form) towards the antecedent (the entity to which it refers), since the anaphor does not convey any semantic meaning on its own. In other words, the anaphor repeats the reference or the sense that the antecedent has already established, such that the two linguistic expressions end up having the same referential value. At this point, a clarification concerning the term ‘anaphor’ is needed. Although the term has traditionally been used for any NP (reflexive or not) that has an antecedent in the preceding discourse, there is another distinct sense of anaphor. Within the generative theory, anaphor is presented as an NP with the features [+anaphor, -pronominal], as opposed to pronouns with the features [-anaphor, +pronominal] (Chomsky 1982); the restrictions of this theory of anaphoric relations will be discussed in Section 1.3.

Anaphora is a very common linguistic device in natural language because, as Blackwell (2003: 2) notes, “anaphoric reference enables speakers to avoid redundancy or repetition by the use of a semantically, lexically, and phonologically attenuated linguistic expression (the anaphor), in place of the full lexical expression (the antecedent)”. According to Huang (2000), anaphora can be syntactically divided into two main categories: NP-anaphora and VP-anaphora; the former can be encoded by gaps (or empty categories), pronouns, reflexives, names and descriptions (Huang 2000: 2). The example in (1) shows the most common type of NP-anaphora, the so-called ‘pronominal anaphora’ in which the coindexed NPs exhibit sameness of reference (or *coreference*).

- (1) After the baroness_i had visited the lord, she_j left the house.

(adapted from Büring 2005: 1)

However, the pronoun *she* in (1) can also be contra-indexed with its antecedent *the baroness*, thus becoming disjoint in reference (*non-coreference*). In such a case, it would refer to an extrasentential referent. This process of assigning an antecedent to a pronoun is known as *anaphora resolution*.

There are three logically possible coreference relations between two NPs¹ in a given sentence (Reinhart 1983: 29):

- (2) a. *Obligatory (stipulated) coreference*: Zelda_i bores herself_{i/*j}.
b. *Obligatory (non-stipulated) non-coreference*: She_i adores Zelda_{*i/j}'s teachers.
c. *Optional (free) coreference*: Zelda adores her_{ij} teachers.

The three-valued system of coindexing illustrated in (2) operates along these lines: (i) NPs such as the R-expression (the 'referring' or referential expression) *Zelda* and the reflexive pronoun *herself* can be positively coindexed (2a), (ii) negatively coindexed (2b) or (iii) neutrally indexed (i.e. neither positively nor negatively coindexed) (2c). In (2a), the reflexive *herself* must be interpreted anaphorically as referring to the subject antecedent *Zelda* due to a syntactic constraint on the distribution of reflexives (see Principle A in Section 1.3). In (2b), the R-expression *Zelda* must be disjoint in reference or non-coreferent with the preceding pronoun *she* (see the following section). Finally, (2c) presents the optional (free) coindexing of a pronoun and a nominal antecedent.

Büring (2005: 2) adopts the following definition of antecedent: "A is the antecedent of B iff (if and only if) (i) A precedes B, and (ii) A and B corefer". This holds for *forward anaphora*² sentences like (1) in which the pronoun follows the antecedent (Guasti 2004: 300). There are, however, cases in which the ordering of the antecedent and anaphor can be switched, resulting in *backward anaphora* (or *cataphora*, as it is also known) in which the pronoun shows reference to something later in the text:

¹ In generative syntax, ever since Abney's (1987) proposal, the structure of NPs has been extended to DPs. Throughout this work, we refer to NPs and DPs interchangeably, depending on the terminology used by the cited researchers.

² In the generative tradition, pronouns in examples like (1) are regarded as *forward anaphora* because one has to move forwards (starting from the antecedent) when the pronoun follows the antecedent and backwards when the former precedes the latter. However, scholars in other traditions working on cohesion (such as Halliday & Hasan (1976)) use *anaphora* for cases in which the antecedent precedes the pronoun and *cataphora* when the antecedent follows it.

- (3) After she had visited the lord, the baroness left the house.

Examples of both forward (1) and backward anaphora (3) can be regarded as anaphora more generally because the term is sometimes used generically to cover both types of anaphoric relations (Dixon 2010: 247).

1.1.1 Clause order

A main clause can be interpreted independently of its subordinate clause; however, the interpretation of a subordinate clause depends on the meaning of the main clause. Consequently, it is plausible to ask whether the antecedent biases of pronouns are likely to be affected by different clause orders during the processing of complex sentences consisting of a subordinate (dependent) and a main (independent) clause. In fact, studies on adult English that do not directly address pronoun interpretation have demonstrated that main and subordinate clauses are processed differently (Bever & Townsend 1979, Clark & Clark 1968), but these differences do not always point in the same direction. For example, better verbatim memory performance has been observed in sentence-initial subordinate clauses (especially for adversatives such as *though*-clauses) than in final subordinates in lexical probe latency tasks, although this tendency is statistically non-significant (Bever & Townsend 1979). However, the same authors report better semantic comprehension for initial main clauses in semantic latency probe tasks. Clause-effect differences have been attributed to the encoding stages of processes: the meaning of a subordinate clause must be kept in memory until the main clause is processed — unlike main-subordinate clauses, where the superficial form will be lost and short-term memory is available for processing subordinate clauses. This suggests that a higher processing load is required in subordinate clauses, especially when they precede main clauses, because they are not semantically complete, and the complete semantic level of processing depends on the content of the main clauses. According to Bever and Townsend's (1979) results, the on-line accessibility of superficial form in initial clauses decreases in this order, with *though*-clauses more easily held in memory than the rest: *though*-clause>*while*-clause>*when*-clause>*since*-clause>*if*-clause. Bever and Townsend's (1979) finding that subordinate clauses are retained longer in short-term memory is in line with Jarvella and Herman's (1972) results from an imitation task and Garnham, Oakhill and Cain's (1998) conclusions on the processing of VP-ellipsis.

Although Bever and Townsend's study does not directly deal with pronoun interpretation, they mention that in contexts such as **He_i ate supper before John_i left town*, syntactic rules (i.e. disallowing coreference) precede listeners' usual strategy of representing asserted main clauses in semantic form as quickly as possible (Bever & Townsend 1979: 211). Attention to constraints on coreferential pronouns depending on clause order had already

been emphasised by, among others, Lakoff (1968), Lasnik (1976) and Reinhart (1976). Although coreference between an embedded DP and the following matrix pronoun is possible, as seen in (4c), a pronoun in an initial main clause that is coreferent with a subsequent embedded DP is not grammatically allowed (4b). Note that anaphora directionality changes from forward (4c) to backward anaphora (4b). In contrast, pronouns that occur in subordinate clauses are not subject to coreference constraints ((4a) and (4d)).

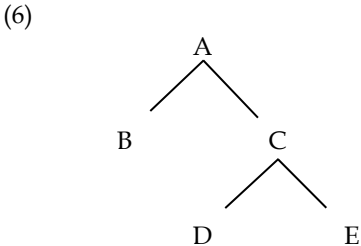
- (4) a. Sarah_i listens to music [when she_i reads poetry].
- b. *She_i listens to music [when Sarah_i reads poetry].
- c. [When Sarah_i listens to music], she_i reads poetry.
- d. [When she_i listens to music], Sarah_i reads poetry.

(examples from Lust 2006: 214)

In backward anaphora in which the main clause precedes the subordinate clause, as in (4b), the structure-dependent notion of c-command comes into play, blocking intrasentential coreference with the subsequent subject antecedent. Hence, the only possible interpretation remaining is the disjoint reference interpretation. C-command, first proposed by Reinhart (1976, 1983), is defined in terms of dominance:

- (5) A node A c-commands a node B if and only if
 - (i) A does not dominate B
 - (ii) B does not dominate A
 - (iii) The first branching node dominating A also dominates B

On this basis, the following set of c-command relations is possible in the structure in (6):



- D mutually c-commands E (D c-commands E, E c-commands D)
- B mutually c-commands C (B c-commands C, C c-commands B)
- B c-commands D
- B c-commands E

(taken from Frank, Hagstrom & Vijay-Shanker 2002: 111)

From this, it follows that anaphoric dependencies are explained in terms of *binding* in the generative tradition, which explicates when a pro-form may be bound or free (i.e. not bound) (Haegeman 1994: 212):

- (7) A binds B if and only if
 - a) A c-commands B
 - b) and A and B are coindexed (i.e. A and B refer to the same entity)

Returning to example (4b), the reason why *she* and *Sarah* cannot be coreferent is that *she* c-commands *Sarah*. That is, following the criteria mentioned in (5), *she* does not dominate *Sarah*, *Sarah* does not dominate *she*, and the first branching node (IP) dominating the DP *she* dominates the DP *Sarah* (see Figure 1. R-expressions like *Sarah* cannot be c-commanded by any antecedent as formulated in Principle C (see Section 1.3 for more details), explaining why *she* and *Sarah* cannot be coreferent.

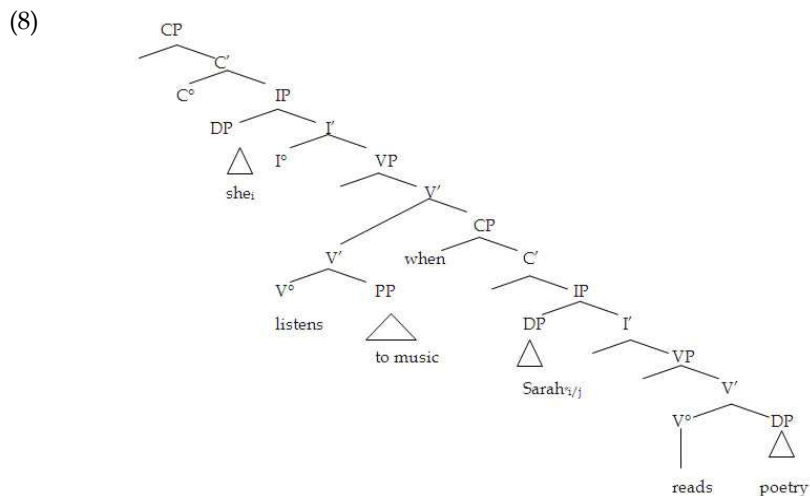


Figure 1. Syntactic tree of example (4b)

In contrast, in the sentence in (4d), where the subordinate clause precedes the main clause, the embedded DP *she* does not c-command the matrix subject *Sarah*: the first branching node dominating *she*, the IP, does not dominate the DP *Sarah*. Thus, as *Sara* is not c-commanded by any antecedent, it can be coreferential with the preceding subject pronoun *she*, as shown in Figure 2.

(9)

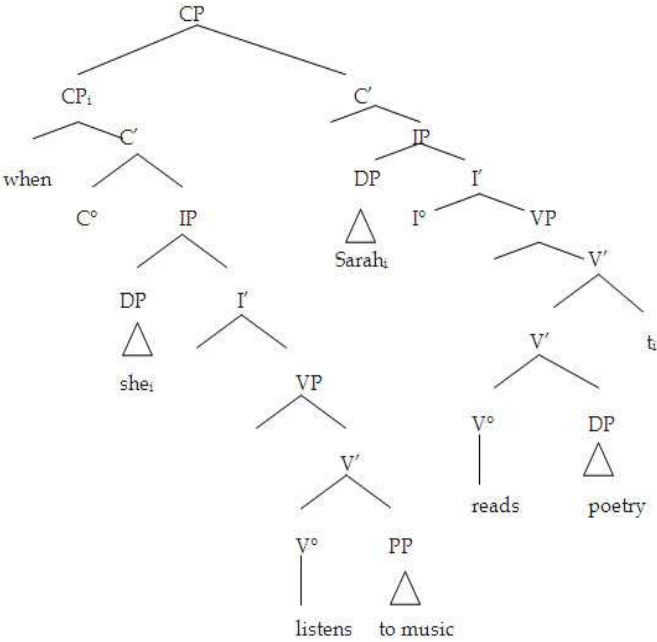


Figure 2. Syntactic tree of example (4d)