Relative clause attachment preferences in second language learners' parsing performance

Guillermo Rodriguez

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/pwpl/vol10/iss1/13
For more information, please contact libraryrepository@pobox.upenn.edu.
Relative clause attachment preferences in second language learners' parsing performance
1 Introduction

Preferred interpretations in the resolution of structural ambiguity may provide empirical evidence to support or reject theoretical models of the processes underlying the parsing of linguistic input (Cuetos and Mitchell, 1988; Frazier and Clifton, 1994; Mitchell, 1994).

One of the grammatical structures in which these preferred interpretations have been identified is the complex noun phrase (NP), such as the doctor of the tycoon who left for India. This construction involves a head noun phrase (NP1, the doctor) followed by a prepositional phrase with an embedded noun phrase (NP2, the tycoon), and a relative clause (RC, who left for India).

The ambiguous nature of this particular structure stems from the fact that the RC, who left for India (a modifier of a NP), could be attached to the first NP, the doctor, or to the second NP, the tycoon.

Previous research claims to have established language specific preferences for the attachment of the relative clause to NP1 or to NP2. In English (and Norwegian, Swedish, Romanian, Brazilian Portuguese, Arabic) the predominant trend is to append the RC to the second NP. This phenomenon is referred to as low or late attachment. On the other hand, claims have been made that Spanish (and German, Dutch, French, Russian) show a tendency to attach the modifying clause to the first NP; this preference is called high or early attachment.

Different explanations have been advanced to account for these cross-linguistic differences between Spanish and English when parsing relative clauses. Some of these accounts deal with the difference between the Norman (the only form used in Spanish) and the Saxon (used in English together with the Norman type) genitive constructions, the length of the relative clause, the type of NPs joined in the complex noun phrase and their semantic relationship, the different prosodic characteristics of English and Spanish, and the frequency of attachment preferences in analyzed corpora, among others.

The goal of this paper is to further explore the reasons behind these cross-linguistic preferences and to assess the influence of a pragmatic constraint on this phenomenon. Another of the objectives is to investigate attachment preferences in second language learners’ parsing performance.
2 Principles, Construal and Pragmatic Constraints

2.1 Principle-Based Parsing Models

In spite of the apparent language-specific strategy for the assignment of structure to the complex noun phrase in English and Spanish, principle-based models attempt to explain parsing processes from a universal standpoint. Most of these models posit a set of general principles shared by all languages that constitute a set of parsing algorithms. The most important representatives of this theoretical approach are the Garden Path theory (Frazier, 1987) and the Construal theory (Carreiras and Clifton, 1993; Gilboy, Sopena, Clifton, and Frazier, 1995). The most relevant principles advanced in this framework are minimal attachment and late closure.

Minimal attachment emphasizes a universal preference for building the simplest structure possible. It states that the parser should not postulate any unnecessary structural nodes, so that the outcome of the process is the simplest structure possible. An example of this parsing tendency appears in sentences (1) and (2):

(1) The teacher told the children the ghost story that she knew would frighten them.
(2) The teacher told the children the ghost story had frightened that it wasn’t true.

Frazier (1978) found that sentence (1) required shorter reading times than sentence (2). In the interpretation of the former sentence NP[the story CP[that she knew would frighten them]] two structural nodes are required. In sentence (2), however, more new nodes are necessary in order to interpret the utterance: NP[the children CP[the ghost story had frightened]] CP[that it wasn’t true]. The preference of the human parser to attach incoming structures minimally would be supported by reading times that suggest ease of processing when following the minimal attachment principle.

Late closure, the second principle, determines that any new incoming item should be attached to the phrase or clause currently being processed, if that is grammatically permissible. This principle allows for the speed with which parsing algorithms unfold and can be seen in action in sentences such as (3) and (4):

(3) Mary kissed John and his brother when she left.
(4) Mary kissed John and his brother started to laugh.

In (3), the second NP, *his brother*, is attached to the first member *John*
as an argument of the verb *kissed*. On the other hand, the same NP can not be attached to the direct object of *kiss* in (4) and has to be reinterpreted as the subject of the following predicate, *started*. In (4) late closure is violated and as a consequence the reading times for that sentence increase when compared to (3).

In the most recent version of the Garden Path model (Frazier 1987), Construal, a classification of structures to be parsed determines whether the universal principles described above are applied (Frazier & Clifton 1994).

### 2.2 Construal

Construal is an updated and revised version of the Garden Path model. In this new version, Frazier & Clifton (1994) claim that there is a degree of underspecification in the structural assignment the parser performs.

Perfetti (1990) proposed a very similar account in which underspecification is understood as a parser working by pieces or chunks. In Perfetti's pieces parser, some syntactic decisions are left open. Input continues to be analyzed, and when a particular trigger is found, the structure is completed by building a new syntactic component. These pieces or chunks are put together by resorting to semantic information.

On the other hand, Construal underspecification is present only when the clauses being parsed belong to the category called non-primary phrases. The universal principles in the Garden Path theory, late closure and minimal attachment, apply only to primary phrases, i.e. the subject and main predicate of any finite (+ or -) clause and their complements and obligatory constituents. Relative clauses are considered non-primary phrases and their addition to the structural parse is ruled by the Construal principle. This principle states that all phrases should initially be analyzed to determine whether they are primary or not. If the latter is true, then the non-primary phrase should be associated, not attached, to the current thematic processing domain (cf. Pritchett, 1992). The actual wording of the principles involved in Construal is shown below in (5):

(5)  

a. Analyze an input, X, as instantiating a primary phrase if possible.

b. Otherwise, associate X into the current thematic processing domain.

In this model, association does not imply attachment. Hence, non-primary phrases are left unattached until they find an appropriate host.

Once the constituent is attached, there is not a need for reanalysis, since there was not an initial commitment to a particular attachment site; the phrase was only associated to the processing domain. The association principle in (6) states that:
Attachment to a node within XP would not bring about any particular cost (longer reading times); however, if the parser does not assign a host to the non-primary phrase within the current processing domain, it would be costly (Frazier and Clifton, 1994). In order to attach non-primary phrases, the parser can make use of non-syntactic information. As argued by Frazier and Clifton (1997), non-primary phrases are optional, which makes them irrelevant in the interaction with lexical analysis. They do not function as complements of lexical items, and can be thus left unspecified without affecting the general and immediate process of structure assignment that primary phrases undergo.

2.3 Modifiability

Modifiability, the amount of modification a certain noun phrase may take (Thornton et al., 1999), has been posited as one of the pragmatic constraints that may exert influence over the attachment site of an ambiguous structure. Thornton et al. elaborated on the idea that the amount of modification needed to distinguish an entity from a set is minimal, and that this complies with the Gricean maxim of quantity (Grice, 1975).

2.3.1 Thornton et al.'s Proposal

Although Thornton and colleagues worked with a different type of complex NP, this notion of minimal modification can be applied to attachment preferences in noun phrases with relative clauses. If a certain NP has been previously modified in the discourse, that same NP will be a less likely candidate for further modification in the same utterance. Consequently, RC attachment preferences may also be influenced by the degree of modification present in the NPs involved in the structure. If NP1 has undergone modification, then the RC will have NP2 as a better candidate for its host-site; low attachment will result. In contrast, if NP2 is the constituent undergoing modification, NP1 will be more likely to be assigned the modifying clause (RC)—high attachment.

2.3.2 Gilboy et al.'s Proposal

The opposite argument regarding the influence of modification has been advanced by Gilboy et al. (1995). They hypothesized that the presence of an
adjective modifying one of the NPs in the complex phrase would make it easier for the modified noun to attach the RC. This is based on the assumption that an N' is postulated when adjectival modification is added to a noun. This N' node will be built in order to allow for the adjective and, as a consequence, it will be available to dominate the RC later. N' nodes are postulated only when necessary, then, because if there is no adjective to modify the noun, the N' would not be present, and would not facilitate RC attachment (Gilboy et al., 1995).

All the former models of sentence parsing and modifiability accounts were developed to explain phenomena related to first languages. As Juffs (2001) points out, experimental techniques designed to assess language processing phenomena have been available since almost the beginning of the twentieth century. Unfortunately, second language researchers have focused mainly on the study of a second language competence, instead of looking at the processes involved in language comprehension and production. This trend seems to be changing and more data on second language processing strategies have recently surfaced.

3 Attachment preferences and transfer

Information about attachment preferences in the parsing performance of second language learners is very limited. There have been only a handful of studies focusing on different attachment ambiguities (relative clauses and prepositional phrase attachment). The different methodologies used in these studies (self-paced reading, questionnaire and eye-tracking) and the groups tested in these experiments, do not really allow one to arrive at a conclusion regarding the influence of the first language on the comprehension processes of a second tongue.

Frenck-Mestre (1997) investigated the RC attachment preferences of beginning French learners whose first languages were Spanish and English. Results of this experiment showed that Spanish speakers behaved like French native speakers (high attachment), unlike English-speaking learners of French, who showed no preference for attachment of the RC.

In another attachment preference study, this time analyzing the bias when joining a prepositional phrase (PP) to a verb phrase (VP) or a determiner phrase (DP), Frenck-Mestre and Pynte (1997) used eye-tracking measures with advanced French-speaking learners of English and English-speaking learners of French. Preferences shown by these groups seem to be based on the different argument structure of the verbs that were used in the stimuli (ditransitive and monotransitive verbs). However, L2 learners had a preference to attach new input locally and sometimes took longer to read verb
phrases that differed in usage between the target and the native languages. The researchers argued that these effects were the result of L1 transfer. As Papadopoulou and Clahsen (2002) point out, these results might also be explained due to a difference in the material to be processed with optional transitive verbs (more options to be processed on-line) and intransitive verbs. As a result, the transfer claim does not receive unequivocal support.

In a study conducted by Fernández (1999), early and late Spanish learners of English were given off-line questionnaires to determine their RC attachment when processing sentences in their L2. Speakers who started learning the language before age 10 (early learners) and those who started after that age (late learners) showed the same high attachment preference, whereas the native speakers’ production in the same test rendered a low attachment preference. Fernández concluded that the results supported the idea of transfer of parsing algorithms from the native language to English.

Papadopoulou and Clahsen (2002) tried to determine attachment preferences in advanced second language learners of Greek. This language shows a preference for high attachment, and the structure of the relative clause is very similar to that of English. Spanish, German and Russian were the native languages of the Greek learners who participated in this experiment. The advanced language learners who took part in the study showed similar preferences as a group, but they did not prefer the first NP as native speakers of Greek do. These highly proficient learners showed no preference at all in the attachment of RCs. Hence, Papadopoulou and Clahsen claim that these results do not provide support for the transfer argument, since if there were any influence from their first languages, these advanced speakers would have had a bias towards NP2 attachment (because in their first languages, they prefer to attach the RC low). As was previously pointed out, studies on attachment preferences in second language research do not really help predict whether transfer will take place or not.

Based on the literature review outlined above, the behavior of native speakers when parsing relative clauses can be predicted using Construal and the conversational maxim of clarity (Be clear and unambiguous) posited by Grice (1975). Because in a complex noun phrase the last theta assigner is the first NP, then NP2, which is an argument of NP1, should also be considered as a possible host for the attachment of the RC. Construal then predicts that both noun phrases would be equally capable of hosting the modifier clause.

When trying to explain the difference found in previous experiments in which English speakers chose NP2 more than Spanish speakers, Gricean maxims of conversation can be of help. In particular, the maxim of clarity is useful since Spanish and English differ in the structural devices at hand to express a genitive relation. If a cooperative English native speaker/writer’s production is analyzed, the unambiguous Saxon genitive—the colonel's
daughter—(unavailable in Spanish) would be used when trying to modify NP1 with a relative clause. Thus, when native speakers of English are exposed to a set of sentences that employ only the Norman genitive—*the daughter of the colonel*—the natural tendency would be to think that modification to NP2 was intended. Therefore, Construal and Grice’s maxim of clarity predict a higher rate of NP2 choices for native speakers of English than for Spanish speakers (Gilboy et al., 1995).

One of the goals of this study was to determine if non-native speakers of English who are learning Spanish show the predicted tendencies, i.e. whether the inherent characteristics of their processing choices in English are transferred to their second language or not.

### 4 Experimental Design

#### 4.1 Non-native speakers

The second-language learners who participated in this study were recruited from conversation and writing courses in the Spanish Department at the University of Pittsburgh. They were all native speakers of American English who had taken Spanish courses for at least five years. There were twenty students in this group who were granted extra credit in their current Spanish course for taking part in this study. The average age was twenty-one years old and eleven of the participants had had experience using their Spanish abroad; eleven of these people had also studied a second foreign language.

The proficiency of the students who participated in the study may cause transfer to occur. These students achieved an overall 61% accuracy rate in the MLA reading comprehension test and 83% in the material used as filler items in this study. The results of these instruments seem to indicate that these students’ proficiency in Spanish might be characterized as incomplete.

#### 4.1.1 Native Speakers

The group of native speakers of Spanish was made up of twenty-one participants who belong to the University of Pittsburgh’s community. They were all from Latin-America and had spent at least one year in the United States. The purpose of their stay in this country was mainly to improve their command of English.

#### 4.2 Materials

There were twenty-four target items taken from previous studies on relative
clause attachment ambiguity and slightly modified to fit the purposes of this work (Carreiras and Clifton, 1993; Gilboy et al., 1995; Fernández, 2000). The target items were divided into three subgroups: unmodified NPs (7), NP1 modified (8) and NP2 modified (9).

(7) Julia habló con la secretaria del abogado que estaba de vacaciones.
*Julia talked to the secretary of the lawyer who was on vacation.*

(8) Carlos conoció al intérprete alemán del embajador que estaba cenando.
*Charles met the German interpreter of the ambassador who was having dinner.*

(9) Pedi prestado el ordenador del secretario japonés que era nuevo.
*I borrowed the computer of the Japanese secretary that was new.*

Modifiability was operationalized by the addition of an adjective (8) to the NP1 or to NP2 (9). The complex noun phrases were classified according to Gilboy et al.’s (1995) criterion. There were five NP-of-NP constructions in which the relationship between the nouns was one of kinship (el primo del artista / the cousin of the artist). A functional relationship (el intérprete del embajador / the interpreter of the ambassador) characterized another set of five target NPs. Finally, fourteen complex noun phrases were joined by a possessive relation (el coche del vecino / the car of the neighbor). In Gilboy et al.’s classification these three semantic relationships between NPs make up what they call Type B NPs. Both noun phrases in this type are equally good candidates to attach the RC, since they are both referential. In this case, referential means that both heads introduce discourse entities or describe entities already present in the discourse model. Gilboy and colleagues further argue that restrictive modifiers such as relative clauses are preferentially attached to referential entities (referentiality is operationalized in Gilboy et al. (1995) with determiners).

Both groups of participants (native and non-native) performed two tasks designed to assess their attachment preferences involving the complex noun phrase structure (with an RC) described in the preceding sections.

There were two conditions per task in which modification to the target items varied in the following way: in group 1, the first twelve NPs were unmodified, NP1 was modified in the next six target items and in the remaining six sentences, modification was added to NP2. The opposite pattern was applied to the second group: the first six NPs received modification on NP1, NP2 was modified in the following six NPs, and the last group of twelve NPs received no modification. Participants who parsed group 1 of target items in the off-line task were assigned to the opposite condition in the on-line task. In this way, participants were not exposed to exactly the same target item in both tasks.
4.3 Procedure

4.3.1 Off-line Task

The first task was an off-line reading comprehension exercise in which participants were presented with a target sentence and then had to choose one of two possible related sentences as the best continuation according to the meaning of the first sentence read.

It has been argued that the ultimate language preference as regards attachment (high or low) manifests itself in this kind of task, since no time constraint is imposed on the participant to process the sentence. This kind of presentation (the whole utterance at once) permits the participant to go back and reanalyze the sentence in order to make a choice. This allows for all possible sources of information (syntactic, semantic, pragmatic and lexical) to be taken into account when processing a particular utterance (Spivey-Knowlton and Sedivy, 1995). All the items in the task were pseudo-randomized so that no participant would find more than two target items presented consecutively. In this exercise, there were twenty-four target items and fifty-six fillers.

4.3.2 On-line task

The second task was implemented on a personal computer using E-prime, Version 1.0 (PST, 2001), software especially developed to create experiments. Participants had to read a sentence that was presented in three fragments in a non-cumulative way. All fragments appeared on the center of the screen, and, in order to advance in the reading of a particular sentence, participants had to press the space bar of the computer keyboard.

After the last fragment of a sentence was read, the participant pressed the space bar again and then two possible related sentences appeared on the screen. The choices were marked with "1" or "2" and the participant chose the answer he/she preferred by pressing the corresponding number key on the keyboard.

As mentioned above, the target items were divided into three fragments. This segmentation was motivated by a need to prevent participants from assigning prosodic structure to the sentences presented. This choice stems from the fact that some scholars have attributed the cross-linguistic differences in attachment between Spanish and English to the different prosodic structures of these two languages (Fodor, 1998).

4.4 Results
4.4.1 Off-line task

The mean for NP1 attachment in the off-line exercise for native speakers was 71.02% for unmodified noun phrases, 63.48% when NP1 was modified and 61.10% when NP2 was modified with an adjective. Non-native speakers’ choices rendered a 54.57% preference for unmodified noun phrases, 59.99% for modified NP1 and 56.66% when modification was added to NP2. As predicted in the literature, native speakers of Spanish had a marked preference to attach the RC to the first noun phrase in the complex construction (NP1 - high attachment) with an overall NP1 preference of 66.7%. On the other hand, English-speaking learners of Spanish did not show a significant bias towards either of the NPs with an overall 56.46% for NP1 (although there seems to be a tendency towards NP1 attachment). When the three different sentence conditions are considered, there is an effect of modification in the attachment of the RC. In the native speakers’ performance, modification to the first NP seems to draw the attachment of the RC away from NP1, and a very similar result obtains when NP2 is modified. It seems then that modification decreases the preference for attachment to NP1, regardless of whether the modified noun phrase is the first or the second. The performance of the non-native group seems to show the opposite tendency: when NP1 is modified the preference for attachment to that NP increases and the same happens when modification is present on NP2.

4.4.2 On-line task

The mean preference for NP1 attachment in the performance of native speakers was 68.25% when the noun phrases were unmodified. The mean for NP1 modified was 67.45% and when the modified noun belonged to NP2 the mean was 56.34%. Non-native speakers’ attachment preferences favored NP1 62.07% of the time when there was no modification. Sentences that contained modification to NP1 presented a 55.83% bias towards the first noun phrase. Finally, when NP2 was modified, non-native speakers preferred to attach the RC to NP1 54.16% of the time.

The behavior of native speakers in this on-line task differs from the pattern presented in the off-line task. These speakers still show a preference for NP1, even when the noun phrases do not contain modification. There does not seem to be a great change when the first NP is modified. The opposite happens when the second NP is modified; in this case, the preference to attach high decreases. The non-native speakers show a similar pattern when unmodified NPs and NP2 modified phrases are considered. There is a difference though in the behavior of these learners of Spanish regarding the target items that received modification on the first NP. Non-native speakers’ choice
of NP1 in this case decreases significantly more than in the native speaker group, approaching the levels shown for NP2 modification.

5 Discussion

The previously described results were submitted to a three way ANOVA in which group (native/non-native), task (on/off-line) and modification (modified/unmodified) were the factors analyzed. The differences identified in the previous outline of the results did not achieve significance. There was neither a significant effect for group at .05 level (F(1,39)=0.12, p=.737), nor a significant effect for task F(1,39)=0.04, p=.834). The only statistically significant effect obtained was the influence of modification at the .05 level (F(2,78)=3.34, p=.041). An additional result that approached significance was the interaction between task, modification and group at .05 level (F(2,78)=1.80, p=.171).

On-line measures of non-native speakers' performance seem to contradict the low attachment bias that could have been predicted for this group. Cross-linguistic differences and the intermediate level of proficiency of these learners may have favored transfer of processing strategies from their first language (English) to their second (Spanish). However, in the on-line task, these non-native speakers have an overall 59% preference for high attachment, and in the off-line task, this tendency achieves almost 57%.

It is a bit puzzling to find that intermediate learners whose proficiency is developing might have acquired the necessary structural knowledge to closely replicate the overall behavior of native speakers. This resemblance is more evident in the off-line task, where the preference for NP1 attachment when the nominal phrase is modified is the closest for native and non-native participants: 63% and 60%; and 61% and 57% respectively. The interesting result is that modification in the non-native group seems to have had a similar effect to the one it exerted in the native speaker group: when the second NP is modified, high-attachment decreases considerably.

The prediction advanced by Gilboy et al. (1995) stated that adjectival modification would open up the syntactic tree and facilitate the further attachment of the RC. This seemed to prove correct in the on-line task, and is evident from the native speakers' data, which show how modification to NP2 decreases the attachment of the RC to NP1 in the complex nominal construction in the on-line task when compared to the unmodified condition. A similar pattern can be identified in the off-line data that show a decrease in the preference for NP1 attachment when the modified entity in the discourse is NP2.

The influence of modification in both groups can be further explained...
by returning to Gilboy et al.'s (1995) claims in relation to the minimal attachment principle advanced by Frazier and colleagues. The human parser attempts to build a structure with the minimum number of possible nodes. Hence when an adjective is added to an NP node, the X-bar 
(Xn > Xn – 1) schema is opened up, and the addition of further modification to that opened node is favored by this principle. If the preference followed Thornton et al.'s (1999) proposal, additional nodes would be needed in order to join the RC to the remaining unmodified NP, which would, in turn, create a more complex structure.

6 Conclusions

The most important finding is again the influence that modification has on the attachment preferences of both native and non-native speakers of Spanish. However, claims about their parsing performance would not be relevant unless supported by an assessment of their preferences in the native language. Furthermore, the influence of modifiability as a pragmatic constraint favoring a simpler structure in both groups would provide evidence in favor of those accounts of parsing in which syntactic (structural) information is of paramount importance.

References


E-prime, version 1.0 [Computer software]. 2001. PST.
