To appear in *Studies in Second Language Acquisition*

Parsing Ambiguous Relative Clauses in L2 English: Learner Sensitivity to Prodosic Cues

Heather Goad, Natália Brambatti Guzzo, Lydia White
McGill University
heather.goad@mcgill.ca, nataliaguzzo@me.com, lydia.white@mcgill.ca

Abstract

We investigate effects of prosodic cues on interpretation of ambiguous sentences containing relative clauses in English by Spanish-speaking learners. English and Spanish differ in default preference for RC attachment: English has a weak low attachment (LA) preference (RC modifies NP2); Spanish has a stronger high attachment (HA) preference (RC modifies NP1). We conducted an interpretation task with auditorily-presented stimuli to examine whether prosodic cues determine attachment. Target items were manipulated for position of break and length of RC, NP1, and NP2. For both groups, break and length are significant. For the learners, proficiency interacts with break suggesting L1 transfer: lower proficiency learners choose HA more when break points to LA; higher proficiency learners choose HA more when break points to HA. Lower proficiency learners are more likely to choose LA overall, suggesting a recency effect. Our results confirm the importance of using aural stimuli when testing interpretation of ambiguous sentences.

Keywords

Sentence parsing; relative clause attachment preferences; ambiguity resolution; prosodic cues; constituent length; second language acquisition.

Introduction

Syntactic ambiguity frequently arises in natural language. A well-known example is provided by constructions containing relative clauses (RCs), where the complex NP includes more than one potential head for the RC. Usually, the semantic or pragmatic content of the RC will bias the interpretation and resolve the ambiguity. For example, in *The boy admired the label of the bottle that was written in Korean*, the RC is understood as modifying the first noun (*label*) in the complex NP, while in *The boy admired the label of the bottle that was filled with orange soda*, the RC modifies the second noun (*bottle*). However, in sentences that are not semantically or pragmatically biased, ambiguity arises: in *Someone shot the servant of the actress who was on the balcony*, the RC can be interpreted as modifying either the first noun (*servant*) or the second noun (*actress*).

Cross-linguistic research has shown that speakers of different languages have distinct default interpretations for ambiguous relative clauses (see, e.g., Cuetos & Mitchell, 1988; Fodor, 2002). For instance, while native speakers of English have a slight preference for low attachment (LA) (i.e., the RC modifies the second noun), native speakers of Spanish prefer high attachment (HA) (i.e., the RC modifies the first noun). In the context of second language (L2) acquisition, a question that arises is whether L2 learners (henceforth L2ers) are able to reset their default interpretation for RC attachment, when the attachment preference of the L2 differs from the L1. Previous research on RC attachment ambiguity in several language pairs has provided mixed answers to this question. Depending on the study, it has been found that L2ers: (i) exhibit no attachment preference in the L2 (e.g., Felser, Roberts, Marinis, & Gross, 2003); (ii) prefer the pattern that is associated with their L1 (e.g., Frenck-Mestre, 2002); or (iii) show the same attachment preference in both of their languages, either the L1 pattern or the L2 pattern (e.g., Dussias, 2003).

Inconsistent findings across previous studies may be due in part to the fact that participants were required to make interpretations based on written stimuli. It has been proposed that native speakers assign prosody in silent reading such that their interpretations reflect the default prosody of the language (Fodor, 2002). Not much is known about the extent to which L2ers are able to use prosody when resolving ambiguities in their second language, and, if so, whether they resort to L1 prosody or make use of target prosody (see Dekydtspotter, Donaldson, Edmonds, Liljestrand Fultz, & Petrush (2008), discussed below). Use of prosody for disambiguation should depend on learners' understanding of the cues to prosodic phrasing in the L2. In this paper, we examine how prosodic cues influence L2ers' interpretation of RCs. Specifically, we investigate whether acoustic cues—indicating prosodic breaks and constituent length—constrain the attachment preferences of L2ers, and whether there are benefits to having both types of cues pointing to the same attachment site. We focus on Spanish-speaking learners of English, given that these languages reportedly have distinct attachment preferences, with the preference for HA in Spanish being stronger than the preference for LA in English (see, e.g., Bergmann, Armstrong, & Maday, 2008).

Relative Clause Attachment Ambiguity

As mentioned, Spanish and English both display ambiguity involving RC attachment. Consider the sentences in (1) (adapted from Cuetos & Mitchell, 1988, p. 77).

(1) (a) El periodista entrevistó a la hija del coronel que tuvo el accidente.
(b) The journalist interviewed the daughter of the colonel who had the accident.

The relative clause (*que tuvo el accidente/who had the accident*) can be interpreted as modifying either the first noun within the constituent that forms the direct object of the verb (*hija/daughter*) or the second noun (*coronel/colonel*).

The bracketing in (2) shows that the syntactic structure of each interpretation is different. In (2a), the RC is the most deeply embedded constituent, sister to the N' that contains only the second noun (*colonel*). In this case, the RC modifies *colonel*, henceforth NP2 (i.e., LA interpretation). In (2b), on the other hand, the relative clause is sister to the N' that contains both nouns (*daughter of the colonel*). In this case, the RC modifies *daughter*, henceforth NP1 (i.e., HA interpretation).[1]

(2) Syntactic structure:

(a) Low attachment (RC modifies NP2):
The journalist interviewed [the daughter [of the colonel [who had the accident]]]

(b) High attachment (RC modifies NP1):
The journalist interviewed [[the daughter [of the colonel]] [who had the accident]]

Much of the research on ambiguity resolution has explored the role of parsing principles in sentence processing, including *late closure* (Frazier, 1978) or *recency* (Gibson, Pearlmutter, Canseco-González, & Hickok, 1996). The idea is that new material should be integrated locally so as not to overburden the processor. When applied to ambiguous RC constructions, such principles lead to the selection of the closest NP, namely NP2, as the head of the RC, resulting in LA. However, as mentioned above, LA is a preference rather than being categorical and many languages in fact have a bias towards HA. In view of this, an additional, competing, principle, *predicate proximity*, was proposed by Gibson et al. (1996) and Gibson, Pearlmutter, and Torrens (1999). This principle favors attachment as close as possible to the head of the predicate phrase, resulting in a preference for NP1. Gibson et al. (1999) hypothesize that the role that this principle plays varies across languages, being stronger in Spanish than in English. Competition between these two parsing principles, late closure/recency and predicate proximity, then, explains, at least in part, variability in RC attachment.

Frazier and Clifton (1996, 1997) proposed an additional principle, *construal*, which makes a distinction between primary phrases (subject, predicate, obligatory arguments) and non-primary phrases (adjuncts and modifiers, such as RCs). The idea is that any non-primary phrase is associated into the thematic domain that is currently being parsed. Where a thematic domain permits association with more than one position, as is the case in the RC construction under consideration, nonstructural information, including semantic and pragmatic factors, may be taken into consideration in deciding where to attach the non-primary phrase.

In the following sections, we explore cross-linguistic differences in RC interpretation, including the ways in which interpretation is constrained by prosody.

[1] Although NP1 and NP2 do not refer to syntactic constituents in (2), these terms are standardly used in the literature on RC attachment; we follow this practice.

Cross-linguistic Differences in Attachment Preferences

As mentioned, languages differ both in their default interpretation for ambiguous relative clauses and in the strength of their attachment bias. Native speakers of English show a LA preference in offline and online tasks (Bergmann et al., 2008; Carreiras & Clifton, 1999; Fernández, 2002, 2003; Hemforth et al., 2015; but see Gilboy, Sopena, Clifton, & Frazier, 1995; Jun, 2010). Native speakers of Spanish, in contrast, show a HA preference, which has been widely demonstrated through offline and online tasks (Bergmann et al., 2008; Carreiras & Clifton, 1993, 1999; Carreiras, Salillas, & Barber, 2004; Cuetos & Mitchell, 1988; Fernández, 2002, 2003; Gibson, Pearlmutter, & Torrens, 1999; Igoa, Carreiras, & Meseguer, 1998; but see Gilboy et al., 1995; Hemforth et al., 2015). Languages other than English with a LA preference include Egyptian Arabic, Farsi, Norwegian, Brazilian Portuguese, Romanian, and Swedish; languages other than Spanish with a HA preference include Afrikaans, Croatian, Dutch, French, German, Japanese, Korean, and Russian (Fodor, 2002; Jun, 2003). Notably, closely related languages can show distinct default preferences (e.g., English vs. Dutch; Spanish vs. Romanian), suggesting that attachment bias may not necessarily relate to syntactic structure.

Turning to cross-linguistic differences in the strength of the attachment bias, the preference for LA in English is relatively weak, only around 60% in offline studies. Indeed, some online studies have found no difference in reaction time between high and low attachment interpretations (Carreiras & Clifton, 1993). And even though the preference for HA in Spanish is often reported as being relatively robust, independent investigations have yielded uneven results. For example, Bergmann et al. (2008) report an overall HA rate of 78%, while Cuetos & Mitchell (1988) have found the default preference to be considerably weaker, around 60%.

Some variability across studies may be attributable to methodological differences in data collection and analysis (see Fernández, 2003), but other factors inherent to the linguistic context are also at play. In accordance with construal, described above, these include the type of conceptual relationship that holds between NP1 and NP2 (e.g., Gilboy et al., 1995; Igoa et al., 1998) or the choice of preposition linking the two NPs, in particular, differences between thematic prepositions like *with*, which carry semantic content, and purely grammatical, case-assigning prepositions like *of* (e.g., Felser et al., 2003; Fernández, 1999). However, while approaches that focus on the influence of semantic factors on RC attachment can explain some of the variability in interpretation preferences, they cannot account for cross-linguistic differences observed in the interpretation of semantically and pragmatically neutral RC constructions.

Prosodic factors also contribute to interpretation, including the position of prosodic breaks and the length of the constituents involved. These factors may interact as well: long relative clauses tend to follow a prosodic break, which makes them more likely to be interpreted as attaching high, even in a LA language like English (Fodor, 1998). In addition, prosody has been proposed to impact interpretation in experimental tasks involving silent reading. Fodor (1998, 2002) advanced the Implicit Prosody Hypothesis, according to which speakers, when reading to themselves, project prosodic structure on sentences, prompting them to choose one interpretation over another.

The next section further discusses the effects of prosody on RC attachment, focusing specifically on the languages of interest for the present study, English and Spanish.

Prosodic Influences on Attachment Site

The syntactic differences in (2) are reflected in the way in which these sentences are parsed in spoken language; prosodic breaks are used by speakers to signal intended meaning and by listeners to interpret potentially ambiguous sentences. In (3a), the presence of a prosodic break (indicated by //) between NP1 and NP2, and the corresponding absence of a break between NP2 and the RC, indicates that the RC attaches low. In (3b), on the other hand, high attachment is signaled through the presence of a prosodic break between NP2 and the RC, and the corresponding absence of a break between NP1 and NP2. The relationship between prosodic break and RC attachment site has been observed for languages with distinct default preferences, including English and Spanish (Jun, 2003).

(3) Prosodic structure:
(a) Low attachment:
The journalist interviewed [the daughter] // [of the colonel who had the accident]

(b) High attachment:
The journalist interviewed [the daughter of the colonel] // [who had the accident]

Cross-linguistically, prosodic breaks are typically signaled through pauses, lengthening of the constituent that immediately precedes the break, and pitch changes (Jun, 2003; see also Fodor, 2002; Jun, 2010). Such pitch changes are typically signaled by boundary tones and/or reset of fundamental frequency (F0) after the break. While boundary tones are used to mark phrasal boundaries regardless of the presence of a pause, F0 reset invariably follows a pause. F0 reset may be employed differently depending on the choice of attachment site. For example, Quinn, Abdelghany, and Fodor (2000; cited in Fodor, 2002) report on a study where native speakers of English exhibit F0 reset at the beginning of a long RC, but only when HA is the desired interpretation. This is consistent with the observation that HA is typically signaled more strongly than LA across languages (Bergmann et al., 2008; Jun, 2003), presumably due to the syntactic discontinuity present in the structure that corresponds to HA; see (2b).

In addition to the position of the prosodic break, it has been shown that speakers' RC attachment preferences are influenced by the length of the relative clause (e.g., Fodor, 1998; Pynte & Colonna, 2000). In general, rates of HA responses are higher when the RC is long, and lower when the RC is short, as exemplified in (4) (see Fernández, 2002, 2003; Hemforth et al., 2015).

(4) Attachment site preferences and relative clause length:

(a) Long RC, HA preferred:

The professor read the review of the poem *that was published at the end of the magazine.*

(b) Short RC, LA preferred:

The professor read the review of the poem *that just came out.*

The effect of RC length on speakers' attachment preferences is a consequence of parsing demands: long RCs tend to be phrased separately, which makes them more likely to be interpreted as attaching high. In contrast, short RCs are less likely to be parsed as a separate prosodic constituent, which forces a local interpretation (e.g., Jun, 2010).

Fodor (1998) has developed an account of ambiguity resolution that takes additional aspects of length into consideration: she notes that prosodic phrasing favors balanced structures "in which sister constituents are roughly equal in prosodic weight" (p. 302). Prosodic weight is defined in terms of length, as well as the number of stressed elements (i.e., phonological words) in a given constituent (see also Fernández, 2003). Fodor (1998) formalizes this observation as the Same-Size Sister Constraint (SSSC) and suggests that relative clause attachment is subject to it: long RCs attach high to balance the whole complex NP whose head is NP1, whereas short RCs favor low attachment to balance only NP2.

We extend Fodor's proposal in two respects: we take the SSSC as implying that a short RC should prefer to go with the shorter of the two NPs, regardless of where in the structure they occur, and similarly for a long RC. Fodor does not discuss potential effects of the length of NP1, independent of the complex NP as a whole. With respect to NP2, she only considers cases where NP2 is short; however, we take the claim that prosodic phrasing favors balanced structures to imply that if both NP2 and the relative clause are long, then the preferred attachment site will remain low rather than high.

Length predictions made on the basis of the SSSC differ from predictions based on RC length alone. Consider the examples in (5), where the NP under focus and the RC are bracketed to highlight their relative lengths.

(5) (a) Long NP1, long RC

The bartender served [the cheerful outgoing cousin] of the actor [that always ordered peanuts with his beer]

(b) Long NP2, long RC

The bartender served the cousin of [the cheerful outgoing actor] [that always ordered peanuts with his beer]

(c) Short NP1, short RC

The bartender served [the cousin] of the cheerful outgoing actor [that ate peanuts]

(d) Short NP2, short RC

The bartender served the cheerful outgoing cousin of [the actor] [that ate peanuts]

Both accounts predict HA for (5a) and LA for (5d). They differ as far as (5b) and (5c) are concerned. Considering only RC length, a long RC will bias towards high attachment, regardless of the relative length of the NPs, while a short RC will tend towards low attachment (see, e.g., Fernández, 2002, 2003; Hemforth et al., 2015). The SSSC, at least on our interpretation, predicts that a long RC will attach low if NP2 is long, as in (5b), while a short RC may attach high if NP1 is short, as in (5c). Essentially, we have chosen to focus on the SSSC rather than on RC length alone because the SSSC is sensitive to the overall rhythmic profile of the sentence.

In summary, both the position of the prosodic break and the length of various constituents have been proposed to affect the interpretation of potentially ambiguous sentences involving RCs. However, in most offline experiments, participants' interpretations are made on the basis of their own silent prosody, which prevents the establishment of broader generalizations about interpretation based on specific prosodic profiles. The experiment we undertake strives to rectify this as stimuli are presented auditorily; both the position of the prosodic break and the length of the RC relative to the NPs are manipulated. Before we describe our experiment, we briefly review findings from previous literature on ambiguity resolution in a second language.

Ambiguity Resolution in Second Language Parsing of Relative Clauses

Previous research on attachment preferences in L2 parsing has examined whether L2ers and bilingual speakers exhibit attachment preferences, and if so, the extent to which the L1, the L2, or other factors, such as L2 proficiency or working memory, shape choices. In addition, there has been work on effects of the L2 on the L1 in heritage speakers (e.g., Jegerski, 2018; Jegerski, Keating, & VanPatten, 2016) and attrited speakers (e.g., Dussias & Sagarra, 2007). Here, we consider research that has focused on L2 learners or sequential (late) bilinguals who acquired the L2 after the L1, as this will be the population of speakers we examine in our experiment.[2]

Studies on L2 parsing of RCs have employed the same offline and online methods used in studies with native speakers. Various L1/L2 combinations have been studied, involving situations where native speakers of one of the languages show a HA preference, while native speakers of the other language prefer LA. Taken together, the studies on RC attachment disagree on the role of transfer of parsing preferences from the L1, as well as on the effects of parsing principles which favor the choice of NP2, like recency (Gibson et al., 1996) or late closure (Frazier, 1978).

In an early study that looked at L2 parsing, Fernández (1999) investigates speakers of Spanish who acquired L2 English after the age of 10, in addition to Spanish–English bilinguals and native speakers of English. The task was an offline questionnaire, where participants were presented with ambiguous sentences similar to (1b), and were required to indicate which NP the RC referred to. At issue was whether the L2ers would give interpretations favoring LA, as

[2] The distinction between L2ers and late bilinguals is often unclear in the literature. Typically, bilinguals are assumed to use both of their languages to a considerable extent, whereas for L2 speakers, this is not necessarily the case.

preferred in English, and also as predicted by recency or late closure. The groups differed significantly; for sentences involving the preposition *of* (e.g., *the review of the play*), the LA preference was 68% for native speakers, 40% for bilinguals, and 30% for late L2ers. All groups made more LA choices when the thematic preposition *with* was involved (e.g., *the singer with the guitarist*), which is, in fact, consistent with construal, although Fernández does not discuss this. L2 proficiency was positively correlated with attachment preference, suggesting that the less proficient learners were strongly influenced by their L1.

Dussias (2003) investigated the interpretations of proficient L2 speakers of Spanish and English, as well as monolingual controls, using an offline questionnaire. She found that Spanish monolinguals overwhelmingly preferred HA (74%), in contrast to English monolinguals (14% HA). Proficient speakers of L2 English dispreferred HA (22%), suggesting no L1 influence from Spanish, while speakers of L2 Spanish chose HA to a greater extent (44%), again suggesting little influence from L1 English. In a Spanish self-paced reading task, with one of the interpretations forced by means of gender agreement, native speakers' reading times were significantly longer when gender indicated LA, whereas Spanish L2ers showed no significant difference in responses to sentences biased toward LA vs. HA.

Using a similar offline task, Felser et al. (2003) found no attachment preference in L2 English when the L1s were German and Greek, both languages with a HA preference. L2ers did, however, show a clear LA preference when the preposition was *with* rather than *of*, similar to the findings of Fernández (1999). In self-paced reading tasks with stimuli disambiguated by means of number, the L2ers showed no significant difference in reaction times to NP1 and NP2 at the disambiguating segment when the preposition was *of*; for *with*, the reaction times to NP2 were significantly faster. Felser et al. conclude that L2ers neither make use of L1 parsing strategies nor rely on a recency effect. When it is available, they make use of lexical-semantic information. Along related lines, Pan, Schimke, and Felser (2015) provided short contexts favoring either NP1 or NP2; they found that native speakers of English, German-speaking learners of English, and Chinese-speaking learners of English used context to determine choice of NP1 or NP2 in an offline task. However, in a self-paced reading task, only the L2ers were affected by the prior context. Following the shallow structure hypothesis (Clahsen & Felser, 2006), these researchers argue that L2ers are less able than native speakers to make use of syntactic structure in parsing. Rather, they rely on non-syntactic cues to determine their interpretations, including word meaning and other contextual information.

In contrast to most of the above studies, Frenck-Mestre (2002) argues for L1 influence, but only at lower proficiency levels. L2 speakers of French (HA) with English L1 (LA) at two levels of proficiency (low, advanced), as well as native speakers, participated in an eye-tracking experiment where they read potentially ambiguous sentences which were in fact disambiguated by agreement on the verb in the RC. The first pass gaze of the lower proficiency group at the disambiguating verb was faster when the verb agreed with NP2 (=LA). In contrast, for the advanced group, as for the native speakers, the gaze was faster when the verb agreed with NP1 (=HA). Frenck-Mestre argues that the lower proficiency group is influenced by the L1, whereas

the advanced group has overcome L1 effects, due to increased experience with the L2. She argues that the behavior of the lower proficiency group cannot be explained as a recency effect.

Another approach to RC ambiguity resolution in L2 focuses on working memory (WM), and the extent to which individual differences in reading span might shape attachment preferences. Hopp (2014) conducted online (eye-tracking) and offline (judgment) tasks with German-speaking (HA) learners of English (LA), testing RC attachment preferences, together with measures of reading span. There was a significant correlation between L2 proficiency and reading span scores.

In the online task, first pass and total reading times on sentences disambiguated by means of verb agreement (e.g., *The director congratulated the instructor of the schoolboys who was writing reports*) were significantly slower for both native speakers and L2ers in the HA condition, suggesting a LA preference. In the offline task, in contrast, the L2ers showed a HA preference, as did the native speakers, perhaps surprisingly. For the L2ers, reading span was negatively correlated with HA responses. In other words, participants with lower reading spans were more likely to choose HA, contrary to what might be expected under a parsing principle like recency. Similar findings were earlier reported by Swets, Desmet, Hambrick, and Ferreira (2007) for adult native speakers of English and Dutch; regardless of language, individuals with the shortest WM span preferred HA, whereas those with higher spans preferred LA. In line with Fodor's implicit prosody hypothesis, these authors hypothesize that low span readers insert a break between NP2 and the RC, leading to a HA preference. Hopp (2014) makes the same argument for L2ers.

These results contrast with those of Felser, Marinis, and Clahsen (2003) who tested English-speaking children (mean age 6;8) on a self-paced listening task, as well as a listening span test. Children in the low listening span group showed faster reaction times for stimuli where LA was involved, in contrast to the high span group who were faster with stimuli favoring HA.

To summarize thus far, studies on L2 parsing of ambiguous relative clauses have suggested a number of different possibilities determining the choices that L2ers make, including L1 influence (or lack thereof), recency effects, reduced reliance on structure compared to native speakers, as well as effects of working memory. We return to these issues in the discussion.

One important methodological concern relating to the studies discussed so far is that the stimuli were not presented auditorily. Assuming that participants impose prosody in silent reading (following Fodor, 2002), this could affect their interpretations of ambiguous sentences. Silent prosody cannot be observed; however, controlling the prosodic profile that participants are exposed to could allow us to determine the extent to which prosodic factors play a role in ambiguity resolution. In other words, it is important to test participants with stimuli that they hear, with the various prosodic factors accounted for.

To our knowledge, studies that have explored the effects of sentence prosody on L2 ambiguity resolution are limited to Dekydtspotter et al. (2008) and Liljestrand Fultz (2007) on L2ers' interpretations of auditorily presented stimuli. In addition, there have been a few studies on prosodic effects manifested in L2ers' productions (e.g., de la Cruz-Pavía & Elordieta, 2015;

Fernández, 2005, 2010). Some research has attempted to test effects of implicit prosody by manipulating RC length but without using auditory stimuli (e.g., Jegerski, 2018; Jegerski et al., 2016).

Dekydtspotter et al. (2008) probe the relationship between syntax and prosody in the parsing of RCs by English speakers who were of intermediate proficiency in French (HA). L2ers were tested on three tasks, one of which involved ambiguous sentences presented aurally, as in (6):

(6) Nous adorons le secrétaire du psychologue qui se promène (au centre-ville).
'We adore the secretary of the psychologist who is taking a walk (downtown).'

Stimuli controlled for prosodic break (after NP1, *secrétaire*, or NP2, *psychologue*) and RC length (long/short, i.e., with or without *au centre-ville*). Results showed no main effects for RC length or position of prosodic break. However, a more detailed examination of the data revealed that about one third of participants were in fact sensitive to the position of the break, in that they chose NP2 as the head of the RC when the break occurred after NP1. In a related study, Liljestrand Fultz (2007) observed that higher proficiency learners were more sensitive to the position of break cues than lower proficiency learners.

As far as we are aware, no study has looked at potential effects on L2 parsing of both the position of the prosodic break and rhythmic factors that go beyond RC length alone. We report on an experiment on RC disambiguation that uses auditory stimuli and manipulates prosodic cues, namely, position of breaks and length of the RC relative to the NPs involved. We hypothesize that native speakers and learners of English will be sensitive to such cues, albeit possibly in different ways, and that performance will be enhanced when both types of cues point to the same attachment site.

Method

In order to test the effects of acoustic cues on native speaker and L2 interpretations of RC attachment, we conducted an experiment in which the stimuli were presented auditorily. Given that English has a weak preference for LA and that LA is typically signaled less strongly than HA, we were interested in determining whether Spanish speakers, whose native language has a stronger preference for HA, can appropriately attend to the prosodic cues to disambiguation used in English and come to mirror native speaker preferences in their second language.

Participants

Participants were native speakers (n = 20) and second language learners (n = 53) of English, between the ages of 18 and 45 at the time of testing. The native speakers were all from North America. The majority of L2ers were originally from Latin America (especially Colombia and Mexico), although a handful were from Spain. (Previous studies on ambiguity resolution in

Spanish RC constructions have found comparable results for both European and Latin American varieties, e.g., Bergmann et al., 2008; Carreiras & Clifton, 1993; Fernández, 2002, 2003.)

Age of first exposure to English was reported by the learners as being between the ages of 4 and 35 (mean age 8), that is, after the age of 3, typically considered to be the cut-off for simultaneous bilingualism (see Lakshmanan, 2009). All participants had moved to an English-speaking country after the age of 13 (mean age 27). All learners had lived in an English-speaking country for a minimum of 6 months, ensuring that they had had exposure to native speaker input.[3]

All participants (both learners and native speakers) were living in Montreal at the time of testing. All reported having had some exposure to French, a HA language, although none of them spoke French (or any other language) at higher than an intermediate level of proficiency.

Learners' proficiency in English was determined by the Versant English Test (Pearson Education Inc., 2008), which examines various aspects of L2 oral language proficiency (sentence mastery, vocabulary, fluency, and pronunciation). The highest possible score is 80; learners who score a minimum of 47 are considered by Pearson to be low intermediate. The scores of the L2ers included in our experiment ranged from 48 to 80; 31 scored in the intermediate range (47-68), and 22 in the advanced range (69-80). As will be discussed below, proficiency was treated as a continuous variable in our statistical model.

Stimuli

Each participant was required to interpret 66 test sentences: 24 target items and 42 fillers. The target items had the shape: *subject verb* NP1 *of* NP2 RC. Unlike some of the studies described above (Felser et al., 2003; Fernández, 1999), we did not include PPs involving *with*. Since *with* favors LA, we wanted to ensure that any preference for LA on the part of the learners could be attributed to their appropriate interpretation of the prosodic cues used in English and not to a possible bias imposed by this particular preposition. All target items were potentially ambiguous, with semantic/pragmatic factors controlled for, such that both NPs were appropriate as potential heads of the RC, as determined by a group of 5 native speaker consultants. The fillers were biclausal; in addition to serving as distractors from the main items, the fillers were included to ensure that participants had sufficient command of biclausal sentences in general without having to contend with ambiguity (see fn. 3).

Ninety-six target items were created according to two prosodic factors: break and constituent length. Although both were tested in every sentence, we discuss them separately here. Considering break alone, the examples in (7) show that the position of the break either falls after NP2, pointing to high attachment, or after NP1, pointing to low attachment.

[3] Twelve additional learners were excluded for one or more of the following reasons: 5 started learning English before age 4; 2 had lived in an English-speaking country for less than 6 months; 6 scored less than 70% on the fillers (see below).

(7) Break:

Break after NP2 (pointing to HA):

(a) The bartender served the cheerful outgoing cousin of the actor // that always ordered peanuts with his beer.

(b) The bartender served the cousin of the cheerful outgoing actor // that always ordered peanuts with his beer.

Break after NP1 (pointing to LA):

(c) The bartender served the cheerful outgoing cousin // of the actor that ate peanuts.

(d) The bartender served the cousin // of the cheerful outgoing actor that ate peanuts.

As far as length is concerned, the RC matches either NP1 or NP2. In the examples in (8), the NP and the RC in question are bracketed to indicate their relative lengths. The sentence types in (8a-b) point to HA; the sentence types in (8c-d) point to LA.[4]

(8) Length:

RC matches NP1 in length (pointing to HA):

(a) Both long: The bartender served [the cheerful outgoing cousin] of the actor [that always ordered peanuts with his beer].

(b) Both short: The bartender served [the cousin] of the cheerful outgoing actor [that ate peanuts].

RC matches NP2 in length (pointing to LA):

(c) Both long: The bartender served the cousin of [the cheerful outgoing actor] [that always ordered peanuts with his beer].

(d) Both short: The bartender served the cheerful outgoing cousin of [the actor] [that ate peanuts].

The RC was always introduced by the complementizer *that*, and was classified as short if it contained four syllables or less, including the complementizer, whereas it was long if it had more than seven syllables; on average, the short RCs were 3-4 syllables in length (two lexical words) while the long ones were 10-11 syllables (four lexical words), consistent with Fernández (2003). The short NPs were 2-3 syllables in length (one lexical word) while the long ones were 7-8 syllables (three lexical words).

[4] The break conditions in (7) combined with the length conditions in (8) add up to four of eight logical possibilities. For example, stimuli following the pattern in (8a), where the RC matches NP1 in length and both are long, always occurred with the break after NP2. Missing from our design are the corresponding cases with the break after NP1. We felt that a longer task, with all eight possibilities, would prove too taxing for lower proficiency L2ers.

The prosodic break after NP1 or NP2 was marked by a pause, a boundary tone on the noun preceding the pause, and F0 reset on the word following the pause. The break cues did not differ depending on whether length favored HA or LA. The use of pauses to signal attachment site has been observed in previous production studies (e.g., Bergmann et al., 2008; Jun, 2003). With regard to pitch, the majority of the test items had either a HL% or HLH% boundary tone on the noun preceding the pause. These contours are possible in pre-pausal position in English (e.g., Gussenhoven, 2004) and were observed by Bergmann et al. (2008) on the NPs in sentences containing relative clauses (see also Fernández, 2005). Regarding F0 reset, although previous literature found this cue only on long RCs following a pause (Quinn et al., 2000; cited in Fodor, 2002), it was present in all of our stimuli, independent of the length of the RC. The duration of the pause differed according to where the break was placed: for breaks after NP1, the average duration of the pause was 370 ms, while for breaks after NP2, it was 500 ms. This is consistent with the finding from production studies that native speakers of English tend to mark the boundary between NP2 and the RC more strongly than the boundary between NP1 and NP2 (Bergmann et al., 2008; Jun, 2003).

Test sentences were produced in a naturalistic way by a native speaker of Canadian English with voice training as well as training in linguistics. The sentences were recorded with a head-mounted microphone, in a sound-attenuated booth. Each target item was recorded with both HA and LA break cues. The prosodic profile of all target items was checked in Praat (Boersma & Weenink, 2019), to ensure that the cues were consistent across stimuli. A list of target items is provided in Appendix A.

Procedure

The experiment involved an interpretation task with stimuli presented only auditorily (based on Hwang, Lieberman, Goad, & White, 2011). Participants listened to the test items through headphones attached to a computer. After each sentence, they were presented with a comprehension question written on the computer screen. For target items, the question asked which NP the RC referred to. Participants had to choose among three possible answers: NP1, NP2, or “I don’t know”. No time limit was imposed.

As mentioned above, each participant heard 24 target items (out of 96 possible target items) and 42 fillers. The target items were distributed across four versions of the experiment in a Latin Square design: each version contained only one sentence out of four possibilities from the 24 different sentence sets. Test items and fillers were pseudo-randomized, and they were preceded by 6 practice items.

The experiment was developed using E-Prime (Schneider, Eschman, & Zuccolotto, 2002). Participants were tested in a sound-attenuated booth. They took approximately 45 minutes to complete the task and background questionnaire; learners took on average an additional 15 minutes to complete the proficiency test. Participants were compensated for their time.

Predictions

Our goal is to test whether prosodic breaks and constituent length (as determined by the SSSC) affect participants' responses, and whether one of them has a larger effect size than the other. Given the preference for HA in the learners' native Spanish, we expect to find a higher proportion of HA responses among learners than native speakers. However, we anticipate that the learners' L1 bias will be mitigated by the prosodic cues to LA (break and length), especially in the case of more proficient learners. In other words, as proficiency increases, learners' responses will be more influenced by the prosodic cues that are present in the stimuli. We also anticipate that the cues will have a cumulative effect, in that both native speakers and learners will have more HA responses when both cues point to HA, and more LA responses when both cues point to LA.

Statistical models

We examined potential differences between native speakers and learners via a logistic regression with GROUP as a simple effect and item and speaker as random intercepts. We then modeled participants' responses for the target items with two mixed effects logistic regressions in R (R Development Core Team, 2019) by means of the glmer() function in the lme4 package (Bates, Maechler, Bolker, & Walker, 2015), whose summary output provides p-values based on asymptotic Wald tests.[5] The maximal converging model containing the native speaker data includes the following predictors: BREAK (whether the position of the break points to high or low attachment), LENGTH (whether the RC matches either NP1 or NP2 in length, pointing to high or low attachment, respectively), an interaction between BREAK and LENGTH, as well as by-speaker and by-item random intercepts. The maximal converging model containing the learner data includes BREAK, LENGTH, VERSANTSCORE, all possible interactions between BREAK, LENGTH and VERSANTSCORE, as well as by-speaker and by-item random intercepts.[6] Both models were initially run with **high** as the reference level for BREAK and LENGTH. To be able to examine the effect of all combinations amongst the levels of BREAK and LENGTH, the same models were refitted with **low** as the reference level for BREAK (see Bates, 2010, for variable releveling in lme4). Below, we report the results for the models where **high** is the reference level for both prosodic variables, unless otherwise noted. In the learners' model, VERSANTSCORE is treated as a continuous variable. Regarding the dependent variable, low attachment was coded as zero and high attachment was coded as 1, which means that the models predict the probability of high attachment. The models disregard "I don't know" responses as they are negligible: 1% of total responses for the native speakers and 3.8% for the learners.[7]

[5] Given that proficiency was a predictor of interest, and Versant Score was obtained only for learners, we modeled native speakers' and learners' responses separately.

[6] Neither the native speaker model nor the learner model converged with random slopes.

[7] In addition, one sentence from one set (item (12b) in Appendix A) was excluded from the analysis as it had the incorrect prosody.

Results

Figure 1 shows the proportion of high attachment responses by the native speakers and learners for the two prosodic variables under examination, namely, BREAK and LENGTH. "High" and "low" indicate whether the prosodic cue in question points to high or low attachment. Learners are grouped together, regardless of proficiency.

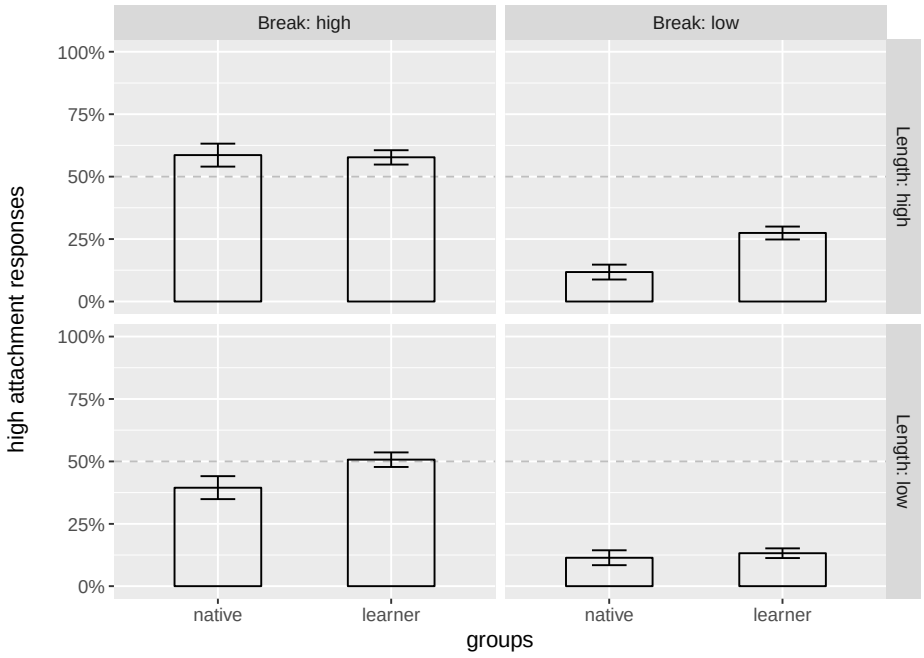

Figure 1. Native speaker and learner responses for BREAK and LENGTH.

As can be seen, the profile of responses is remarkably similar for the native speakers and Spanish-speaking L2ers. Although the overall rate of HA is not much higher for the Spanish speakers (the mean HA response rate is 37.2% for the L2ers vs. 30.3% for the native speakers), the model comparing the two groups of participants indicates that the learners choose significantly more HA responses than the native speakers ($\hat{\beta}$ = 0.37, p = 0.01). Given the relatively small effect size obtained in the model and the proportions in Figure 1, it does not appear that the learners are transferring their L1 preference for HA into English. We discuss the possibility of L1 transfer in light of the results of the statistical models in the next section.

Figure 1 suggests that BREAK has a greater effect on participants' responses than LENGTH: proportions for both groups are very similar for the two facets where BREAK points to HA, as well as for the two facets where BREAK points to LA, and there are seemingly more HA responses when BREAK suggests HA than when it suggests LA. We examine the statistical effects of BREAK and LENGTH below.

As previously mentioned, we expect L2ers' responses to be affected by their proficiency. The trend lines in Figure 2 appear to support this, as the more proficient learners show greater sensitivity to the prosodic cue profile used to signal attachment preferences. Consistent with Figure 1, the trend lines seem to be affected by BREAK: with increasing proficiency, learners

have more HA responses when BREAK points to HA, and fewer HA responses when it points to LA.

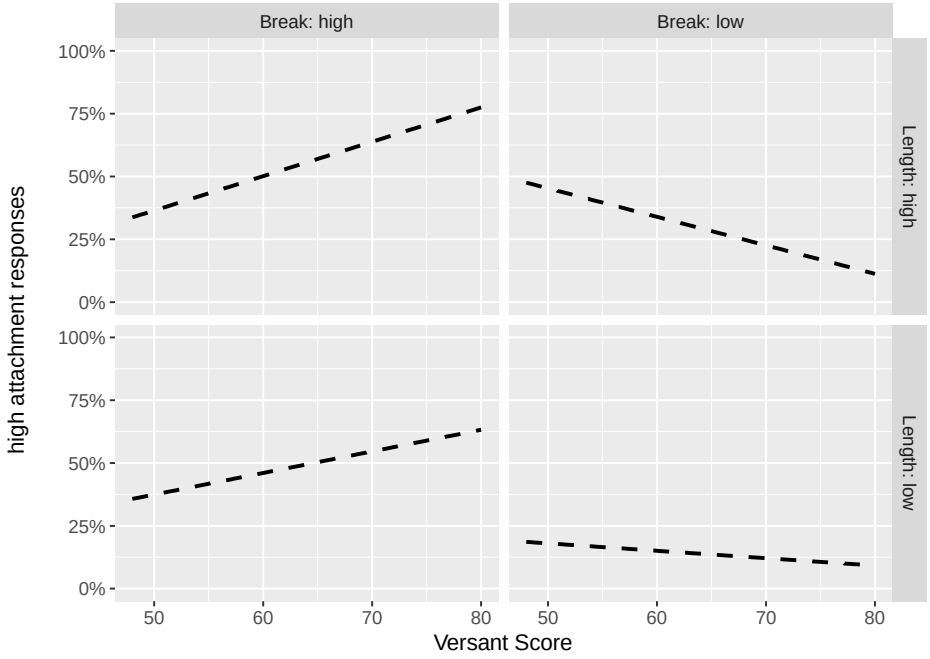

Figure 2. Learner responses by proficiency score for BREAK and LENGTH.

The statistical model with the native speaker data indicates that both BREAK and LENGTH are significant, but the interaction between them is not. The effect size of BREAK ($\hat{\beta}$ = -3.00, $p < 0.0001$) is greater than the effect size of LENGTH ($\hat{\beta}$ = -0.99, $p = 0.02$), consistent with what is observed in Figure 1. The negative estimates for these predictors indicate that when BREAK points to LA (with LENGTH held constant) and when LENGTH points to LA (with BREAK held constant), native speakers have fewer HA responses. In the refitted model, LENGTH is not statistically significant, indicating that when BREAK is held constant at **low**, LENGTH pointing to LA does not affect responses. BREAK, however, remains significant. A table with the statistical model estimates for the native speaker data can be found in Appendix B.

The statistical model indicates that signaling LA either through BREAK or through LENGTH disfavors HA responses, with BREAK having a stronger effect than LENGTH. The absence of a BREAK * LENGTH interaction indicates that the effect of one predictor does not depend on the effect of the other one. We discuss potential cumulative effects of cues in the next section.

Turning to the model with the learner data, BREAK has a significant effect ($\hat{\beta}$ = -1.46, $p < 0.0001$), indicating that when BREAK points to LA (with LENGTH held constant), learners also have fewer HA responses, consistent with the overall proportions shown in Figure 1. VERSANTSCORE ($\hat{\beta}$ = 0.63, $p < 0.0001$) is also significant: the positive estimate indicates that as proficiency increases, learners choose more HA when BREAK and LENGTH are held constant at their reference levels. As the trends in Figure 2 suggest, the interaction between BREAK and

VERSANTSCORE ($\hat{\beta}$ = -1.21, p < 0.0001) is also significant. Here, for each unit increase in proficiency, BREAK pointing to LA lowers the odds of HA by a factor of 3.35.[8] This indicates that as learners' proficiency increases, they become more sensitive to the BREAK cues used to signal RC attachment in the L2. The interaction between BREAK, LENGTH, and VERSANTSCORE is also significant ($\hat{\beta}$ = 0.58, p = 0.04). The positive estimate indicates that each unit increase in VERSANTSCORE leads to more high attachment responses either when BREAK points to HA (and LENGTH points to LA) or when LENGTH points to HA (and BREAK points to LA), compared to when both BREAK and LENGTH point to LA. This is consistent with the trend lines in Figure 2 (compare the Break low/Length low panel with the Break high/Length low panel and with the Break low/Length high panel).

LENGTH is not significant ($\hat{\beta}$ = -0.36, p = 0.13) for the learners in the model where **high** is the reference level for both BREAK and LENGTH. It is significant, though, in the refitted model ($\hat{\beta}$ = -0.95, p = 0.0006), indicating that when BREAK is held constant at **low**, LENGTH pointing to LA yields fewer HA responses, consistent with the proportions in Figure 1. As in the native speaker model, the effect size of LENGTH obtained for the learners is smaller than the effect size of BREAK. In the refitted model, BREAK, VERSANTSCORE, the interaction between BREAK and VERSANTSCORE, and the interaction between BREAK, LENGTH, and VERSANTSCORE remain significant. The other interactions included in the model are not statistically significant. A table with the statistical model estimates for the learner data can be found in Appendix B.[9]

Discussion

In this paper, we have suggested that two different types of prosodic cue—position of prosodic break and length relationship between the RC and potential heads—influence the interpretations of native speakers and L2ers when they are faced with ambiguous RCs in English. Previous research has mostly relied on written stimuli. While issues relating to length can be (and have been) tested with written stimuli, issues relating to breaks cannot. The results of our interpretation task indicate that both native speakers and L2ers show sensitivity to break and length cues, although the effects of length cues are less robust than the effects of break cues. L2ers become more sensitive to break cues with increasing proficiency, with responses

[8] These values correspond to $\exp(|\hat{\beta}|)$. Given that VERSANTSCORE was scaled and centered, each unit is equivalent to one standard deviation in learners' proficiency score (SD = 9.51).

[9] An anonymous reviewer suggested that age of acquisition and amount of exposure to English might play a role in L2ers' preferences. To examine this possibility, we ran another logistic regression with the following predictors: BREAK, LENGTH, VERSANTSCORE (continuous variable), AGEOFACQUISITION (continuous variable), and YEARSOFEXPOSURE (continuous variable). The model also included three-way-interactions among BREAK, LENGTH, and each of the continuous variables, as well as by-speaker and by-item random intercepts. No effect was found for AGEOFACQUISITION, YEARSOFEXPOSURE, or the interactions involving these variables. The effects reported for BREAK, LENGTH and VERSANTSCORE reported in this section were also observed in this model. We compared the model including AGEOFACQUISITION and YEARSOFEXPOSURE with the model that does not include these variables via an ANOVA. Although the two models are not statistically different, the model that does not include these variables has a lower Akaike Information Criterion (AIC; Akaike, 1974; 1356.7 compared to 1360.8) and is thus the model we have reported in the paper.

exhibiting two trends: an increase in HA responses when break points high and an increase in LA responses when break points low. Our finding that sensitivity to break cues correlates with proficiency mirrors what was observed earlier by Liljestrand Fultz (2007). Her study examined attachment preferences in a HA language (French) by learners whose L1 is LA (English), the opposite profile to what was examined here. Our results demonstrate that even a weakly cued preference can be acquired.

Turning to length effects, recall that we have investigated length in terms of the SSSC, rather than in terms of RC length alone. On other approaches, long RCs have been argued to favor HA while short RCs favor LA (e.g., Fernández, 2002, 2003; Hemforth et al., 2015; Jun, 2010). On our approach, a long RC could in fact favor LA—provided that NP2 is also long. Because of the design of our stimuli, the effect of LENGTH that we have found cannot be attributed solely to the RC but must, instead, be due to the SSSC, that is, to the length of the RC relative to NP1 or NP2. In the two sentence types where LENGTH points to HA, NP1 differs in length and the corresponding RC matches it in length; similarly, in the two sentence types where LENGTH points to LA, NP2 differs in length, as does the corresponding RC.

Recall that, in our stimuli, break and length cues either match (both signaling HA or LA) or mismatch (one signaling HA and the other LA) (see Figure 1). We might thus expect the highest proportion of HA responses to be found when both BREAK and LENGTH point high and the lowest proportion of HA responses to be found when both cues point low, while cue mismatch should lead to proportions that are intermediate between these two. Examination of the model estimates reveals that cues do indeed have a cumulative effect in both native speaker and L2 responses. However, this effect is manifested differently for the two groups, as detailed below.

In the model for the native speakers, the significant result for LENGTH when **high** is the reference level for BREAK can be seen in the lower left panel in Figure 1, where BREAK points high and LENGTH points low. Here, native speakers have fewer HA responses than when both cues point high. In the case of the learners, the significant result in the refitted model (with **low** as the reference level for BREAK) can be seen in the lower right panel in Figure 1, where both BREAK and LENGTH point low. Here, the learners have fewer HA responses than when BREAK points low and LENGTH points high. As previously mentioned, the effect size of LENGTH in both models is not as strong as the effect size of BREAK, explaining (i) why the profile of native speaker responses in the lower left panel is more similar to the panel above it, where both cues point high, rather than to the panel to its right, where BREAK points low, and (ii) why the profile of learner responses in the lower right panel is more similar to the panel above it, where BREAK points low, rather than to the panel to its left, where BREAK points high but LENGTH points low. Although previous studies have looked at the effects of break and constituent length (e.g., Dekydtspotter et al., 2008), they have not examined a potential conspiracy of cues. Our experimental design has allowed us to probe whether cues may work together or against each other, as well as to potentially determine the relative strength of each type of cue. On the whole,

our results provide evidence for cumulative effects, even though these are manifested in opposite directions for the two groups under analysis.

An additional point needs consideration, however, namely how we have determined length effects. We have suggested that looking at the SSSC as opposed to RC length alone is advantageous as it takes account of the rhythmic structure of the entire sentence. However, our design does not allow us to explore the effects of the SSSC, RC length, and the location of the break to their full potential. As mentioned in footnote 4, we included 4 out of 8 logically possible sentence types so as not to overburden lower proficiency learners. By including all 8 options, the relative contribution of each cue could be more thoroughly investigated, including the possibility of augmented or diminished effects when a subset of cues point to the same or different attachment sites, as observed by Hwang et al. (2011) for native speakers of English.

One notable finding in our study is that the proportion of HA preferences was not particularly high overall for the L2ers, only 37.2%. Given that the preference for HA in Spanish is reported to be about 70% across studies, and given that our stimuli were designed such that in 75% of cases, the break cue and/or the length cue points to HA, our expectation was that L1 transfer would play a role, especially for less proficient learners, leading to a proportion of HA choices that mirrors what has been observed in earlier studies on Spanish. The results in Figure 1 appear to be inconsistent with this: the L2ers perform similarly to the native speakers, suggesting little L1 influence. However, when considering the results in Figure 2, where L2ers' responses are plotted by proficiency, a different picture emerges. It can be seen that when the break cue points low (and the length cue points high), the slope is negative, indicating that lower proficiency learners are more likely to choose HA than higher proficiency learners, which suggests a possible L1 effect. Conversely, when the break cue points high, the slope is positive; indeed, the proportion of HA responses for the higher proficiency learners well exceeds that of the native speakers (see the left panels in Figure 1 where the number of HA responses for the native speakers is only 58.6% (top) and 39.5% (bottom)), indicating that transfer is likely playing a role for higher proficiency learners as well. If this is indeed the case, this finding differs from the results of Frenck-Mestre (2002) who observes L1 effects only for lower proficiency learners. This may be due, in part, to the language combination examined. Frenck-Mestre focused on English-speaking learners of French, a HA language, in contrast to our focus on Spanish-speaking learners of English, a LA language. Given that, as previously mentioned, LA is less strongly cued across languages, L1 effects may be more persistent in the language combination we have examined.

In the language pair under focus in Frenck-Mestre's study, both L1 transfer and a recency effect would lead to a preference for LA; in the latter case, LA would be favored because it is less demanding on working memory (Dussias, 2003), though Frenck-Mestre dismisses this possibility in the case of her data. In contrast, in the language pair under focus in our study, L1 transfer and recency make opposite predictions, the former leading to a HA preference and the latter to LA. We have found little evidence of L1 transfer in the high break conditions for less proficient learners in our study. We speculate that their behavior may indeed be attributable to

recency effects: when the profile of their responses across all combinations of BREAK and LENGTH is examined (see Figure 2), we find that the proportion of HA responses rarely rises above 50%. If LA is generally favored by less proficient learners because of working memory constraints, this will impede detection of prosodic cues to HA present in the stimuli.

Recall that Hopp (2014), in his investigation of WM effects on parsing of relative clauses, found that L2 participants with lower reading spans were more likely to choose HA in an offline task, apparently at odds with our proposal here. Hopp suggested that low span L2 readers insert a break between NP2 and the RC, leading to a HA preference, following a similar proposal by Swets et al. (2007) for native speakers. Crucially, Hopp's offline task required participants to read the stimuli, so it is indeed conceivable that they were inserting a break at that point in their silent reading. Our stimuli, on the other hand, were auditory, with breaks predetermined by us, in accordance with the hypotheses being tested. It is possible, then, that participants with low WM span resorted to recency to determine their preferred responses, since they could not control the segmentation of the stimuli. Consistent with this proposal is the fact that Felser, Marinis, and Clahsen (2003), using a self-paced listening task, found that children with low listening spans responded faster to LA stimuli than to HA. This was in spite of the fact that all of their stimuli appeared in predetermined groupings which in fact favored HA.

Since we tested the proficiency of our participants but not their WM, our proposal can only be speculative. Although Hopp reports a correlation between WM and proficiency, he did not find effects for proficiency alone. A direction for further research would be to include tests of WM and to compare written and auditory stimuli to see whether or not participants show different attachment preferences depending on task modality and memory span. Additionally, it would be worth testing L2ers in their L1 as well, in order to determine the extent to which our assumptions about potential transfer from the L1 are justified.

A final issue to consider is whether the comparative lack of HA responses overall indicates that the L2ers are adopting an astructural approach to syntactic parsing. Felser and colleagues (Clahsen & Felser, 2006; Felser et al., 2003; Pan et al., 2015) argue that L2ers fail to show consistent preferences for either NP1 or NP2 and that this supports the shallow structure hypothesis. We concur with Hopp (2014) and Cunnings (2017) that LA responses are not astructural; rather they indicate a choice made on structural grounds, including recency. Furthermore, the fact that the L2ers in our study can make use of break cues, aligning them with syntactic boundaries, suggests that their LA responses are in fact structurally based rather than simply linear.

In sum, we suggest that the L2ers' relatively low rate of HA responses may be attributable to two factors: (i) working memory constraints, principally for the lower proficiency learners, and (ii) the finding that break cues to LA were detected in the stimuli, principally for the higher proficiency learners.

In conclusion, previous investigation of cross-linguistic differences in RC interpretation for L2ers has largely focused on syntactic, semantic, or pragmatic factors, and has mostly relied on written stimuli, in some cases yielding inconsistent results. We have highlighted the

importance of also investigating prosodic cues to sentence interpretation, by means of auditorily-presented stimuli. When such factors are taken into consideration, a more nuanced picture of native speaker and learner behavior is obtained.

Acknowledgments

We would like to thank three anonymous reviewers for their helpful comments. We would also like to thank our collaborators on earlier related work, Hyekyung Hwang and Moti Lieberman, as well as Alex Friesen, Dan Goodhue, Jennifer Morehouse, and Adèle-Elise Prévost for research assistance, and Guilherme Garcia for help with the statistics. This work was supported by grants from SSHRC 410-2011-0809, 435-2015-0490 and FRQSC 2010-SE-130727, 2016-SE-188196. All authors contributed equally.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, AC-19,* 716–723.

Bates, D. M. (2010). *lme4: Mixed-effects modeling with R*. Unpublished manuscript, Department of Statistics, University of Wisconsin, Madison. Retrieved from: https://www.researchgate.net/publication/235709638.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67(1),* 1–48.

Bergmann, A., Armstrong, M., & Maday, K. (2008). Relative clause attachment in English and Spanish: A production study. *Proceedings of Speech Prosody 2008.* Campinas, Brazil (pp. 505–508).

Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer* (Version 6.0.46) [Computer software]. http://www.praat.org

Carreiras, M., & Clifton Jr, C. (1993). Relative clause interpretation preferences in Spanish and English. *Language and Speech, 36(4),* 353–372.

Carreiras, M., & Clifton Jr, C. (1999). Another word on parsing relative clauses: Eyetracking evidence from Spanish and English. *Memory & Cognition, 27(5),* 826–833.

Carreiras, M., Salillas, E., & Barber, H. (2004). Event-related potentials elicited during parsing of ambiguous relative clauses in Spanish. *Cognitive Brain Research, 20,* 98–105.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27,* 3–42.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition, 30,* 73–105.

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition, 20,* 659–678.

De la Cruz-Pavía, I., & Elordieta, G. (2015). Prosodic phrasing of relative clauses with two possible antecedents in Spanish: A comparison of Spanish native speakers and L1 Basque bilingual speakers. *Folia Linguistica, 49,* 185–204.

Dekydtspotter, L., Donaldson, B., Edmonds, A. C., Liljestrand Fultz, A., & Petrush, R. A. (2008). Syntactic and prosodic computations in the resolution of relative clause attachment ambiguity by English-French learners. *Studies in Second Language Acquisition, 30,* 453–480.

Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition, 25,* 529–557.

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism: Language and Cognition, 10,* 101–116.

Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition, 11,* 127–163.

Felser, C., Roberts, L, Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics, 24,* 453–489.

Fernández, E. M. (1999). Processing strategies in second language acquisition: Some preliminary results. In E. C. Klein & G. Martohardjono (Eds.), *The development of second language grammars: A generative approach* (pp. 217–239). Amsterdam: John Benjamins.

Fernández, E. M. (2002). Relative clause attachment in bilinguals and monolinguals. In R. R. Heredia & J. Altarriba (Eds.), *Bilingual Sentence Processing* (pp. 187–215). Amsterdam: North-Holland/Elsevier.

Fernández, E. M. (2003). *Bilingual sentence processing: Relative clause attachment in English and Spanish*. Amsterdam: John Benjamins.

Fernández, E. M. (2005). The prosody produced by Spanish-English bilinguals: A preliminary investigation and implications for sentence processing. *Revista da Associação Brasileira de Linguística, 4(1/2),* 109–141.

Fernández, E. M. (2010). Reading aloud in two languages: The interplay of syntax and prosody. In B. VanPatten & J. Jegerski (Eds.), *Research in second language processing and parsing* (pp. 297–317). Amsterdam: John Benjamins.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research, 27,* 285–319.

Fodor, J. D. (2002). Prosodic disambiguation in silent reading. In M. Hirotani (Ed.), *NELS 32: Proceedings of the Thirty-Second Annual Meeting of the North East Linguistic Society* (pp. 113–132). University of Massachusetts, Amherst, MA: GLSA.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies* (Doctoral dissertation). University of Connecticut, Storrs, CT.

Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.

Frazier, L., & Clifton, C. (1997). Construal: Overview, motivation and some new evidence. *Journal of Psycholinguistic Research, 26,* 277–295.

Frenck-Mestre, C. (2002). An on-line look at sentence processing in the second language. In R. R. Heredia & J. Altarriba (Eds.), *Bilingual sentence processing: Advances in psychology* (pp. 217–236). Amsterdam: North-Holland/Elsevier.

Gibson, E., Pearlmutter, N., Canseco-González, E., & Hickok, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition, 59,* 23–59.

Gibson, E., Pearlmutter, N., & Torrens, V. (1999). Recency and lexical preferences in Spanish. *Memory & Cognition, 27(4),* 603–611.

Gilboy, E., Sopena, J.-M., Clifton Jr., C., & Frazier, L. (1995). Argument structure and association preferences in Spanish and English complex NPs. *Cognition, 54,* 131–167.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.

Hemforth, B., Fernandez, S., Clifton Jr, C., Frazier, L., Konieczny, L., & Walter, M. (2015). Relative clause attachment in German, English, Spanish and French: Effects of position and length. *Lingua, 166,* 43–64.

Hopp, H. (2014). Working memory effects in the L2 processing of ambiguous relative clauses. *Language Acquisition, 21,* 250–278.

Hwang, H., Lieberman, M., Goad, H. & White, L. (2011). Syntactic ambiguity resolution: Effects of prosodic breaks and prosodic length. In M. Byram Washburn et al. (Eds.), *Proceedings of WCCFL 28* (pp. 267–274). Somerville, MA: Cascadilla Proceedings Project.

Igoa, J. M., Carreiras, M., & Meseguer, E. (1998). A study on late closure in Spanish: Principle-grounded vs. frequency-based accounts of attachment preferences. *The Quarterly Journal of Experimental Psychology, 51A(3),* 561–592.

Jegerski, J. (2018). Sentence processing in Spanish as a heritage language: A self-paced reading study of relative clause attachment. *Language Learning, 68(3),* 598–634.

Jegerski, J., Keating, G. D., & VanPatten, B. (2016). On-line relative clause attachment strategy in heritage speakers of Spanish. *International Journal of Bilingualism, 20(3),* 254–268.

Jun, S.-A. (2003). Prosodic phrasing and attachment preferences. *Journal of Psycholinguistic Research, 32,* 219–249.

Jun, S.-A. (2010). The implicit prosodic hypothesis and overt prosody in English. *Language and Cognitive Processes, 25(7/8/9),* 1201–1233.

Lakshmanan, U. (2009). Child second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 377–399). Bingley, UK: Emerald.

Liljestrand Fultz, A. (2007). Prosody in syntactic disambiguation in English-French interlanguage. In H. Caunt-Nulton, S. Kulatilake & I.-H. Woo (Eds.), *BUCLD 31: Proceedings of the 31st annual Boston University Conference on Language Development* (pp. 394–405). Somerville, MA: Cascadilla Press.

Pan, H.-Y., Schimke, S., & Felser, C. (2015). Referential context effects in non-native relative clause ambiguity resolution. *International Journal of Bilingualism, 19(3),* 298–313.

Pearson Education Inc. (2008). Versant English Test. https://www.versanttests.com/

Pynte, J., & Colonna, S. (2000). Decoupling syntactic parsing from visual inspection: The case of relative clause attachment in French. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 529–547). Oxford: Elsevier.

Quinn, D., Abdelghany, H., & Fodor, J. D. (2000). More evidence of implicit prosody in silent reading: French, English and Arabic relative clauses. Poster presented at the 13th Annual CUNY Conference on Human Sentence Processing, La Jolla, CA.

R Development Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh: Psychology Software Tools Inc.

Swets, B., Desmet, T., Hambrick, D., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, *136*, 64–81.

Appendix A. Target items.

1. a. The bartender served the cheerful outgoing cousin of the actor that always ordered peanuts with his beer.
   b. The bartender served the cousin of the cheerful outgoing actor that always ordered peanuts with his beer.
   c. The bartender served the cheerful outgoing cousin of the actor that ate peanuts.
   d. The bartender served the cousin of the cheerful outgoing actor that ate peanuts.
2. a. The detective pursued the fashionably dressed bride of the prince that watched tennis at Wimbledon last year.
   b. The detective pursued the bride of the fashionably dressed prince that watched tennis at Wimbledon last year.
   c. The detective pursued the fashionably dressed bride of the prince that watched tennis.
   d. The detective pursued the bride of the fashionably dressed prince that watched tennis.
3. a. The boy climbed the ornate old-fashioned fence of the house that was partly covered with grape vines and ivy.
   b. The boy climbed the fence of the ornate old-fashioned house that was partly covered with grape vines and ivy.
   c. The boy climbed the ornate old-fashioned fence of the house that was painted.
   d. The boy climbed the fence of the ornate old-fashioned house that was painted.
4. a. The architect sketched the splendid candlelit foyer of the mansion that was filled with gorgeous statues.
   b. The architect sketched the foyer of the splendid candlelit mansion that was filled with gorgeous statues.
   c. The architect sketched the splendid candlelit foyer of the mansion that had statues.
   d. The architect sketched the foyer of the splendid candlelit mansion that had statues.
5. a. The director discussed the well-known beautiful set of the movie that cost a large sum of money.
   b. The director discussed the set of the well-known beautiful movie that cost a large sum of money.
   c. The director discussed the well-known beautiful set of the movie that was costly.
   d. The director discussed the set of the well-known beautiful movie that was costly.
6. a. The crewman cleaned the extremely heavy oar of the boat that was crafted by local artisans.
   b. The crewman cleaned the oar of the extremely heavy boat that was crafted by local artisans.
   c. The crewman cleaned the extremely heavy oar of the boat that was yellow.
   d. The crewman cleaned the oar of the extremely heavy boat that was yellow.

7. a. The housekeeper searched for the flighty scatter-brained maid of the duchess that was looking out the window at the car accident.
   b. The housekeeper searched for the maid of the flighty scatter-brained duchess that was looking out the window at the car accident.
   c. The housekeeper searched for the flighty scatter-brained maid of the duchess that was eating.
   d. The housekeeper searched for the maid of the flighty scatter-brained duchess that was eating.
8. a. The journalist interviewed the sarcastic witty agent of the author that knew a lot of celebrities in town.
   b. The journalist interviewed the agent of the sarcastic witty author that knew a lot of celebrities in town.
   c. The journalist interviewed the sarcastic witty agent of the author that knows everyone.
   d. The journalist interviewed the agent of the sarcastic witty author that knows everyone.
9. a. The therapist treated the extremely rude patient of the nurse that was at the walk-in clinic yesterday.
   b. The therapist treated the patient of the extremely rude nurse that was at the walk-in clinic yesterday.
   c. The therapist treated the extremely rude patient of the nurse that was Dutch.
   d. The therapist treated the patient of the extremely rude nurse that was Dutch.
10. a. The cop greeted the talented clever son of the cook that helped solve the murder case.
    b. The cop greeted the son of the talented clever cook that helped solve the murder case.
    c. The cop greeted the talented clever son of the cook that solved murders.
    d. The cop greeted the son of the talented clever cook that solved murders.
11. a. The visitor toured the beautiful stately garden of the palace that was used as a set in many well-known movies.
    b. The visitor toured the garden of the beautiful stately palace that was used as a set in many well-known movies.
    c. The visitor toured the beautiful stately garden of the palace that was famous.
    d. The visitor toured the garden of the beautiful stately palace that was famous.
12. a. The journalist photographed the fiery short-tempered daughter of the king that had exceptional taste in coffee.
    b. The journalist photographed the daughter of the fiery short-tempered king that had exceptional taste in coffee.
    c. The journalist photographed the fiery short-tempered daughter of the king that loved coffee.
    d. The journalist photographed the daughter of the fiery short-tempered king that loved coffee.

13. a. The tourist disliked the horribly ugly roof of the church that was severely in need of some work.
    b. The tourist disliked the roof of the horribly ugly church that was severely in need of some work.
    c. The tourist disliked the horribly ugly roof of the church that needed work.
    d. The tourist disliked the roof of the horribly ugly church that needed work.
14. a. The chef hired the highly accomplished friend of the waiter that knows a lot about hockey history.
    b. The chef hired the friend of the highly accomplished waiter that knows a lot about hockey history.
    c. The chef hired the highly accomplished friend of the waiter that plays hockey.
    d. The chef hired the friend of the highly accomplished waiter that plays hockey.
15. a. The reporter questioned the surprisingly old coach of the gymnast that always got nervous before a competition.
    b. The reporter questioned the coach of the surprisingly old gymnast that always got nervous before a competition.
    c. The reporter questioned the surprisingly old coach of the gymnast that looked nervous.
    d. The reporter questioned the coach of the surprisingly old gymnast that looked nervous.
16. a. The fan despised the wildly popular mascot of the team that was renamed after the selling of the franchise.
    b. The fan despised the mascot of the wildly popular team that was renamed after the selling of the franchise.
    c. The fan despised the wildly popular mascot of the team that was renamed.
    d. The fan despised the mascot of the wildly popular team that was renamed.
17. a. The professor read the overly wordy review of the poem that was published at the end of the magazine.
    b. The professor read the review of the overly wordy poem that was published at the end of the magazine.
    c. The professor read the overly wordy review of the poem that just came out.
    d. The professor read the review of the overly wordy poem that just came out.
18. a. The lawyer contacted the quiet serious clerk of the judge that took copious notes at the last trial.
    b. The lawyer contacted the clerk of the quiet serious judge that took copious notes at the last trial.
    c. The lawyer contacted the quiet serious clerk of the judge that took notes.
    d. The lawyer contacted the clerk of the quiet serious judge that took notes.
19. a. The singer called the excitable tense sister of the groom that liked going to expensive restaurants.

b. The singer called the sister of the excitable tense groom that liked going to expensive restaurants.
c. The singer called the excitable tense sister of the groom that liked parties.
d. The singer called the sister of the excitable tense groom that liked parties.

20. a. The sculptor finished the elegantly carved base of the statue that was made of mahogany from South America.
b. The sculptor finished the base of the elegantly carved statue that was made of mahogany from South America.
c. The sculptor finished the elegantly carved base of the statue that was oak.
d. The sculptor finished the base of the elegantly carved statue that was oak.

21. a. The emperor described the famed elaborate painting of the crown that was destroyed in the fire.
b. The emperor described the painting of the famed elaborate crown that was destroyed in the fire.
c. The emperor described the famed elaborate painting of the crown that was stolen.
d. The emperor described the painting of the famed elaborate crown that was stolen.

22. a. The publisher rejected the shockingly racy cover of the book that was sent to him for approval.
b. The publisher rejected the cover of the shockingly racy book that was sent to him for approval.
c. The publisher rejected the shockingly racy cover of the book that was soon leaked.
d. The publisher rejected the cover of the shockingly racy book that was soon leaked.

23. a. The teenagers spied on the amazingly cute niece of the teacher that was walking dogs in the park.
b. The teenagers spied on the niece of the amazingly cute teacher that was walking dogs in the park.
c. The teenagers spied on the amazingly cute niece of the teacher that was jogging.
d. The teenagers spied on the niece of the amazingly cute teacher that was jogging.

24. a. The dean looked at the charming historic emblem of the college that was pictured on the front of the admissions brochure.
b. The dean looked at the emblem of the charming historic college that was pictured on the front of the admissions brochure.
c. The dean looked at the charming historic emblem of the college that was on TV.
d. The dean looked at the emblem of the charming historic college that was on TV.

Appendix B. Statistical models.

1. Native speaker model.

Reference levels: Break = high, Length = high

	Estimate	Std. error	*z* value	*p* value
Intercept	0.44	0.33	1.34	0.18
Break (low)	-3.00	0.55	-5.42	< 0.0001
Length (low)	-0.99	0.45	-2.18	0.02
Break (low) : Length (low)	1.00	0.73	1.36	0.17

Reference levels: Break = low, Length = high

	Estimate	Std. error	*z* value	*p* value
Intercept	-2.55	0.45	-5.62	< 0.0001
Break (high)	3.00	0.55	5.42	< 0.0001
Length (low)	0.005	0.57	0.01	0.99
Break (high) : Length (low)	-1.00	0.73	-1.36	0.17

2. Learner model.

Reference levels: Break = high, Length = high

	Estimate	Std. error	*z* value	*p* value
Intercept	0.37	0.17	2.08	0.03
Break (low)	-1.46	0.25	-5.83	< 0.0001
Length (low)	-0.36	0.24	-1.49	0.13
VersantScore	0.63	0.14	4.48	< 0.0001
Break (low) : Length (low)	-0.59	0.36	-1.61	0.10
Break (low) : VersantScore	-1.21	0.19	-6.16	< 0.0001
Length (low) : VersantScore	-0.26	0.18	-1.43	0.15
Break : Length : VersantScore	0.58	0.28	2.01	0.04

Reference levels: Break = low, Length = high

	Estimate	Std. error	*z* value	*p* value
Intercept	-1.09	0.18	-5.79	< 0.0001
Break (high)	1.46	0.25	5.83	< 0.0001
Length (low)	-0.95	0.27	-3.41	0.0006
VersantScore	-0.57	0.15	-3.82	0.0001
Break (high) : Length (low)	0.59	0.36	1.61	0.10
Break (high) : VersantScore	1.21	0.19	6.16	< 0.0001
Length (low) : VersantScore	0.32	0.22	1.43	0.15
Break : Length : VersantScore	-0.58	0.28	-2.01	0.04