**Effects of Working Memory Capacity, Bilingualism, and Language Proficiency**

# on Ambiguous Relative Clause Attachment

A major focus of research on language comprehension is the resolution of linguistic ambiguity. At any point in a sentence, we may come across ambiguity at one or more levels: Lexical, syntactic, phonological, or morphological. Consider an example of syntactic ambiguity, for instance, in the following sentence: *The man saw someone on the hill with a telescope.* This sentence is globally ambiguous – there are two associated interpretations. One interpretation is the person on the hill had the telescope while the alternative interpretation is that the man used the telescope to watch the person.

Here, we explore individual differences in working memory capacity and language experience, two widely investigated areas of ambiguity resolution that have largely been studied separately up until this point. Considering recent findings concerning language exposure and syntactic parsing strategies of bilinguals (e.g., Dussias, 2003; Dussias & Sagarra, 2007; Jegerski et al., 2016), we aim to follow up on recent research to provide a more representative account for bilinguals from varying language backgrounds.

# Theories of Relative Clause Attachment

The attachment of ambiguous relative clauses (RC) has been at the center of discussions regarding syntactic ambiguity resolution for decades (e.g., Frazier, 1978; Frazier & Rayner, 1982; Ferreira & Clifton, 1986; Cuetos & Mitchell, 1988; Frazier & Clifton, 1997). For example, in the globally ambiguous sentence *The friend of the moviestar who was sitting on the balcony was under investigation*, the RC can be syntactically attached to either noun: *The friend (N1)* or *the moviestar (N2)*. Although both interpretations are possible, it seems L1 English speakers prefer to syntactically attach the RC to the second N2 *(e.g., the moviestar)*, which is commonly referred to as a low attachment preference (in reference to the location of the phrase in the syntactic tree). In the past, this disambiguation preference has been captured by the garden-path theory, which proposes readers prefer low attachment because it is the most recently processed and accessible attachment site (Frazier, 1979). From this perspective, the parser uses two specific, universal heuristics for ambiguity resolution: Minimal Attachment and Late Closure. Under the strategy of Minimal Attachment, the parser can structure new material more quickly, thus leading to greater processing advantages. The theory of Late Closure pushes the reader towards a low attachment preference as the relative clause should be attached to the word or phrase “currently being processed” (Frazier, 1979).

However, it is worth noting that although there is a low attachment preference, this preference is usually mild (e.g., 55/45) so not all relative clauses get attached low. And importantly, the garden-path model was originally based on the processing of English, and later a great number of studies showed that RC attachment preferences vary cross-linguistically (e.g., Brysbaert & Mitchell, 1996; Carreiras & Clifton, 1993; Carreiras & Clifton, 1999; Cuetos & Mitchell, 1988; De Vincenzi & Job, 1993; Gibson et al., 1996; Gilboy et al., 1995; Hemforth et al., 2000a; Mitchell et al., 1995; Mitchell & Brysbaert, 1998; Zagar et al., 1997). For example, Spanish has been identified as a high attachment language (e.g., Cuetos & Mitchell, 1988; Gilboy, Sopena, Clifton, & Frazier, 1995; Carreiras & Clifton, 1999; Carreiras, Salillas, & Barber, 2004). A list of additional high attachment languages includes (but is not limited to) German, Dutch, French, and Russian. Low attachment languages include Mandarin (Shen, 2006), Arabic (Bidaoui, Foote, & Abunasser, 2016), Basque (Gutierrez-Ziardegi et al., 2004) and Swedish (Ehrlich, 1999). Again, although these preferences are statistically attested, it is also important to appreciate that they are not absolute, with preference values ranging from 55% to 65%.

But why would these disambiguation preferences vary from one language to another? One theory to explain this variation is The Tuning Hypothesis (Mitchell & Cuetos, 1991) which posits that an individual’s attachment preference is simply a reflection of the frequency of exposure to relative clauses at the possible attachment sites, with the parser “tuning” itself so that its preferences mirror those frequencies.

Although the Tuning Hypothesis was not originally proposed as a theory of multilingualism, it has influenced exposure-based accounts of RC attachment. In the case of bilinguals, the less dominant language could be influenced by the language with the greatest exposure under the assumption that greater exposure to one language would also entail greater exposure to specific linguistic forms (such as relative clauses). A series of studies have observed such trends. For example, Fernandez (2003) reported effects of bilingual language dominance in the offline processing of two bilingual groups with opposing dominance – Spanish-dominant Spanish-English bilinguals and English-dominant Spanish-English bilinguals. The Spanish-dominant sample generally preferred high attachment in both Spanish and English whereas the English-dominant sample showed a more-neutral preference in both languages (recall that high attachment is more common in Spanish and low attachment in English). It is important to note that the study did not account for additional potentially important language experience variables – for example, age of onset of bilingualism and exposure—that could have influenced both offline preferences and online reading behavior.

Several recent studies have closely examined the impact of these variables on L2 processing. For example, Dussias (2003) explored offline preferences in two different groups of highly proficient bilinguals: Spanish-English and English-Spanish bilinguals. The Spanish-English sample was immersed in an English environment (their L2) for an average of 7.5 years at the time of participation. Dussias observed that both groups preferred low attachment, a preference consistent with monolingual speakers of English. Similarly, two follow-up studies that measured online processing observed a similar low-attachment strategy for Spanish-English bilinguals who had been immersed in the United States for a shorter period (mean of 3.7 years) than in Dussias (2003) (Dussias, 2004; Dussias & Sagarra, 2007). A possible explanation for these results appeals to the notion of language exposure. Exposure to RC English constructions in Spanish-English bilinguals’ day-to-day immersion in English could be a reason for the low attachment preference observed in these studies.

However, recent studies indicate that heritage bilingual users of Spanish may not be sensitive to ongoing language exposure despite significant exposure to their second language (Jegerski et al., 2016; Jegerski, 2018). Heritage speakers are commonly defined as early bilinguals whose first language is a language acquired in an informal setting (e.g., household), while their second is learned in a more formal setting (e.g., school; Montrul, 2005). Jegerski et al. (2016) found the late-bilinguals’ preferences were like those observed in previous studies on bilinguals (i.e., a low-attachment preference in Spanish), but heritage speakers showed a distinct ‘monolingual-like’ preference (69%) for high-attachment. From an exposure-based account’s perspective, heritage speakers should show a large effect of exposure to English because they are early bilinguals with significant second language exposure, and thus would be expected to prefer low attachment even in Spanish. However, in a follow-up study, Jegerski (2018) found that even heritage speakers with an intermediate L1 proficiency behave similarly to the highly proficient bilinguals from Jegerski et al. (2016). One possible explanation for this finding, taken from Dussias (2003; 2004), Dussias & Sagarra (2007), and Jegerski, VanPatten et al. (2016), is that although heritage speakers have significant exposure to English, they may be more equipped at handling such interference due to their early language experience with both languages. Rather surprisingly, the two languages appear to act as two separate systems that do not experience competition with one another, or at the very least, RC attachment may be an area of sentence processing that early onset of bilingualism and extensive exposure creates an advantage.

The finding that syntactic parsing strategies are guided by relative language exposure also suggests that such strategies could be influenced by individual differences in proficiency. An exception is recent work that has shown there to be an effect of L2 proficiency on L1 processing strategies in L2 learners (Frenck-Mestre, 2002; Miyao & Omaki, 2006; Karimi, Samadi, & Babaii, 2021). Given what we know about the variability of proficiency in early bilinguals (e.g., Jegerski et al., 2018) and recent work that has shown there to be an effect of L2 proficiency on L1 processing strategies, one aim of the current study is to examine whether individual differences in L2 proficiency impact RC attachment in the L1.

# Theories of Working Memory and Language Comprehension

Although language experience is one factor that may help to explain strategies for ambiguity resolution, another important influence is an individual’s working memory capacity. This idea is influenced by decades of research appealing to working memory as an explanation for different sentence processing phenomena. For example, the garden-path model invokes memory constraints as the primary explanation of incremental processing and the need to make attachment decisions before all potentially relevant information becomes available (e.g., Frazier, 1979). This preference for the low attachment site appeals to the role of recency, which might lead to greater processing advantages because of the decrease in the amount of information held in a parser’s short-term memory (see Frazier, 1983; Frazier & Fodor, 1978; Frazier & Rayner, 1982; & Rayner, Carlson, & Frazier, 1983 for further information).

Another example of a theory that appeals to working memory constraints to explain processing phenomena is Gibson’s (1998) dependency locality theory (DLT), which views sentence processing as highly constrained by the finite amount of information that a processor can maintain and integrate over time (e.g., Gibson, 1998; Gibson, Desmet, Watson, Grodner, & Ko, 2005; Grodner & Gibson, 2005). A good example of this is in instances of syntactic prediction. DLT assumes that maintaining these predictions places a burden on the reader’s working memory resources because the prediction is taking up memory units that are proportional to the distance that the prediction must be maintained. However, other theories have advocated for separate, linguistically dedicated working memory resources. For example, Caplan & Waters (1999) and Waters & Caplan (1992) posit two separate working memory resources that support both preinterpretive (e.g., syntactic parsing) and post interpretive operations that are shared with a variety of different cognitive systems.

According to the theories discussed so far, working memory is viewed as a distinct cognitive system, as assumed in influential models such as Baddeley and Hitch (Baddeley at al., 2021). However, several theorists argue that working memory is simply an outcome of processing fluency that relates to the extent of language experience. One example is MacDonald & Christiansen’s (2002) connectionist framework. From this perspective, sentence parsing is a function of activation within neural networks, so individual differences in comprehension stems from an interaction of biological factors and language experience.

Although we have discussed multiple theories of working memory that are potentially at odds with one another, the working memory account most relevant to the current study is the shared resource account, which assumes the existence of a distinct, domain-general working memory system where such resources are shared across processing and storage within tasks (e.g., Just and Carpenter, 1992; Just & Varma, 2002). As storage demands increase, less working memory will be available for such processes. An example from this approach assumes that individual differences in working memory capacity are associated with constraints on sentence processing (e.g., Just & Carpenter, 1992;

MacDonald, Just, & Carpenter, 1992; Pearlmutter & MacDonald, 1995).

# Working Memory Contributions to Relative Clause Attachment Processing

In part for the reasons already discussed, the relative clause ambiguity itself is an ideal structure for studying differences in working memory resources. Under the assumption of Late Closure, traditionally, it was assumed that readers with lower working memory spans should prefer low attachment more than those with higher working memory spans (Frazier, 1979). And although languages have statistical preferences for either attachment site (high or low), these preferences are relatively mild (e.g., typically 55/45), and as such could shift given other sources of information, which has been the center focus of several studies (Desmet, Brysbaert, & De Baecke, 2002; Desmet, De Baecke, Drieghe, Brysbaert, & Vonk, 2006; Thornton, MacDonald, & Gil, 1999; Traxler, Pickering, & Clifton, 1998).

The first study to directly investigate the relationship between relative clause attachment ambiguity and the expectation for recency-based attachments used a self-paced listening task to examine the difference between adult and children’s preferences (Felser et al. 2003). Children with lower working memory spans adopted what would appear to be a more memory-friendly recency-based strategy and preferred to attach the relative clause to the low attachment site, while children with higher spans preferred high attachment. The authors found no relationship between the adult’s working memory and their attachment preference. The authors interpreted such results as indicative of cognitive differences outside of the language system, due to children’s limited working memory spans.

As a follow-up, Traxler (2007) examined whether working memory capacity influenced adult readers’ online parsing preferences. Consistent with the children’s preferences discussed in Felser et al. (2003), the eye movement data revealed that adult readers with higher working memory spans preferred high attachment. In comparison to low-span readers, the high-span readers experienced an increase in reading time when reading sentences where the relative clause modified the second noun. The author suggested the results are in support of the continuity hypothesis (e.g., Crain & Thornton, 2006) because the results from both Felser et al. (2003) and the 2007 study suggest children and adults adopt similar strategies to resolve syntactic ambiguity.

However, more recently, findings with larger sample sizes, updated working memory measures, and multiple language samples, have challenged those from Felser et al. (2003) and Traxler (2007). Among these studies is Swets et al. (2007), which will be a large focus of our current study. Swets et al. (2007) examined the offline preferences of two samples of adults whose first language was either Dutch or English. This study was the first to examine the effects of domain-general and domain-specific working memory on ambiguity resolution given prior evidence suggesting a domain-general factor acts alongside domain-specific factors due to a high correlation between these factors (e.g., Kane et al., 2004). Participants were given two working memory tasks: A reading span task to measure verbal working memory and a spatial span task to measure nonverbal spatial working memory. Despite the replicated slight preference for high attachment in Dutch and low attachment in English, participants from both samples preferred high attachment as their working memory capacities decreased, which runs counter to the results found in children from Felser et al. (2003) and adults in Traxler (2007).

As an explanation for these findings counterintuitive to recency, the authors appealed to implicit prosody. To test this idea, they conducted a second study where, instead of displaying each experimental item as one unit, the sentences were displayed in three segments: First, the complex noun phrase (The maid of the princess), followed by the relative clause (who scratched herself in public) and then last, the matrix verb phrase (was terribly embarrassed). This manipulation would force participants to insert a prosodic break between the second noun and the relative clause. And presumably, the insertion of this break before the relative clause would increase the tendency to attach high because the break creates a “structural discontinuity in the syntactic tree” (Fodor, 2002, p. 4).

As hypothesized, the variance in attachment preferences across both samples was dramatically reduced, and both samples showed a strong preference for high attachment. Based on this ‘chunking’ assumption, the relationship between working memory and attachment preferences are due to differences in the way high vs. low span readers internally chunk the constituents during silent reading.

Traxler (2009) followed up on Swets et al. (2007) to determine whether working memory modulates the relationship between implicit prosodic phrasing and online processing. Consistent with the results from Swets et al. (2007), the findings suggest readers preferred to attach the relative clause to the high attachment noun and there was no observable effect of working memory on such parsing, which is almost identical to the results found in Study of Swets et al. (2007). Recall that in the second study of Swets et al., when participants were forced to insert a prosodic break between the complex noun phrase and the relative clause, any effect of working memory was diminished. A majority of the participants began to attach the RCs high, mimicking the chunking strategies of low span readers.

More recently, James et al. (2018) found that lower verbal working memory span was associated with a stronger preference for high attachment. Although the evidence cannot speak directly to implicit prosody as suggested by Swets et al. (2007), the authors of James et al. (2018) support the theory that individuals with limited working memory resources adopt a chunking strategy during silent reading that encourages high attachment. A similar relationship was observed in older adults, as discussed by Payne et al. (2014). A series of mixed-effects models indicated that high attachment sentences were harder to process than low attachment sentences and verbal working memory modulated this relationship.

While earlier studies reported conflicting results on the directionality of the relationship, evidence from all later studies (e.g., Swets et al., 2007; Payne et al., 2014; James et al., 2018) has indicated that a higher working memory span leads to a preference for low attachment. Those studies also suggest the relationship is modulated by segmentation strategies low span readers adopt during silent reading that affect the implicit prosodic form generated for the experimental sentences.

# The Current Study

In the present study, we explored these issues further by examining three language groups that differ significantly from the ones examined in previous work and making use of the same paradigm reported as the first study in Swets et al. (2007). Swets et al. was the first to use two span tasks aimed to examine the effects of domain-general and domain-specific working memory decisions on relative clause attachment processing; prior investigations had examined only the impact of verbal working memory. For this reason, we included almost identical versions of the original tasks. The slight modifications will be discussed in detail in the Method section. Each participant completed two working memory span tasks (reading span and spatial span) and a relative clause attachment task. Additionally, we included a language history questionnaire (Li et al., 2020) to examine proficiency, exposure, and frequency of use in the first and second languages, so we could use these as covariates in our analyses. Consistent with the 2007 study, we analyzed the relationship between our variables of interest via structural equation modeling (SEM).

Our study addresses two core questions about syntactic ambiguity resolution. First, we build upon previous work that examined the directionality of the relationship between working memory and preferences for attaching ambiguous relative clauses. Given earlier work suggesting both high and low attachment preferences for those with smaller working memory spans, and the importance of replication in psychological research, we ask whether the pattern shown in Swets et al. (2007) will replicate nearly two decades later and in different languages. That is, we predict participants with low working memory spans will show a greater tendency to attach to the high attachment site, as previously reported. Second, given that syntactic parsing strategies are influenced by relative language exposure (e.g., Dussias, 2003; Dussias, 2004; Dussias & Sagarra, 2007), we ask whether proficiency predicts ambiguity resolution. To this end, we predict the following: Bilinguals with greater proficiency in their second language will begin to attach RCs in their first language the same way they do in their second.