University of Reading

Combining behavioral and neurolinguistic methodologies to investigate Spanish scalar indefinites among mono- and bilinguals: An event-related potential study

For the degree of Doctor of Philosophy

School of Psychology and Clinical Language Sciences

by

David Miller

15/07/2017

Contents

To SarahJane, Owen, Leo and my parents

Acknowledgements

From start to finish, the PhD experience is one that requires maximum effort from the student, love from his/her family and the support of the mentors who sacrifice so much of their time to make it all possible; I would not be where I am without those people. Many thanks go to my best friend—my wife—for moving with me to a foreign land, leaving behind the comforts of home and family, to share in this experience …and for dealing with my mid-life crisis at age 30. Te amo! To Owen, my son and little champion, for making every day bright and joyful, more worth living. To Leo, my most loyal dog-son, for greeting me with love and kindness each and every late night return home from university. To my mom and dad for their overseas love and support and for teaching me the value of hard work. To Jason, my beloved friend and advisor, for sacrificing time, money, and limitless energy to ensure my general progression and academic development. To Doug Saddy, our trusted neuroscientist and advisor, for helping fund this experience and providing invaluable advice for projects now and to come. I am very grateful also for the European Union's Seventh Framework Programme (grant number 613465), as well as for the University of Reading in providing an International Studentship and other necessary funding for this research.

I also want to give special thanks to José Alemán-Bañón (Pepe). You taught me a great deal about ERPs and data acquisition that I would not have been able to learn on my own. I'm grateful for the time you took to mentor me.

Of course, thank you to my department and other lab members who helped me along the way. The small things make a difference and the department and its employees made this a less painful experience than it would have otherwise been.

I am certainly grateful for the wonderful friends I have made along the way: Eloi, Fatih, Vince, Jorge, Ian, Ludo, and many more. You made it worthwhile to come to university every day and you were there to put a smile on my face when things got tough. I would love nothing more than to remain close moving forward. Thank you also to those who have been a part of this work from the very beginning: David Giancaspro, Mike Iverson, and Roumyana Slabakova. The first project we did together was enough for me to get hooked. Each of you has played an integral role in the development of my character, both personally and professionally. Thank you.

I have learned rewarding lessons of trust, strength, commitment, and sacrifice in the last three years at Reading. I will always look back on this time as one of the most rewarding

and satisfying experiences of my life. Traveling the world, presenting research from Tokyo to good ol' USA and seemingly everywhere in between, the wonderful people around me made it all possible and I hope I can always show you how grateful I am for the opportunity to be part of the team. Here's to the future; may it be full of loads of high-impact journal articles, books and book chapters, festschrifts for our advisor(s), and a great deal of life success. I love you all.

Abstract

Quantifiers like *some, most, many* and *all* are said to form part of a scale on which the relative informativity of each quantifier is weighted against the others. Weak quantifiers like *some* generate an implicit meaning beyond their literal semantic meaning. For example, ***some*** *PhD students like writing* can be taken to mean that ***not all*** *PhD students like writing*. Such an implied meaning is called a scalar implicature (SI).

Researchers have examined SIs in child and adult populations using a variety of methodological designs. While children are more inclined to retrieve lower bounded interpretations than adults concerning SI derivation, particularly without explicit instruction, both groups show considerable variability. A growing body of SI work aimed at explaining this variability has examined implicit brain responses while individuals perform various linguistic as well as cognitive tasks and methodological design factors that may give rise to SI interpretative variation. Some of this research has found that SI derivation is associated with scores on cognitive tests such as the Autism Spectrum Quotient (ASQ), working memory (WM) as cognitive load is increased, and the Systematizing Quotient-Revised (SQ-R) questionnaire, and other task design factors. One implication drawn from this research is that lack of SI derivation among adults in particular may be due to a differences in inherent cognitive or psychological mechanisms, such that higher or lower scores on the above-named tasks are apparently correlated with rates of SI derivation and that methodological design can create issues in data interpretation if confounds are not carefully controlled.

The purpose of this dissertation is to further shed light on specific methodological and population (bilingualism) factors that contribute to differential SI derivation, in this case among adult native Spanish speakers and Spanish-English bilinguals. I maintain herein that research must fully consider its own role in how SIs are treated experimentally, whether in design, connections made to cognition, or population choice, before generalizing results. Once such variables have been controlled, we may more fully be able to understand inter-individual variation in SI derivation in healthy populations.

Declaration: I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

David Miller

List of Tables

List of Figures

Chapter 1: General introduction

Whether one is interested in language as a topic of scientific inquiry in its own right (e.g., linguistics) or as a window into the human mind (e.g., psychology, cognitive science), studying its various domains continues to shed light on the essence of what it means to be human. Language—as distinct from other forms of complex communication[1]— distinguishes us from every other species of animal life on the planet and yet it is mostly taken for granted by its every day users. For example, most language users are not aware that, linguistically speaking, language has several levels or distinct units by which it is measured and analyzed: phonology and phonetics, morphology and syntax (morphosyntax), semantics, and pragmatics. Historically, and especially following the advent of generative linguistics (Chomsky 1957, 1959, 1965, 1982, 1995), the linguistic level that has received the most attention in research is morphosyntax, that is, the hierarchical structure of language. Arguably, this has been because the study of morphosyntax was understood from early on to be a ripe domain through which one could probe characteristics of the linguistic computational system, such as modularity and hierarchical structure. The generative turn in linguistic thinking, which was a reaction to the behavioralist approaches to human cognition (Skinner, 1935, 1938, 1963), turned out to be a *prima facie* catalyst of the cognitive revolution in psychology in the 1950s propelled in part by Chomskyan linguistics. Notwithstanding, the heavy bias of research on the morphosyntactic structure of language, its acquisition and its processing may have contributed to the comparatively understudied subfield of semantics. Semantics is necessary as an ingredient for communication and it is crucial for the emergence and evolution of language (e.g., Culicover & Jackendoff, 2005; Jackendoff, 1983; 1991; 2002; Pinker & Jackendoff, 2005).

Though not its sole purpose, language is predominantly used as a means to convey meaningful messages to other language users. We use language to talk about events, situations and other objects in the physical world, which we do easily and in a systematic way by creating representations or associations between the expressions of our language and those objects in

[1] There is no question that other animals, such as bees, whales, great apes, and dolphins, have complex systems of communication; however, language they do not have. "Communication" is defined as transfer of information from one or a group of animals to one or more other animals that is systematic enough such that senders and receivers can reliably intuit the intended message. This can be conveyed through sound, smell, gesture and more. A skunk uses its infamous smell to ward off predators. Elephants use their trunks to send messages over great distances. Bees use the patterns of movement relative to the sun to relay information about location of food sources. None, however, has a complex system of communication that express thoughts, desires, or emotions. Whereas animal communicative systems can convey messages in the present and potentially the near future, only human language can express anything related to the past. Communication is functional and indeed a precursor to language, but language requires much more.

the world on a large scale. One characteristic of semantics that distinguishes it from the other levels of linguistics is that semantic mappings to particular overt morpho-phonological forms are overwhelmingly arbitrary. In other words, the relationship between the phonological shape of any word and what its semantics are is a matter of convention, a concept first discussed in detail by de Saussure more than 100 years ago in what is known as the arbitrary nature of the linguistic sign (Saussure, 1915). Whether the phonological form of the concept (word) happens to be tree [ˈtriː], árbol [ˈaɾ.βol] or arbre [ˈaɾ.βɾə], the thing in the world it represents (prototypically) has a brown wooden trunk with branches and leaves growing out from it. Although trivial to any pedestrian language user, these form/sound-to-meaning correlations are remarkably interesting. Irrespective of such arbitrariness, however, speakers within the same language communities are able to communicate effectively—for the most part—due to internalizing the same rules for pairing language-specific (morpho)phonological forms onto universal concepts from conceptual structure (which is external to language). Speakers use the same operations for computing meaning of complex syntactic structures by means of the semantics of individual lexical units from which such structures are formed (see Chierchia & McConnell-Ginnet, 2000; Kratzer & Heim, 1998). Notwithstanding, we are not immune breakdowns in communication.

Even when we "know" the lexical and functional meanings of specific words, their interpretation in any given context can be largely ambiguous (Kempson, 2003; Lappin & Fox, 2015). Speakers assign an intended meaning to the words, phrases, and expressions they use, and the recipients of these messages are free to assign either a parallel (i.e., literal) or a divergent (i.e., inferred) meaning to them. Thus, while lexical and compositional semantics seem to adhere to unwritten rules or conventions, interpretation can take place on an individual level and independent of any intended meaning by the speaker. Though this will be discussed in greater detail in later sections, it is worth noting here that this notion of literal versus implicated and/or inferred meanings was formalized by Grice (1975, 1989), who postulated several conventional guidelines that interlocutors tend to adhere to in order to carry out meaningful conversation. Of most relevance for the purposes of this dissertation, I will discuss only the Cooperative Principle, within which a specific Maxim of Quantity was addressed, which proposes the following: (i) be as informative as needed, but not more so and (ii) say only what is relevant to the purposes, timing and direction of the exchange. While such conversational standards often guide communication towards an ultimate mutual understanding between what is meant and what is interpreted, they are not always strictly followed. In this respect, Grice was the first to systematically examine cases in which the

semantics of a word or utterance transcends their conventional meaning due to conversational discourse, identifying two types meaning: that which is explicitly stated and that which is both implicated and possibly inferred. Take the exchange between a mother and her toddler in (1) for example, where material in brackets is optional:

1. Owen: Can I have a lollipop?
   SarahJane: You need to eat your dinner (first).

Anyone who can empathize with Owen knows that what is really being said by SarahJane is something like "No, at least until after dinner." However, neither the individual lexical items nor the entirety of the utterance state that. SarahJane has literally stated that Owen needs to eat his dinner first. This example highlights that while there is an apparent mismatch between conventional semantics and what is actually inferred on the basis of pragmatics, there is also a need for such an interaction between our linguistic and non-linguistic systems. This interaction is what ultimately leads Owen to understand that though his mom's answer contained no specific "yes" or "no" in response to his question, he can infer on the basis of pragmatics that she means something like "While you can't have a lollipop now, you can have one after dinner". The interaction between semantics and pragmatics, thus, is one that seems to more fully appeal to our rationale as humans.

The goal of this work, given the above discussion, is to synthesize the (relevant) work on semantics and pragmatics over the decades and offer a panoramic perspective of the impact this work has had not only on linguistic theory but also on the nature of the brain and human cognition. I begin by setting the stage with relevant concepts such logic and truth, implicature (content in an utterance that is implied and/or inferred and not explicitly stated), and then connect these concepts to language acquisition, processing and cognition. The experimental studies herein make a strong appeal to give credence to the already established observation that methodological design in work on pragmatics plays an important role in how individuals treat implicatures (e.g., Degen & Tannenhaus, 2011; 2015; Guasti, Chierchia, Crain, Foppolo, Gualmini & Meroni, 2005; Papafragou & Musolino, 2003). In this vein, Studies 1 and 2 highlight some crucial elements of task designs typically employed in implicature research that have been found to modulate interpretations of the scalar terms. I argue that such design variables present potential confounds that, in some cases, may have produced either unclear results or led to tenebrous interpretations of results. Study 3, on the other hand, presents findings related to an understudied population within implicature work, namely adult sequential bilinguals. However, rather than study L2 development or acquisition of native-like

interpretations of implicature generating items, Study 3 offers an entirely new data set within the domain of L1 attrition. Together, these three studies offer insight into the experimental issues inherent to implicature research, as well as a fresh perspective on the integral role of the population of testing for the development of semantic and pragmatic theory.

Because a significant portion of the background and relevant literature is described similarly across the three independent studies comprising this dissertation, I will present herein a more panoramic view than is (could be) appropriate in each of the individual articles. This will be especially true when presenting work on implicature. To finish, I present the research problem and a more detailed explanation of the goals of the dissertation as viewed within the larger research program of ongoing work on human cognition and pragmatic language processing. Thus, the goal of this section is to bring the three articles together to tell a coherent story of how they form part of my nascent research program.

1.1 Implicatures

Logic, the study of all things pertaining to deductions and inferences, is inherently related to semantics. The connection of logic to semantics lies in the concept of truth, whereby the rules of logical inference are centred on the preservation of truth. We can think about (propositional) logic as *if-then* proposals where the supposed truthfulness of any premise or premises within such proposals will dictate the truthfulness of the conclusion (i.e., truth conditions). When it is understood that the meaning of any utterance or sentence is related to its truth-conditional content, the connection between semantics and logic becomes ever clearer. However, it is important to note that when speaking of truth there is a distinction between that which is factually true and that which is logically true:

2. A. All Trump supporters are bigots.
   B. Charles is a Trump supporter.
   Conclusion: Charles is a bigot.

While the above logical inference leads to a valid conclusion, it is not the case that all Trump supporters are indeed bigots. Whether one takes them to be rational or not is a separate question. In language, this distinction between factual and logical truth is made clear by playing on the interactions and meanings of logical words such as *if, then, and* and *not*, for example. Thus, by changing (2) to ***If*** *all Trump supporters are bigots* ***and*** *Charles is a Trump supporter,* ***then*** *Charles is a bigot,* it is clearer to see that the truthfulness of the conclusion would entirely depend on the factual truth of each premise. Logic, thus, is independent of facts and entirely dependent on formal propositions of the following form:

3. *If p, then q*
   p
   Therefore q

Formal logic, then, is concerned with encoding valid inferences by formalizing them in logical form (e.g., Keenan, 1973). Again, the connection to semantics lies in the relationship between meaning and truth whereby knowing the meaning of a sentence is knowing the conditions under which the sentence would be true. In other words, the meaning of a sentence is identifiable through its truth-conditions. For example, to verify the meaning of a sentence such as *It's raining right now*, one must simply go outside and determine whether it is indeed raining. If it is raining, then the sentence was understood on the basis of truth-conditional semantics, whereby meaning is derived on the basis of logical operations.

Language, however, is not used in isolation of other non-linguistic factors that give rise to additional interpretations beyond the literal semantics of words or utterances. Such non-linguistic factors are said to form part of the domain of pragmatics, or language use in context. Because humans are social beings, language forms part of our everyday interactions with the outside world. Such interactions take place in a variety of situations that, in consideration of social norms, expectations, relationships and even cognitive economy (i.e., our brain's inherent inclination to be maximally efficient in processing strategies), dictate how we use language. Ultimately, such interactions contribute to both the richness and subtleness of an utterance's overall potential meaning. In spite of such vast and rich potential, however, contextual use of language often leads to breakdowns in communication. Simply put, we do not compute language in the absence of any additional information that is outside the domain of structure or logical operations (i.e., syntax). On the contrary, language comprehension is partially dependent on cues within the surrounding context.

As introduced above, Grice's (1975, 1989) work is particularly crucial to our modern understanding of meaning that extends beyond conventional and compositional semantics. He referred to such as *implicature* because it is *implied* and not explicitly stated. Grice further distinguished between types of implicatures, such as *conversational (generalized* and *particularized*) and *conventional* implicatures. Conversational implicatures are those, like the one calculated in (1) above, that are dependent on the conversational context in which language is being used. Within this group, however, there are also generalized (i.e., those derived on the basis of a proposition made in ordinary contexts of utterance) and particularized (i.e., those that are derived on the basis of a proposition made in particular contexts of utterance) implicatures.

Conventional implicatures as a class are derived on the basis of the composed conventional meaning(s) of the individual lexical items comprising a sentence:

4. Leo is a redhead, but he's handsome.
   a. Leo is a redhead and he's handsome.
   b. Redheads are not typically handsome.

According to Gricean accounts, (a) above is what is actually said by (4) and (b) is what is conventionally implicated. Because any perceivable truth-values associated with that which is conventionally implicated are mostly irrelevant regarding the truth-evaluation of (4) itself, Grice posits that content like (b) is merely implicated and not necessarily part of what is said (see also Bach, 1999; Blome-Tillmann, 2013).

Although *implicature* was first coined by Grice, the concept was also discussed by Horn (1972) and later extended by him (e.g., 1989; 2004) as he detailed implicatures that arise from certain quantifiers. These quantifiers are said to belong to a so-called scale that determines their relative informativity when evaluated against one another. Implicatures emerging on the basis of scaled informativity are termed *scalar implicatures* (SIs), which are the focus of this dissertation.

1.2 Scalar implicature

Scalar implicature (SI) refers to an interpretation of an informationally weak proposition (i.e., one that is less informative) as meaning that a stronger one is not applicable or is not true. Recall from the previous section that the meaning of a statement is derived from its truth conditions using logical operations. In the case of an SI, the supposed truthfulness of a statement is inferred on universal principles of pragmatics rather than logic alone. Quantification, for example, is pertinent to the domain of predicate logic as opposed to propositional logic described in the previous section. Predicate logic breaks down propositions into predicates such as verb phrases (VPs) and adjective phrases (APs), as well as arguments such as nouns or noun phrases (NPs). Quantified expressions have peculiar logical properties precisely because they do not refer to specific individuals, rather they refer to sets or quantities of individuals:

5. a. #Fatih$_1$ is Turkish and Fatih$_1$ is not Turkish.
   b. Some linguists$_i$ struggle with logic and some linguists$_{\#i/j}$ do not struggle with logic.

The contradiction of (5a) stems from the fact that the subject of the propositions is the same. In predicate logic, this is written $p \,\&\, \neg p$, where ¬ denotes negation, and will always come out

false. Quantifiers do not share this characteristic. Therefore, (5b) does not present any contradiction, rather it expresses a proposition that very well might be true given our knowledge of the world. Crucially, this is the case because *some* does not refer to any specific individual, rather to a set of one or more individuals. Across the propositions of (5b), then, *some* can refer to distinct sets (i.e., disjoint reference).

Quantification is important to the study of scalar implicature in the sense that weak quantifiers like *some* can be vague, ultimately leading to a pragmatic inference (i.e., one that might be true or one that is probably true given the discourse of the utterance). The associated SI is often said to be derived on the basis of certain quantifiers belonging to a scale in which each term is ordered according to its informativity, such as <*some, many, most, all*>, where the stronger terms naturally entail the weaker ones but not vice versa. Thus, the use of an informationally weaker term like *some* implies that a stronger term like *all* is not applicable (e.g., Grice, 1975; Horn, 1972; Levinson, 2000; Noveck & Sperber, 2007). In order to distinguish the semantically stable meaning of *some* from its pragmatic enriched meaning, let us consider the following exchange:

6. Cory: How's Niko doing?
   Kristina: He ate some rocks and now his belly hurts.

In this case, it is impossible for any specific number of rocks to be associated with *some*. Although *some* means 'more than zero', in this example plural morphology on the subsequent noun automatically generates a meaning of 'more than one'. So, Niko could very well have eaten two rocks or 22 rocks. In essence, he could have eaten *any* number of rocks greater than or equal to two. In scenarios like this, *some* takes an existential reading whereby it can refer to any number but zero (e.g., Newstead, 1988). However, this is not always the case. While *some* remains indefinite irrespective of the context in which it is used, it can be interpreted variably by narrowing or enlarging the set to which it refers:

7. Sean: Did your students do well on the final exam?
   Tom: Some did.

Here, although we are not aware of how many students there are, our world knowledge suggests that, if this is a state university in an average class, there are probably between 20-30. Logically, Tom's response is taken to mean any number of students greater than or equal to one. Pragmatically, however, when Tom says that *some* students did well on the exam, he may actually mean any number greater than zero but less than *all* possible students in his class.

Gricean principles suggest that any implicature Tom wished to convey, and subsequent inference made by Sean based on Tom's statement, would be made on the basis of the speakers generally adhering to the Cooperative Principle and Maxim of Quantity. If Tom had wanted to say that all of his students did well, he would have said so. Thus, Sean is left to infer on the basis of the implicated content of the expression (i.e., the choice of the speaker to not use *all*) that *not all students did well*.

Some is not, however, the only expression or word that gives rise to both implicated and inferred meanings due to the lexical scale. For example, A *or* B can be interpreted to mean not [A *and* B], and an adjective such as *cool* can mean 'not cold'. Notwithstanding, it is worth mentioning that while other expressions also give rise to implicatures, different scales dictate the strength of the scalar implicature they give rise to (see Doran, Ward, Larson, McNabb, & Baker, 2009; Doran, Baker, McNabb, Larson, & Ward, 2012).

While the distinction between the semantic and pragmatically enriched meanings of these words appears nuanced at first glance, they can actually be reliably distinguished by a simple test of defeasibility. In the case of *some,* if the content of the expression can be revised without contradiction, then such content was pragmatically enriched. If the content cannot be revised without contradiction, then it was semantic in nature. For example, compare the following:

8. a. Some linguists are smart. Actually, they all are.
   b. Some linguists are smart. #Actually, none of them are.

In (8a) the SI can be cancelled to mean *all* without resulting in a contradiction and in (8b) it cannot, delineating a clear contrast between semantic entailments (8b) and pragmatically enriched implicature (8a). As discussed throughout this introduction so far, an implicature is derived by the comprehender on the basis of a set of assumptions about what she thinks the speaker has meant. While these assumptions can be inaccurate, ultimately leading to breakdowns in communication, they also allow for fluid and efficient communication. One open question receiving considerable attention in the literature is what the role of the comprehender really is. Is it the case that interlocutors are actively making decisions to infer in the face of vague information, or is the process more automatic? This question then leads to that of why some individuals do and some do not derive implicatures, especially in contexts that promote their derivation.

1.3 Scalar implicatures in Spanish

Given the language of testing pertinent to this dissertation, it is prudent to discuss early on how SIs work in Spanish. Spanish is interesting because it has two scalar quantifiers that roughly translate to English *some*, *algunos* and *unos*. Despite having similar meanings and overlapping semantic and pragmatic distribution (Alonso-Ovalle & Menéndez-Benito, 2013; Gutiérrez-Rexach, 2001; López Palma, 2007; Martí, 2008), these two quantifiers differ concerning their felicitousness and semantic meaning in various contexts. It is worth mentioning that *algunos* and *unos* can be further distinguished by their idiosyncratic distributional interactions with verb class, predicate type, and contrastive focus (e.g., Gutiérrez-Rexach, 1999; Laca & Tasmovski, 1996; Vargas-Tokuda et al., 2009; Villalta, 1994), which we considered in the construction of our methodology to avoid potential confounds. For the purposes of this dissertation, however, I expand only upon that which is argued to distinguish *algunos* from *unos* as far as SI derivation is concerned.

The relevant difference between *algunos* and *unos* for SIs is the ability of each to combine (or not) with NPs to trigger scalar implicatures. According to the literature, only *algunos* triggers an upper-bound interpretation (see Alonso-Ovalle & Menéndez-Benito, 2010, 2013; Gutiérrez-Rexach, 2001; Martí, 2008; Vargas-Tokuda et al., 2009). As evidenced in (9a) and (9b), however, the distinction is not always easily discernible. Below, both quantifiers refer to some members of a group, in this case some children performing a given action and the meaning that is derived is that *some children are crying*.

9. a. *Unos niños lloran.*
"Some children cry."

b. *Algunos niños lloran.*
"Some children cry."

Notwithstanding, if we examine the (in)felicitousness of each item in subset and whole set contexts, differences with respect to their ability to trigger an implicature become more apparent. Imagine that there are four children together; however, only three of them cry. In this "subset context", both (9a) and (9b) above are felicitous. However, the differences between the two items, as demonstrated below in (10a) and (10b), only emerge in "whole set" contexts, i.e., when all four children are crying. This is because *algunos* is claimed to be presuppositional in that it presupposes a discourse-relevant referent, while *unos* is semantically non-presuppositional except for in certain contexts such as topic/focus marking.

10. Context – All children within the group cry.
    a. *Unos niños lloran.*
    "Some children cry."

    b. *#Algunos niños lloran.*
    "Some children cry."

As shown in (10), only *unos*, which denotes the logical 'at least one and possibly all' interpretation, is felicitous in a context where all members of the set are performing the same action. *Algunos*, on the other hand, is not felicitous in such a context because it triggers a pragmatically enriched quantity implicature. In one prominent theory, this implicature is argued to arise due to a partitive feature linked to *algunos* (Gutiérrez-Rexach, 2001, 2004). Thus, *algunos* is said to adhere to a D-linking constraint (Pesetsky, 1987) that allows it but not *unos* to refer back to an already expressed (salient) referent within the discourse. Thus, when a Spanish native speakers hear an utterance like (9b/10b), they interpret *algunos* as meaning that *not all children cry*.

The pragmatic status of *algunos* can also be seen when, in certain contexts—such as downward entailment[2]— its associated implicature does not arise (Chierchia, 2001, 2004).

11. Context – All four children cry.
    a. *Si unos niños lloran, no van a dormir.*
    "If some children cry, they will not sleep."

    b. *Si algunos niños lloran, no van a dormir.*
    "If some children cry, they will not sleep."

In (11b), the scalar implicature associated with *algunos* does not arise, meaning that *algunos*, like *unos*, can subsequently refer to the entire set.

While there is a great deal of SI work across languages, there is comparatively very little work done in Spanish. The work that does exist in Spanish further establishes that individuals treat scalar terms with variability. In this respect, there have been discrepancies across studies concerning the nature of Spanish's scalar quantifiers and how individuals treat them, monolinguals and bilinguals alike. One of the purposes of this dissertation is to shed light on some potential reasons for such disaccord and to highlight some ways of potentially overcoming it.

[2] Downward entailment inverts the entailment relations from [subset to superset] to [superset to subset]: e.g., *nobody likes fruit (superset)→ nobody likes apples (subset)*.

1.4 Scalar implicature processing

Default models of implicature processing (e.g., Levinson, 2000) propose that SIs arise automatically when *some* is encountered. If needed, later reanalysis of the implicature takes place when additional information becomes available. Context-dependent models, on the other hand (e.g., Carston, 1990; Sperber & Wilson, 1995), argue that SIs arise only when the context is sufficiently appropriate for their derivation. Moreover, other research has suggested that because *some* belongs to a set of quantifiers whose meaning is more or less informative than the other(s), the possible pragmatic meaning of *some* would be derived on the basis of scaled informativity (e.g., Chierchia, Fox, & Spector, 2012; Horn, 1972; Levinson, 2000; Sauerland, 2012). Researchers have also argued that SIs arise by contributing to the truth-conditional content within an expression rather than its implied content (e.g., Carston, 1998; Horn, 1992), or potentially as a result of lexical calculations (Levinson, 2000) and higher-level structure (Chierchia, 2006; Chierchia et al., 2012; Geurts & Pouscoulous, 2009; Katsos, 2008).

Whatever the mechanisms underlying SI derivation, children are thought to be more logical at a group level and adults more (readily) pragmatic (though see Syrett et al. 2017a, 2017b), a difference that has been argued to emerge as a result of a pragmatic or experiential deficit among young children (e.g., Chierchia, 2005, 2006). More contemporary work however, shows that even adults do not always opt for pragmatic interpretations when dealing with implicature generating items (e.g., Bott & Noveck, 2004; Hunt, Politzer-Ahles, Gibson, Minai, & Fiorentino, 2013; Noveck & Posada, 2003; Politzer-Ahles, Fiorentino, Jiang, & Zhou, 2013; Spychalska, Kontinen, & Werning, 2016; Tavano, 2010). In these cases, still, individual variation is said to be the potential effect of a type of psychological and/or cognitive deviation from normal behavior. The general assumption from the above work appears to be that because one *can* derive pragmatic meanings, one *should* derive them, especially when the experiment is designed to elicit them. When pragmatic meanings are not derived, then, there must be some reason for it. As I will discuss in the experimental studies presented herein, the general assumption that individuals *should* derive SIs merely because it is an available option may inadvertently lead research on a quest to find a problem within individual speakers at the level of cognition. However, given the fact that individual variation also emerges as a by-product of experimental design and, potentially, of bilingualism, the above claims may be premature.

Much current research, especially that which tackles the above issues from a psycho- or neurolinguistic approach, tends to question the status of the underlying cognitive mechanisms thought to be responsible for drawing pragmatic inferences, particularly among people who do not (by majority) derive them from ambiguous language. For example, recent studies have

shown that pragmatic inferencing, or the lack thereof, is potentially associated with high or low working memory (Marty & Chemla, 2013), scores on the Autism Spectrum Quotient (ASQ) (Nieuwland, Ditman, & Kuperberg, 2010; Zha et al. 2015), as well as scores obtained on the Systemizing Quotient-Revised (SQR) questionnaire (Barbet & Thierry, 2016). Furthermore, it has also been suggested that adults, at least, may pick an arbitrary solution to dealing with ambiguity presented in quantification and stick with that strategy throughout the task (Katsos & Bishop, 2011). Although there is still considerable debate as to the nature of what leads to the derivation of SIs, it is generally accepted that there is something, whether cognitive, linguistic or some combination thereof, that gives rise to variability in their interpretation.

To this end, much of the contemporary work examining SIs from a psycholinguistic perspective has attempted to hone in on the processing of SI calculation/derivation in real time, using technologies that permit a more precise capturing of the SI calculation (e.g., eye-tracking, ERPs, MEG) under the assumption that SI derivation should happen under ideal circumstances. As mentioned, however, one potential problem is that normal/ideal circumstances vary from person to person, from task to task, and from experience to experience. To be sure, there has been a stream of SI work in other languages, although by comparison much smaller than the body of work on English. Given the trend in the literature to examine SIs in English mainly among monolingual adults, such research may inadvertently complicate generalizations relating to other populations and descriptions of other languages.

1.5 Acquisition of scalar implicatures

There is a considerable body of research examining both the acquisition and interpretation of SIs in child first language (L1) acquisition (e.g., Breheny, Katsos & Williams, 2005; Guasti, Chierchia, Crain, Frappolo, Gualmini & Meroni, 2005; Noveck, 2001; Papafragou & Musolino, 2003; Su, 2013; see also Crain, 2012 and Noveck & Reboul, 2008 for review), as well as children's capacity for logic and reasoning more generally (e.g., Ackerman, 1983; Bever, Mehler, & Epstein, 1968; Bloom & Wynn, 1997; Braine & Rumain, 1981; Cassidy, 1998; Gelman, 1993; Gelman & Gallistel, 1978; Greeno, Riley, & Gelman, 1984; Hurewitz, Papafragou, Gleitman, & Gelman, 2006; Lidz & Musolino, 2002; Wynn, 1992). This research has shown that children can derive pragmatic meanings from scalar terms. However, they do not *always* have access to such pragmatically enriched interpretations, nor do they spontaneously compute SIs with the same frequency as adult speakers, especially at young ages.

For example, Su (2013) examined Mandarin-speaking children's ability to both calculate SIs and cancel them in DE environments. Rather than examine *some,* the study examined scalar (disjunctive) *or* and *and,* which are scalar in meaning. The results showed that children accepted *or* and *and* in both DE environments and upward entailing (UE) environments, lacking the same sensitivity to the entailment patterns of Mandarin-speaking adults. Moreover, Papafragou and Musolino (2003) tested child and adult Greek monolinguals and also found that children are largely insensitive to the SI reading as compared to adults, unless they are provided with enriched instruction or contextual cues. The authors argued that children do not treat all scalar terms alike and their (non)derivation of SIs can be linked to the complexity of the experimental manipulation and/or enriched instruction. In other words, children can calculate scalar implicatures but may show insensitivity to context–dependent information (sometimes) due to the demands imposed by certain experiments. Consistent with this conclusion is the work of Guasti et al. (2005), which showed that Italian-speaking children as young as seven years of age were able to compute SIs in certain experimental conditions. Like Papafragou and Musolino (2003), Guasti et al. (2005) argue that an experimental design can condition a child's understanding of certain scalar terms and that such design effects can be attenuated by enriched instruction and/or additional task training.

Although there are often differences between one study and the next, one can posit on the basis of common trends within these studies that young children behave differently than adults in experiments involving the realization of pragmatically enriched interpretations of SIs. That is, children appear to successfully derive SIs, although not to the same degree as adult speakers, and especially so under demanding conditions. However, this does not necessarily mean that the mental representations in young children are qualitatively different from adults', but rather that something about the demands of certain experiments or SI derivation more generally makes them perform differently. While the aforementioned trend is true of young children tested in virtually all languages, Vargas-Tokuda et al.'s (2009) study of monolingual Spanish children stands out in sharp contrast. This study is of particular interest given the language of testing in this dissertation.

Vargas-Tokuda et al. (2009) examined monolingual Spanish-speaking children's ability to successfully derive the pragmatic implicature associated with *algunos* and generate the alternative sets associated with the lexico-semantic meaning of the counterpart *unos*. Participants in this study completed a Truth Value Judgment Task (TVJT) in which they saw various sets of 2, 3 or 4 farm animals (out of a group of 4) jumping over a fence. After seeing the animals jump over the fence, the participants were asked to judge the truthfulness of follow-

up sentences, which included either *unos* or *algunos*. If children were calculating the SI linked to *algunos*, they would be expected to reject it in whole set conditions while accepting *unos* in either the whole set or subset conditions. Results showed that these children derived the quantity implicature associated with *algunos* much like adults, thus rejecting it when referring to whole sets and accepting it when referring to subsets.

While SI research among monolingual children has proven to be a rich source of investigation in the last several decades, comparatively little work has been done on bilingual children (e.g., Siegal, Iozzi & Surian, 2009; Siegal, Matsuo & pond, 2007; Syrett et al. 2017a,b). While the same general trends in monolingual populations hold true in bilingual child populations, that is, that bilingual children appear to be more logical in their interpretations of scalar terms, the bilinguals in the above-mentioned studies were seemingly at a marginal advantage in other areas of pragmatic reasoning, which was ascribed to more general strengthened cognitive functions as a result of bilingualism (see Bialystok, 2009; Bialystok et al. 2009; Bialystok 2017 a,b). However, these so-called advantageous pragmatic abilities appear less concerned with SI interpretation than they were with general reasoning skills in that bilingual children, like their monolingual counterparts, are more or less at chance levels of SI derivation (Syrett et al. 2017a,b for review). Though not entirely specific to SI derivation in child populations, research in bilingualism more generally has shown that cross-linguistic influence plays a role in semantic and pragmatic abilities in children, especially when such abilities are attenuated by phenomena at the so-called interface between syntax and pragmatics (e.g., Hulk & Müller, 2000; Müller & Hulk, 2001; Serratrice, 2007, 2013; Kupsich and Rothman, 2016). Thus, one open question is whether bilinguals will have distinct interpretations of scalar terms as compared to monolinguals due to an effect of bilingualism, especially when the lexical items being tested differ across languages with respect to features that differentially affect their pragmatic distribution (such as the case of Spanish and English). Syrett et al. (2017a,b), for example, are two recent studies that shed light on the nature of SI interpretation among bilingual children. These studies provide evidence that monolingual and bilingual children sometimes, but not always, derive upper-bounded interpretations of *algunos* in Spanish. However, the bilingual groups show a peculiar insensitivity to the quantification of *all* and they do not distinguish *algunos* from *unos*. The authors argue that such an insensitivity did not reflect an inability to calculate implicatures, rather that the type of implicature dictated its specific interpretation. That is, bilinguals exhibited no difficulty deriving upper-bounded interpretations with particularized conversational implicatures (see also Austin, Sánchez, Syrett, Lingwall & Pérez-Cortes, 2015).

In summary, evidence from L1 acquisition of SIs is somewhat mixed. While most studies show that children are less capable of reliably calculating SIs than adults, research from L1 Spanish acquisition shows that children can pattern with adults in their interpretation of scalar quantifiers, at least roughly at the age of 5 and older. Furthermore, as relates to bilingual children, the little work that does exist seems to corroborate monolingual findings with some minor caveats pertaining to the quantification of *all.* In the next subsection, I will briefly outline adult L2 learner performance with respect to the interpretation of SIs.

1.6 Adult L2 acquisition of scalar implicatures

Work on the adult L2 acquisition of SIs is nascent (e.g., Hasley, 2010; Lieberman, 2009; Miller Giancaspro, Iverson, Rothman & Slabakova, 2016; Slabakova, 2010). One prominent study, Slabakova (2010), has suggested that mastering SIs in a given L2 is difficult, which may be related to the complexity of the lexical mapping from the L1 to the L2 (e.g., Lardiere, 2009). When both the L1 and the L2 have equal distributional patterns for the calculation of SIs, as is the case with Korean to English, then the acquisition process should be less difficult. Slabakova (2010) tested L1 monolingual Korean speakers learning English as an L2 on their ability to derive SIs both with and without enriched context. There were also control groups of English and Korean native speakers. In the first experiment, participants read sentences such as (12) and (13) and were asked to judge their appropriateness.

12. Some elephants have trunks.
13. Some books have color pictures.

The critical items in this experiment were the sentences like (12), which, while logically true, are also pragmatically infelicitous (i.e., underinformative) because indeed all elephants have trunks (Noveck, 2001). If participants rejected sentences like (12), it was assumed they were acting *pragmatically*, calculating the 'some but not all' implicature. Acceptance of such sentences, on the other hand, was taken to mean that they were acting *logically,* treating *some* as an existential quantifier. Surprisingly, results showed that L2 learners acted more pragmatically than both the monolingual English and monolingual Korean control groups, rejecting underinformative *some* more than either of the control groups. The second experiment provided participants with contextualized sentences and again asked them to judge their appropriateness. Results from this experiment indicated that L2 learners behaved more pragmatically than the monolingual control groups. Slabakova claimed that SIs "present no problem to L2 learners" (p. 2444), probably because SIs are presumed to involve universal

pragmatic principles (see also Dekydstpotter & Hathorn, 2005).

While Slabakova (2010) showed that L2 learners do not have difficulty deriving implicatures in an L2 when only one lexical item (*some*) is involved in the SIs of interest, another group's pilot study examining the L2 acquisition of Spanish SIs indicates that the addition of a second lexical item (*algunos* and *unos*) may make the learning task more difficult for L2 learners. For example, Vargas-Tokuda et al. (2009) discussed in brief the results from a pilot study of advanced L2 Spanish learners' interpretation of Spanish SIs. Unlike the L2 learners in Slabakova's (2010) study, those in Vargas-Tokuda et al.'s pilot study seemed to be "utterly unaware" (p. 114) of the distinction between the two quantifiers in Spanish. However, the methodologies used in each of the studies were different and, thus, not comparable. In Vargas-Tokuda et al. (2009) the L2ers were tested using the same procedure they used for the children described in detail above, a live presentation and interaction between experimenter and participant. Slabakova tested the acceptance of underinformativity with a pen-and-paper TVJT.

Moreover, Hasley (2010) presented results that run in sharp contrast to Vargas-Tokuda et al's. Hasley replicated the same methodology as Slabakova (2010)—adjusted by necessity to Spanish—and showed roughly the same results as Slabakova. His Spanish L2 group successfully performed much like the adult monolingual controls. Given the stark differences in methodology between Vargas-Tokuda et al. (2009), Slabakova (2010) and Hasley (2010), the disparity in performances might be more indicative of confounds in methodology than anything else.

Following the above studies, Miller et al. (2016) tested both highly advanced L1 English-L2 Spanish speakers, as well as native L1 Spanish-L2 English bilinguals as controls. They found that while both groups distinguished between the quantifiers appropriately, they did so to a lesser extent than the child and adult monolinguals reported in Vargas-Toukda et al. (2009) and the controls and L2 learners of Hasley (2010).[3] The behavior of the bilingual controls was, at first glance, explained in terms of a possible general effect of bilingualism, such as representational or processing difficulties, or of unexpected L1 attrition having taken place. In order to determine how monolingual controls would perform on the same task as Miller et al., at the start of this dissertation's research I used the same experiments to test a

[3] It is interesting to note that the control group in Hasley (2010) was very similar in make up to the bilingual control group in Miller et al. (2016), especially considering the differences between them. In both cases, the control group comprised dominant native L1 speakers of Spanish who were in the US at the time of testing enrolled in graduate studies (so their English was of a high proficiency level, despite being clearly dominant in the L1).

group of 30 native Spanish-speaking adults from Spain who had no or very low English proficiency upon arriving to the United Kingdom. In this pilot data, the monolinguals, like the L2 learners and the Spanish-English bilinguals of Miller et al. (2016), did not calculate any upper-bounded meaning of *algunos* nor did they distinguish *algunos* from *unos*. Because previous reports on Spanish SIs show that even young children reliably distinguish the quantifiers and calculated the implicature of *algunos*, Miller et al. (2016) determined that there was either an issue pertaining to the methodology they used, which differed from other experimental designs, or the theory of how SIs arise as related to Spanish's scalar quantifiers was not entirely accurate. It is on this basis that the present dissertation is testing both a revised methodology using event-related potentials (ERPs) as well as native Spanish-speaking adults and bilingual attriters.

1.7 ERPs and scalar implicatures

Since event-related potential (ERP) evidence will be crucial for two of the three articles that comprise this dissertation, it is worth explaining here what ERPs are and why the methodology lends itself nicely to SI research.

ERPs provide a non-invasive method to investigate electrophysiological correlates of mental processes. They emerge as a result of small voltages (measured in microvolts) that are generated in the brain when large groups of neurons fire in synchrony due to the onset of specific cognitive, sensory or motor events. ERPs are typically divided into two types: early and late waves (e.g., Peterson et al., 1995). The early waves, which peak within the first 100 milliseconds post stimulus, are called 'sensory' or 'exogenous' because they depend largely on the physical characteristics of the stimulus (e.g., auditory vs. visual). Termed 'cognitive' or 'endogenous' waves because they are characterized by information processing, the later waves are those that reflect how the stimulus is being evaluated in real-time.

A primary advantage of using ERPs is that the brain responds uniquely to different aspects of language processing, giving rise to reliably distinct ERP signatures and thus enabling one to tease apart various types of linguistic processing (e.g. grammatical repair versus failed expectation). For example, the brain responds differently to a morphosyntactic violation as instantiated in a gender or number violation in Spanish than it does to a semantic incongruency or a failed expectation of upcoming information. ERPs are definitively not the preferred neurological methodology for pinpointing brain locations implicated in specific types of processing—MRI and MEG, for example, would be much better for such questions—however, ERPs allow for a better examination of a participant's' sensitivity to a given stimulus at its

precise onset and computation. That is, they provide high temporal resolution in the measure of milliseconds. Therefore, if our bilingual and control groups show qualitatively different brain responses to the same stimuli, this can be taken as a reflection of differences at a so-called level of linguistic representation that ultimately leads to the use of distinct processing mechanisms (see Alemán Bañón, Fiorentino & Gabriele, 2014; Alemán Bañón, Miller and Rothman, 2017; Phillips & Ehrenhofer, 2015).

I now provide a brief explanation of the most relevant ERP component to SI research, the N400. The N400 is said to reflect lexico-semantic processes, strength of lexical associations and implausibility and it emerges as a negative going wave roughly between 250-500ms post stimulus (e.g., Kutas & Hillyard, 1980; see also Lau, Phillips, & Poeppel, 2008 for review). It is worth noting that while specific ERP components associated with language processing of one kind or another (e.g., P600 or N400) arguably may not reflect linguistic representation per se, they do emerge reliably in monolingual and advanced bilingual datasets examining any given linguistic property. Thus, if such effects emerge in the third study herein differentially based on learner type, we can say minimally that the controls and bilingual attriters have distinct electrophysiological processing responses (mechanisms) to the same stimuli.

Due to its connection to content stimuli, specifically semantic (im)plausibility and incongruency, probing for an N400 effect has featured prominently in ERP studies examining underinformative contexts with scalar quantifiers (e.g., Nieuwland et al. 2010; Noveck & Posada, 2003). In such studies, participants make acceptability judgments of underinformative sentences based on their knowledge of the world. Violations of world knowledge have been shown to be associated with more robust N400 effects (see Hagoort et al., 2017; Hald, Steenbeek-planting, & Hagoort, 2007). Thus, if *some* is interpreted with its upper-bounded meaning, predicates containing underinformativity violate one's expectations based on world knowledge and should, therefore, result in increased amplitude for the N400 effects than predicates without such a violation (e.g., *some crows have feathers* versus *some people are funny*). Caution is given, however, not only to the overall interpretation of N400 effects in these studies, but also to methodological design. This is because the N400 can also be modulated by the lexico-semantic relationship of words, specifically a target word and the main noun phrase in a given sentence. This relationship is often measured with respect to the frequency with which words co-occur in specific contexts (e.g., cops and robbers versus cops and dinosaurs) and is referred to as the *latent semantic analysis* value or LSA (Landauer et al., 1998). Larger N400 effects are expected in designs that employ underinformative *some* in world-knowledge contexts compared to informative ones, an effect that may be attenuated by the LSA value of

the main noun phrase and the target predicate in informative contexts whose noun phrase and predicate have a stronger lexico-semantic relationship. Attenuation can also be found in false sentences such as *some crows have radios* (see Noveck & Posada, 2003).

While somewhat different stimuli have been used to examine SIs using ERPs, most of these studies have found an N400 or an N400-like effect such as the Nref (negative reference), that is, a sustained late negativity (e.g., Politzer-Ahles et al. 2013; but see also Hartshorne et al., 2014 for positivity). Thus, we hypothesize that the N400 will be the component most likely to emerge in our data given the reliable findings of N400 effects in related research.

1.8 First language attrition

Because language is a social phenomenon, the multifaceted experiences of individual people, groups, and societies with language complicate its characterization (e.g., Schmid, Köpke, & de Bot, 2013). Studying such an array of variables and their effect on language, however, haS been at the center of work on L1 *attrition*—"the non-pathological decrease in a language that had previously been acquired by an individual" (Schmid & Kopke, 2004:5)—which challenges both the assumption that an L1 remains stable after development and calls for revisions in the methodologies used to test emerging patterns in bilingual development.

There is no question that various cognitive approaches to language acquisition and processing have progressed our understanding of the nature of the human mind-brain. However, still relatively little is known about the specifics of internal and external factors that characterize language maintenance and loss across the lifespan. Considering cross-linguistic influence, most bilingual studies to date focus their attention on the influence of the L1 on the L2 (see e.g., Foley & Flynn, 2013 for review). Other important questions, such as those pertaining to language maintenance and loss, or whether there is more or less linguistic interference in the inverse direction, have not been so frequently addressed. Historically, much of this work has taken the position that a fully acquired first language (L1) remains stable—that it, is impervious to change in at least some areas of grammar in spite of either the acquisition of additional language(s) or isolation from continued accessibility to the L1. The goal of the third and final study in this dissertation, thus, is to increase our knowledge of this understudied cross-linguistic direction, that is, L1 attrition and what its implications are for SIs in bilingualism more generally.

Schmid (2013) and Schmid & Köpke (2017) argue that changes to a first language go hand-in-hand with general bilingual development. Thus, studying L1 effects on L2 development and changes in the L1 as a result of L2 acquisition are both crucial to understanding larger questions

in language and cognitive sciences, particularly in a world that is becoming more and more multilingual. Perhaps the tradition of using a monolingual baseline as a control comparison in bilingual studies perpetuates the idealized notion of monolingual L1 competence, even within bilingual studies themselves (e.g., Rothman, 2008; Ortega, 2013). In turn, this could be a contributing factor underlying the erroneous assumption that an L1, once acquired, is more or less impervious to change, providing a non-moving target for comparison purposes.

The questioning of the so-called stability of an L1 system in L1 attrition research stems from the observation that most people live most of their life in an L1 environment where fluctuation in accessing native L1 input over the lifespan varies relatively little (see Iverson & Miller, 2017 for discussion). It is not counterintuitive, therefore, that early work assumed that a developing L1 becomes impervious to (major) changes by late childhood. Of course, we continue to add new lexical items throughout the lifespan, but grammatical properties (e.g., L1 syntax, morphology and phonology) are seemingly more resilient to change in the average person. Recent research shows that when the default balance of exposure to L1 input and opportunity to use it are thrown off, as in immigrant immersion contexts, the L1 is much more susceptible to change (under the right contexts dramatically so) than previously thought (e.g., Iverson, 2012). There is also a correlation between proficiency in the L2 in an immersion context, continued exposure to/use of the L1 in an L2 immersion context, age of acquisition of the L2, and degree of so-called attrition (see e.g., Schmid, 2011; Schmid & Köpke, 2017 for review). This is especially true at the levels of grammatical knowledge that are cognitively more costly, such as in real-time application of processing strategies (e.g., Dussias & Sagarra, 2007) and/or for use of linguistic properties subject to felicity conditions at linguistic interfaces (e.g., Gürel, 2007, 2011; Serratrice and Sorace, 2009; Sorace, 2011; Tsimpli, Heycock & Sorace, 2004; Tsimpli & Sorace, 2006). Interfaces are defined as abstract points at which two submodules of grammar (e.g., internal interfaces, syntax-semantics) or one submodule of language and one external to language (external interfaces, e.g., syntax-discourse), integrate information (e.g., Chomsky 1981; Jackendoff, 1983, 2006; Ramchand & Reiss, 2007; Reinhart, 2006).

Current developments in theories of bilingualism and in the methodologies to test them indeed show that diachronic and synchronic language change are standard at a societal and individual level, for both monolingual and bilingual populations (see Schmid and Köpke, 2017 for review); however, the vast majority of research on language attrition places an emphasis on the conditions that give rise to language loss. Alternatively, Iverson & Miller (2017) suggest that the reconceptualization of a steady-state grammar in light of evidence that an L1 can

undergo significant changes across the lifespan, and reframing the question around the conditions that foster language maintenance is necessary and would enrich current theories of attrition.

In this vein, there have been many theories in various linguistic paradigms aimed at explaining the so-called selectivity of attrition and why it might unfold the way it does. For example, Paradis (2004) put forward the Activation Threshold Hypothesis (ATH) as part of a neurolinguistic approach to bilingualism. The ATH claims that the most frequently used forms or languages are 'activated', while those used less frequently are 'inhibited'. Inhibition of a form or language raises its so-called 'activation threshold'. Forms and languages with a high activation threshold are more difficult to reactivate than those with a lower threshold. In cases where two similar forms in two different languages are in competition with one another, the form in the language that is more frequently used will be activated while the form in the language that is less frequently used will be inhibited; such inhibition is argued to happen simultaneously with the activation of the other form/language. According to the ATH, then, the less frequent and, therefore, less activated of two competing forms will be more susceptible to attrition. Attrition from this perspective is seen as inhibition of a linguistic form such that if there is no competing form, there will be no frequency-induced inhibition. It stands to reason from the ATH that the longer one lives in L2 immersion, but crucially the (greater) extent to which one comes to replace the L1 with the L2 in daily use and thus progressively becomes more isolated from the L1 (not necessarily guaranteed in L2 immersion per se), the deeper the effects of L1 attrition will take hold. In this sense, time in L2 immersion is potentially a necessary—but not sufficient—proxy for modeling L1 attrition over time. Finally, the ATH does not claim that any sub parts of grammar have such a special or entrenched status that they will be impervious, but rather the relative activation of the languages determines the degree of vulnerability to change.

Another theory regarding the nature of attrition was put forward by Jakobson (1941) in the form of The Regression Hypothesis (RH). The RH claims that attrition unfolds in the inverse order of the L1 acquisition process: the last things acquired in a language will be the first things that are lost. Originally, Jakobson's claims were formed on the basis of data from attrition caused by aphasia. However, the RH can be extended to other forms of non-pathological attrition (e.g., Keijzer, 2004 see also Iverson, 2012 for review). For example, because patterns of general development in L1 acquisition have been established, the RH can be tested straightforwardly. Considering that L1 acquisition is a process by which a speaker begins with a set of unmarked universal linguistic properties and later acquires marked

language-specific properties, the RH predicts that the marked values should be those that are more susceptible to loss. Furthermore, if some developmental sequences in first language acquisition are universal, the RH predicts that some attrition sequences might also be universal. It is worth mentioning, especially in light of significant changes to how language has become conceptualized since 1941—the very beginning of the cognitive revolution in psychology and linguistics was more than a decade away at this time—that the general observations seemingly covered by the RH might be complicated in 2017. For example, many properties that are late acquired relate to domains where linguistic and non-linguistic information must be integrated (so-called external interfaces) and/or display a high level of formal complexity. Therefore, it is possible that "last learned in the L1" is an inadvertent proxy for something else. For many properties that have been tested, as a result, data consistent with the RH are also inevitably consistent with other theories, such as the Interface Hypothesis, to which I now turn.

The Interface Hypothesis (IH) predicts that properties that are associated with or dependent on contextual appropriateness (discourse-pragmatics) are the most likely to undergo change (e.g., Sorace, 2000, 2011; Tsimpli & Sorace, 2006; Tsimpli, Sorace, Heycock, & Filiaci, 2004). It also claims that properties that are internal to the grammar, such as narrow syntax and so-called internal interfaces where submodules of the linguistic system interact such as syntax-semantics, are significantly more durable. Born from these strong predictions, a great deal of work has shown variability at the external interfaces and somewhat convincingly claimed that it arises at least in part due to limitations in working memory, processing capacity or efficiency, and attentional resource allocation, which are already stressed in bilinguals as compared to monolinguals for obvious reasons (see Rothman & Slabakova, 2011; Sorace, 2011, 2012 for discussion). In this respect, production consequences stemming from the activation of multiple linguistic systems, which requires constant inhibition of one grammar while using/processing the other, are the claimed locus of variability, even in highly proficient bilinguals, and not necessarily disparate (attrited) linguistic representations (see Serratrice & Sorace, 2009; Sorace, 2011, 2012; Wilson, Sorace & Keller, 2009).

Indeed, language processing for bilinguals is argued to be more costly than for monolinguals given that the former must necessarily deal with additional mental tasks; these include but are not limited to (inhibition) from either language, correction of automatic reflexes from language systems that are not relevant for use, dealing with multiple lexicons, and more (e.g., Abutalebi, Della Rosa, Green, Hernandez, Scifo, Keim, Cappa & Costa, 2011; Bialystok, 2009; Green, 2011). Bilinguals also have to maintain processing-specific routines based on comparatively less input than monolinguals, which may contribute to linguistic features and

lexical items—in either language—that are more weakly represented or less activated than those of monolinguals (e.g., Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Hopp, 2013). Within the greater discussion of bilingual development, however, crosslinguistic competition and decreased activation of language-specific properties is not only a reality in the direction of the L1 to L2, but also from L2 to L1. One goal of current attrition research, then, is to measure beyond performance alone and determine the extent to which differences in production, processing and competence reflect real modifications to specific mental L1 representation(s). It has thus been argued that in order to better understand not only the essence of bilingual development but also to shed light on open questions regarding the language capacity of humans, we must develop a clear method for testing whether and how the mechanisms constraining L2 acquisition can impact linguistic knowledge already developed in the L1 (Schmid & Köpke, 2017).

In the third study herein, we address the above issues by administering the same experiments from articles 1 and 2 to a group of native L1 Spanish speakers who have resided in the UK (thus an English L2 immersion context), commencing in adulthood, for a significant amount of time. Recall that SIs are recognized as a property that arise due to pragmatics, and whose language-specific instantiations in Spanish and English display distinct distributional patterns. SI derivation is dependent on an individual's association of a numerical quantity with the lexical item that gives rise to the SI. If on the right track, then, we hypothesize that scalar (quantity) interpretations may be susceptible to attrition in a different way than pragmatically-induced changes in other frequently examined phenomena are (i.e., null/overt subjects or objects see article 3 for review). That is to say, the semantic and conceptual mappings of scalar interpretations onto specific linguistic items, in this case quantifiers, are hypothesized to be what undergoes attrition, not the conceptualizations of quantity themselves. It is not enough, however, to simply state that attrition occurs and what might give rise to it. To get at the question of whether attrition is a case of representational change (in our case conceptual representation as mapped on to linguistic knowledge) or processing challenges in the bilingual mind, we test scalar interpretations using a combination of offline acceptability judgments and online neurolinguistic measures as exhibited via ERPs.

I must concede that there is no such thing as a true smoking gun for a discipline that is forced by circumstance to test the object of query indirectly via various types of behavioral performances. The best we can do is offer generalized conclusions based on a preponderance of converging evidences. Thus, I submit that simultaneously testing for L1 attrition by combining a battery of offline and online methodologies will be in an optimal position to

adjudicate between various levels at which the captured changes sit. That is, do they reflect a (momentary) processing burden, bona fide changes to previously acquired mental representations, or something else entirely?

1.9 Research problem

While it would be interesting to understand *why* some individuals *do* and some *do not* derive SIs, taking as a given that pragmatic interpretations are how one *should* behave ultimately leads to searching for an answer in some inherent problem within the individual, rather than acknowledging that variation can arise from external factors. Inferencing is largely dependent on pragmatics. The question, then, is what external variables (e.g., experimental design and task discourse) affect the derivation of SIs. In this respect, the issue is not *only* whether there are mitigating underlying cognitive/psychological issues at play during pragmatic inferencing, but also what the parameters used to make such a claim in empirical work should first be in order to most reliably tap individuals' knowledge of the properties in question. If the task design considers potential confounding variables, removes them and individuals still show the same pattern of variable interpretation, we can then more soundly associate cognitive metrics to individuals' rates of deriving pragmatic inferences.

A significant contribution of this dissertation rests in the discussion we offer as it relates to some methodological issues commonly applicable in SI research we maintain affect the way SIs are derived in experimental contexts (especially articles 1 and 2 of this dissertation). That is, we question the appropriateness of some commonly employed task types and stimuli designs in contemporary SI research, specifically regarding the division of subjects into those who are logical and those who are pragmatic. We outline these methods and their potential pitfalls and we put forward specific data sets from a set of combined behavioral and neurolinguistic (EEG/ERP) studies that actively avoid these potential methodological issues. While the studies focus on inferencing in Spanish, an understudied language in the SI literature, we maintain that the results inform well beyond anything specific to Spanish and are indeed beneficial for SI research in the broader sense. We are careful to highlight specifically what methodological changes we made as compared to previous studies in order to capture a more accurate snapshot of how SIs work in Spanish, which may prove useful for how SIs are tested in other languages and would be equally applicable to adults and children subjects, as well as monolingual and various types of bilingual and multilingual speakers. These studies show how having complementary behavioral and neurolinguistic measures are crucial to a better understanding of how SIs work in Spanish); that is, how insights from the behavioral task led us to do a more

revealing analysis with the ERP data that would not have followed otherwise. Once this is done at the level of the typical control baselines in SI work, namely for the adult native speaker living in her L1 environment, we then get a more accurate picture of what is going on in L1 Spanish (at the group and crucially individual speaker levels). This then allows us to extend such a methodology to L1 attrition (article 3), revealing variation in control groups that are inadvertently obscured otherwise to see how L1 attrition groups and indeed individual performances truly differ from the range of performances of properly vetted controls.

Furthermore, showing that there is a correlation between ASQ scores and individual differences in SI judgment/computation, for example, would only be truly revealing to the extent that one understands what underlies the ASQ scores themselves, which is not well understood at present. Thus, however robust the correlation is, it does not necessarily solve the problem of understanding the mechanisms underlying individual differences, much less the actual causation loop. It is worth mentioning that in these data such a distinction between logicians and pragmatists is at least superficially just a matter of preference. We argue that the seemingly less common cases of extreme polarization of interpretations, such as those where a logical interpretation is preferred to the exclusion of a pragmatic one, are more reflective of a potential inability to derive pragmatic meaning when supported by context.

We also know that bilinguals treat scalar *some* differently than monolinguals. As we will see in the experimental studies in this dissertation, in stark contrast to our L1-dominant control data (articles 1 and 2), bilingual attriters (article 3) reveal no distinction between logical and pragmatic responses, either in the offline or the online tasks. Because *algunos* is also treated in much the same way as *unos*, there is no meaningful way to divide the responses that would be revealing of the pragmatic capacities of our attriters. This is not to say they are incapable of being pragmatic or that there is indeed some deficiency regarding the cognitive mechanism responsible for pragmatic abilities, but that bilingualism muddies the water for a specific description of these abilities, particularly as they relate to the derivation of SIs.

Chapter 2: Article 1

Abstract

Studies on the derivation of the implied meaning of the scalar term *some* (scalar implicature henceforth SI) have reported variation in interpretations. Early work suggested that such variation may have been the result of the type of task used (e.g., underinformative *some* in world knowledge contexts, lack of instruction, etc.) and more recent work both confirms this and posits additional task-related effects in the way individuals treat scalar *some*. However, not much is known about SIs in adult Spanish and whether or not there is similar individual variation among individuals in Spanish. Spanish is interesting in that it has multiple lexical items whose semantic meanings overlap considerably but whose pragmatic use varies.

Herein, we follow up with the idea that both methodology and individual variation are relevant, implementing two tasks on the interpretation of the scalar quantifiers *algunos* 'some$_A$' and *unos* 'some$_B$' in Spanish among monolingual Spanish-speaking adults. Compared to previous work in Spanish, we show that calculation of SIs in Spanish can increase by nature of the tasks used and that an analysis of whether SIs are calculated or not can be informed by the way in which individual participants interpret scalar terms. We argue that some variation in previous research on SI in Spanish specifically might be best explained as a confluence of both methodological and psycholinguistic factors, though such a confluence may have considerable knock-on effects for how SIs are examined in any population and in any language.

Keywords: scalar implicature, logician, pragmatist, methodology

2.1 Introduction

The understanding of language requires interlocutors to rapidly make use of both the semantics of words and the non-linguistic context in which they are used. Thus, we must understand that which is entailed by a statement and incorporate any meaning that extends beyond that which is explicitly stated or entailed by the utterance (e.g., Bach, 1999; Grice, 1975; Morris, 1938). Consider the following statement:

1. It's hot in here.

The only information one has upon hearing such a statement is there is a place that is presumably hot. We know what was said, but we do not know why or where or to whom. In

order to understand why things are said and how to interpret them, we must necessarily have at least some idea of the context in which a given utterance is expressed. Now, imagine (1) being used in the following context: a couple is driving down a desert road during the middle of the summer and the air conditioning is off. The passenger says to the driver, "it's hot in here". With more context we can see that the driver has at least two ways to understand what is being said. The first would be an understanding that is "logical" or literal in which the passenger merely intended to comment on her observation of the temperature. In this case, the driver may leave the air conditioning off and continue driving. The second would be an understanding that is "pragmatic" or implicit, whereby the message is understood as an indirect speech act that is in fact a request for the driver to remedy the temperature situation by turning on the air conditioning. Processing such an implicit meaning would require the hearer to make an inference above and beyond that which was explicitly stated, that is, a pragmatic inference.

Contemporary research examining pragmatic inferencing has paid particular attention to a subtype of inference known as *scalar implicature* (SI) resulting from a number of lexical items, though research tends to focus on quantifiers such as *some*. In the case of *some*, the associated SI is said to be derived on the basis of it belonging to a class of quantifiers that form part of a scale in which each term is ordered according to its informativity, such as <*some, most, all*>. On this scale, the stronger terms naturally entail the weaker ones but not vice versa. Thus, the use of an informationally weaker term like *some* implies that a stronger term like *all* is not applicable (Grice, 1975, 1989; Horn, 1972).

Over the last three decades, the majority of SI-related research work has focused either on monolingual children by examining both their capacity to calculate SIs and their general reasoning skills (e.g., Ackerman, 1983; Bever, Mehler, & Epstein, 1968; Bloom & Wynn, 1997; Braine & Rumain, 1981; Cassidy, 1998; Gelman, 1993; Gelman & Gallistel, 1978; Greeno, Riley, & Gelman, 1984; Hurewitz, Papafragou, Gleitman, & Gelman, 2006; Lidz & Musolino, 2002; Noveck, 2001; Papafragou & Musolino, 2003; Smith, 1980; Syrett et al., 2017; Vargas-Tokuda, Grinstead, & Gutiérrez-Rexach, 2009; Wynn, 1992), or monolingual adults (e.g., Bergen & Grodner, 2012; Bott, Bailey, & Grodner, 2012; Breheny, Katsos, & Williams, 2006; De Neys & Schaeken, 2007; Dieussaert, Verkerk, Gillard, & Schaeken, 2011; Hartshorne & Snedeker, 2013; Huang & Snedeker, 2009; Katsos, 2008; Nieuwland, Ditman, & Kuperberg, 2010; Politzer-Ahles, Fiorentino, Jiang, & Zhou, 2012; Sauerland, 2012, Zhao et al. 2015). Notwithstanding the breadth and depth of the research into the comprehension of scalar terms in any population, it is fair to say that there has been an equally diverse range of methodologies employed to examine them and a varied collection of results. While—in recent

work—adults are often the focus of extensive online experimental work (i.e., eye-tracking, self-paced reading, event-related potentials (ERP), and more), children, for obvious reasons, are predominantly tested offline (but see Huang & Snedeker, 2009).

In general, the research shows that while children tend to be more logical than adults in their treatment of scalar terms, that is, for children *some* is more frequently compatible with *all* than for adults (but see Syrett et al. 2017a, 2017b), individual adults, too, can be either logical or pragmatic. For children, such variation early on was thought to arise for a variety of developmental reasons. For example, it was posited that either young children lacked the prerequisites necessary for adult-like pragmatic inferencing (pragmatic limitation hypothesis) or children did calculate pragmatically-based meanings but less often than adults (pragmatic delay hypothesis) (see Chierchia, 2005; Guasti, et al. 2005). Current research shows that children indeed can calculate SIs when provided with enriched instruction or when the nature of the task is more explicit (Guasti et al. 2005; Papafragou & Musolino, 2003; Syrett et al. 2017a, 2017b).

For adults, online investigations of scalar interpretations have shown, for example, a relationship between either an increase, decrease, or null effect in pragmatic inferencing and working memory (e.g., Marty & Chemla, 2013), scores on the Autism Spectrum Quotient (ASQ) (e.g., Heyman & Schaeken, 2015; Nieuwland, Ditman, & Kuperberg, 2010; Zhao et al. 2015), as well as a possible relationship between scores obtained on the Systemizing Quotient-Revised (SQR) questionnaire and other pragmatizing tests. Regarding the SQR, it has been said that "systemizing and intolerance to pragmatic violations…would tend to increase with SQ-R score" (Barbet & Thierry, 2016). However, all of these studies show evidence that, in spite of any psychological underpinnings to the complexities of deriving implicit, pragmatically induced meanings, individuals treat scalar *some* with a great deal of variability, interpreting it either logically (i.e., *some* can be 'some and possibly all') or pragmatically (*some* is only 'some but not all'). Interestingly, however, in the whole of SI research, inter-individual variation seems to neatly categorize participants as either logical or pragmatic in only a handful of studies (e.g., Bott & Noveck, 2004; Hunt, Politzer-Ahles, Gibson, Minai, & Fiorentino, 2012; Noveck & Posada, 2003; Politzer-Ahles et al., 2012; Spychalska, Kontinen, & Werning, 2016; Tavano, 2010). This begs the question of whether or not individuals are indeed either one or the other, or whether the answer to such a question depends on other variables both internal and external to individual speakers/listeners.

The question of why there is such lack of conformity to any predictable behavior among individuals seems so pervasive in SI work was first addressed explicitly by Papafragou &

Musolino (2003) and Guasti, Chierchia, Crain, Foppolo, Gualmini & Luisa (2005) who demonstrated that the manipulation of experimental demands can affect task performance, such that training and providing rich, naturalistic context can increase the rate with which the pragmatically enriched meaning of scalar terms is derived, even in child populations. However, these (and other) studies have shown that adult performance is not without considerable variation, with as much as 30%+ of adult control groups accepting underinformative uses of *some* such as *Some elephants have trunks.*

The main goal of this paper is not to take sides on any one theoretical stance about the psycholinguistic mechanisms that underlie pragmatic inferencing per se. Rather, our aim is twofold: (1) to shed further light on the importance of methodological considerations when testing subtle linguistic properties that require the collective intake and comprehension of linguistic cues and the larger context in which those cues are given. (2) To understand how and why only some individuals calculate implicatures in real time when presented with the same contexts? To help make the points alluded to throughout this paper, we present experimental behavioral data from monolingual adult interpretation of SIs in Spanish via the quantifiers *algunos* and *unos* 'some' paying particular attention to inter-individual variation, something that has not been explicitly addressed in previous on Spanish. The research on SI interpretation in Spanish at any age has yielded disparate results across studies (e.g., Miller, Giancaspro, Iverson, Rothman, & Slabakova, 2016; Syrett et al., 2017a,b; Vargas-Tokuda, Grinstead, & Gutiérrez-Rexach, 2009). That is, child and adult populations treat Spanish *some* with a range of interpretations. In light of the above discussion, then, the present study made several methodological modifications to how SIs were investigated in the Spanish literature on adult SI calculation in order to determine what affect—if any—simple modifications might have on individual performance.

2.2 Scalar Implicatures

Grice (1975, 1989) proposed various principles and maxims of communication that guide communication and the understanding of language. The *Cooperative Principle*, for example, states that a speaker should make his or her contribution only as is required given the timing of the utterance and the direction and/or purpose of the exchange. Thus, at any given moment during a conversation, some contributions will be appropriate and others will not. The *Maxim of Quantity* states that a speaker's contribution to a given exchange should neither be more nor less informative than is required for the purpose(s) of the exchange. Horn (2005) elaborated on characteristics of this maxim stating that it is arranged in a way that automatically promotes an

upper-bounded meaning (inference-based) from scalar terms allowing for conversational implicatures. Consider the following statement:

2. Some linguists like pragmatics.

Most native speakers of English, upon reflection, would be able to arrive at a meaning where (2) implies that *not all linguists like pragmatics.* This 'not all' interpretation, however, is not explicitly expressed nor entailed by the utterance. The course by which a hearer interprets a sentence like (2) as meaning *not all* is referred to as a pragmatic enrichment process, through which she comes to the conclusion that the speaker will be maximally cooperative by using the most informative term(s) possible. However, the speaker's choice to use the weaker term must mean that the stronger term *all* is not true. Horn (1972, 1989) and Levinson (1983) proposed that such lexical items fall on a scale that quantifies their informativity (e.g., <*some, many, most, all*>) where the entailment of weaker lexical items by stronger ones is referred to as calculating an *upper-bounded* interpretation. This stems from the notion that the weaker item being used to trigger the SI is naturally entailed by and compatible with stronger ones of the same scale. It is true, therefore, that if all linguists like pragmatics, then the same holds for most, many and some.

An important hallmark of SIs as opposed to semantic entailments is the fact that they are defeasible or revisable, thus leading *some* in the following exchange to be reinterpreted as *all*.

3a. What happened to the cookies?
3b. Owen ate **some** last night. Actually, he ate **all** of them.

The pragmatic status of the SI associated with *some* in 3b is evident in that the *not all* interpretation can be cancelled without resulting in a nonsensical utterance. The same is not true of the purely semantic meaning of *some*, which is "more than zero":

4. Owen ate **some** (cookies) last night. #Actually, he ate **none** of them.

Recent years have seen a sharp increase in work on both the experimental and theoretical investigation of pragmatic inferencing (see Crain, 2012 and; Noveck & Reboul, 2008 for reviews). Indeed, many proposals have been put forward in an attempt to explain the specific nature of SIs, as well as the mechanism(s) that underlie their interpretation, comprehension, and computation. For example, one open question seems to be whether SIs are context-dependent conversational inferences (Carston, 1990; Hirschberg, 1985) or inferences that

appear by default any time a scalar term like *some* is encountered (Levinson, 2000). Moreover, there is an ongoing discussion as to the derivation of SIs being associated with either scaled informativity more broadly (Chierchia, Fox, & Spector, 2012; Gazdar, 1979; Horn, 1972; Levinson, 2000; Sauerland, 2012) or an apparent relevance-based mechanism (Carston, 1998; Sperber & Wilson, 1995). Finally, much research on SIs has shown both how some scalar inferences contribute (or not) to the truth-conditional content expressed in an utterance rather than the unexpressed content (Carston, 1998; Horn, 1992) and whether they are the potential result of lexical (local) calculations (see Levison, 2000) as opposed to higher-level structure/informativity (Chierchia, 2006; Chierchia et al., 2012; Bart, Geurts & Pouscoulous, 2009; Katsos, 2008).

Irrespective of any of these theories, it is safe to say that there is considerable variation with respect to how individuals interpret scalar terms more generally. As noted in the introduction, we maintain that at least some of this variation might be accounted for as an artefact of experimental tasks, in addition to any underlying psychological characteristic that arbitrarily determines the calculation of upper-bounded meanings in scalar terms. In an attempt to validate this claim, we examine SIs in Spanish, which have neither been as extensively investigated as in other languages, nor are they immune to similar interpretative variation. Some enlightening work on the role methodology plays on SI calculation makes a strong case for great care to be given to the design of the methodology and the type(s) of stimuli used (e.g, Degen & Tanenhaus, 2011, 2015). With this work in mind, we implement a revised methodology compared to those employed in Halsey (2010), Lieberman (2009) and Miller, Giancaspro, Iverson, Rothman & Slabakova (2016), all of which examined SIs in adult Spanish.

2.3 Scalar Implicatures in Spanish: Algunos and Unos

Spanish presents a relatively unique and interesting test case to examine SIs experimentally in that it has two plural scalar quantifiers that are a rough equivalent to *some* in English: i) *algunos/as* ‘some-pl’ and ii) *unos/as* ‘some-pl’. It has been argued that these determiners have “a uniform denotation as indefinite (existential) expressions of type <<*e*,*t*>, <<*e*,*t*>,*t*>>” and, as such “they denote intersective (existential) functions” (Vargas-Tokuda et al. 2009, p.101; see also Alonso-Ovalle & Menéndez-Benito, 2002; Keenan, 1987). In other words, in order for their inherent semantic conditions to be satisfied, it is only necessary that both be a referent of a plurality of entities greater than or equal to two.

5. *Algunos/unos niños están durmiendo.*
‘Some/some children are sleeping.’

Despite their intersective functions, however, *algunos* and *unos* are quite distinct, especially regarding their pragmatic distributions, and there are several theoretical approaches aimed at explanating said distinctions. In all approahces, *algunos* is argued to be compatible with a quantity implicature while *unos* is not. For Gutiérrez-Rexach (2001, 2003), *algunos* is associated with a semantic “no linking” constraint (see also Vargas-Tokuda et al., 2009), which allows it to refer back to an already expressed referent in the discourse, while *unos* cannot. Martí (2008), on the other hand, posits that *alg-*, while it has no truth-conditional content, attaches to *unos* and inherits all of its properties (i.e., it is non-generic, it has existential import, it is semantically plural and it is a quantifier), also adding an impicature. Finally, Alonso-Ovalle & Menéndez-Benito (2002, 2013) hold that the two most common approaches to explaining the behavior of weak determiners (i.e., the pragmatic approach, see de Hoop, 1992; Diesing, 1992; Partee, 1989, and the ambiguity approach, see Buring, 1996) are needed to explain *algunos* and *unos*. For example, Alonso-Ovalle and Menéndez-Benito claim that *unos* can be pragmatic and therefore presuppositional, but in so being it must be marking topic or contrastive focus. On the other hand, they claim that *algunos* is both presuppositional and non-presuppositional where either reading is not dependent on the topic/focus articulation of the sentence in which it appears.

Thus, while both quantifiers are said to allow for a subset reading, it is only *unos* that can refer to all members of an entire set. However, taking Gutierrez-Rexach’s no-linking constraint as an example, *unos* should not be able to relate back to a specific referent that is already present in the discourse because it does not have a specific/partitive interpretation like *algunos*.

6. *Los libros de matemáticas están en el cajón, los de física debajo de la cama y hay ??unos/algunos de lingüística sobre la mesa.*
‘The math books are in the drawer, the physics ones are under the bed and there are some linguistics ones on top of the table.’
[Vargas-Tokuda et al. 2009, p.103, example (37)].

Thus, *unos* contributes only a non-linked referent to the discourse, while *algunos* is D(iscourse)-linked and, therefore, partitive in nature. In this respect, *algunos* is said to be the determiner that gives rise to the quantity implicature. Again, however, *unos* can have a ‘some but not others’ reading when contrastively focused as in (7). The pragmatic status of *algunos*

can be seen in (8) where the 'not all' interpretation associated with it can be cancelled or defeated.

7. *Unos atletas están corriendo, otros no.*
   'Some athletes are running, others are not.'

8. *Llegaron algunos/#unos estudiantes, de hecho todos.*
   'Some students arrived, in fact all of them did.'

In sum, while these quantifiers are similar on the surface and share overlap in meaning, specific properties of *algunos* make it, and not *unos,* the implicature generating item. It is interesting to note that irrespective of any of the proposed constraints on these determiners, it is in fact some semantic feature (or morpheme) of either that predicts the generation of the implicature, not pragmatics per se. In this sense, it seems logical that there have not been results in the Spanish SI literature to date indicating a division between those who might respond logically or those who respond pragmatically as a general tendency given that each term has a contextually relevant interpretation. We will see that our results come in stark contrast to previous reports in Spanish.

2.4 The experimental study of SIs

Herein, we highlight some traditional tasks used to examine scalar *some,* specifically variables within them that have been highlighted as variables that can contribute to task-induced variability (e.g., Degen & Tannenhaus, 2011, 2015; Guasti et al. 2005; Papafragou & Musolino, 2003; Syrett et al. 2017a).

2.4.1 World knowledge and plausibility

A large number of studies testing SIs have examined the acceptance of *some* in underinformative sentences, requiring individuals to call upon their knowledge of the world. In cases where comprehension depends on such variables, it can be difficult to determine whether or not an individual makes a specific judgment because of how extensive her knowledge of the world is, what she determines is or is not plausible or how she interprets *some* strictly speaking. For example, Guasti et al. (2005) note that an individual might accept or reject a statement such as *some giraffes have long necks* based on her ability to evaluate any number of possible alternatives (e.g., Could there be giraffes without long necks? What about baby giraffes? What about a giraffe with a congenital neck shortening disease?), rather than

what she determines is the actual meaning of *some*. Tasks of this type, then, offer little control over whether or not all the participants are using the same criteria (i.e., prototypical entities or best exemplars) to evaluate the critical statement. For example, it has been shown in some psycholinguistic work on SI calculation that though adults seem to exhibit a slow-down effect when reading underinformative uses of *some,* thus showing a sensitivity to its (in)felicity, they also show a general acceptance of such conditions in their offline judgments (e.g., Nieuwland et al. 2010). In other words, the implicit brain response measured via reading times suggests that such sentences are at a minimum odd or unexpected, but the offline judgments reveal the opposite. Moreover, Barbet & Thierry (2016) highlight that differences in the strength of the underinformative statements used may play a role in how individuals interpret them. For example, a statement such as *some infants are young* is underinformative because, by the very definition of what it means to be an infant, all infants are young. Contrastively, a statement such as *some hammers have a handle* allow individuals to more easily envisage counter-examples to the alternative *all* statements. Thus, because additional context is needed for the latter (weaker) example but not the former, using such statements to judge whether an individual does or does not calculate the SI associated with it may fail to entirely explain why some individuals are more tolerant of pragmatic violations than others. It is worth noting that some Spanish SI work has used world knowledge-dependent stimuli (Halsey, 2010; Lieberman, 2009) and found equally disparate results.

2.4.2 Lack of ambiguity

Some recent studies on implicature have underscored the importance of distinguishing partitive *some of* from simple *some* (e.g., *some of; summa→some of; alla→all of; only some*, etc.) (e.g., Degen & Tanenhaus, 2011, 2015; Grodner, Klein, Carbary, & Tanenhaus, 2010; Hartshorne et al., 2014; Hunt et al., 2013; Nieuwland et al., 2010; Papafragou & Musolino, 2003). To understand how the use of *some of* might bring about complications, we examine Hartshorne et al. (2014), an event-related potential (ERP)[4] study that created a novel task to test whether SIs were calculated or not by examining *some of* in both upward (UE) and downward entailing (DE)[5] contexts. It has been well established that there are certain contexts in which an implicature is expected not to arise. One such context is that of downward entailment.

[1] Event-related potentials (ERPs) entail the time-locking of the raw electroencephalogram (EEG)—mapping of the electrical activity produced by the brain in real time—to the onset of a specific stimulus of interest in an experimental design, most commonly via rapid serial visual presentation (RSVP) of the stimuli. RSVP involves presenting the stimuli (in this case sentences) one word at a time in the center of a computer monitor which the participant is viewing. The EEG is then time-locked to the stimulus of interest within the sentence in both grammatical and ungrammatical (felicitous and infelicitous) conditions. These conditions are compared such that the difference in amplitude (voltage differences produced by the brain)

UE→ *Sally ate **some of** the cookies, and **the rest** are on the table*
DE→ *If Sally ate **some of** the cookies, then **the rest** are on the table*

Hartshorne and colleagues tested both contexts with the expectation that the implicature would not arise in DE contexts and, thus, neither would any specific electrophysiological component known to emerge with implicature generation (e.g., N400). This design is revealing in that, as opposed to picture-sentence verification tasks, for example, the authors were able to see when the implicature itself was or was not calculated by focusing on the portion of the conditional statements following the quantifier (i.e., *the rest*). What the authors showed differed from other ERP studies examining SIs in the emergence of a positivity (e.g., P600) and not a negativity associated with the calculation of the SI at *the rest* in contexts that did not support implicature (i.e., DE). No such effect was associated with *some* in either the UE or DE contexts:

> Perhaps the most intriguing finding was the absence of an effect at the scalar implicature trigger *some*. We found parallel results in five self-paced reading experiments involving similar stimuli: *some* was read no slower (or faster) in contexts where the scalar implicature was calculated. Nonetheless, analyses at *the rest* – both here and in the self-paced reading experiments – show that our manipulation affected scalar implicature calculation. (Hartshorne et al. 2014, p. 631).

One reason Hartshorne and colleagues may not have seen an effect at *some*, however, is the use of *some of* rather than just *some* in both the above-mentioned study and the parallel self-paced reading experiments. Using these partitive constructions in place of simple *some* is known to generate robustly pragmatic interpretations (e.g., Crain, 2012; Politzer Ahles et al. 2013). Thus, when *some of* is seen, even when presented separately within a rapid-serial visual presentation (RSVP) design common to ERP studies, one cannot determine whether what prompted a given electrophysiological response downstream of the quantifier was due to the

between a grammatical sentence and an ungrammatical one at the moment of the critical stimulus is the resulting ERP effect. ERPs provide high temporal resolution of a given brain response in milliseconds, offering a view of language processing as it unfolds over time. Distinct ERP effects have been shown to emerge as a direct result of different types of language processing (i.e., morphosytactic versus semantic). The N400, for example, is a negative going wave which peaks at around 400 ms—hence N(egative) 400—and is a component shown to emerge as a result of either a semantic or, in the case of SIs, pragmatic violation. The N400 is produced reliably in ERP work studying SI calculation and is thus taken to be the expected component in such studies (but see Hartshorne et al. 2017).

[2] Downward entailment environments invert the entailment relations such that [subset to superset] becomes [superset to subset]: e.g., *nobody likes science (superset) → nobody likes linguistics (subset)*. An example of a DE environment would be a conditional statement.

participant's interpretation of *some* alone or whether it was the phrase *some of*. As with any ERP task, there are necessarily many trials (hundreds, in fact) and the result we are presented with is the grand average of all such trials. After several trials, then, a participant might come to realize that every time she sees *some* in the context of the experiment, *of* will follow, thus potentially priming a given response by lessening or removing the ambiguity of *some* by itself. Thus, it would be no surprise to see that when testing *some of* there would be either near-ceiling rejection of it offline in underinformative contexts or robust electrophysiological effects in the ERP precisely because *some of* is not ambiguous like *some*. Therefore, one cannot meaningfully conclude on the basis of data testing only *some of* that the same patterns would necessarily hold for *some* tested in isolation.

A potential reason no effect was seen at *some* may have been due to the lack of context in addition to the target sentences. That is, the semantic and/or pragmatic status of *some* could not yet have been determined at its onset (see Degen & Tannenhaus 2011, 2015). Deriving a pragmatic inference from a scalar term is said to require—minimally—extralinguistic cues from the discourse which one can use to resolve a possible ambiguity, otherwise a first pass reading may well be merely existential (e.g., Newstead, 1988). However, decontextualized sentences such as those used in the above study present no discourse or context from which an individual could resolve an ambiguity nor could the quantifier itself be ambiguous at the moment *some's* onset. In these cases, *some* would carry an existential meaning with no reference to any specified set of cookies of which *some* can form a part. The portion of these sentences from which *some* can get additional reference and from which it can be inferred to mean anything other than simply a set of cookies is the subsequent conjoining phrase "and/then the rest". Thus it is reasonable that no effect was found at *some* and only at *the* rest. Again, these arguments would hold true of Spanish, too. Simple *algunos* as compared to *algunos de* 'some of' does not elicit as robust a partitive reading and, thus, should be used in experiments whose goal it is to assess the status of the quantifier without additional linguistic or contextual support (Halsey, 2010).

2.4.3 Live presentations and audio/video recordings

Some SI studies have used tasks in which critical stimuli are presented either in a live scenario or in an audio-video recording. This design has been particularly prevalent in the study of SIs in Spanish (Maatman, 2009; Miller et al. 2016; Syrett et al. 2017a; Vargas-Tokuda et al., 2009) but also found in other work (e.g., Guasti et al. 2005; Papafragou & Musolino, 2003; Politzer Ahles et al. 2013 among others) adopting a standard Truth Value Judgment Task design (Crain

& Thornton, 1998). Performing a task live, however, can bring in a new set of confounding variables, one of which is prosody (e.g., Fodor, 2002). It has been shown that vocally focusing or emphasizing a scalar item like *some* increases the likelihood that it will be interpreted pragmatically (e.g., Crain, 2012; Miller et al. 2005; Vargas-Tokuda et al. 2009). Consider the following scenario of a live presentation quite common to Spanish SI work: an experimenter has four horse dolls, moves all four horses over a fence, then asks a participant if the statement *some horses jumped over the fence* is true. It is true in this case that some (a set of) horses jumped over the fence, thus fulfilling *some*'s semantic conditions. It would then be up to the participant to respond either logically or pragmatically. Recall that either response is semantically appropriate, but one is pragmatically infelicitous. However, if the experimenter were to ask if *SOME horses jumped over the fence*—where SOME indicates vocalized emphasis—the reading is more clearly partitive.

If done with audio-video recordings, the experimenters can control prosodic factors and guarantee an equal delivery of stimuli to each individual participant by using software designed to manipulate audio recordings. However, by performing the trials live it would be difficult to ensure that all of the trials throughout the task were performed in exactly the same way, and much less so across all participants, irrespective of experimenter training. This leaves open the possibility that emphasis on certain words or disparities between the delivery and presentation of individual trials might favor certain judgments.

2.4.4 Priming the lexical scale

A considerable body of research has examined whether SI calculation occurs by default or whether it occurs as the result of calculation over the broader propositional content of an utterance (Feeney, Scrafton, Duckworth, & Handley, 2004; Grodner, Klein, Carbary, & Tanenhaus, 2010; Huang & Snedeker, 2009, 2011; Levinson, 2000). A long-standing assumption has been that lexical-level processing is fast and more or less automatized; however, there is a growing body of research showing that SI calculation is not only very fast, but also flexible in certain contexts. For example, it has been shown that SI calculation can depend on whether listeners believe a speaker has no reason to be more informative (Bonnefon, Feeney, & Villejoubert, 2009) or if speakers are uncertain that a more informative statement is true or false (Bergen & Grodner, 2012). Research on the flexibility of SI calculation was recently questioned by Hartshorne et al. (2014) who highlighted that while the mechanisms responsible for SI calculation may be rapidly and flexibly utilized, they can still operate over a lexical representation rather than the whole proposition of an utterance. Hartshorne and

colleagues highlight that studies showing immediate sensitivity to contexts that result in implicature could have primed the lexical scale by almost always providing contexts containing *all* or a similar term such as *each*, while the contexts that did not support implicature never did so (e.g., Bergen & Grodner, 2012; Breheny, Katsos & Williams, 2006). Thus, if a participant sees *some* in an *all* context and *all* in an *all* context within the same task, the latter would likely stand out as the more felicitous option, pushing individuals to opt for the more conservative 'but not all' interpretation given the strong contrast between *some* and *all*.

2.4.5 Overrepresentation of English

Finally, to date, the vast majority of experimental work on implicatures has been carried out either using English or, to a lesser extent, another western language as the language of testing, such as German, Greek, Italian, and French, among others. While there is less experimental implicature work in Spanish, the focus has been more on children, where adults are used mainly as control groups (Syrett et al. 2017a,b; Vargas-Tokuda et al. 2009). As noted, Spanish offers an interesting test case for experimental work on implicature given its two indefinite scalar quantifiers as opposed to one like many other languages tested. Their unique distributional pattern could allow us to tease apart the comprehension and use of scalar quantifiers that are purely semantic in nature versus those that are discourse-dependent, where both share a similar general meaning, offering insight into the differnces between semantic and pragmatic processing (see e.g., Politzer-Ahles et al. 2013).

Examining other languages that do not work like English with respect to SIs has the potential to elucidate contemporary semantic and pragmatic theories, broaden our view of how these properties are dealt with more generally and inform, say, any instance of non-monolingual (e.g., bi-multilingual) language (Ln) acquisition /processing/etc., an area of research which has seen scant work concerning SIs.

2.5 Methodology

The present study employed two experiments. The first was a picture sentence verification task (PSVT) modified from Politzer-Alhes et al. (2012), Spychalska et al. (2016), Tavano, (2010) and Wu & Tan, (2009). This task examined offline judgments comparing pragmatically infelicitous or underinformative contexts to semantically correct ones via pictures. The second was an attempt to create a novel task that did not force individuals to dichotomize judgments as good or bad, but one that would allow for an explicit association of a specific numerical quantity to each of the quantifiers in pre-defined contexts.

Data for both experiments were collected from 30 native Spanish speakers (23 females, age range 21-47, M = 26.5, SD = 4.1). All participants were in the UK for work, school, or personal travel and were right handed (Oldfield, 1971). Prior to beginning the experimental tasks, participants filled out the LEAP-Q questionnaire to gather linguistic and biographical information (Marian, Blumenfeld, & Kaushanskaya, 2007). Two participants had been in the UK for one year, while the rest—on average—only 4.7 months at the time of testing. Thus, it is reasonable to suspect that none of the controls were in fact bilingual at the time of testing. Participants were recruited from a tightly knit group of Spaniards in the greater London area and were either very low proficiency in English or had none at all. Given the proximity of Spain to the UK, it is nearly impossible to find a Spaniard who has absolutely no experience with or exposure to English, either through schooling or the media; however, all participants herein were as inexperienced with English as one could hope to find, thus avoiding any issues with the recruitment of bilinguals given that such participants have been shown to treat scalar quantifiers in Spanish uniquely to monolingual participants (e.g., Miller et al. 2016). All participants were right handed as measured by the handedness questionnaire (Oldfield, 1971).

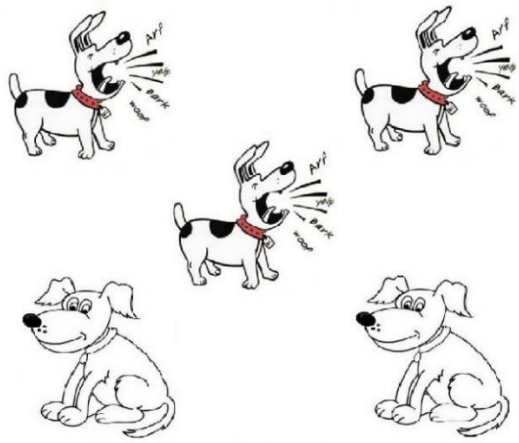
As participants did both tasks in one session, the order of the tasks was randomized between all participants such that half were administered Experiment A first, followed by Experiment B, and the other half were administered the experiments in the reverse order. After the testing session, a short debrief was carried out to ensure that participants were not aware of the specific aims of the study. All participants provided their informed consent. Ethical approval was granted by the University of Reading Ethics Committee.

2.5.1 Experiment A – Picture-sentence verification task

For each trial, an image was presented in which either five characters are all engaged in the same activity or a subset was engaged in one activity while others were engaged in another. Images were preceded by the carrier phrase *En esta imagen* 'In this image' and followed by a sentence that described the image correctly, incorrectly (semantics) or infelicitously (see Figure 1) by using the target quantifiers *algunos*, *unos* and *todos* 'all'. The critical quantifiers *algunos* and *unos* were presented in simple form without *de* 'of'. In each trial, the target determiner phrase (DP) was followed by a prepositional phrase (PP) such as "in the park" or "at the beach", for example, to ensure the critical DP did not appear in sentence final position. [6]

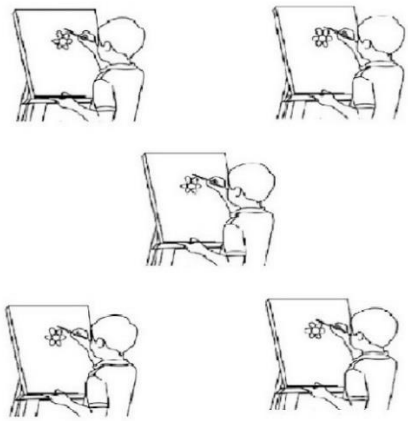
[6] You will note the non-canonical word order in the Spanish sentences: fronting the infinitive form (and therefore nominalizing) of the verb. This was done for the purposes of using these stimuli as part of our ERP experiment. Recent work on SIs using picture tasks such as this in an ERP design have noted that having the implicature generating item at the front (e.g., *some girls are sitting on blankets*) in rapid serial visual presentation display (i.e., one word at a time) does not offer individuals enough information in reference to the context provided in the images at the time the quantifier appears on the screen to make an

Algunos Felicitous



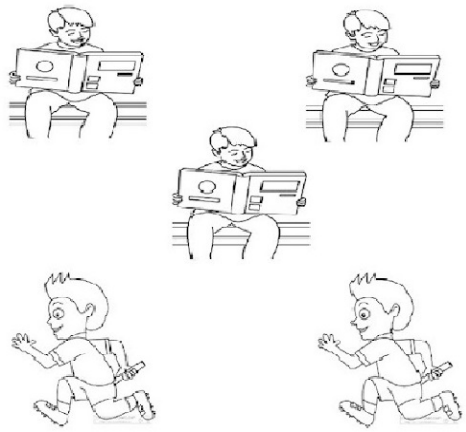
Ladrar es lo que hacen algunos perros
"Barking is what some dogs are doing"

Algunos Infelicitous



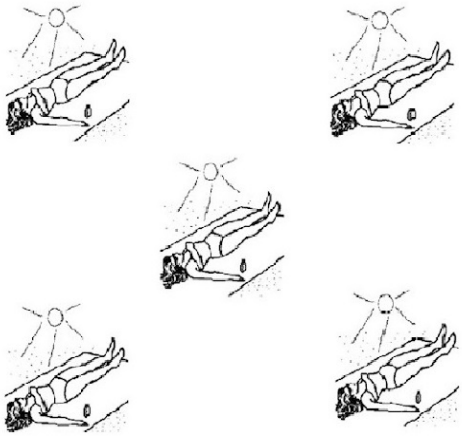
Pintar es lo que hacen algunos chicos
"Painting is what some boys are doing"

Unos felicitous



Leer es lo que hacen unos chicos
"Reading is what some boys are doing"

Unos Felicitous



Broncearse es lo que hacen unas chicas
"Sunbathing is what some girls are doing"

accurate judgment (Politzer-Ahles et al. 2013). Thus, here we place the verbal predicate at the front so that when the quantifier appears one has all the pertinent information to make a judgment. Participants performed various practice trials prior to beginning the experiment in order to get used to the nature of the non-canonical syntax used in the experimental trials.

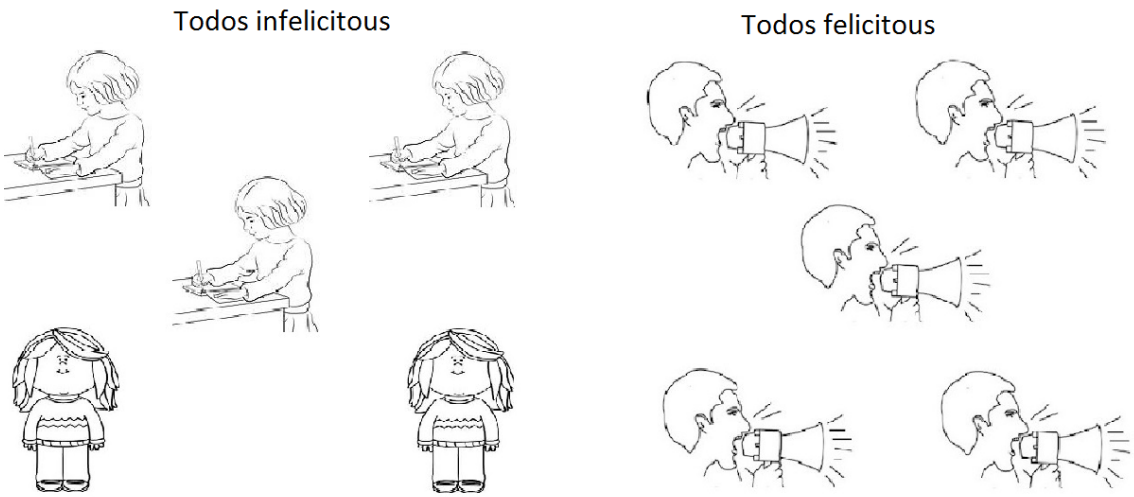


Figure 1: Sample pictures and sentences from Experiment A

Following the presentation of the image and the sentence, the participant was instructed to indicate with a computer mouse whether the sentence described the image well or poorly. To serve as a control within Experiment A, we opted to include both a correct and incorrect *all* condition. This provides a 3 (Quantifier) x 2 (Type) design where inconsistent and/or incorrect *some/all* conditions were compared with lexically matched controls.

This task takes into account the methodological considerations offered herein in the following ways: First, we are using a language other than English that, itself, has seen less experimental research with respect to SIs. In fact, of the available studies on SIs in Spanish, this is the first to focus on monolingual adults in an effort to have a truly accurate and in-depth baseline description for how SIs are computed under experimental conditions in monolingual Spanish. As mentioned, the studies examining Spanish in monolingual child development or in bilingualism indeed use adults as controls. However, the control data—for obvious reasons—is not the focus of the discussion and as such inter-individual variation in native adult monolingual populations has not been properly unpacked, contextualized or meaningfully discussed in Spanish despite being present in many other studies. Second, this task does not rely on world knowledge or plausibility as the target quantifiers are being evaluated only against the images within the task. Third, we used only the quantifiers themselves rather than *algunos de/unos de* 'some of' in order to avoid biasing a pragmatic response. Fourth, we did not present the stimuli in an auditory or live fashion so as to avoid any possibility that explicit prosody or differences across presentations would affect the responses. It is worth noting that

participants are still free to assign an implicit prosody to the target sentences (Fodor, 2002). Finally, though we did have an *all* condition, this was done in order to compare Experiment A to Experiment B, the latter of which did not have an *all* condition.

2.5.2 Materials and procedure

The critical trials were created from 180 sets of black and white pictures arrays. An additional 180 sets of picture arrays were designed for the fillers. All images were obtained freely on Google and were modified with Microsoft Paint or Adobe Photoshop. It is worth noting again that *unos* can refer to whole sets and subsets given that it lacks semantic features that do not allow it to have a specific partitive interpretation. Thus, while *algunos* 'some' and *todos* 'all' have both a felicitous/true and infelicitous/false condition, *unos* has two conditions that are both felicitous. To make *unos* infelicitous, one would have to create a picture context in which none of the characters are performing the target activity or only one of them is and the others are not. Because Spanish expresses gender and number morphologically, however, if the image were to mismatch the target sentence, the latter would necessarily contain morphological cues that would render the context true or false, thus becoming a grammaticality judgment and not a truth value judgment. Furthermore, ungrammaticality for these conditions would become detectable at the onset of the preceding verb due to its plural morphology. Pilot data for this experiment, which contained such trials, did not show modulation of overall offline performance by these trials; given that they ultimately would have served as a control condition for *unos* only, showing that it cannot be 'one' or 'zero', they were removed.

Filler items required participants to judge the sentences against pictures in which the location of one or more objects either matched or did not match the sentence itself. In this respect, we avoided using quantifiers for fillers while still remaining consistent regarding task type. Each of the critical picture sets included five characters performing one activity or another, providing contexts in which each quantifier would be both felicitous and infelicitous. 30 trials were created to form each of the critical conditions, which were divided as in Table 1. Each quantifier was used in each context of felicitousness with a unique picture, never appearing more than once for any one condition. In other words, infelicitous and felicitous *algunos* conditions shared the same sentence, though an all-type picture was used for infelicitous conditions and a Some-type picture for felicitous conditions. However, unique predicates were used for each of the other quantifiers such that no sentence or image was ever repeated in the counterbalanced trials throughout the task. We designed the experiment this way in order to control for the possibility of associative comparisons across trials. That is, were

a participant to see a picture and sentence for *algunos* in infelicitous conditions *unos* felicitous conditions within the same task, she may more easily develop a response pattern based on "better" options and/or discover the property under examination, especially given the nature of forced judgment tasks (e.g., Katsos & Bishop, 2011).

Condition	Algunos	Unos	Todos	Filler
Felicitous/True	n= 30	n= 60	n= 30	n= 90
Infelicitous/False	n= 30	n/a	n= 30	n= 120
Total	n= 60	n= 60	n= 60	n= 210

Table 1: Total number of conditions

Seven lists were created for the presentation of the stimuli. We obtained the lists by counterbalancing the order in which the stimuli appeared and in which form (felicitous or not). Each list began with one of the target condition sentences and the remainder of the list was presented in a randomized fashion.

2.5.3 Results Experiment A

In order to examine whether the experimental group differentiated between the four target conditions in similar ways, data were analyzed with a binary logistic regression model and pairwise contrasts with Bonferroni correction where appropriate using SPSS. Within the mixed model, the Satterthwaite Approximation was used to get the degrees of freedom in order to run F-tests on the individual variables. The dependent variable for the experiment was *acceptance* (bien 'good' versus mal 'poor'). The model also included variables of Condition (partitive versus whole), Word (*algunos* versus *unos*) and the interactions of these variables. Predictor variables were numerically coded within the model, being assigned a unique variable number to distinguish them from the others. The random effects structure was the maximal structure supported by the data and included random by-subjects intercepts and slopes (e.g., Barr, Levy, Scheepers, & Tily, 2013).

Figure 2 provides an overview of the mean acceptance for each condition. The model revealed significant main effects of Word ($F(2,5932) = 18.24$, $p < .001$, Condition ($F(1,5932) = 191.41$, $p < .001$, and a Word*Condition interaction ($F(2,5932) = 660.01$, $p < .001$. Additional pairwise comparisons revealed that the effect of Word was mainly driven by the *algunos*/*unos* distinction such that *unos* was markedly more acceptable in whole conditions (*M*

= .542, SE = .048) than *algunos* (*M* = .129, *SE* = .021) estimate =.413, *t*(5,932) = -11.14, *p* = .003. Furthermore, *algunos* partitive conditions were significantly more acceptable: estimate = .068, *SE* = .027, *t*(1,747) = 2.6, *p* < .001. This provides evidence that a distinction is not only being made between the the two lexical items themselves, but crucially that the (in)felicitousness of *algunos* in partitive versus whole contexts is systematically affected by the context of the pictures.

However, Figure 2 also interestingly shows that the whole *unos* conditions are not being accepted as often as one might expect given that *unos* is said to be equally felicitous in partitive and whole contexts. This discovery was followed up with additional post hoc analyses such that the data were divided between participants who responded "logically" (*some and possibly all*) by accepting *unos* in whole contexts and those who responded "pragmatically" (*some but not all*) by rejecting *unos* in whole contexts. The division was calculated on the basis of any given individual showing a majority preference of one of the two *unos* conditions. Though surprising that such a distinction is being made on *unos*, a logical/pragmatic split in responses is consistent with findings from previous research in other languages (e.g., Bott & Noveck, 2004; Hunt et al., 2013; Noveck & Posada, 2003; Politzer-Ahles et al., 2013; Tavano, 2010 among others), showing clear inter-individual variation between "logical" and "pragmatic" responders. While in the present study there were some pragmatic individuals who accepted *unos* in whole conditions, as a whole they largely rejected them as compared to the logical individuals (see Figures 2 and 3).

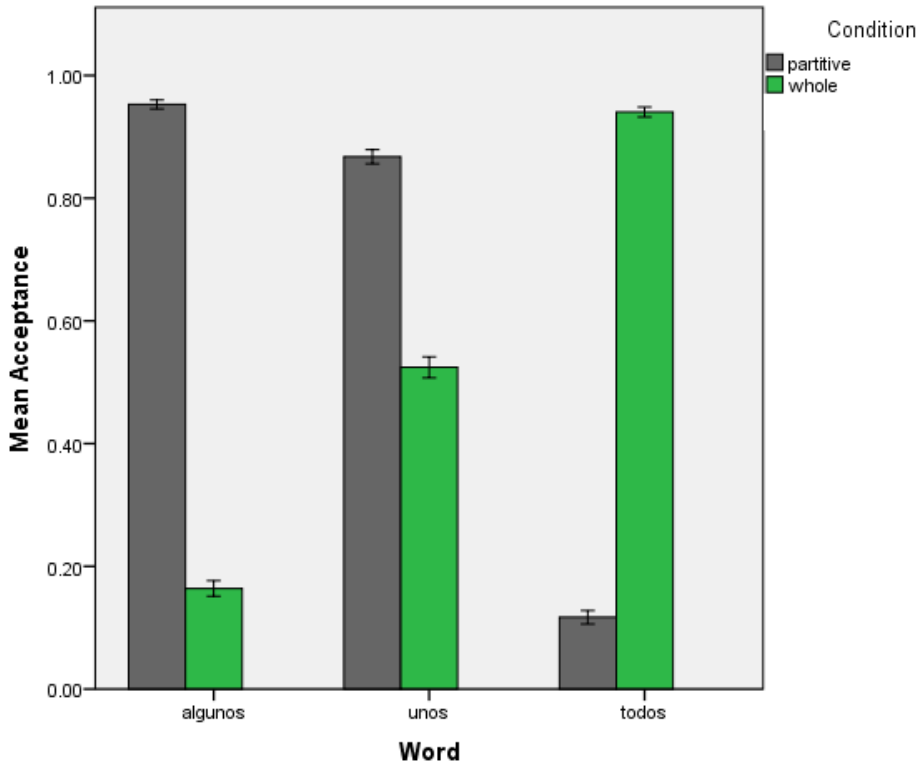


Figure 2: Percent acceptance of target conditions (+/- 1 SE)

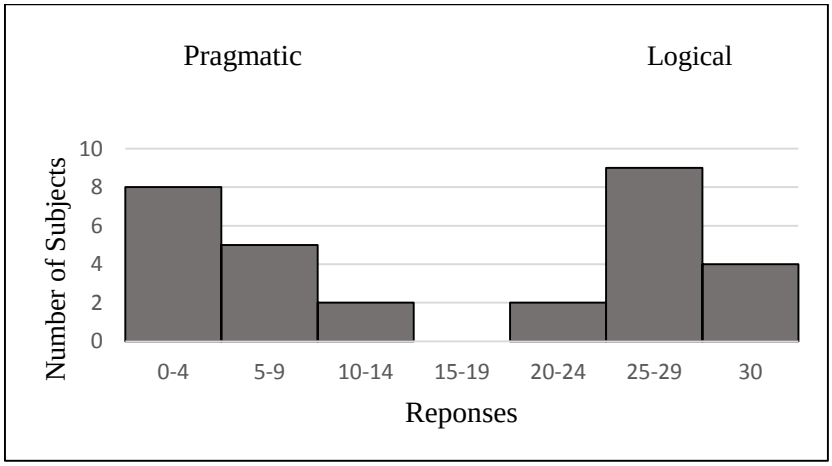


Figure 3: Histogram of logical versus pragmatic responses on unos

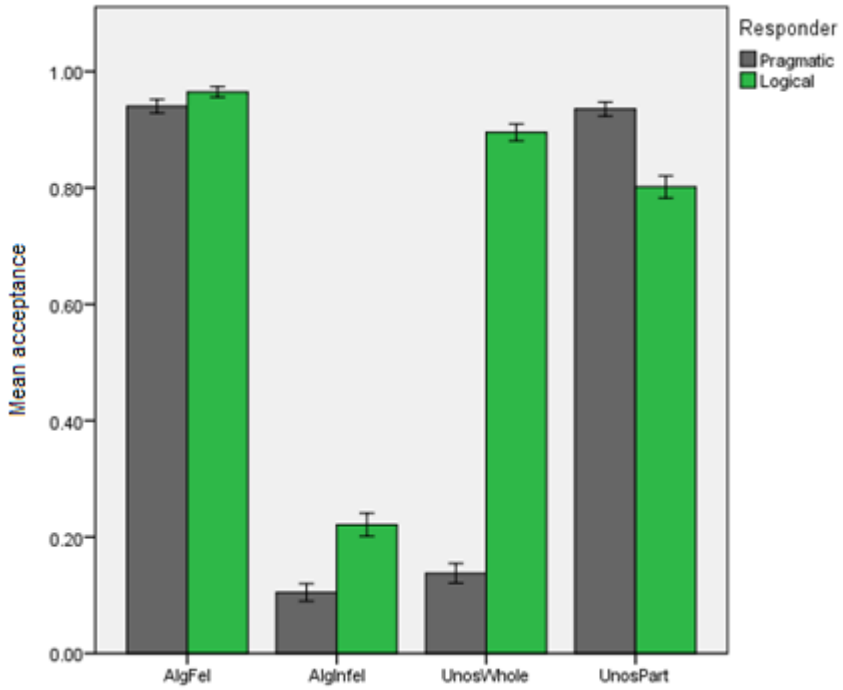


Figure 4: Pragmatic versus logical responders' acceptance rates (+/- 1 SE)

It is interesting to note that the distinction between *algunos* in whole contexts and *unos* in whole contexts is lost within the pragmatic responder group, indicating that for these individuals both quantifiers are equally infelicitous.

2.5.4 Experiment B – Non-binary free interpretation task

Deriving a pragmatic inference can be manipulated by experimental conditions. In many SI studies, participants' responses are manipulated purposefully by instruction (e.g., Barbet & Thierry, 2016), enriched context (e.g., Guasti et al. 2005; Papafragou & Musolino, 2003), high cognitive demands (e.g., Marty & Chemla, 2013), prosodic focus (see Crain, 2012 for overview) and by the limitation of response time (Bott & Noveck, 2004, De Neys & Shaeken, 2007). In these cases, however, one does not know how a participant would interpret SIs in a more neutral environment without such experimental manipulations and when they are free to give a non-coerced, non-binary judgment and explicit judgment. In order for a comprehender to derive an SI without any bias or pressure from task demands, she must not be primed by any explicit methodological factor within those tasks. She must make her judgment as she would in as close to a normal context as possible. With this in mind and following up from Experiment A, we designed a task in an attempt to avoid such task bias.

Within the task, each condition defines a set of characters and asks the participant to choose all possible quantities from 0-*X* (where *X* is pre-defined) associated with the scalar term

used to refer to the pre-defined set of characters. For each trial, the participant was provided with a short context containing one or the other scalar term and was asked to choose all possible quantities associated with the scalar term given in the context (see Figure 6). Both *algunos* and *unos* were used in conditions in which there were four, five, or six possible referents in order to determine if Number differentially affected the overall interpretation of each scalar term (e.g., Degen & Tanenhaus, 2011, 2015). It should be noted here that the 4 out of 4 condition falls within the subitizing range (i.e., 1-4), but the other conditions reflect a set size for *some* to be more naturally acceptable as meaning 'not all'. Thus, we might expect that individuals would differentially treat the scalar terms based on the set size to which they refer, particularly the smaller set size as compared to the larger ones. It would be preferable to use even larger set sizes; however, a pilot study revealed that when—in a task such as this—participants are free to tick any number within the answers, the more possible answers there are leads to more time involved in the experiment. Given the length of Experiment A, it became prudent to minimize the set sizes with Experiment B while retaining the distinction between the subitizing and larger ranges.

In order to remain consistent across both filler and target trials, all answers contained the values between '0' and the highest possible number within the set. Because '0' and '1' were options in some of the filler trials, they remained so in the target trials so as not to draw any attention to differences between targets and fillers even though they were not felicitous options.

Word	**Four entities**	**Five entities**	**Six entities**
Algunos	n = 3	n = 3	n = 3
Unos	n = 3	n = 3	n = 3

Table 2: Condition types for Task B

Gloria fue a una fiesta con unas amigas. Si Gloria tiene 4 amigas, con cuántas amigas posiblemente haya ido a la fiesta? Por favor, escoge todas las posibles respuestas.

0 ☐
1 ☐
2 ☐
3 ☐
4 ☐

Figure 5: Sample of Experiment B Unos condition

In Figure 5, the context is translated as follows: "Gloria went to a party with *unas* 'some' friends. If Gloria has four friends, how many friends possibly went to the party? Please choose all possible answers."

This experiment also takes into account the methodological considerations in section 3 in the following ways. First, we do not prime the context with already existing infelicity, allowing the participant to make an unbiased judgment. Second, we used the quantifiers by themselves without *of* or 'de'. Finally, participants had the opportunity to decide themselves what the scalar term meant in a given context as it relates to quantity. In other words, a participant could calculate—or not—freely the SI associated with the scalar term in each context.

2.5.5 Materials and procedure

To make the target contexts in this task, we used each scalar term three times in unique contexts of four, five or six characters performing an action, totalling nine targets for *algunos* and nine for *unos*. We also created 36 unique fillers to give a ratio of one target to two fillers. Fillers were made with similar short contexts, but rather than using a scalar quantifier, we made simple arithmetic problems such as "John needs to write 35 letters by Friday. If today is Monday, how many letters must John write each day until Friday? Please choose the best answer(s)." In this respect, though we did not ask for judgments of scalar quantifiers per se in the fillers, we did manipulate quantification as a variable in order to maintain a parallel between targets and fillers and to further mask the nature of the experiment. This experiment was completed using SurveyGizmo on a laptop with internet in the laboratory where Experiment A was administered.

2.5.6 Results Experiment B

Data for Experiment B were also analyzed with a binary logistic regression model and pairwise contrasts with Bonferroni correction where appropriate using SPSS. For the purposes of making inherently non-binary responses amenable to a binary regression, those responses which constituted a partitive-type judgment were assigned one number (i.e., 1) while those responses that were whole-type were assigned another number (i.e., 2), thus collapsing raw responses into binary responses within the statistical model. The dependent variable for the model was Response Type, this being either partitive or whole. Furthermore, we included variables of Word (*algunos* versus *unos*), Condition (*x* of 4, *x* of 5, *x* of 6), ResponderType (logical or pragmatic) and all higher-order interactions of these variables. As with Experiment A, the predictor variables were coded by assigning each a unique number within the model.

Figure 6 presents the mean responses for all conditions without dividing the participants based on response type.

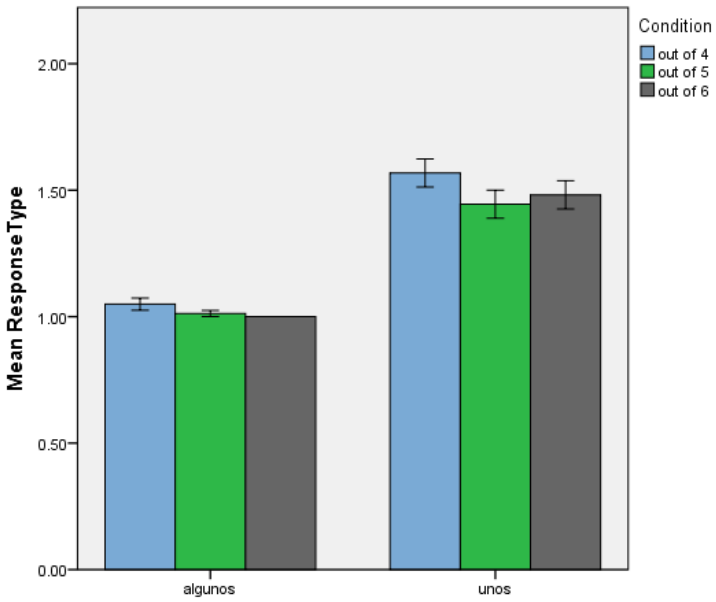


Figure 6: Mean response for all conditions (+/- 1 SE)

Descriptively, these data show that while *algunos* is largely interpreted as partitive, *unos* shows more variation, similarly to Experiment A. Figure 8, on the other hand, shows mean responses for word between "logical" and "pragmatic" responders, a division which was done only after analyzing data from Experiment A where the usefulness of such a division was first noticed. This will be discussed in more depth in the following section. However, suffice it to say for now that the division of pragmatic and logical responders at the individual level overlapped completely across the two experiments such that all and only those who demonstrated a "pragmatic" or "logical" response bias in Experiment A also did so in Experiment B.

Furthermore, because Spanish, like English, has specific words for *a pair* or *a few*, we might have expected that—particularly in the 4 out of 4 contexts—'2' and '3' would be avoided, forcing participants to choose '4'. In this case, we would have seen an over acceptance of *algunos* in whole 4 out of 4 conditions, but we did not. Participants reliably judged *algunos* as felicitous in subset contexts and infelicitous in whole set conditions. Figure 7 provides the frequencies for individual responses in each of the relevant conditions across words.

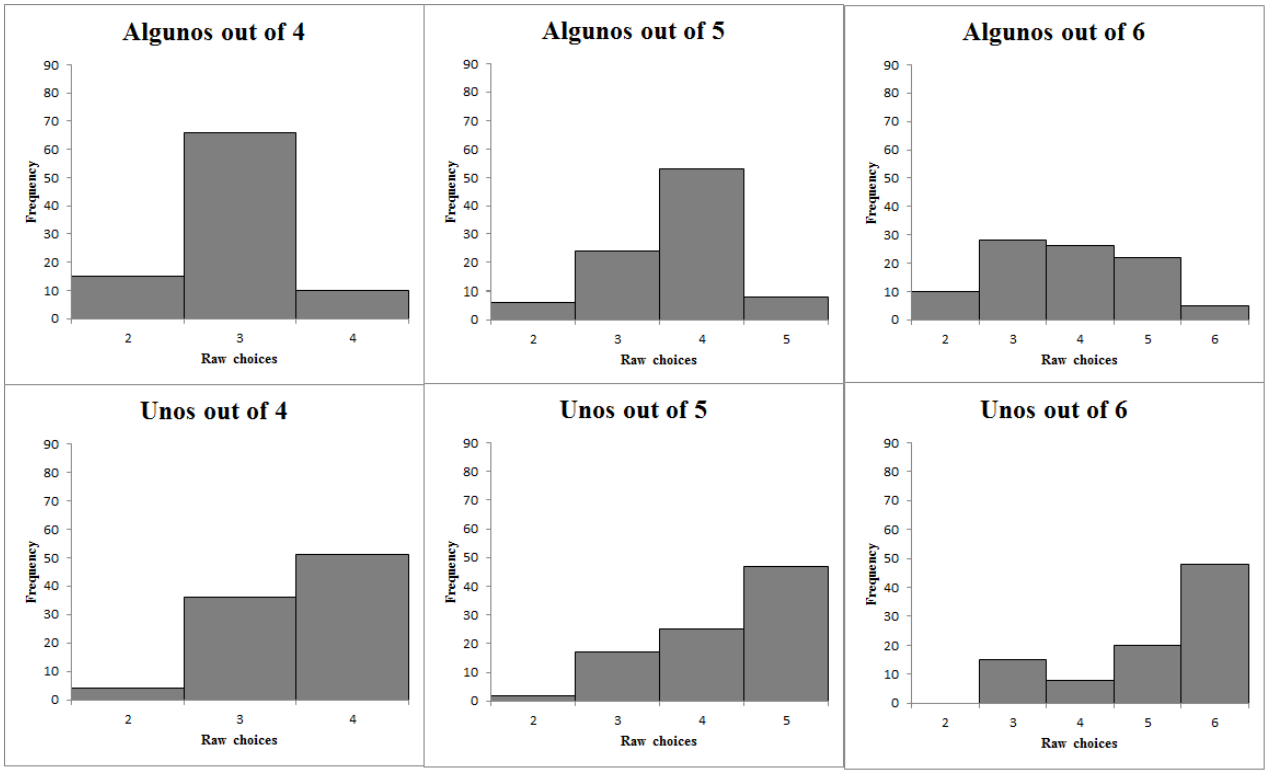


Figure 7: Histograms of raw responses for all conditions

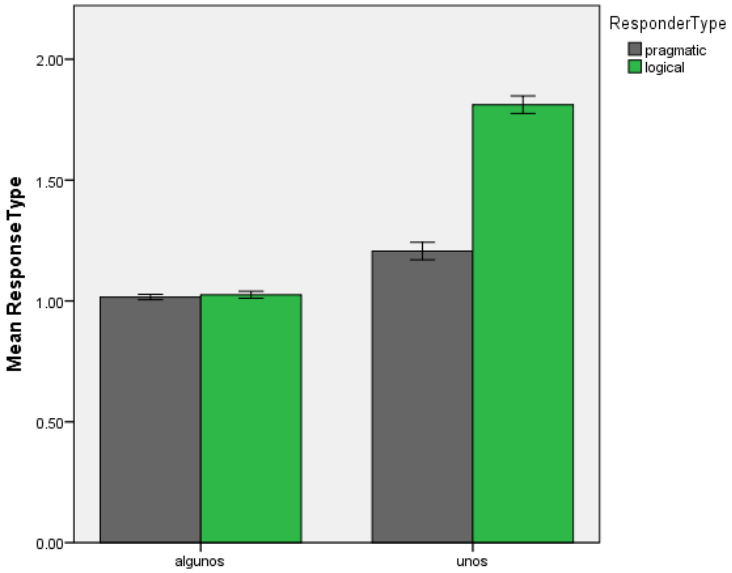


Figure 8: Mean response for words between responder type (+/- 1 SE)

The model revealed no significant main effects other than a Condition*Word*ResponderType interaction: ($F(6,474) = 7.05$, $p < .001$. Further pairwise contrasts crucially show that there is a distinction between *algunos* and *unos* as far as acceptability in certain contexts. While none of the *algunos* conditions is different from the other, each is significantly different from its counterpart *unos* conditions such that *algunos* is more clearly partitive and *unos* is either partitive or whole: *algunos* out of 4 to *unos* out of 4: estimate = .289, $SE = .113$, $t(470) = 3.10$, $p = .019$; *algunos* out of 5 to *unos* out of 5: estimate = .317, $SE = .088$, $t(470) = 3.6$, $p = .007$; *algunos* out of 6 to *unos* out of 6: estimate = .437, $SE = .104$, $t(470) = 4.2$, $p = .001$.

2.6 General discussion

This paper was concerned with investigating whether the observed variability in SI calculation among (adult) native speakers of Spanish—and potentially other languages and/or other subject populations more generally (e.g., children, bilinguals, attriters)—can be reduced by specifically controlled experimental designs and, if so, what the ramifications would be for previous and future research. Results from both a picture-sentence verification task and a non-binary free choice interpretation task provide evidence that individuals reliably derive SIs in Spanish without some of the aforementioned tasks biases. Additionally, some of the variation in individuals' interpretations can be reduced when subjects are divided into categories of responder type based on pragmatic and/or logical responses. It is worth noting, however, that

unlike, say, English, that only has one word for *some* thus leaving its interpretation ambiguous, Spanish presents a case where there are specific lexical properties of the scalar quantifiers under examination that govern their use in context. It is claimed that *algunos* is discourse linked and *unos* is not, which is the primary characteristic that leads *algunos* to be associated with a quantity implicature. In this respect, one might expect less variation in Spanish than in English concerning the interpretation of *algunos* and *unos*. However, previous research has demonstrated that Spanish *some* is just as susceptible to variable interpretation as English *some*. I should clarify that while the present study has shown, as compared to previous reports on adult SI interpretations in Spanish, that the variable interpretation of *algunos* has decreased, a novel finding also arose with respect to *unos*. In other words, it is the ambiguity of *unos* that appears to be driving the variability in its interpretation, where some individuals are logical and some are pragmatic. In the case of *unos*, thus, we have not seen a decrease in variability, rather an increase, though the difference is categorical.

Different from previous investigations on SI calculation in Spanish (see Miller et al. 2016 for adults; see Syrett et al. 2017b for bilingual children; see Vargas-Tokuda et al. 2009 for monolingual children and adults), our subjects reliably distinguish between *algunos* and *unos*. In Experiment A, images denoting all-type contexts used with *algunos* are much less acceptable than their *unos* counterparts in the same contexts. That is, *unos* is more compatible with *all* than *algunos*.

Results from Experiment B showed much of the same. Participants reliably distinguished between *algunos* and *unos*, even without an *all* condition to potentially prime the lexical scale. Overall, acceptance of *algunos* in whole contexts in Experiment B are less frequent than in Experiment A. Moreover, recall that Experiment B allowed participants to make their own determination as to the quantifiable meaning of each scalar term without systematically manipulating felicity or infelicity via pictures. In other words, Experiment B did not examine what happens when a sentence mismatched prior knowledge, rather it tested what the maximum quantity each scalar term allowed for when the reference set was predefined. Results indicated that participants almost never equated *algunos* with *all*. However, the logical/pragmatic division based on responses to *unos* conditions still arose.

The balance between logical responders and pragmatic responders in these data accounts for much of the variation within the individual responses. It is important to remind the reader that this division was made on the basis of individuals remaining consistent in their responses both within and across the experimental tasks with minimal noise. Such a division has been revelatory in at least two ways, the first of which is the possibility that the description of *unos*

in the theoretical literature may not tell the entire story as we can see a clear divide in the way it is interpreted. The second is that *algunos* also may need further clarification given that what gives rise to its so-called quantity implicature is in fact contained within its lexical semantics.

The scalar quantifiers of relevance in Spanish, unlike English, are said to be constrained by specific features or morphemes (e.g., [+/- partitive], *alg*) that make them either partitive or not. So, while *algunos* is the said to be the item associated with the quantity implicature, *unos* can refer to partitive and whole contexts. While it is claimed that *unos* may not behave like *algunos* in that the 'not all' meaning cannot be cancelled in DE contexts, for example, in our data there is minimally a clear interpretation bias such that any one individual can interpret it as either whole or partitive without further constraining it by contrastive focus as claimed by Gutierrez-Rexach (2001, 2003). Irrespective of specific linguistic features that are said to constrain the interpretation of *unos*, in these data it is being treated in a manner than is inconsistent with the theoretical literature. Moreover, if what gives rise to the implicature for *algunos* is in fact a semantic feature, it may not be an SI generating item, rather *unos* seems a better candidate. Future research should consider contrasting these quantifiers in both upward and downward entailing environments to examine when the SI does and does not arise for both.

Finally, in our data infelicitous *algunos* is widely rejected by both types of responders but more so by pragmatic responders. However, the division of participants provides evidence that it is more likely that a generally logical individual will treat *algunos* and *unos* as compatible with 'all' and a pragmatic individual will treat them as meaning 'not all'. When viewed as a whole, our participants performed only marginally "better" than those of previous investigations on Spanish scalar terms. When divided based on individual responses, however, the data become clearer in that the partitive term *algunos* becomes much less compatible with *todos* 'all'. This pattern suggests that, to further answer the question posed by Guasti et al. (2005) of why children and adults sometimes (but not always) compute SIs, simply put (potentially both children and) adults seem to be intrinsically variable in their interpretation of scalar terms, even when the tasks control for confounding variables. The fact that both populations are able to derive pragmatic meanings of scalar terms with and without additional support—though performance improves with support—goes to show that the answer might very well be as simple as *sometimes*: sometimes children calculate SIs and sometimes adults do, too. And sometimes their performance increases in some situations and decreases in some others. In contrast to Miller et al. (2016), we have shown that Spanish speaking adults reliably calculate the pragmatically enriched meaning associated with *algunos* in Spanish. We believe

and will maintain that controlling specific methodological variables was crucial to such a finding.

In sum, the question has never been whether or not scalar terms can be ambiguous, rather it is how and why certain individuals calculate implicatures in real time and others do not when they are presented with the same contexts. Our overall goal is to understand pragmatic inferencing and how it develops, but before attributing psychological or cognitive factors to observed variation, we need clearer parameters for testing SIs experimentally that reliably tap representation while observing how ambiguous domains of language unfold for comprehension. If careful experimental design can reduce some of the variability in the interpretation of scalar terms in Spanish, then we know that, minimally, said variability cannot be characterized in its entirety by a cognitive or developmental deficit in any one population.

2.7 Conclusion

We find that task design and task type can affect monolingual adult Spanish speaker's interpretation of the scalar terms *algunos* and *unos*. In particular, we find that while the pragmatically enriched interpretation is the most likely one for *algunos* in partitive contexts, especially among pragmatic responders, such an interpretation is not impossible with *unos*, contrary to the theoretical description of these terms. Furthermore, we maintain that a careful examination of individual responses can affect the interpretation of the results given that some individuals seem to be inherently more semantic in their responses and others more pragmatic, a finding that may prove insightful when examining other populations (i.e., children). We do not challenge or support any psycholinguistic theory as to the nature and/ or mechanisms that underlie SI calculation, rather we provide evidence that subtle properties such as these first require appropriate methodological design in order to reliably tap individuals' interpretation and /calculation of SIs more generally.

Chapter 3: Article 2

Abstract

This study investigates the interpretation and processing of scalar implicatures (SIs) in native L1 Spanish using offline judgments and event-related potentials (ERPs) via a picture-sentence verification task. We examine *algunos* and *unos* which ostensibly translate to 'some' in English, but have distinct distributional patterns based on contextual restrictions. *Algunos* gives rise to a contextually induced or pragmatically enriched quantity implicature and *unos* does not. A great deal of psycholinguistic work has examined the processing and interpretation of scalar quantifiers in a number of languages and has shown that their associated SIs arise due to distinct methodological and linguistic factors, and have been shown to correlate with general cognitive functions. In spite of copious SI work, however, relatively little explanation (other than the variables already mentioned) has been put forward for why some individuals seem to more readily derive SIs and others do not, all things being equal. Because *algunos* and *unos* give rise to distinct interpretations, the former being constrained by semantic features and the latter not, previous work has shown that in Spanish there is no significant inter-individual variation regarding SI derivation. So-called "logical" and "pragmatic" responses arise only because *algunos* is contextually restricted and, therefore, pragmatic while *unos* is not, providing a natural dichotomy independent of interlocutor interpretation. Our data, however, reveal that *unos* is differentially interpreted based on context in much the same way *algunos* is but that each gives rise to distinct ERP signatures. Our data prompt semantic theory of Spanish scalar quantifiers to consider inter-individual variation regarding the interpretation of *unos* and show the benefit of using offline and online measures simultaneously.

Keywords: ERP; scalar implicature; pragmatics; logic

3.1 Introduction

Even when we "know" the lexical meanings of specific words, their interpretation in any given context can be remarkably variable (e.g., Lappin, 2003). For the most part, this is because meaning to sound correlations (semantic mappings to particular overt morpho-phonological forms) are arbitrary. Speakers intend the words, phrases, and expressions they use to carry specific meaning, and the recipients of these messages are free to assign either a parallel or a divergent meaning to them. Thus, interpretation takes place on an individual level and independent of any intended meaning by the speaker. Pragmatic principles, however, guide conversation towards an ultimate understanding between what is meant and what is interpreted

as speakers cooperatively communicate. Still, such principles are not always followed (Grice, 1975; 1989). Take for example the following exchange between two siblings:

1. Younger sibling: Can you pass the salt?
2. Older sibling: Yes (but crucially doesn't pass the salt).

Conventionally, *can* means 'to be able to' or 'to be permitted to'. The response (without action) in (2), then, is an acceptable one semantically speaking. However, language is not used in isolation of other factors that contribute to the richness of its overall meaning. A pragmatically more appropriate reaction from the older sibling, then, would have been to physically pick up the salt and give it to her younger sibling. Why else would someone simply ask if another individual is merely capable of passing salt?

Language users rely heavily on the context in which language is used in order to successfully convey and interpret their intended messages. However, not all interlocutors are equally aware of such factors nor do they always cooperate, which can lead to (un)intentional breakdowns in communication. In the above example, the older sibling is choosing to ignore pragmatics on the account of being logical, yet were she pragmatic, she would have simply given the salt to her sibling. This process by which a listener infers such a pragmatic meaning is called pragmatic inferencing.

Pragmatic inferencing as a domain of scientific and philosophical inquiry has a rich history and has been studied at a purely linguistic (i.e., features of lexical items themselves or the grammatical structure in which they sit that change meaning), philosophical (Grice1975) as well as a psycholinguistic level (see Noveck & Reboul, 2008). Across the board, research has paid special attention to certain quantifiers such as <*some, many, most, all*> in English whose meanings are subject to more than one interpretation on the basis of entailments and inferences. Using *some* as an example, most native speakers of English would be easily able to derive a meaning beyond *some*'s conventional semantics as in (3).

3. Owen likes some dogs. → *Owen does not like all dogs.*

The inferred meaning in italics is called a scalar implicatum[7], henceforth SI. It is scalar because it belongs to a scale of other similar quantifiers whose informativity is greater or lesser than that the other members of the scale, and implicatum or implicature because the inferred meaning is implicit rather than explicit (e.g., Grice, 1975, 1989, Horn, 1972, 1992). Several explanations of the nature of SIs and their derivation have been put forward over the years. One such explanation deals with whether or not SIs are derived automatically. On the one hand, default models (e.g., Levinson, 2000) propose that SIs arise any time *some* is encountered and, if needed, later reanalysis (i.e., cancellation) of the implicature takes place when additional information becomes available. On the other hand, context-driven models (e.g., Carston, 1990; Sperber & Wilson, 1995) assume that SIs arise only when the context is sufficiently appropriate for their derivation.

Furthermore, because *some* belongs to a set of quantifiers whose meaning is more or less informative than the other(s), the possible pragmatic meaning of *some* may be derived simply on the basis of scaled informativity (e.g., Chierchia, Fox, & Spector, 2012; Hirschberg, 1991; Horn, 1972; Levinson, 2000; Sauerland, 2012). For example, Grice (1975, 1989) proposed that speakers follow a set of communicative guidelines in order to be maximally cooperative during communication. These guidelines, specifically the Maxim of Quantity within what is known as the Cooperative Principle, dictate that interlocutors be as informative as needed, but not more so and say only what is relevant to the purposes, timing and direction of the exchange. In this sense, if a speaker chooses to use the weaker quantifier *some* as opposed to the stronger one *all*, there must be some reason for it, which is that *all* is not applicable. Thus, *some* is taken to mean *not all*.

Moreover, research has also shown that SIs arise by contributing to the truth-conditional content within an expression rather than the implied content (e.g., Carston, 1998; Horn, 1992) or as a result of lexical calculations (Levinson, 2000). However, there is also evidence that SI derivation does not take place merely at the lexical level, rather it is contingent upon higher-level structure whereby, for example, one does not expect an SI to arise when embedded in downward entailing (DE) or non-monotonic environments, both of which capture the defeasibility of inferences (Chierchia et al., 2012); or that scalar items optionally activate scalar alternatives that are subject to a recursive compositional procedure by way of an operator similar to *only* thus giving rise to the upper bound meaning (Chierchia, 2006). While there is

[7] H.P. Grice used *implicatum* to refer to the implied content of an utterance. Throughout this paper, however, we will use *implicature* to refer the implied content given post Gricean research conventions.

at least some evidence in support of most of these explanations regarding the nature of pragmatic inferences, the source of inter-individual variation in their interpretation—especially among adult populations—is unclear. However, as Katsos (2008) highlights, investigating the optionality of SIs can benefit from methodologies that are appropriate for examining the role of contextual informativeness in SI derivation, such as eye-tracking, self-paced reading or event-related potentials (ERPs).

Recent psycholinguistic work aimed at explaining SI optionality among individuals has shown that adults can be divided into distinct groups based on their treatment of *some* experimentally. In other words, some adults have been shown to accept *some* in underinformative contexts (i.e., logical responders) while others reject it (i.e., pragmatic responders) (e.g., Bott & Noveck, 2004; Hunt, Politzer-Ahles, Gibson, Minai, & Fiorentino, 2013; Noveck & Posada, 2003; Politzer-Ahles, Fiorentino, Jiang, & Zhou, 2013; Spychalska, Kontinen, & Werning, 2016; Tavano, 2010). While it has yet to be determined that such a dichotomy can be made between truly "logical" and truly "pragmatic" responders in the case of interpreting scalar quantifiers (all things being equal), it is clear that there is considerable variation in the way individuals treat them. The above research, thus, has at least partially catalyzed work examining both the cognitive differences among individuals and differences in task design that promote individual variation in SI derivation. Online investigations of adults' interpretation of scalar terms, for example, have shown that pragmatic inferencing is associated with working memory when cognitive load is increased (Marty & Chemla, 2013), a subset of scores on the Autism Spectrum Quotient (ASQ) (Nieuwland, Ditman, & Kuperberg, 2010; Zhao et al. 2015), scores obtained on the Systemizing Quotient-Revised (SQR) questionnaire (Barbet & Thierry, 2016) and even differences in methodological design across studies (e.g., Degen & Tanenhaus, 2011, 2015; Guasti et al. 2005; Papafragou & Musolino, 2003; Pouscoulous, Noveck, Politzer, & Bastide, 2007). It is worth noting that the implementation of cognitive exams like the ASQ in implicature work are not used as diagnostic metrics nor do researchers explicitly claim that so-called responders who fail to derive SIs are cognitively or pragmatically amiss, rather these metrics are used as tools to better understand why it may be that some people are naturally inclined to derive upper-bound meanings to scalar terms (which is a pragmatic language ability, making some of the ASQ relevant) and others are not, especially when provided with the same experimental contexts.

3.2 Aim of the present study

The effort to understand why some individuals do and some do not derive implicatures, particularly in contexts that prompt their derivation, is worthwhile and interesting. While there is a great deal of experimental research examining SI derivation in a variety of languages, there has been a strong tendency to examine English, with other western languages examined to a lesser degree such as Greek, Italian, French and Dutch, to name a few. By comparison, work examining Spanish-speaking adults is virtually non-existent. Spanish, given the nature of its scalar quantifiers, can help us better understand how natives actually compute and process SIs in a language that presents differently from most others that have been previously examined in the literature. What does exist for Spanish are results from adult Spanish speakers that serve as a control group within a child L1 study or a non-native second language L2 study (Austin, Sanchez, Syrett, Lingwall, Pérez-Cortes, 2015; Syrett, Austin, Sanchez, Germak, Lingwall, Pérez-Cortes & Baker, 2017a; Syrett, Lingwall, Pérez-Cortes, Austin, Sanchez, Baker & Arias-Amaya, 2017b; Vargas-Tokuda, Grinstead & Gutiérrez-Rexach, 2009). And so, it is fair to say this is the first study of its kind; a study specifically designed to understand the dynamics and intricacies of SI computation and processing for the sake of capturing the baseline of Spanish. Moreover, Spanish is a particularly interesting language to study in the case of SIs because it has two plural indefinite quantifiers instead of one (as does French, *certains* and *quelque*, see Pouscoulous et al. 2007): one with an associated quantity implicature whose meaning is constrained by sematic features and one without. Thus, examining Spanish SIs can shed light on the distinction between purely semantic processing/interpretation and that which must take place in consideration of pragmatics and discourse. Thus, it may be the case that what on the surface appears to be inter-individual variation can be explained in cases where the language provides quantifiers whose readings are constrained by semantic features rather than the freer interpretation of just one quantifier. In such a case, differential readings would not be expected to arise for one quantifier or the other in Spanish precisely because each relevant quantifier already has an assigned meaning due to the inherent semantic features constraining its interpretation. If there happens to be a clear division in Spanish speakers for one or the other term, inter-individual variation would not be as straight-forwardly explained on the basis of working memory, ASQ scores, methodology or other pragmatic language abilities alone; a closer look at the quantifiers themselves would be warranted.

Moreover, in light of research highlighting the advantages of testing linguistic properties using a variety of methodologies, particularly online and offline tasks, (e.g., Jegerski, Keating & VanPatten, 2016; Marinis, 2003; Villegas, 2014), we examine the processing and interpretation of SIs among adult native Spanish-speakers using offline

interpretation tasks and event-related potentials (ERPs), the latter offering implicit brain responses of real-time processing. Our purpose in using offline and online methodologies was to examine whether there would indeed be inter-individual variation among our speakers and, if so, whether the offline responses would be corroborated or brought in to question by the ERP data (Nieuwland et al. 2010; Spychalska et al. 2016). For example, Nieuwland and colleagues found that while adults exhibited a slow-down effect in reading times associated with underinformative sentences like *some people have lungs*, their offline acceptance of these sentences was quite high, revealing a disparity in the way the brain implicitly processed underinformativity and how individuals made conscious judgments of it offline (e.g., Nieuwland et al. 2010). Still, in this study the ERP responses corroborated offline judgments. Additionally, Spychalska and colleagues found a complete corroboration of ERP responses with offline judgments such that participants' categorization of "logical" or "pragmatic" was consistent for both data sets. Thus, as online methods may provide evidence for processing strategies and representations that offline methods, by their very nature cannot, using both together can significantly increase the probabilities that we are capturing not only what we seek related to comprehension and real-time processing, but also a potentially more accurate snapshot of the property of interest.

Finally, the majority of current SI studies in Spanish has demonstrated a partiality to examining offline speaker judgments among Spanish speakers either of the Mexican variety or a mixed pool of Latin American participants. It is of interest to the present study, therefore, to investigate a homogenous group of speakers of Peninsular Spanish, particularly among adults not forming part of a control group for a child L1 or adult L2 study. As we cannot and do not assume that all dialects of Spanish will work the same, having a homogenous group allows us to offer a baseline for how this particular dialect of Spanish operates and might reduce potential noise introduced in studies that have combined speakers of various dialects in their control groups. While there is nothing inherently imprudent about generalizing results from groups not meant to be the focus of investigation (the data are the data), it may limit what can be claimed. Different priorities must take precedence in studies that have distinct main target groups. For example, in child L1 studies the tasks are delimited by what a young child can do (length of test, type of test, etc.). Thus, methodologies with a focus on children as the main group are necessarily distinct from what one can do with a study that focuses on adults. And so, this study fills a gap in the literature as it pertains to monolingual Spanish interpretation and processing of SIs.

3.3 Scalar quantifiers in Spanish: Algunos and unos

Unlike most other languages tested to date, Spanish has two plural scalar quantifiers that share meaning with *some* in English: i) *algunos/as* 'some-pl' and ii) *unos/as* 'some-pl'. While both determiners must be a referent of a number of entities greater than or equal to two, sharing overlap in semantic distribution as shown in (4), they have distinctive pragmatic distributions (see (5a and b)).

4. *Algunos/unos niños están durmiendo.*
   'Some/some children are sleeping.'

5. a. **Context**: 3/5 children are sleeping:
   Algunos/unos niños están durmiendo.
   'Some/some children are sleeping.'

   b. **Context**: 5/5 children are sleeping:
   #Algunos/unos niños están durmiendo.
   'Some/some children are sleeping.'

For the purposes of this paper, the conditions in which *algunos* and *unos* demonstrate distinct behaviors are too great in number to explain each in detail (see Vargas-Tokuda et al. 2009 for review) though the experiment was designed to sidestep any potential confounding variables related to their differences, such as their differential relationships with verb class, predicate type, and contrastive focus. Below, however, we focus on the aspects of the quantifiers that are most relevant for the empirical study, that is, what has been argued to be the characteristic related to SI derivation that uniquely identifies them.

Gutiérrez-Rexach (2001, 2010) argues that *algunos* is compatible with a pragmatically enriched quantity implicature and *unos* is not. This distinction is argued to arise due to a "no linking" constraint on *unos*, whereby *unos* is not discourse-linked and *algunos* is. For example, though each quantifier can take a subset reading, only *unos* can refer to both a set and subset of referents as seen in (5). Due to the so-called no-linking constraint, *unos* should not, however, be able to relate back to a salient referent already mentioned in the discourse precisely because it does not have a specific/partitive interpretation as seen in (6).

6. *Los libros de matemáticas están en el cajón, los de física debajo de la cama hay ??unos/algunos de lingüística sobre la mesa.*

"The math books are in the drawer, the physics ones are under the bed and there are some linguistics ones on top of the table."

[Vargas-Tokuda et al. 2009, p.103, example (37)].

While *unos* can have a subset reading, this is most common in cases where its meaning is constrained by topic/focus marking characteristics of the utterance in which it appears, taking a 'some but not others' reading as in (7). *Algunos* does not require the use of any additional material to be interpreted as 'some but not others' as it is naturally partitive. To further elucidate the pragmatic status of *algunos*, (8) shows that the 'not all' interpretation associated with it can be cancelled or defeated—a hallmark of (scalar) inferences as opposed to semantic entailments—while the same is not true for *unos*.

7. *Unos chicos están nadando, otros no.*
   'Some kids are swimming, others are not.'

8. *Salieron algunos/#unos huelguistas, de hecho todos.*
   'Some protesters left, in fact all of them did.'

Martí (2008) offers a slightly different explanation, positing that the difference between the two quantifiers is more related to their unique morphology. *Algunos*, for example, can be broken into two parts: *alg-* (derived from *algo* 'something'), and *unos*. In this respect, *alg-* is simply added to *unos*, thereby inheriting the relevant features of *unos* (i.e., [+quantifier] [+indefinite] [+plural]) to disambiguate partitive and whole readings. Again, *algunos* is the determiner that gives rise to the quantity implicature.

Moreover, Alonso-Ovalle & Menéndez-Benito (2002, 2010, 2013) hold that both the pragmatic (de Hoop, 1992; Diesing, 1992; Partee, 1989) and the ambiguity approaches (Buring, 1996) to explaining the behavior of weak determiners are needed to explain the differences between *algunos* and *unos*. For example, they state that *unos* can be presuppositional, but in order for it to take this reading it must mark a change in topic or induce contrastive focus, such as in example (7) (also true in cases of prosodic focus). On the other hand, they state that *algunos* is both presuppositional and non-presuppositional and that neither reading is at all

dependent on the topic/focus marking characteristics of the sentence in which it appears (just the context). Again, *algunos* is the item associated with an implicature and *unos* is not.

Semantic theories of *algunos* and *unos*, however, have not predominantly relied on experimental data for their conclusions. In large part, theoretical descriptions of Spanish SIs rely heavily on intuition and informal native speaker judgments, and most of the experimental work that does exist has collected data from Spanish speakers of the Mexican (or Latin American) variety. Interestingly, young (monolingual) children seem to appropriately distinguish *algunos* from *unos*—especially with supportive contextual cues (Vargas-Tokuda et al. 2009; Syrett et al. 2017a,b for bilingual children)—while adults have shown a slightly higher degree of variation. That is, at least one study has shown that adult native speakers of Spanish readily accept *algunos* in whole contexts and others do not, though in that study the controls were native L1 Spanish-late L2 English bilinguals (Miller et al. 2016).

With the exception of some slight variation across individuals and populations, however, Spanish has thus far shown no evidence of optionality in an interpretation preference and, therefore, no "logical" or "pragmatic" dichotomization of responders. Such a finding is sensible given that *algunos* is said to always be partitive due to its morphology or inherent semantic features constraining its meaning—except in certain contexts such as downward entailment relations—and *unos* is either partitive or whole, effectively creating a natural distinction between "logical" and "pragmatic" readings via the quantifiers themselves and not necessarily due to inter-individual interpretative variation. Crucially, available data in the Spanish experimental literature show that *unos* is treated as either partitive or whole in equal measure with no apparent preference for one reading over the other, supporting the ubiquitous claim that *unos* is not associated with a pragmatic implicature.

3.4 ERP research and SIs

In ERP research examining SIs, the N400 has played a large role. Given this component's association with content stimuli, specifically semantic (im)plausibility and/or its more general correlation to prediction or (un)expectedness, probing for an N400 effect has featured prominently in both early and more contemporary ERP studies examining underinformative contexts with scalar quantifiers (e.g., Nieuwland et al. 2010; Noveck & Posada, 2003). In these studies, participants make acceptability judgments of underinformative sentences which are based necessarily on their knowledge of the world and violations of this type have been shown to be highly associated with more robust N400 effects (see Hagoort et al., 2007; Hald,

Steenbeek-planting, & Hagoort, 2007). Thus, if *some* is interpreted with its upper-bounded meaning, predicates containing underinformativity violate one's expectations based on world knowledge and should elicit more robust N400 effects than predicates without such a violation. However, caution is given not only to the overall interpretation of emerging N400 effects in these studies, but also to methodological design given that the N400 has also been shown to be modulated by the lexico-semantic relationship of words, specifically a target word and the main noun phrase in a given sentence. This relationship is generally measured with respect to the frequency with which words co-occur and is referred to as the *latent semantic analysis* value or LSA (Landauer et al., 1998). In essence, while larger N400 effects are expected in designs that employ underinformative *some* in world knowledge contexts compared to informative ones, this effect may be attenuated by the LSA value of the main noun phrase and the target predicate in informative contexts whose noun phrase and predicate have a stronger lexico-semantic relationship. Attenuation can also be found in universally false sentences such as *some crows have radios* (see Noveck & Posada, 2003).

While somewhat different stimuli have been used to examine SIs using ERPs, most of these studies have found an N400 or an N400-like effect such as the Nref (negative reference), which is a sustained late negativity typically distributing over anterior—as opposed to posterior—electrodes (e.g., Politzer-Ahles et al. 2013). However, a couple of studies have found positivities as responses either to pragmatically infelicitous contexts (see Spychalska et al. 2016 for frontal positivity) or contexts in which there is no apparent mismatch between a target sentence and prior information, rather the more precise moment an SI is derived in a phrase downstream of the quantifier (Harthshorne et al. 2015).

3.5 The present study

3.5.1 Research questions and predictions

The research questions that guide the empirical study are as follows:

1) Do native Castilian Spanish speakers distinguish between felicitous and infelicitous *algunos* conditions? If so, will this effect be found online, offline or both?

   Prediction: In accordance with previous offline work on SIs in Spanish, we expect minimally to find a tendency for our participants to accept *algunos* in subset conditions and reject it in set conditions. Regarding the ERP measures, we have no basis on which to

form an expectation for how SIs will be processed in Spanish, but given the strong findings of previous ERP experiments examining SIs in other languages we suspect that there will be some difference in the way *algunos* is processed in subset and set contexts. Should this come to bear, it will likely do so in the form of an N400.

2) Do native Castilian Spanish speakers distinguish between infelicitous *algunos* and felicitous (whole set) *unos* conditions? If so, will this effect be found online, offline or both?

   Prediction: We expect that there will be a tendency to reject *algunos* in set conditions but to accept *unos* in whole ones. All previous Spanish data sets (barring bilingual children, see Syrett et al. 2017a) have revealed that participants distinguish the two quantifiers and accept *unos* in both set and subset conditions equally.

3) Will the ERP effects for the *algunos* conditions be distinct from those of the *unos* ones?

   Prediction: *Algunos* carries an implicature as part of its inherent semantic features while *unos* does not. Thus, we expect that the processing of infelicitous *algunos* will give rise to more robust negative ERP effects (i.e., N400) than *unos*, if *unos* does at all. Given previous work showing that *unos* is equally acceptable in set and subset contexts, it may be the case that it is not differentially processed based on the context in which it is used.

3.5.2 The Experimental Tasks

The present study offers two experiments. The first is a picture sentence verification task (PSVT) modified from Politzer-Alhes et al. (2013) and in consideration of Spychalska et al. (2016), Tavano, (2010) and Wu & Tan, (2009), which was used for the ERP responses but also provided an acceptability judgment after each trial in order to compare pragmatically infelicitous or underinformative contexts to semantically correct ones via pictures. The second task was created so as not to force individuals to dichotomize judgments as acceptable or not, but one that would allow for an explicit association of a specific numerical quantity to each of the quantifiers in pre-defined contexts.

Data for both experiments were collected from 30 native Spanish speakers (23 females, age range 21-47, M = 26.5, SD = 4.1). All participants were in the UK for work, school, or personal travel. Prior to beginning the experimental tasks, participants filled out the LEAP-Q

to gather linguistic and biographical information (Marian, Blumenfeld, & Kaushanskaya, 2007). Two participants had been in the UK for one year, while the rest (on average) only 4.7 months at the time of testing. Participants were recruited from a tightly knit group of Spaniards in the greater London area and were either very low proficiency in English or had none at all. Given the proximity of Spain to the UK and the ubiquity of English worldwide, it is nearly impossible to find a Spaniard who has had absolutely no exposure to English, either through schooling or more passively the media sources; however, all participants herein were as inexperienced with English as one could hope to find, thus avoiding any issues with the recruitment of bilinguals given that such participants have been shown to treat scalar quantifiers in Spanish uniquely to monolingual participants (e.g., Miller et al. 2016). All participants were right handed as measured by the handedness questionnaire (Oldfield, 1971).

As participants did both tasks in one session, the order of the tasks was randomized between all participants such that half were administered Experiment A first, followed by Experiment B, and the other half were administered the experiments in the reverse order. After the testing session, a short debrief was carried out to ensure that participants were not aware of the specific aims of the study. All participants provided their informed consent. Ethical approval was granted by the University of Reading Ethics Committee.

3.5.3 Experiment A – Picture-sentence verification task

Each trial presented one of two possible scenarios: a SOME-type picture in which some but not all the characters were in involved in a target activity, and an ALL-type picture in which all characters were involved in the target activity using the quantifiers *algunos*, *unos* and *todos* 'all', where *todos* served as a control condition (see Figure 1). Recall that *unos* can refer to both a set and a subset, though the subset reading is generally triggered (and preferred) in cases where it contrasts with another available set. Thus, while *algunos* 'some' and *todos* 'all' have both a felicitous/infelicitous and or true/false condition respectively, *unos* has two conditions (set and subset) that are both felicitous. In order to make *unos* false, one would have to create a picture in which none of the characters are performing the target activity or only one of them is and the others are not. In either case, because Spanish expresses gender and number morphologically, if the image were to mismatch the target sentence, the latter would necessarily contain morphological cues that would render the context ungrammatical, thus becoming a grammaticality judgment rather than a truth value judgment. Furthermore, ungrammaticality for such trials would become detectable at the onset of preceding verb due to plural morphology, which would add unwanted noise to the EEG signal prior to the onset of

the critical words. Pilot data for this experiment, which contained such trials, did not show modulation of overall offline performance by them and because they ultimately would have served as a control condition only, they were removed from the present study.

Each image was preceded by the phrase *En esta imagen* 'In this image' and followed by a sentence that described the image correctly, incorrectly (semantics) or infelicitously by using the quantifiers *algunos, unos* and *todos* 'all'. The target quantifiers *algunos* and *unos* were presented without *de* 'of' to avoid favoring a pragmatic response. Each quantifier phrase (QP) was followed by a prepositional phrase (PP) such as "at the school" or "in the tree" to ensure both that the critical QP did not appear in nor would the EEG signal be time-locked to the sentence final position, thus avoiding unwanted extraneous noise in the EEG signal (i.e., wrap-up effect).

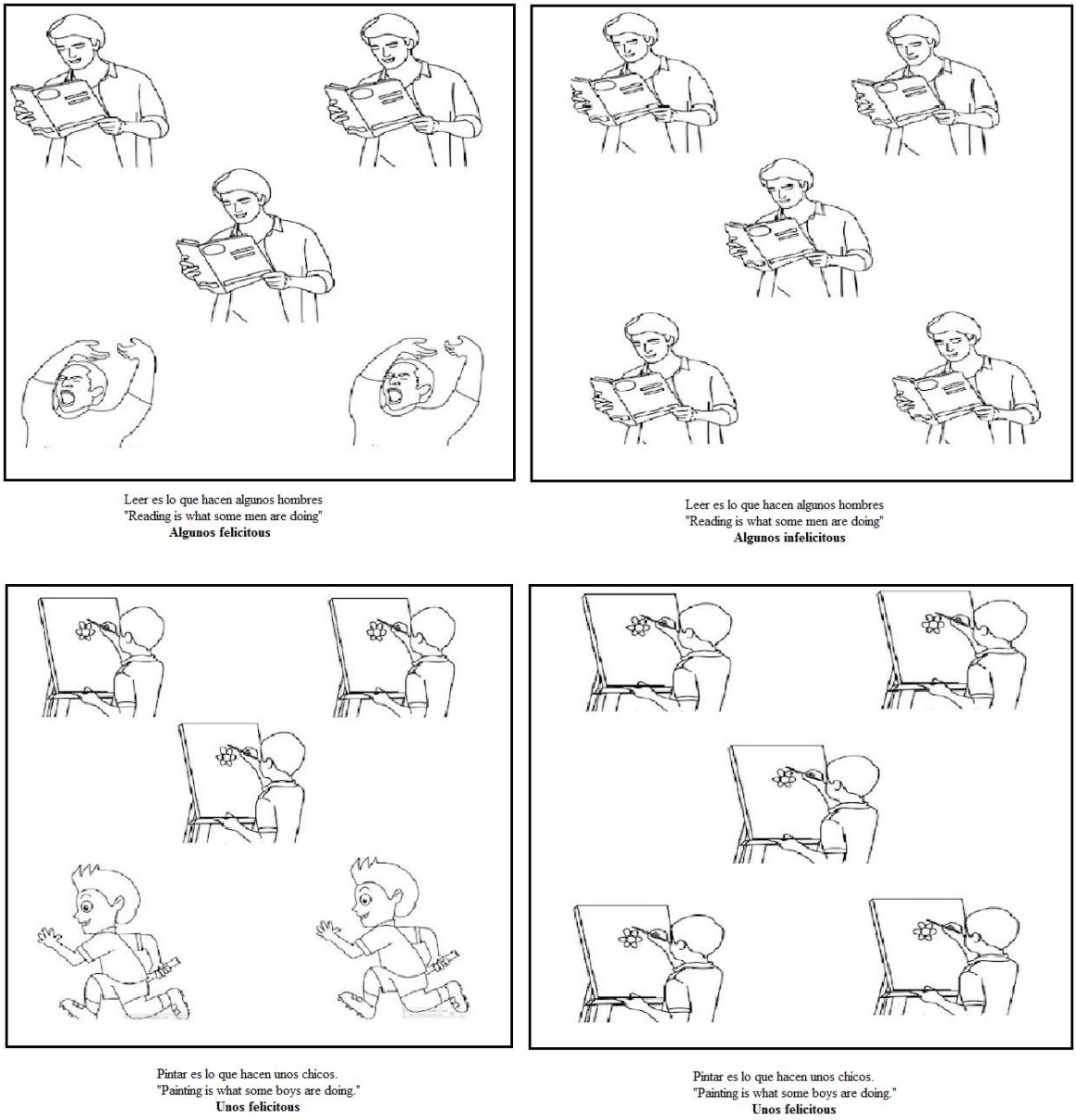


Figure 9: Sample pictures and sentences from Experiment A

Recent work on SIs using picture tasks such as the present study has noted that having the implicature generating item at the front (e.g., ***some*** *girls are sitting on blankets*) does not offer individuals enough prior information to make an accurate judgment when presenting the stimuli in rapid serial visual presentation (RSVP) (see Politzer-Ahles et al. 2013). Fortunately, in Spanish the infinitive verb, like the gerund verbal form in English, can be nominalized and thus appear as the subject of a sentence such as *Bailar es aburrido* 'Dancing is boring'. To avoid having the quantifier (*unos, algunos, todos*) appear too early, we place the nominalized verb at the front and chunked the critical QP (e.g., *algunos perros*) so that one could accurately

judge the status of the QP at its onset later in the sentence. Participants performed eight practice trials prior to beginning the experiment in order to get used to the nature of the non-canonical (yet completely licit) syntax used in the experimental trials. Following each trial, the participant indicated whether the sentence described the image well or not, using only their natural intuitions of Spanish. Recall that we included both a correct and incorrect *all* condition to serve as a control condition. This provided a 3 (Quantifier) x 2 (Type) design where inconsistent and/or incorrect *some/all* conditions were compared with lexically matched controls.

3.5.4 Materials and procedure

We designed 180 sets of black and white pictures arrays for the target trials. Fillers were made from distinct sets of similar image arrays. All images were obtained freely through Google and were modified with Microsoft Paint or Adobe Photoshop. In order to mask the target trials, we avoided using quantifiers for fillers while still remaining consistent regarding task type. That is, responses to fillers required participants' judgments of the sentences against pictures in which the position of relevant objects either matched or mismatched the sentence. All target picture arrays included five characters performing an activity with contexts that would render the use of each quantifier either felicitous/infelicitous or true/false in the case of *todos*. Target conditions consisted of 30 individual trials as divided in Table 1. No target quantifier appeared in tandem with the same image more than once for any individual trial. Infelicitous and felicitous *algunos* conditions, for example, shared the same sentence, though an ALL-type picture was used for infelicitous conditions and a SOME-type picture for felicitous ones. Felicitous and infelicitous conditions were matched in number.

Condition	**Algunos**	**Unos**	**Todos**	**Filler**
Felicitous	n= 30	n= 60	n= 30	n= 90
Infelicitous	n= 30	N/A	n= 30	n= 120
Total	n= 60	n= 60	n= 60	n= 210

Table 3: Total number of conditions

Seven lists were created for the presentation of the stimuli, which were made by counterbalancing the order and type (i.e., felicitous or not) of each quantifier. Each of the lists began with one of the target quantifier sentences and the rest was randomized. Stimuli were

presented in 10 blocks of 39 sentences with a break between each block. Images were preceded by a fixation cross lasting 250 ms and were presented in the center of the screen for 2000 ms. Each image was followed by a blank interval lasting 250 ms before beginning the presentation of the sentences. Individual words were presented for 350 ms with a 250 ms stimulus onset asynchrony (SOA). A randomized blank inter-trial interval between 500-1000 ms appeared prior to each subsequent trial.

3.5.5 Results Experiment A - offline judgments

In order to examine whether the experimental group differentiated between the four target conditions in similar ways, data were analyzed with a binary logistic regression model and pairwise contrasts with Bonferroni correction where appropriate using SPSS. Within the mixed model in SPSS, the Satterthwaite Approximation is used to get the degrees of freedom in order to run F-tests on the individual variables. The dependent variable for the experiment was *acceptance* (bien 'good' versus mal 'poor'). The model also included variables of Condition (partitive versus whole), Word (*algunos* versus *unos* versus *todos*) and the interactions of these variables. The variables were numerically coded within the model, being assigned a unique variable number to distinguish them from each other.

Figure 2 provides an overview of the mean acceptance for each condition. The model revealed significant main effects of Word ($F(2,5932) = 18.24$, $p < .001$, Condition ($F(1,5932) = 191.41$, $p < .001$, and a Word*Condition interaction ($F(2,5932) = 660.01$, $p < .001$. Additional pairwise comparisons revealed that the effect of Word was mainly driven by the *algunos*/*unos* distinction such that *unos* was markedly more acceptable in whole conditions ($M = .542$, SE = .048) than *algunos* ($M = .129$, $SE = .021$) estimate =.413, $t(5,932) = -11.14$, $p = .003$. Furthermore, *algunos* subset conditions were significantly more acceptable than whole conditions: estimate = .068, $SE = .027$, $t(1,747) = 2.6$, $p < .001$. This provides evidence that a distinction is not only being made between the two lexical items themselves, but crucially that the (in)felicitousness of *algunos* in subset versus whole contexts is systematically affected by the context of the pictures.

However, Figure 2 also interestingly shows that the whole *unos* conditions are not being treated as previous reports have shown, that is, equally in whole set and subset contexts. This discovery was followed up with additional post hoc analyses such that the data were divided between participants who responded "logically" (*some and possibly all*) by accepting *unos* in whole contexts and those who responded "pragmatically" (*some but not all*) by rejecting *unos* in whole contexts. The division was calculated on the basis of any given individual showing a

majority preference of one of the two *unos* conditions whereby only participants who showed a 70% acceptance of one or the other were included in the analysis (Spychalska et al. 2016). Only two participants from each response type made closer to chance judgments as low as 57% (13/30) for pragmatic responders and 66% for logical responders (20/30) and were removed from the post hoc descriptive analyses (see Figure 3). Though surprising that such a distinction is being made on *unos*, a logical/pragmatic split in responses is consistent with findings from previous research in other languages (e.g., Bott & Noveck, 2004; Hunt et al., 2013; Noveck & Posada, 2003; Politzer-Ahles et al., 2013; Tavano, 2010 among others), showing clear inter-individual variation between "logical" and "pragmatic" responders.

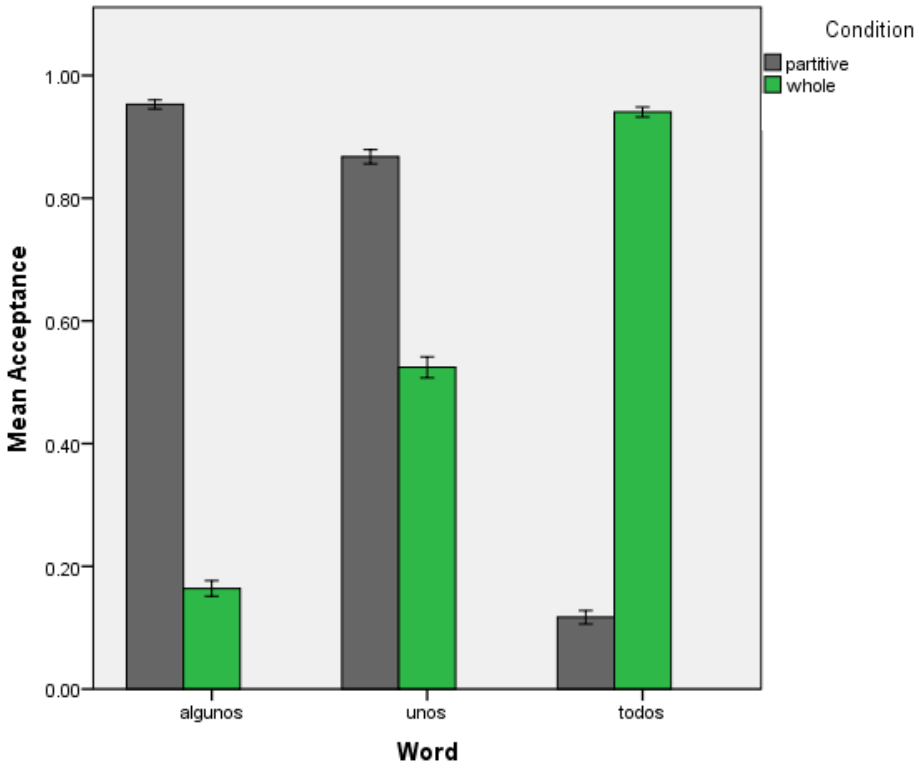


Figure 10: Percent acceptance of target conditions (+/- 1 SE)

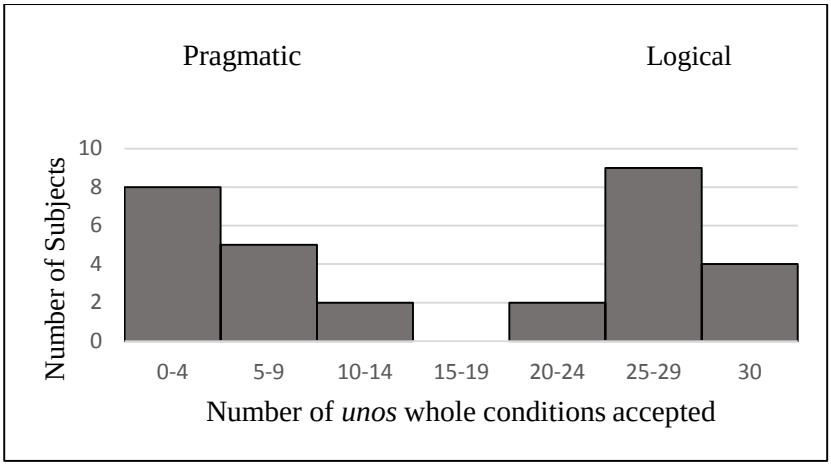


Figure 11: Histogram of overall logical versus pragmatic responses on unos

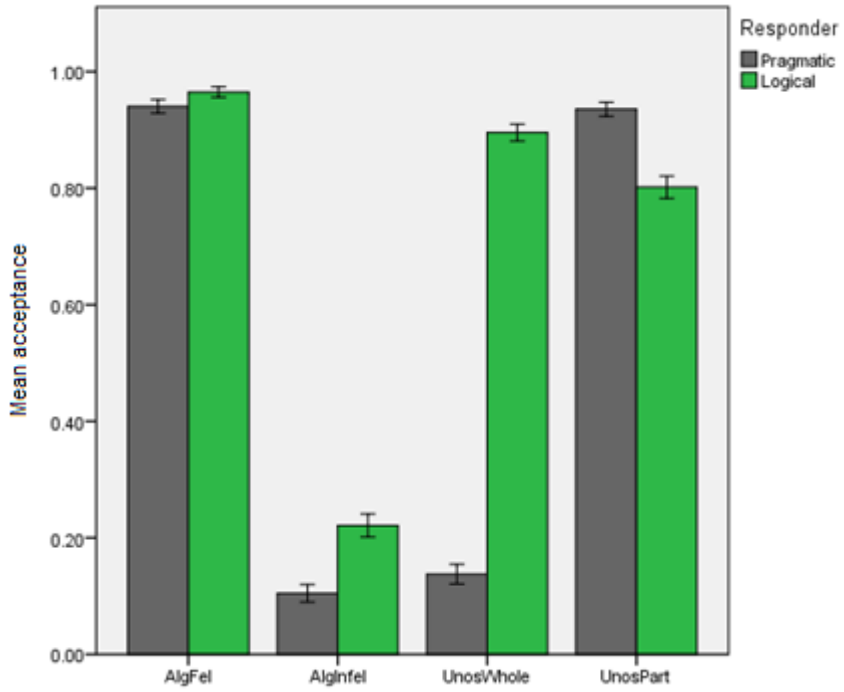


Figure 12: Pragmatic versus logical responders' acceptance rates (+/- 1 SE)

It is interesting to note from Figure 4 that the distinction between *algunos* in whole contexts and *unos* in whole contexts is lost within the pragmatic responder group, indicating that for these individuals both quantifiers are equally infelicitous.

3.6 Experiment B – Non-binary free interpretation task

In many SI studies, participants' responses are manipulated purposefully by instruction (e.g., Barbet & Thierry, 2016), enriched context (e.g., Guasti et al. 2005; Papafragou & Musolino,

2003), high cognitive demands (e.g., Marty & Chemla, 2013), prosodic focus (see Crain, 2012 for overview) and by the limitation of response time (Bott & Noveck, 2004, De Neys & Shaeken, 2007). In these cases, however, one does not know how a participant would interpret SIs in an environment that calls for explicit reasoning and when they are free to give a non-coerced, non-binary and explicit judgment. In order for a comprehender to derive an SI without any pressure from task demands, she must not be primed by explicit methodological factors within those tasks. With this in mind and following up from Experiment A, we designed a task in an attempt to avoid such task conditions.

Within the task, each condition defines a set of characters and asks the participant to choose all possible quantities from 0-*X* (where *X* is pre-defined) associated with the scalar term used to refer to the pre-defined set of characters. For each trial, the participant was provided with a short context containing one or the other scalar term and was asked to choose all possible quantities associated with it given in the context (see Figure 5). Both *algunos* and *unos* were used in conditions in which there were four, five, or six possible referents in order to determine if Number differentially affected the overall interpretation of each scalar term (e.g., Degen & Tanenhaus, 2011, 2015). It should be noted here that the 4 out of 4 condition falls within the subitizing range (i.e., 1-4), but the other conditions reflect a set size for *some* to be more naturally acceptable as partitive. Thus, we might expect that individuals would differentially treat the scalar terms based on the set size to which they refer, particularly the smaller set size as compared to the larger ones.[8]

In order to remain consistent across both filler and target trials, all answers contained the values between '0' and the highest possible number within the set. Because '0' and '1' were options in some of the filler trials, they remained so in the target trials so as not to draw any attention to differences between targets and fillers even though they were not felicitous/true options in the targets.

[8] Though it would be preferable to use even larger set sizes, our pilot study revealed that when participants were free to tick any number within the answers, the more possible answers there were led to more time involved in the experiment. Given the length of Experiment A, it became prudent to minimize the set sizes within Experiment B while retaining the distinction between the subitizing and larger ranges.

Word	Condition Type		
	Four entities	**Five entities**	**Six entities**
Algunos	n = 3	n = 3	n = 3
Unos	n = 3	n = 3	n = 3

Table 4: Condition types for Task B

Gloria fue a una fiesta con unas amigas. Si Gloria tiene 4 amigas, con cuántas amigas posiblemente haya ido a la fiesta? Por favor, escoge todas las posibles respuestas.

0 ☐
1 ☐
2 ☐
3 ☐
4 ☐

Figure 13: Sample of Experiment B Unos condition

In Figure 5, the context is translated as follows: "Gloria went to a party with *unas* 'some' friends. If Gloria has four friends, how many friends possibly went to the party? Please choose all possible answers."

To make the target contexts in this task, we used each scalar term three times in unique contexts of four, five or six characters performing an action, totaling nine targets for *algunos* and nine for *unos*. We also created 36 unique fillers to give a ratio of one target to two fillers. Fillers were made with similar short contexts, but rather than using a scalar quantifier, we made simple arithmetic problems such as "John needs to write 35 letters by Friday. If today is Monday, how many letters must John write each day until Friday? Please choose the best answer(s)." In this respect, though we did not ask for judgments of scalar quantifiers per se in the fillers, we did manipulate quantification as a variable in order to maintain a parallel between targets and fillers and to further mask the nature of the experiment. This experiment was completed using SurveyGizmo on a laptop with internet in the laboratory where Experiment A was administered.

3.6.1 Results Experiment B

Data for Experiment B were also analyzed with a binary logistic regression model and pairwise contrasts with Bonferroni correction where appropriate using SPSS. For the purposes of

making inherently non-binary responses amenable to a binary regression, those responses which constituted a subset-type judgment were assigned one number (i.e., 1) while those responses that were set-type were assigned another number (i.e., 2), thus collapsing raw responses into binary responses within the statistical model. The dependent variable for the model was ResponseType, this being either partitive or whole. Furthermore, we included variables of Word (*algunos* versus *unos*), Condition (*x* of 4, *x* of 5, *x* of 6), ResponderType (logical or pragmatic) and all higher-order interactions of these variables. As with Experiment A, the predictor variables were coded by assigning each a unique number within the model.

Figure 6 presents the mean responses for all conditions without dividing the participants based on response type.

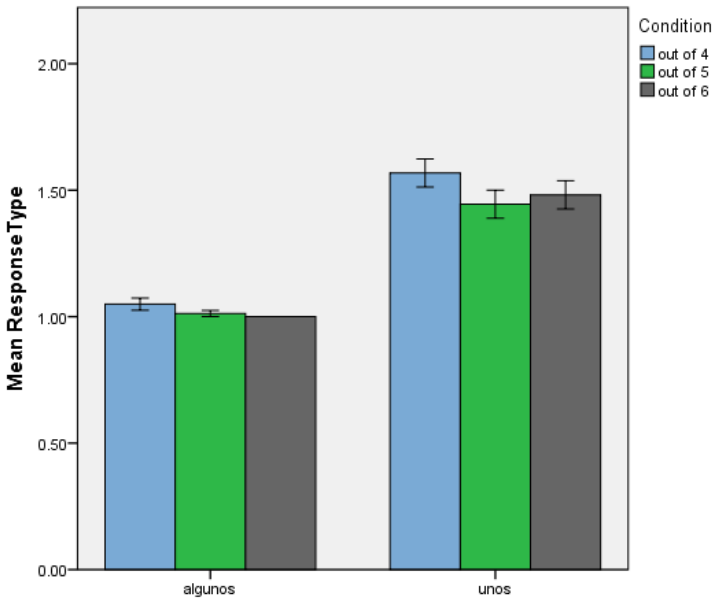


Figure 14: Mean response for all conditions (+/- 1 SE)

Descriptively, these data show that while *algunos* is largely interpreted as partitive, *unos* shows more variation. Figure 7 shows mean responses for word between "logical" and "pragmatic" responders and Figure 8 provides the raw responses reported for each condition. It is worth noting that the division of pragmatic and logical responders at the individual level overlapped completely across the two experiments such that all and only those who demonstrated a "pragmatic" or "logical" response bias on *unos* in Experiment A also did so in Experiment B.

Furthermore, because Spanish, like English, has specific words for *a pair* or *a few*, we might have expected that—particularly in the 4 out of 4 contexts—'2' and '3' would be avoided as raw responses, forcing participants to choose '4' more often. In this case, we would have

seen an over acceptance of *algunos* in 4-out-of-4 conditions, but we did not. Participants reliably judged *algunos* as felicitous in subset contexts and infelicitous in whole set contexts.

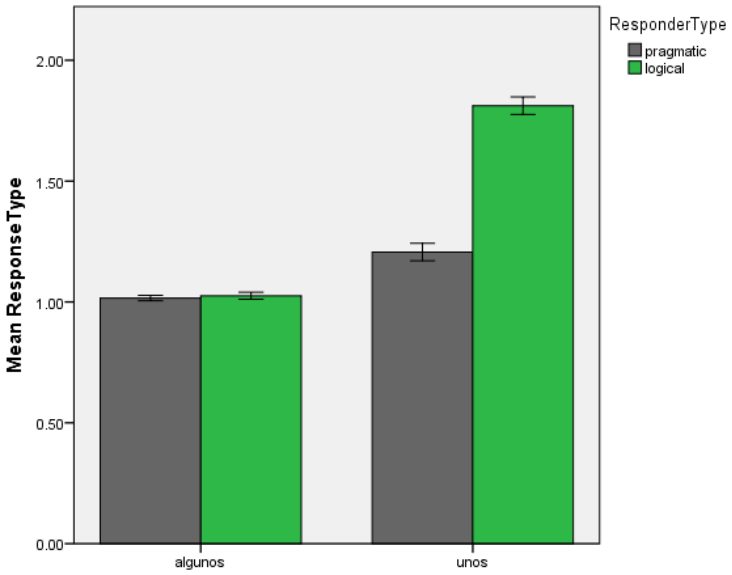


Figure 15: Mean response for words between responder type (+/- 1 SE)

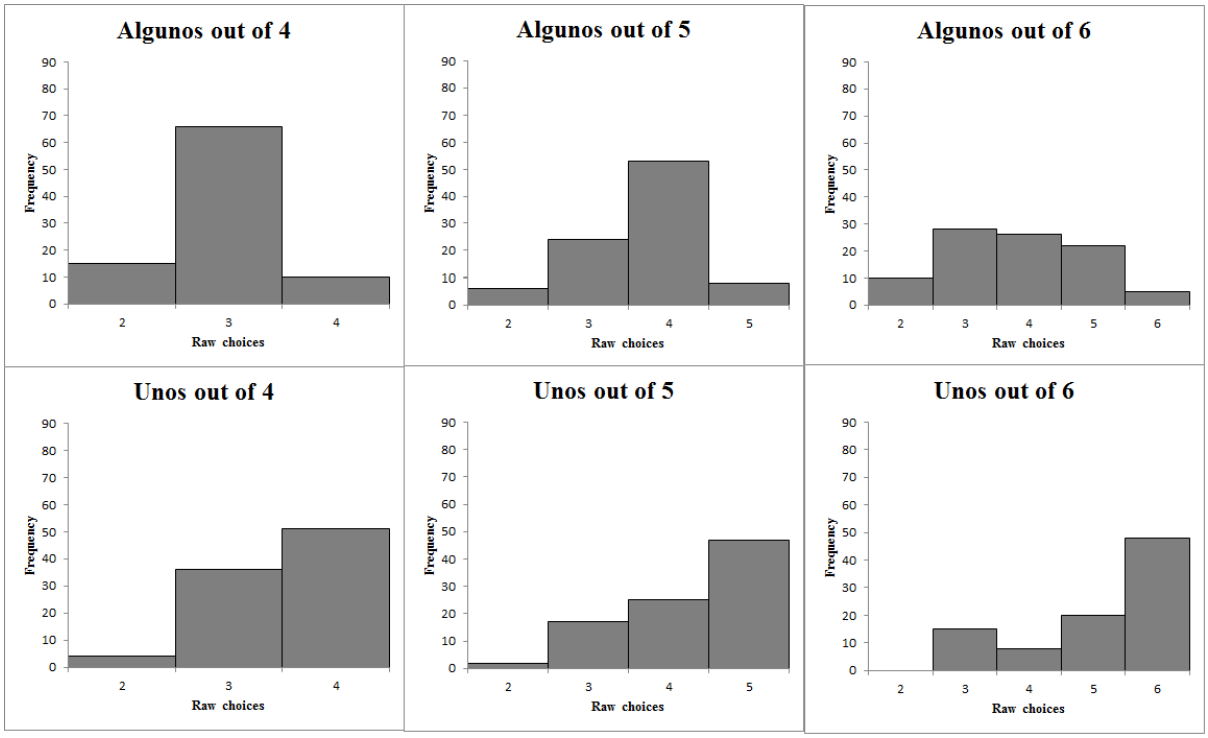


Figure 16: Histograms of raw responses for all conditions and all participants

The model revealed no significant main effects other than a Condition*Word interaction: ($F(3,474) = 7.05$, $p < .001$. Further pairwise contrasts crucially show that there is a distinction between *algunos* and *unos* as far as acceptability in certain contexts. While none of the *algunos* conditions was different from the other, each was significantly different from its counterpart *unos* conditions such that *algunos* is more clearly partitive and *unos* is either partitive or whole: *algunos* out of 4 to *unos* out of 4: estimate = .289, SE = .113, $t(470) = 3.10$, $p = .019$; *algunos* out of 5 to *unos* out of 5: estimate = .317, SE = .088, $t(470) = 3.6$, $p = .007$; *algunos* out of 6 to *unos* out of 6: estimate = .437, SE = .104, $t(470) = 4.2$, $p = .001$.

3.7 EEG recording and data acquisition

Continuous EEG was recorded from 64 sintered Ag/AgCl electrodes on an Easycap (BrainProducts GmbH, Gilching, Germany) as subjects performed the task silently. An additional external electrode (IO) was placed on the outer canthus of the right eye to monitor (horizontal) eye movements. EEG was recorded relative to FCz and later re-referenced offline to the average of the mastoid channels. Impedances were maintained below 10 Ω for all channels. Signals were amplified by a BrainAmp MR Plus amplifier (BrainProducts GmbH), applying a bandpass filter of 0.01 to 200 Hz, digitized at a sampling rate of 1 kHz and filtered offline using a 30 Hz low-pass filter. Epochs of 1500 ms were selected following the critical phrase and were corrected with a 300 ms pre-stimulus baseline. Rejection of artefacts was done automatically, removing trials due to artefacts of an absolute amplitude difference over 100 V/100 ms, or with activity lower than 0.5V in intervals of at least 100 ms. The minimum number of trials within each condition was 20 (66%). Overall, artefact rejection lead to the exclusion of 15% of the data prior to grand averaging of the waveforms.

3.7.1 EEG results

Following provisos from Luck & Gaspelin, (2017), we utilized a collapsed localizer analysis in order to remain conservative with our interpretation of results. Because Spanish *algunos* and *unos* had not previously been investigated using ERPs, we could only hypothesize that an N400-like effect would emerge in infelicitous compared to felicitous conditions in accord with previous work on other languages. Though we have no *a priori* reason to expect hugely different patterning results, it is possible that languages such as Spanish whose quantifiers differ semantically and pragmatically to other languages tested to date would show differences

in ERP signatures. Therefore, we explored the data with no firm commitment to time windows and/or specific regions of interest given that analysis parameters could not be entirely set on the basis of previous work.

The collapsed localizer dictates that we first averaged the waveforms across the conditions that would be used for comparison (i.e., *algunos*), and we used the timing and scalp distribution from the collapsed/averaged waveforms to define the analysis parameters that would ultimately be used for the data that had not been collapsed (i.e., *unos*) (see Luck & Gaspelin, 2017). After averaging across the relevant conditions, the time window and electrode sites showing the greatest negative activity were used to measure the effect in all of the relevant conditions separately (e.g., *algunos part. vs algunos whole, algunos whole vs unos whole, unos part vs unos whole, etc).* The electrode sites showing the greatest negative effect for *algunos* comparisons in the 250-450 ms time window were: Pz, P1, P2, P3, P4, P6, P8, CPz, CP2, CP4, PO2, PO4, PO8, C2, C4, consistent with the N400 latency and scalp distribution.

In order to examine whether the experimental group differentiated between the four target conditions in similar ways, data were analyzed with a linear mixed effects model and pairwise contrasts with Bonferroni correction where appropriate in SPSS. Waveforms and topographical maps were produced using the EEGLab graphics user interface (GUI) within MATLab.

250-450 ms

The dependent variable for the model was Mean Voltage. Within the model, we also included variables of Condition (e.g., partitive versus whole), Word (*algunos* versus *unos*) and an interaction of these variables. Overall, the model revealed only a main effect of Condition ($F(1,438) = 5.70$, $p < .001$. Additional pairwise comparisons revealed that *algunos* was associated with markedly more negative activity in whole conditions ($M = .807$ $SE = 4.4$) than in partitive conditions ($M = 11.37$, $SE = 4.4$), $p = .018$ ($MD = 9.95$). The comparison of whole *algunos* conditions to their *unos* counterparts ($M = 1.16$, $SE = 4.4$) revealed no significant differences. However, the *algunos* subset to *unos* set comparison was significant: $p = .011$ ($MD = 10.44$). Finally, *unos* whole conditions to their partitive counterparts ($M = 9.53$, $SE = 4.4$) revealed no differences in spite of showing an offline preference for partitive conditions.

700-900 ms

This effect was only present in the *algunos* conditions; thus we did not include a comparison of Word with *unos*. The electrode sites showing the greatest positive effect for *algunos*

comparisons in the 700-900 ms time window were: Cz, CPz, Pz, POz, PO4, P2, CP2, CP4, P4, P6, C6. The model revealed a significant main effect of Condition ($F(1,438) = 1.88$, $p = .041$, thus revealing a significant distinction between *algunos* whole ($M = -1.36$, $SE = 2.77$) and partitive conditions ($M = 6.38$, $SE = 2.77$), $p = .041$ ($MD = 7.74$). Though an unexpected result, it is consistent with previous implicature findings of a late positivity after N400 effects (Spychalska et al. 2016). While this effect has various names and, potentially various underlying causes, it has been referred to as a "semantic P600" and has been attributed to processes generally related to reanalysis or repair and/or an attempt to revise an initial parse (see Friederici et al. 1996; Hahne & Friederici, 1999; Kuperberg, 2013; Van Petten & Luka, 2012). Moreover, this effect is known to emerge as a response to sentences with plausible constituents arranged in ways that render them implausible as a whole (Kim & Osterhout, 2005; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003) or prolonged analysis of problematic sentences (Kolk, Cwilla, van Herten & Oor, 2003; Kuperberg, 2007). Delong et al. (2014) also found late positivity effects from strong semantic violations (i.e., contextual implausibility and failed predictions), but, as is typical with the "semantic P600", such effects emerged in the anterior scalp regions as opposed to the typical centro-parietal regions of the standard P600. The positivity in our data emerges over centro-parietal electrode sites (see Figure 10), a point to which we return below.

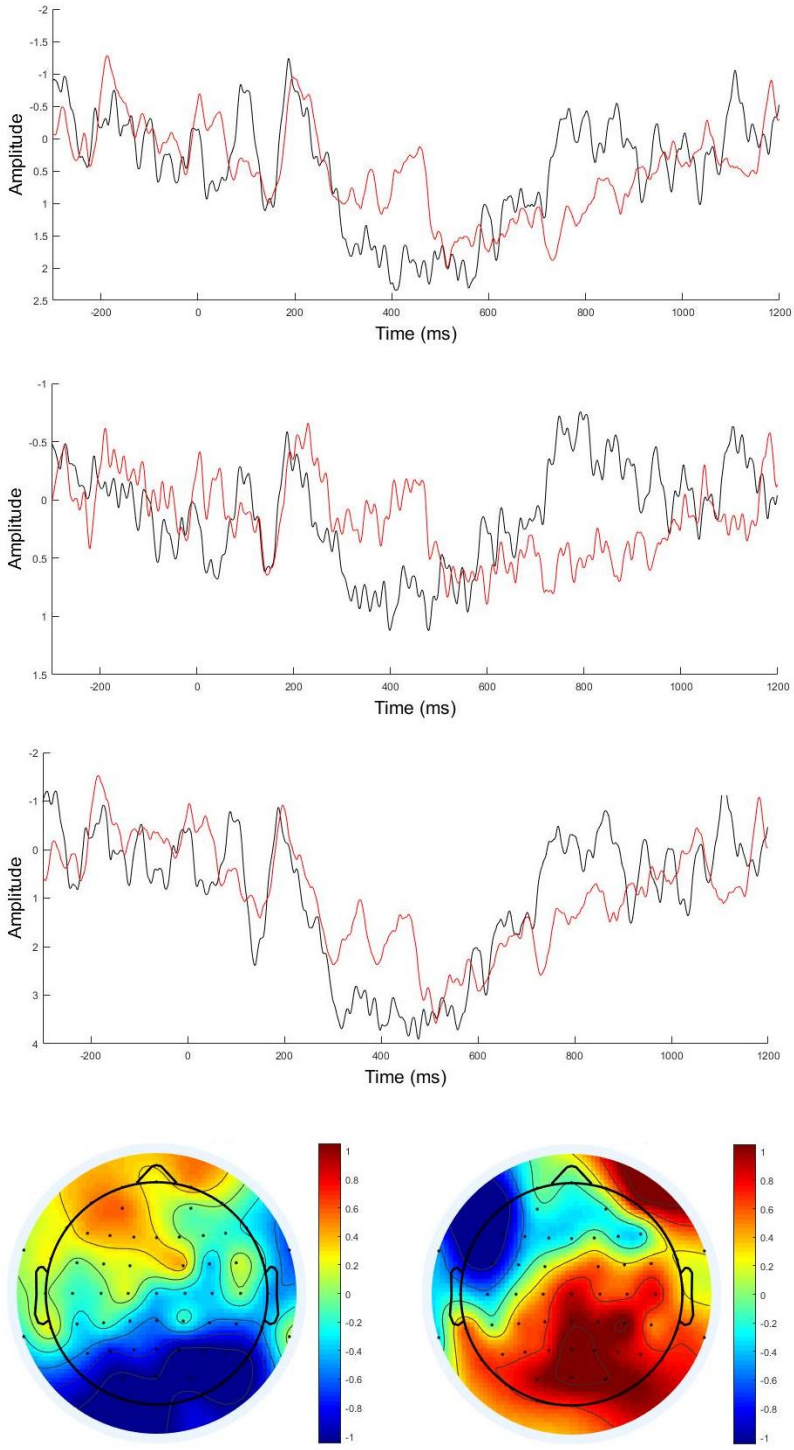


Figure 17: Grand average ERPs elicited by *algunos* partitive versus *algunos* whole shown at electrode sites CPz, Cz and Pz (top to bottom) for all participants. Topographical distribution by subtracting felicitous from infelicitous conditions: **Left:** 200-600 ms. **Right:** 700-900 ms.

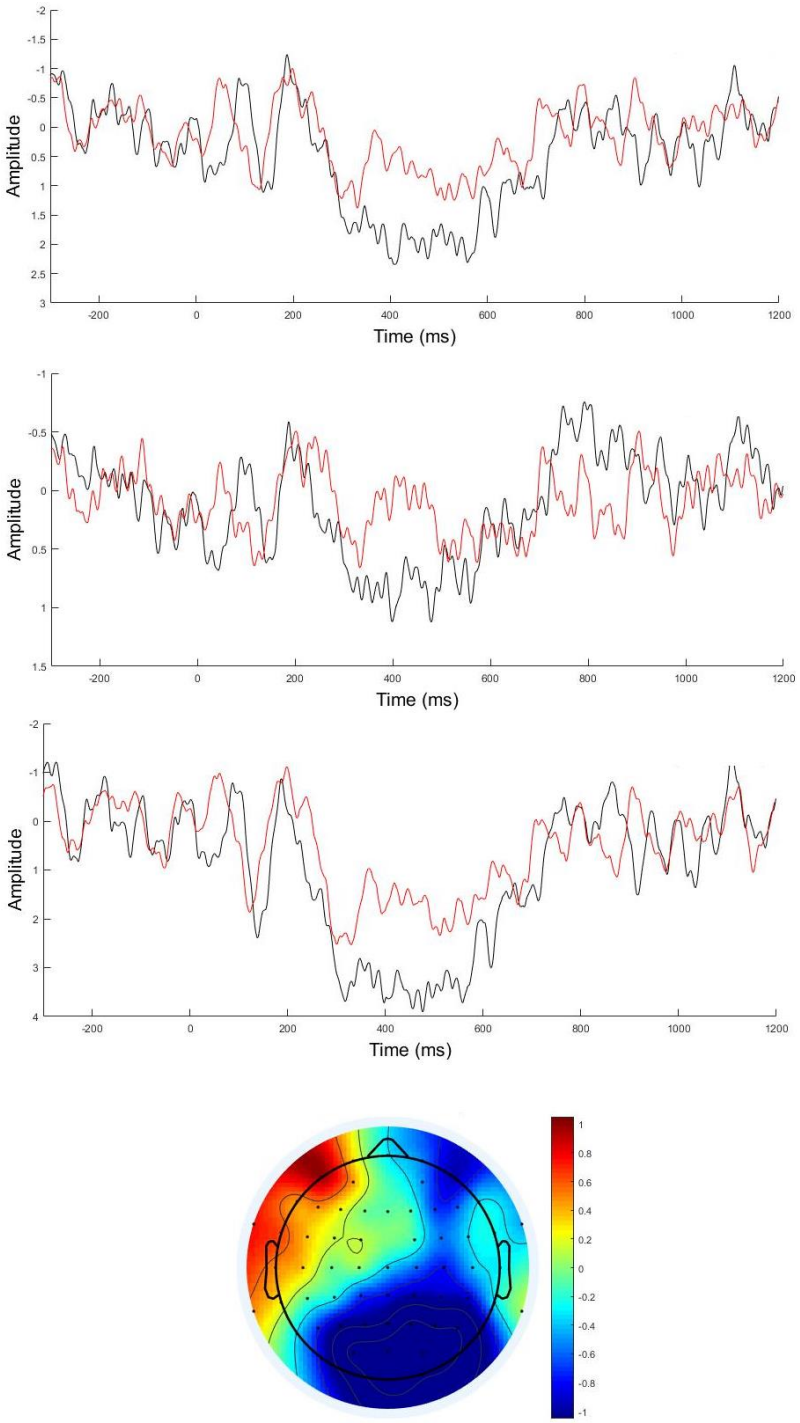


Figure 18: Grand average ERPs elicited by *algunos* partitive versus *unos* whole shown at electrode sites CPz, Cz and Pz (top to bottom). Topographical distribution between 200-600 ms by subtracting felicitous from infelicitous conditions.

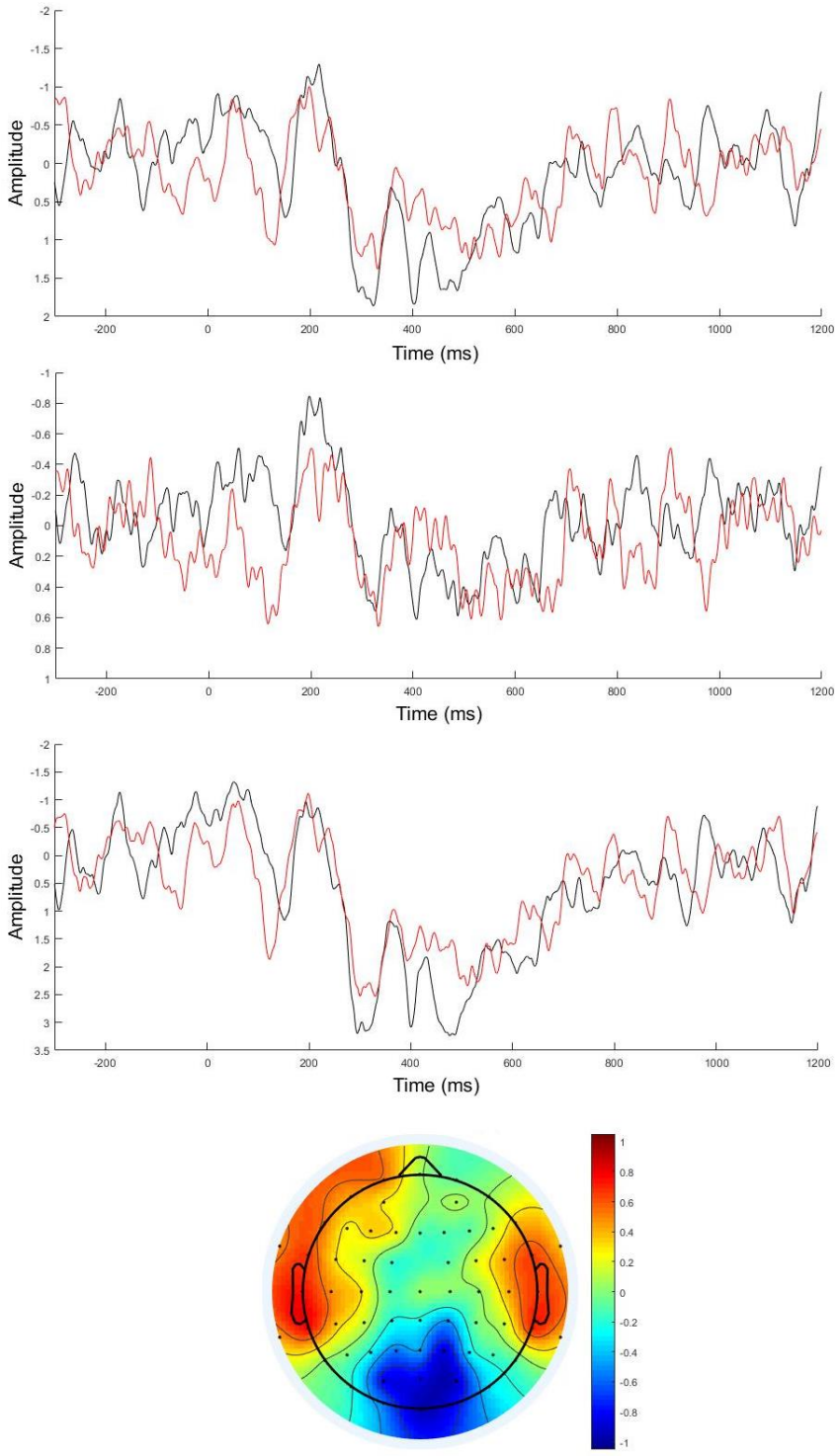


Figure 19: Grand average ERPs elicited by *unos* partitive versus *unos* whole shown at electrode sites CPz, Cz and Pz (top to bottom. Topographical distribution between 200-600 ms by subtracting felicitous from infelicitous conditions.

3.8 Discussion

This study examined the electrophysiological activity associated with variability in SI calculation among (adult) native speakers of Spanish. In the context of the methodological discussion that formed the initial part of this paper, we aimed to test following research questions:

1) Do native Castilian Spanish speakers distinguish between felicitous and infelicitous *algunos* conditions? If so, will this effect be found online, offline or both?
2) Do native Castilian Spanish speakers distinguish between infelicitous *algunos* and felicitous (whole set) *unos* conditions? If so, will this effect be found online, offline or both?
3) Will the ERP effects be modulated/corroborated by offline response type?

In line with the few offline studies in Spanish to date, our aim was to determine whether or not Spanish native speakers would distinguish between pragmatically felicitous and infelicitous conditions with *algunos* and whether they would distinguish between pragmatically infelicitous *algunos* and felicitous *unos* in whole set contexts. Given the surprising results of an offline interpretation preference on *unos,* we compared the ERP results of the two *unos* conditions. Visual inspection of the relevant waveforms for each condition, however, revealed a distinction only between *algunos* subset and set conditions and the *algunos* subset and *unos* whole set conditions. There was no difference in the processing of the two *unos* conditions, in spite of such an effect in the offline data. We also found that in the comparision of *algunos* set and subset conditions, a late positivity emerged following the N400 effect. While less is known about the functional status of this type of effect, it is interesting that it appears in the P600 timeframe, possibly indicating reanalysis or repair. While the P600 does emerge reliably for morphosyntactic violations, it is not syntactic in nature (Tanner, Grey, VanHell, 2017). For example, some work has shown that the P600 emerges in animacy violations (Kuperberg et al, 2003, 2006, 2007), thematic role reversals (Kim & Osterhout, 2005), conflicting information between local coherence and global plausibility (Van Herten et al. 2006), strong semantic violations (Delong et al. 2014), and even incongruous emoji usage (Weissman & Tanner, submitted). Moreover, though the N400 and P600 reliably emerge as responses to specific linguistic stimuli, the characterization of ERP responses as either N400 or P600 (but not both) is not accurate. For example, Tanner (2015) and Tanner et al. (2017) point out that it is not the

case that the cognitive mechanisms underlying N400 or P600 effects are mutually exclusive. As they determine, in fact, these effects can interact and co-occur in time and space, and individual brain responses can differ.

It is possible that the distinction between *algunos* partitive and whole conditions constitutes a "strong semantic violation" or a failed lexical prediction, though the topographical distribution of the effect in our data is not frontally biased rather it emerged over centro-parietal sites typical of the standard P600. Notwithstanding, the proposed theoretical descriptions of these quantifiers all posit that there is something about the lexical semantics of the quanitifers themselves, whether it is a D-linking constraint on *algunos*, partitivitiy associated with *alg-* when it attaches to *un*(*os*) or the presuppositionality of *algunos*, that gives rise its associated implicature. In this sense, the implicature is driven by semantic features. While it is true that the context in which *algunos* appears affects its interpretation, meaning that pragmatics plays a role, such an interpretation cannot be entirely pragmatic given the lexico-semantic constraints to which *algunos* adheres. Thus, the individual who derives an upper-bound meaning of *algunos* is not doing so without the help of the semantic features of the quantifier. It is possible that the infelicity of *algunos* gives rise to what has been described in the literature as a post-N400-positivity associated with failed prediction/lexical disconfirmation and/or strong semantic violations due to its semantic features constraining its use. If this is the case, one might wonder how *algunos* gives rise to a pragmatic implicature at all. Whatever the answer to that question is, it is worth highlighting again that *algunos* does indeed behave like other SI generating items with respect to the defeasibility of its upper-bound meaning in downward entailing environments such as in conditionals and/or canceling the upper-bound meaning with the addition of a stronger quantifier like *todos* 'all'. Still, little experimental work exists to validate these assertions (e.g., Vargas-Tokuda et al. 2009).

Additionally, according to the theoretical proposals already mentioned, *unos* does not carry any implicature, it is unaffected by downward entailment, its semantic 'at least one' meaning can be cancelled when followed by a stronger quantifier like *todos* 'all' and it is equally acceptable in set and subset contexts. However, our offline data do not corroborate the assertion that *unos* does not give rise to an SI. First, across both the picture-sentence verification task and the non-binary interpretation task, the same 50% of our participants largely reject *unos* at a rate of 80% in whole set contexts. The other 50% of participants, termed "logical" readers, accept *unos* in both set and subset contexts equally as previous reports in Spanish have shown. So, what is to be said about those who reject it, the same as they do *algunos*, in whole set contexts? If *unos* indeed does not carry an implicature and it can be both

whole and partitive in its interpretation, one would not expect to find any differences between its readings in set and subset contexts, such has been the case thus far in the literature (Miller et al. 2016; Halsey, 2010; Syrett et al. 2017a,b; Vargas-Tokuda et al. 2009). However, in our data we do find such a difference. Yet the difference between so-called "logical" and "pragmatic" responders on *unos* does not emerge in the ERP data, rather it is restricted to behavioral judgments alone. Thus, had we not run both methodologies together, this pattern of acceptability in juxtaposition to the ERP responses would not have been captured.

Given such strong offline interpretation preferences for *unos*, one might have expected the ERP results to reflect differential treatment of the quantifier based on the context in which it appeared. However, although such a difference was not found, *unos* did produce more negative amplitudes in the wholes set conditions than in the subset conditions (though not significant) and, when compared to *algunos* partitive, *unos* whole conditions did reveal a significant difference. This latter comparison shows that while there may be no interpretable processing difference between *unos* whole and subset readings, the whole context at least appears odd when juxtaposed against a condition which is unambiguously felicitous. The effort of much of the recent experimental SI work to distinguish between responder-type has made it clear that future work in this domain will benefit from examining individual differences in the way participants treat scalar terms. However, it is not always the case that offline judgments modulate ERP results as in Spychalska et al. (2016) or Nieuwland et al. (2010), for example.

To our knowledge, this is the first ERP study of *algunos* and *unos*. Although we had an inclination that the N400 or an N400-like modulation would apply here due to previous SI studies, there was no precedence from which to form firm expectations of what the latency, amplitude or distribution patterns would be for the effects associated with quantifiers whose properties dictate their use and, therefore, interpretation uniquely. It has been argued that responses to a picture-sentence verification task such as this are actually more suited at targeting the effect of a scalar implicature-enriched interpretation mismatching prior knowledge than measuring an effect of the SI itself being calculated (Hartshorne et al. 2015). However, in our experiment, unlike Politzer-Ahles et al. (2013), for example, we did not display the target quantifiers in sentence-initial position, rather we presented them together with the noun phrase and in the middle of the sentence, examining the precise moment the quantifier is interpreted after its status is assessed. While most ERP SI studies have found an N400 effect (e.g., Hunt, Politzer-Ahles, Gibson, Minai, & Fiorentino, 2013; Noveck & Posada, 2003; Nieuwland et al. 2010; among others) associated with pragmatically inconsistent conditions, others have found different effects, such as an Nref (Politzer-Ahles et al. 2013),

positivity (e.g., P600 see Hartshorne et al. 2015) and an N400 in followed by a post-N400 Positivity (e.g., Spychalska et al. 2016). Upon analysis of our waveforms, we see that the effects produced by the quantifiers in Spanish are in line with N400 effects while also producing a post-N400-positivity more closely related to the P600 given its timing and scalp distribution.

Regarding Spanish, previous work has shown a wide range of variability with respect to the interpretations derived from *algunos* and *unos*. Miller et al. (2016), for example, used a video-acceptability judgment task showing that adult native speakers of Spanish readily allow for non-partitive readings of *algunos*. In Experiment B herein, a novel task type was created that allowed participants to freely choose any reasonable number associated with *algunos* or *unos* within a given context that defined the possible number of referents. The results from this non-binary task allowed us to first notice a distinction being made in the *unos* conditions. Thus, we take this as evidence that allowing individuals to choose freely what a scalar term means without increasing cognitive load, requiring the use of world knowledge, or priming a response will provide a more explicit picture of how these terms are treated and understood at an individual level in Spanish.

It is also clear that offline and online methodologies can combine to be more explanatory together than each is in isolation (Bott & Noveck, 2004; Nieuwland et al. 2010; Villegas, 2014). Had we relied solely on the behavioral data, we would have arrived at the conclusion that Spanish presents a case in which Spanish-speakers can be divided neatly into "logical" and "pragmatic" responders based on judgments to *unos*. In contrast, had we relied solely on the ERP data, such a conclusion could not have been reached. Together, the data are slightly puzzling. Whereas participants show a clear offline preference for one reading of *unos* over the other, the electrophysiological responses reveal only that *unos* is different from *algunos* and that *algunos* is preferred in subset contexts. Given the differences between the methodologies, it is plausible, as per Katsos & Bishop (2011), that our participants came up with a strategy early on to which they adhered with great reliability across the behavioral tasks, but that the ERP data reflect an automatic and more implicit response for which a strategy cannot be developed. Moreover, De Neys & Shaeken (2007) and Bott & Noveck (2004) both showed that the rate at which SIs are derived is affected by the time allotted to the participants to make judgments (see also Marty and Chemla, 2013). While no time constraint was relevant for either of the offline tasks herein, the nature of the picture-sentence verification task (ERP task) called for much faster judgments than those in the offline tasks given its online nature. For example, whereas the non-binary judgment task allowed participants to reason freely and

for as much time as needed about the meaning they assigned to each quantifier, the picture-sentence verification task prompted a response within a few seconds at most. Consistent with other work, however, it is not uncommon for disparate results to emerge as byproducts of distinct online and offline methodologies (Jegerski eet al. 2016; Nieuwland et al. 2010; Villegas, 2014). Thus, had only one or the other methodology been employed, we would not have the entire picture of performance. As it stands, we propose that *unos* may in fact carry an implicature in spite of current theoretical descriptions and we maintain that in order to capture a more generalizable description of Spanish SIs one must consider the many dialects of Spanish that may give rise to differences in interpretation. Thus, the theoretical work can benefit from the empirical work.

3.9 Conclusion

The results of the present study—especially as informed by the offline data with the same subjects—show a clear difference in the way individual participants process scalar terms in Spanish. On the one hand, this experiment was exploratory, aiming to examine both how the brain treats Spanish scalar quantifiers and whether those responses would be modulated by offline responses. On the other hand, our data suggest that testing whether individuals calculate implicatures, or whether minimally one is pragmatic or logical, can be more thoroughly informed by employing more than one type of methodology, if possible. No experimental study will match naturalistic language use with respect to ecological validity; however, future research (ours included) should consider experimental factors, not just design but choice of overall method, that can affect individuals' interpretations of words or language that are highly susceptible to extra-linguistic variables. We posit that using online and offline measures when appropriate benefits the overall strength of the data and its subsequent interpretation in work examining implicatures.

Chapter 4: Article 3

Abstract

The tradition in bilingualism studies has been to focus on the role of the first language (L1) in second language (L2) acquisition, assuming that the L1 remains stable across the lifespan. Recent work on the non-pathological erosion or loss of previously acquired linguistic knowledge (i.e., attrition), however, has called in to question the assumption that an L1, once acquired, is impervious to major change. However, relatively little is known about the factors that contribute to L1 loss and/or maintenance across the lifespan. While the cognitive tasks of bilinguals are more complicated than those of monolinguals, which can be evidenced in differential performances on cognitive tasks such as those related to executive functioning, it is an open question whether bilingual attriters truly lose previously acquired mental representations or simply experience performance burdens due to bilingualism more generally. The present study examines a property that is argued to be ripe for attrition (i.e., scalar quantifier interpretations) using both online (processing) measures in the form of ERPs and offline judgments (interpretations) in an attempt to disentangle processing burden from the erosion of mental representations among Spanish-English bilinguals in long-time L2 immersion. Our data indicate that though bilinguals' offline performance is not monolingual-like, their ERP responses are similar in some conditions to the monolinguals'. The data show that automatic (implicit) brain responses (i.e., processing) in these bilinguals indicate marginal sensitivity to pragmatic violations in quantifier interpretations even after prolonged exposure to an L2, but their explicit responses (i.e., representation) show no such sensitivity.

Keywords: attrition, scalar implicature, ERP, bilingualism, processing

4.1 Background and introduction

Concerning crosslinguistic influence, bilingual studies to date show an overwhelming partiality towards examining transfer from the first language (L1) to the second language (L2) (see Foley & Flynn, 2013 for review). Other important questions, such as those pertaining to language loss, erosion or the extent to which there truly is less linguistic interference in the inverse direction (i.e., L2 to L1) have been examined less. The present study contributes to increasing our knowledge of this understudied direction, that is, L1 attrition. Schmid (2013) and Schmid & Köpke (2017) argue that changes to a first language go hand-in-hand with general bilingual development. Thus, changes in the L1 as a result of L2 acquisition as well as L1 effects on L2 development are both crucial to understanding larger questions in language and cognitive sciences, particularly in a world that is becoming more and more multilingual. The tradition of using a monolingual baseline as a control comparison in bilingual studies, however, perhaps perpetuates the general idealized notion of monolingual L1 competence (Rothman, 2008; Ortega, 2013). In turn, this could be a contributing factor underlying the erroneous assumption that an L1, once acquired, is more or less impervious to major change and thus provides a non-moving target for comparison purposes.

No one (or theory) denies that language is a social phenomenon. As such, the complexities of individual people, groups, and societies along with their independent experience(s) with language complicate its characterization (e.g., Schmid, Köpke, & de Bot, 2013). The heterogeneity of variables that affect the development, maintenance and even loss of language has delimited traditional experimental approaches designed to measure it. These questions have been at the center of work on L1 *attrition*, "the non-pathological decrease in performance (comprehension and production) in a language that had previously been acquired by an individual" (Schmid & Kopke, 2004:5). Research on L1 attrition challenges both the assumption that an L1 remains stable after development and calls for revisions in the methodologies used to test emerging patterns in bilingual development in an unbiased way.

Conventional assumptions related to L1 stability find their genesis in observations from the vast majority of people worldwide who spend their lives in their L1 environment where fluctuation in accessing high quality L1 input over the lifespan varies very little. It is not surprising, then, that early work assumed that an L1 becomes impervious to (major) changes by late childhood. However, an increasing amount of research in fact shows that when the default balance of exposure to L1 input and opportunity to use it are thrown off, as in immigrant immersion contexts, the L1 is much more susceptible to change than one might have thought

possible. There is also a correlation between proficiency in the L2 in an immersion context, continued exposure to/use of the L1 in an L2 immersion context, age of acquisition of the L2, and degree of so-called attrition (see e.g., Schmid, 2011; Schmid & Köpke, 2017 for review). This is especially true at the levels of grammatical knowledge that are cognitively more costly, such as in real-time application of processing strategies (e.g., Dussias & Sagarra, 2007) and/or for use of linguistic properties subject to felicitousness conditions at linguistic interfaces (Sorace, 2011; Tsimpli, Heycock & Sorace, 2004; Tsimpli & Sorace, 2006).[9]

Thus, developments in theories of attrition and the methodologies to test them show that both diachronic and synchronic language change are standard at a societal and individual level, for monolinguals and bilinguals alike (see Schmid and Köpke, 2017 for review). Nevertheless, the vast majority of research on language attrition places an emphasis on the conditions that actuate language loss. Perhaps, as suggested by Iverson & Miller (2017), reconceptualizing the idea of a steady-state grammar in light of evidence that an L1 can undergo significant changes across the lifespan and reframing the question around the conditions that foster language maintenance rather than those that catalyse its loss, could enrich current theories of attrition. For example, if specific domains of language attrition are predictable, then the precise nature of this selectivity is a matter that should stand at the forefront of attrition research (Iverson, 2012). Is attrition the result of affected or eroded mental representations (i.e., unstable features in the L1 being affected by opposing features in the L2, Lardiere, 2009) or is it a matter of a more superficial, momentary bleeding over of the competing L2 grammar from a processing burden arising as a result of bilingualism more generally (e.g., multiple active linguistic systems in the mind could engender variability in specific domains of the grammar, e.g., (de Bot & Clyne, 1994; Schmid, 2002, 2009; Serratrice & Sorace, 2009; Sorace, 2011)?

In this vein, there have been many theories in various linguistic paradigms aimed at explaining the so-called selectivity of attrition and why it might unfold the way it does (e.g., the Threshold Hypothesis, Paradis, 2007; Regression Hypothesis, Jakobson, 1943; L2 induced selectivity, Gürel, 2004, see Gürel, 2007 for review). Herein, we assess the predictive and explanatory power of one such theory, the Interface Hypothesis (henceforth IH) (Sorace, 2000, 2011; Tsimpli and Sorace, 2006; Tsimpli, Sorace, Heycock, & Filiaci, 2004). We focus on the IH given the linguistic domain under investigation being one at internal and external interfaces.

[9] Interfaces are defined as abstract points in language computation at which two submodules (internal interfaces) of grammar (e.g., syntax-semantics) or one submodule of language and one external to language (external interfaces, e.g., syntax-discourse) integrate information.

Moreover, the IH makes strong predictions and it can shed a new light on discrepant results to date stemming from research examining its claims. The IH predicts that: i) properties that are associated with or dependent on contextual appropriateness (discourse) are the most likely to undergo attrition and ii) properties that are internal to the grammar, such as narrow syntax, are comparatively more durable, if not impermeable to change in mental representation. A great deal of work has shown that the variability found at the external interfaces is due to limitations in working memory, processing capacity or efficiency, and resource allocation (see Rothman & Slabakova, 2011; Sorace, 2011 for discussion). Additionally, other work has suggested that variability in these domains may also be linked to burdened attention allocation that is typical of bilingualism. In this respect, the activation of multiple linguistic systems requiring constant inhibition of one grammar while processing the other is the claimed locus causing variability in even the bilinguals of exceedingly high proficiency, not necessarily disparate (for attrition truly eroded) linguistic representations (see Serratrice & Sorace, 2009; Sorace, 2011; Wilson, Sorace, & Keller, 2009).

In spite of inconsistencies in replicability across all bilinguals and/or bilingual environments (see Papp, Johnson & Sawi, 2015), the evidence showing that the impact of multiple activated linguistic systems on cognition, including the ensuing neuroanatomical effects, remains persuasive (Bialystok, 2016). For example, language processing for bilinguals is argued to be more costly than for monolinguals given that the former must necessarily deal with an array of additional mental tasks, such as resisting unhelpful incoming information (inhibition) from either language, correction of automatic reflexes from language systems that are not relevant for use, dealing with multiple lexicons with potentially drastically different lexical items, and more (Abutalebi et al. 2011; Bialystok, 2009; Green, 2011). Furthermore, bilinguals are challenged by the charge to maintain processing-specific routines that are founded on comparatively less input, which may ultimately contribute to linguistic features and lexical items—in either language—that are more weakly represented or less activated than those of monolinguals (Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Hopp, 2013).

Within the greater discussion of bilingual development, however, crosslinguistic competition and decreased activation of language specific properties is not a reality in the direction of the L1 to L2 only, but also the L2 to L1. One goal of current attrition research, then, is to measure beyond performance alone and determine the extent to which differences in production, processing and competence reflect potential modifications to specific mental representation(s) of the L1. It has been argued that in order to better understand not only the essence of bilingual development but also to shed light on open questions regarding the

language capacity of humans, we must develop a clear method for testing whether and how the mechanisms constraining L2 acquisition can impact on linguistic knowledge already developed in the L1 (Schmid & Köpke, 2017). These claims not only extend to the IH's application to attrition, they also motivate a broader question than that of merely asking where attrition is likely to occur, that is, why does it happen at all? (Iverson, 2012).

4.2 Aim of the present study

Notwithstanding the uncontroversial findings regarding why bilingual attriters show variability in specific modules or sub-modules of grammatical knowledge, much of the work on the IH's application to the vulnerability of certain interface-conditioned properties has led to discrepant findings. It is possible that these disparities have emerged due to a strong partiality towards work examining knowledge of topic- and focus-marking strategies in the use of null and overt subjects or object clitics in Indo-European languages. However, in spite of a recent growth in research on other properties relevant to the internal-external interface claim (Donaldson, 2012a, 2012b; Ivanov, 2012; Slabakova, Kempchinsky, & Rothman, 2012; Valenzuela, 2006), these results have also been discordant with respect to their claims about the vulnerability of such properties in both adult language acquisition and attrition (Belletti, Bennati, & Sorace, 2007 compared to Rothman, 2009). We argue that at least part of this discordance could stem from the types of properties being examined. While the use of null versus overt pronouns in morphologically rich languages like Spanish, for example, indeed depends on discourse factors, the misuse of these pronouns is actually determined by an already present instantiation of one (or not) in the linguistic discourse and not as much by factors outside the scope of the submodules of language. We maintain that in order to truly test the vulnerability of an external interface phenomenon, one must test properties whose felicitous use is not determined by previous instantiations of a related item in the linguistic discourse, rather those that are more dependent on variables outside language.

In the present study, we address these questions by testing interpretations of scalar quantifiers in the Spanish of Spanish-English bilinguals. Spanish scalar quantifier interpretations are highly associated with pragmatics (non-linguistic discourse) and have distinct distributional patterns as compared to English. Unlike the case of subject pronoun anaphora, scalar quantifier interpretation does not depend on whether the lexical item that gives rise to the interpretation was previously used in the linguistic discourse or not; rather, their interpretation is dependent on an individual's conceptual association of a numerical quantity

with the lexical item itself. If on the right track, then, we expect that scalar (quantity) interpretations will be susceptible to attrition in a way that pragmatically-induced changes in null/overt subjects or objects are not. That is to say that the semantic and conceptual mappings of scalar interpretations on to specific linguistic items, in this case quantifiers, are hypothesized to be what undergoes attrition, not the conceptualizations of quantity themselves. It is not enough, however, to simply state that attrition occurs and what might give rise to it. To get at the question of whether attrition is a case of representational changes, in our case conceptual representation as mapped on to linguistic knowledge, or processing challenges in the bilingual mind, we test scalar interpretations using a combination of offline acceptability judgments and online neurolinguistic metrics via event-related potentials (ERPs).

4.3 Scalar implicatures

Deriving meaning beyond the literal or conventional semantics of a given word or sentence is called pragmatic inferencing. Traditionally, research has examined pragmatic inferencing at a linguistic and a psycholinguistic level (see Noveck & Reboul, 2008). In either case, the dominating phenomena of testing are quantifiers such as <*some, most, all*> in English that have multiple meanings. For *some,* most—if not all—native speakers of English would be able to arrive at a meaning beyond its conventional semantics:

1. Some Catalonians want independence. → *Not all Catalonians want independence.*

This non-semantic meaning derived from *some* is called a scalar implicatum (henceforth SI).[10] It is *scalar* because the lexical item that gives rise to it sits on a scale (like the one above) with other similar quantifiers whose inherent informativity is stronger or weaker than that of the other members of the scale. It is an *implicatum* or *implicature* because it is implicit rather than explicit (e.g., Grice, 1975, 1989, Horn, 1972, 1992). We say that most if not all native English speakers could get this implied meaning because there are some that do not, at least not without (experimental) manipulation. It has been the focus of SI research, both the purely linguistic and the psycholinguistic, to explain the derivation of upper-bounded 'not all' interpretations of *some.* One line of thought posits that SIs are either automatically derived or context sensitive.

[10] H.P. Grice used *implicatum* to refer to the implied content of an utterance itself and *implicature* to the concept of implicated content more generally. In this paper, we will use *implicature* to refer the implied content given post Gricean research conventions.

These are the default and context-driven models respectively. Default models (e.g., Levinson, 2000) propose that SIs emerge automatically any time *some* is encountered and, if needed, later reanalysis of the implicature occurs when disambiguating information becomes available. Alternatively, context-driven models (e.g., Carston, 1990; Hirschberg, 1991; Sperber & Wilson, 1995) claim that SIs arise only when the context, sufficiently stocked with information, calls for them.

In addition to such models, other research has claimed that because *some* is scalar and, therefore, has a weaker status compared to other members of the scale, its possible pragmatic meaning is derived on the basis of scaled informativity (e.g., Chierchia, Fox, & Spector, 2012; Horn, 1972; Levinson, 2000; Sauerland, 2012). This is nicely exemplified in example (1) above. Grice (1975, 1989), for example, proposed that speakers adhere to a communicative protocol in order to be maximally efficient and cooperative during communication. This protocol, specifically the Maxim of Quantity as a constituent of the larger Cooperative Principle, postulates that language users will be as informative as needed, but not more so and they will say only what is relevant to the purposes, timing and direction of the exchange. Thus, if a speaker chooses to use the word *some* as opposed to *all,* there must be some reason for it, which is that *all* is not applicable or is not true.

Furthermore, other relevant research has concluded that SIs are the result of a particular contribution, or lack thereof, to the truth-conditional content within an expression and not necessarily the implied content (e.g., Carston, 1998; Horn, 1992) and there is also evidence that SIs emerge on the basis lexical calculations (Levinson, 2000) but also higher-level structure (Chierchia, 2006; Chierchia et al., 2012; Katsos, 2008). The preponderance of available data considered together seems to favor the context-dependent models to SI derivation, though some evidence does support the default accounts. However, the explanation of individual differences in interpretation of SIs is not easily explained by any approach and thus an open question.

In general, young children are found to be less pragmatically inclined at a group level than adults when confronted with underinformative sentences (though see Syrett, Lingwall, Perez-Cortes, Austin, Sánchez, Baker, Germak, & Arias-Amaya, 2017 for results in Spanish among bilingual children). In early work, this difference was attributed either to young children lacking the pragmatic prerequisites to derive pragmatic interpretations, a cognitive skill that develops with time, or young children do in fact interpret pragmatically, but they do so less frequently than adults (Chierchia 2006; Guasti et al. 2005 for review). However, more recent research stemming from findings from Guasti et al. (2005), Papafragou & Musolino (2003),

and Degen & Tanenhaus (2011, 2015)—to name a few—shows that enriched instruction and/or specific methodological design can increase the rate with which an upper bound (pragmatic) interpretation is derived among both children and adults, thus challenging the early idea that lack of SI derivation in healthy populations was attributable to potential cognitive disparities between young children and adults. The root source of individual variation among adults, however, is still relatively unclear.

In the pursuit of exploring such variation, many researchers have probed the status of the underlying cognitive mechanisms responsible for drawing pragmatic inferences, particularly among those who do not (by majority) derive such meanings from ambiguous language. Recent neurolinguistic investigations, for example, have shown that the ability to call upon pragmatics when inferencing is related to working memory (WM) capacities (Marty & Chemla, 2013), scores on the Autism Spectrum Quotient (ASQ) (Nieuwland, Ditman, & Kuperberg, 2010; see also Heyman & Schaeken, 2015; Zhao et al. 2015), or even scores based on the Systemizing Quotient-Revised (SQ-R) questionnaire. In this final case it has been said that pragmatism scores would tend to increase in proportion to higher SQ-R scores, though only if the scores were taken as an index for participants' ability to discover the nature of the experiment and, thus, distinguish informative from underinformative sentences (Barbet & Thierry, 2016).

While the above studies offer somewhat mixed results, it is generally accepted that there is something, be it cognitive, linguistic, task-related, or all of the above which motivates the interpretative variability of SIs across individuals. It is important to highlight that no researcher, to our knowledge, is claiming that because SIs are not derived by some individuals when they are by others, those who do not must have deficient cognitive abilities. Rather, scores on cognitive metrics like those mentioned above are merely used as tools to determine why there might be such high rates of inter-individual variation in SI derivation. While the present study does not test the predictions of the above hypotheses, it does offer new insight into the workings of the bilingual brain. We are particularly interested in understanding the effects of prolonged exposure to an L2 that works differently from the L1 regarding SIs and how these effects come to bear on pragmatic interpretations. We paraphrase results from AUTHORS (submitted) for a an analysis of our L1 control data. We will present a summary of the L1 data in the results section here for ease of comparison to the bilingual data.

4.4 Scalar implicatures in Spanish

Spanish, in the case of SI research, is distinct from most other test cases in the literature to date in that it has two plural scalar quantifiers that mean 'an indefinite and non-specific plurality of things': i) *algunos/as* 'some-plural' and ii) *unos/as* 'some-plural'. Both determiners must be a referent of a plural entity greater than or equal to two, which causes an overlap in semantic distribution, while having idiosyncratic pragmatic distributions.[11]

2. *Algunos/unos perros están jugando.*

'Some/some dogs are playing.'

Gutiérrez-Rexach (2001), Martí (2008), Vargas-Tokuda, Grinstead & Gutiérrez-Rexach (2009) and Alonso-Ovalle & Menéndez-Benito (2002), for example, argue that *algunos* (being presuppositional) is compatible with a quantity implicature and *unos* is not. For Gutiérrez-Rexach, this difference may be due to a semantic "no linking" constraint on *unos* ([+D-linked]) or feature pertinent only to *algunos*. While each quantifier is able to take a subset reading, it is only *unos* that is said to be able to refer to whole and partial sets equally. Due to this no-linking constraint, however, *unos* should not be able to relate back to a referent already expressed in the discourse because it does not have a specific/partitive interpretation. *Algunos*, on the other hand, is unaffected by previously instantiated discourse-related referents.

3. *Los libros de matemáticas están en el cajón, los de física debajo de la cama hay ??unos/algunos de lingüística sobre la mesa.*

"The math books are in the drawer, the physics ones are under the bed and there are some linguistics ones on top of the table."

[Vargas-Tokuda et al. 2009, p.103, example (37)].

For Martí (2008), the difference is one of morphology, whereby *algunos* is merely the partitive prefix *alg-* (from *algo* 'something') attached to *unos* 'some'. Thus, *algunos* is the only determiner of the two that is associated with a quantity implicature. *Unos* can, however, have

[11] The conditions in which *algunos* and *unos* demonstrate distinct behaviors are both too great in number to explain in detail here (e.g., interaction with verb class, predicate type, and contrastive focus. See Vargas-Tokuda et al. 2009 for review) and irrelevant for the purposes of this paper. However, the methodology was designed in a way as to avoid variables that could inadvertently interfere with an objective analysis. For the remainder of this section, we focus only on the aspect of the quantifiers that is argued to be related to SI derivation.

a ‘some but not others’ reading if it is contrastively focused as in (4). The pragmatic status of *algunos* is seen in (5) where its ‘not all’ interpretation can be cancelled or defeated.

4. *Unos perros están ladrando, otros no.*

‘Some dogs are barking, others are not.’

5. *Vinieron algunos/#unos amigos, de hecho todos.*

‘Some friends came, in fact all of them did.’

Finally, Alonso-Ovalle & Menéndez-Benito (2002) hold that the pragmatic approach (de Hoop, 1992; Diesing, 1992; Partee, 1989) and the ambiguity approach (Buring, 1996) to explaining the behavior of weak determiners are both needed to explain the differences between *algunos* and *unos*. For example, Alonso-Ovalle and Menéndez-Benito claim that *unos* can be presuppositional, but when taking this reading it generally marks a change in topic or induces contrastive focus. On the other hand, they claim that *algunos* is both presuppositional and non-presuppositional and neither reading is at all dependent on the topic/focus marking characteristics of the sentence in which it appears.

Semantic theory of *algunos* and *unos*, however, has not predominantly relied on experimental data for its conclusions. By in large, theoretical descriptions of Spanish SIs rely heavily on intuition and non-experimental native speaker judgments. Moreover, a great deal of the experimental work that does exist has collected data from Spanish speakers of the Mexican and/or Latin American varieties. Interestingly, (monolingual) Spanish-speaking children seem to appropriately distinguish *algunos* from *unos* (Vargas-Tokuda et al. 2009; Syrett et al. 2017 for bilingual children), while adults have shown a slightly higher degree of variation. That is, some studies have shown that adult native speakers of Spanish readily accept *algunos* in whole contexts and others do not (Miller et al. 2016). With the exception of some slight variation across individuals and populations, however, Spanish has thus far shown no evidence of optionality towards an interpretation preference of *unos* and, therefore, no “logical” or “pragmatic” split between responders as some other work in other languages has shown (Bott & Noveck, 2004; Hunt, Politzer-Ahles, Gibson, Minai, & Fiorentino, 2012; Noveck & Posada, 2003; Politzer-Ahles et al., 2012; Spychalska, Kontinen, & Werning, 2016; Tavano, 2010). This is sensible given that *algunos* is said to always be partitive due to its morphology or inherent semantic features constraining its meaning—except in certain contexts such as downward entailment relations—and *unos* can be either partitive or whole. Crucially, available data in the Spanish experimental literature show that *unos* can be either partitive or whole in

equal measure with no apparent preference for one reading over the other, supporting the idea that *unos* is not associated with an implicature. As we will see below, however, the present data embody a stark contrast to other data sets available for Spanish.

4.5 ERP research and scalar implicatures

Event-related potentials (ERPs) provide a non-invasive method to investigate electrophysiological correlates of mental processes. They are small voltages that are generated in the brain when large groups of neurons fire in synchrony, which can be elicited by an array of cognitive, sensory or motor events and, in general, can be divided into two categories: early and late waves (see Kok, 1997 for review). The early waves, which peak typically within the first 100 milliseconds post stimulus, are called 'sensory' or 'exogenous' because they depend largely on the physical characteristics of the stimulus (e.g., auditory vs. visual). The late waves are those that reflect the manner in which the subject evaluates the stimulus and are termed 'cognitive' or 'endogenous' because they examine information processing (e.g., how the stimulus is being evaluated in real-time). All waveforms are described based on latency and amplitude, which can be seen distributed across the scalp via an electrode cap.

There are many advantages to using ERPs to investigate linguistic phenomena, but one of the main advantages is that we know the brain responds uniquely to different aspects of language processing and these distinct reactions give rise to distinct ERP signatures. As an example, the brain responds differently to morphosyntactic violations as instantiated in a gender or number violation than it does to a semantic incongruency or a failed expectation of upcoming information. Though not ideal for spatial resolution due to the inherent complexity in the layout of what are called dipoles—an oriented flow of current—another advantage of using ERPs is that they provide high temporal resolution, allowing for a better examination of a participants' sensitivity to a given stimuli at its precise onset and computation in real time. Therefore, if our bilinguals and control group show qualitatively different brain responses to the same property, this may be a reflection of differences at a so-called level of linguistic representation that ultimately leads to the use of distinct processing mechanisms (see Alemán Bañón, Fiorentino & Gabriele, 2014).

We will provide a brief explanation of two common ERP components associated with, though not specific to, linguistic processing. First, morphosyntactic violations among native speakers (generally) elicit a P600, which is a positive going waveform that emerges roughly between 500-900ms in central-parietal electrodes of the EEG cap (e.g., Hagoort, Brown, & Groothusen, 1993). This effect has been argued to arise due to various aspects of

morphosyntactic processing such as reanalysis (e.g., Osterhout & Holcomb, 1992), repair (see Molinaro, Barber, & Carreiras, 2011 for review) and integration (e.g., Kaan, Harris, Gibson, & Holcomb, 2000). It is important to note that while the P600 does emerge reliably for morphosyntactic violations, it is not syntactic in nature. For example, some work has shown that the P600 emerges in animacy violations (Kuperberg et al, 2003, 2006, 2007), thematic role reversals (Kim & Osterhout, 2005), conflicting information between local coherence and global plausibility (Van Herten et al. 2006), strong semantic violations (Delong et al. 2014), and also incongruous emoji usage (Weissman & Tanner, 2017).

Another component related to linguistic processing is the N400. This component is said to reflect lexico-semantic processes and it emerges as a negative going wave roughly between 250-500ms. There is a great deal of research showing that the N400 is an index of the strength of lexical associations (e.g., Kutas & Hillyard, 1980; see Lau, Phillips, & Poeppel, 2008 for contemporary review). It is worth noting that while specific ERP components associated with language processing of one kind or another (e.g., P600 or N400) arguably may not reflect linguistic representation per se, they emerge reliably in monolingual and advanced bilingual datasets examining any given linguistic property. Thus, if such effects are found here to emerge differentially based on learner type, we can say minimally that the controls and bilingual attriters have distinct electrophysiological processing responses to the same stimuli. Again, it is important to highlight that while the N400 and P600 reliably emerge as responses to specific linguistic stimuli, the characterization of ERP responses as either N400 or P600 (but not both) is not accurate. Tanner (2017), for example, points out that the cognitive mechanisms underlying N400 or P600 effects are not mutually exclusive, rather they can interact with each other and co-occur in time and space. Furthermore, individual brain responses can differ.

Because the N400 is associated with content stimuli, particularly semantic (im)plausibility and/or its more general correlation to prediction, it has has been the focus of early and contemporary ERP studies examining the processing of underinformative sentneces with *some* or quantifiers like it (e.g., Nieuwland et al. 2010; Noveck & Posada, 2003). In these types fof studies, participants are prompted to make acceptability judgments of underinformative sentences using their knowledge of the world. Violations of world-knowledge contexts are associated with more robust N400 effects (see Hagoort et al., 2004; Hald, Steenbeek-planting, & Hagoort, 2007). Researchers are careful with their interpretation of emerging N400 effects in these studies, as well as the methodological design, however, as the N400 can be modulated by the lexico-semantic relationship of words. One way of measuring this relationship is to examine the frequency with which words co-occur in specific

contexts (e.g., dogs and bones versus dogs and airplanes) and is referred to as the *latent semantic analysis* value or LSA (Landauer et al., 1998). While larger N400 effects are expected in designs that employ underinformative *some* in world-knowledge contexts compared to informative ones, this effect can be attenuated by the LSA value of the main noun phrase and the target predicate in informative contexts whose noun phrase and predicate have a stronger lexico-semantic relationship. Attenuation also occurs in thoroughly false sentences such as *some crows have radios* (see Noveck & Posada, 2003).

Though not all SI studies using ERPs have probed world knowledge violations, most of these studies have still found an N400 or related effects like the Nref (negative reference), that is, a sustained late negativity (e.g., Politzer-Ahles et al. 2013; but see also Hartshorne et al., 2014 for positivity). With this in mind, we hypothesize that the N400 will be the component most likely to emerge in our data.

4.6 The present study

4.6.1 Specific research questions and hypotheses

In an effort to address the macro-question of whether or not the effects of attrition emerge at the level of processing only or reflect bona fide changes to previously acquired grammatical representations, we first entertain the following micro-questions:

1) Will scalar implicature interpretations on Spanish *algunos* show optionality in attrited bilinguals as compared to the control group?

Hypothesis: If the IH is on the right track, we expect to find some degree of optionality with respect to the interpretation of *algunos* in infelicitous contexts as compared to controls. Thus, we predict that attriters will more readily accept infelicitous uses of *algunos* across tasks.

2) If the hypothesis is confirmed for question one, at what level will we find the optionality (i.e., online or behavioral or both)?

Hypothesis: Given the mixed results from available data, we expect to find deviation from monolingual norms for bilinguals at least in the offline tasks. Whether or not we will find associative effects in the ERP results is an open question. However, it is plausible that we will find corresponding brain responses to pragmatic infelicity in bilinguals attriters as is the case for controls (see AUTHORS; Spychalska, Kontinen, & Werning, 2016).

3) Will bilingual attriters show a similar degree of response preference for Spanish *unos* as compared to controls?

Hypothesis: The L1 Spanish-L2 English bilinguals in Miller et al. (2016) showed equal acceptance of *unos* in both partitive and whole contexts. While the authors attributed some of the variation of their results to potential methodological confounds, it is possible that bilingual attriters as compared to controls will not show any distinct preference for *unos* in either condition. However, AUTHORS presents a contrast to Miller et al. (2016) in that the controls in fact do show variable interpretations on *unos* in a systematic way. Thus, this is an open question.

4) If the hypothesis is confirmed for question three, at what level will we find the preference (i.e., online or behavioral or both)?

Hypothesis: Minimally, if there is a preference, we expect to find it at the offline level. Given the scant research in this domain with this methodology and among this population, it is an open question whether the offline data will be corroborated by the ERP results. However, AUTHORS did show that controls' offline data was corroborated by the online data. Thus, we hypothesize that attriters' offline and online performance could be similar. However, it might also be the case that so-called deviation from controls shows up in one or the other methodology, in which case one would need to ponder what such an asymmetry means in terms of what exactly has attrited (mental representations vs. (processing-induced) momentary crosslinguistic influence).

4.7 Methodology

4.7.1 Participants

Data were collected from 30 (peninsular) Spanish-English bilinguals (18 females, age range 24-48, M = 28.7, SD = 3.3) living in the UK for lengths between 5 and 24 years. We compare this group to a control group of previously acquired data focusing on L1 processing (see AUTHORS submitted). Participants filled the LEAP-Q questionnaire, which is commonly used to gather relevant language experience and language use data, as well as biographical information (Marian, Blumenfeld & Kaushanskaya, 2007). This was done to measure daily use of each language, in which contexts each language was used, time spent per year in the home country, age of acquisition (AoA), etc. Biographical information affirmed that bilinguals were either married to English spouses and/or full-time employees in the UK. The controls, on the other hand, recently arrived to the UK from Spain at the time of testing.

All bilingual attriters were either in the process of finishing or had completed a university degree in the UK, arriving on average around the age one finishes A-levels (high

school). They reported that they did not return home more than twice per year and for no longer than three weeks total. All of the recently arrived controls, conversely, had at least completed an undergraduate (some postgraduate) education in their native country. All participants provided their informed consent.

Variable	Descriptives	L1 Controls	Attriters
Number		30	30
Stay in L2 environment	min.	1 week (.25 months)	60 months
	max.	12 months	24 years
	Mean	4.7 months	11.93 years
	SD	1.41 months	4.18 years
Daily Use Spanish	min.	8 hours	2 hours
	max.	12 hours	5 hours
	Mean	10.6	2.6
	SD	1.47 hours	1.69 hours
Daily Use English	min.	2	7
	max.	6	13
	Mean	2.6	10.4
	SD	1.29 hours	1.71 hours
Daily Exposure Spanish	min.	5 hours	1 hours
	max.	9 hours	8 hours
	Mean	7.5 hours	3.5 hours
	SD	1.32 hours	1.98 hours

Daily Exposure English	min.	1 hours	5 hours
	max.	6 hours	15 hours
	Mean	3.9 hours	11.8 hours
	SD	.98 hours	1.33 hours
AoA	Min.	21 years	15 years
	Max.	29 years	33 years
	Mean	26.2 years	22.2 years
	SD	2.38 years	4.27 years

Table 5: Biographical information for attriter and control groups

4.7.2 The Experimental Tasks

Data were collected from three experimental tasks. The first was a picture sentence verification task (PSVT) fashioned after Politzer-Alhes et al. (2013), Spychalska et al. (2016), Tavano, (2010) and Wu & Tan, (2009). This task was designed to have two parts: (i) an untimed offline judgment following each trial and (ii) an ERP measure within each trial. Together, the goal of this first task was to examine both the implicit electrophysiological response differences and offline judgment differences between pragmatically infelicitous contexts and semantically correct ones. Hartshorne et al. (2014) point out that though this methodology may not necessarily measure the precise calculation of an SI at its onset, it is a useful metric for examining the onset of a stimulus that mismatches prior context, whether felicitous or not. Thus, we expect to see unique ERP signatures for those deriving the upper-bounded meaning of *algunos* and or *unos* and those who do not, as well as judgments differentially affected by pragmatic context.

The third experiment was a non-binary free interpretation task. The goal of this experiment was to allow participants to freely indicate the quantity associated with each scalar term without being biased by cues in the experimental discourse and without being constrained by time. Bott & Noveck (2004) and De Neys & Shaeken (2007) both found that with increased

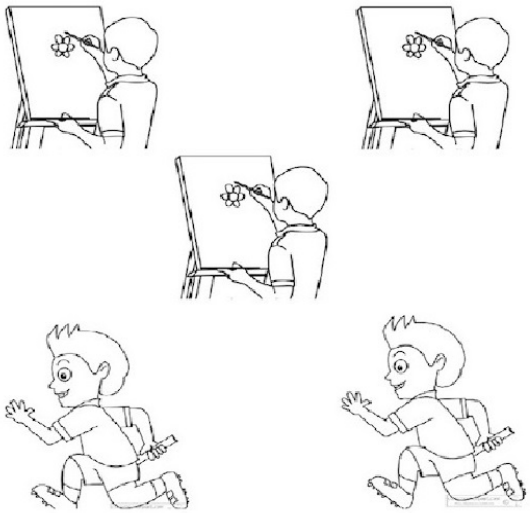
cognitive load (e.g., time constraints), individual interpretation patterns changed. Thus, in order to allow our participants to reason without relying on world knowledge or being constrained by time, the third experiment prompts comparatively explicit responses. We begin the following sections by presenting the nature of the tasks separately, followed by in-depth analyses of each experiment both within and across groups.

4.7.3 Experiment one – Picture-sentence verification task

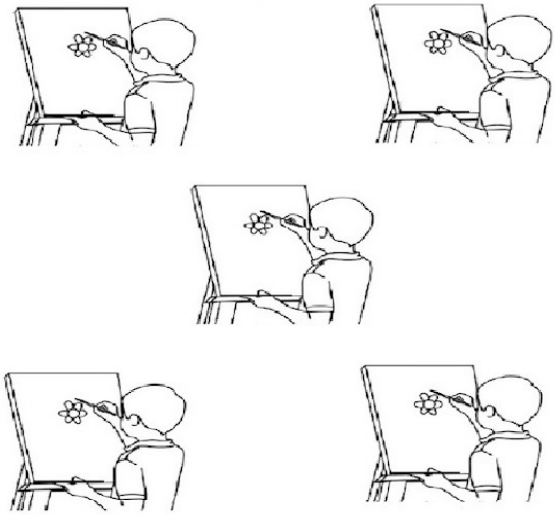
In each trial, either a SOME-type picture was presented in which some but not all the characters were in involved in a target activity, or an ALL-type picture was presented in which all characters were involved in the target activity. This was done using the quantifiers *algunos, unos* and *todos* 'all', where *todos* served as a control condition (see Figure 13). We take the opportunity again to highlight that, in general, *unos* can be both partitive and whole given that it lacks semantic features that would otherwise lead it to have a specific partitive interpretation. Thus, while *algunos* has only one felicitous and one infelicitous condition and *todos* has a correct and incorrect condition, *unos* has two (partitive and whole) felicitous conditions, a finding that is uncontroversial in the Spanish experimental literature so far. *Unos,* thus, should have no better or worse answer depending on context. In order to make it infelicitous, we would have needed to design pictures in which either none of the characters were performing the target action or only one of them was and the others were not. However, Spanish expresses gender and number morphologicall and, thus, a mismatch between the critical images and sentences would emerge primarily due to morphological cues in the linguistic contexts preceding the critical quantifier phrase (QP) rather than the pragmatic context. Where ungrammaticality/infelicitousness for such trials becomes detectable at the onset of the verb preceding the QP, unwanted noise would be added to the EEG signal. Our pilot data for the experiment, which contained such trials, did not show differences in overall offline performance, thus they were removed from the present study.

The carrier phrase *En esta imagen* 'In this image' preceded each image and was followed by a sentence that described the image correctly/felicitously or incorrectly/infelicitously. The QPs were followed by a prepositional phrase (PP) such as "at the house" or "in the street". This was done so that critical QP did not appear sentence finally nor would the EEG signal be time-locked to this position, thus avoiding noise in the EEG signal caused by a general wrap-up effect.

Algunos felicitous

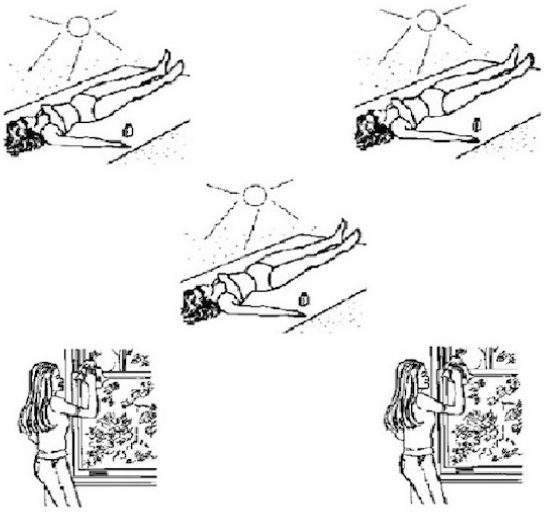


Algunos infelicitous

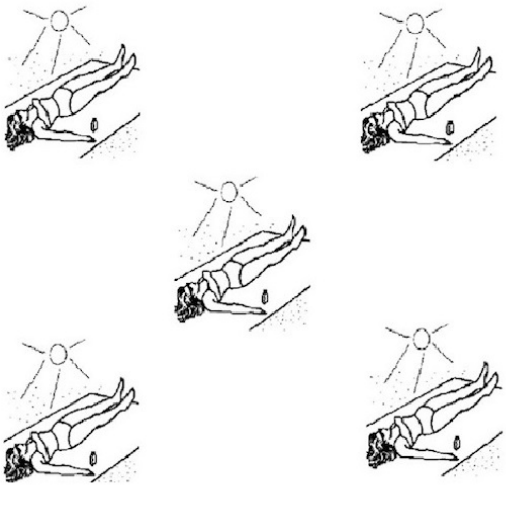


Pintar es lo que hacen algunos chicos.
'Painting is what some boys are doing.'
"Some boys are painting."

Unos felicitous



Unos felicitous



Broncearse es lo que hacen unas chicas
'Sunbathing is what some girls are doing.'
"Some girls are sunbathing."

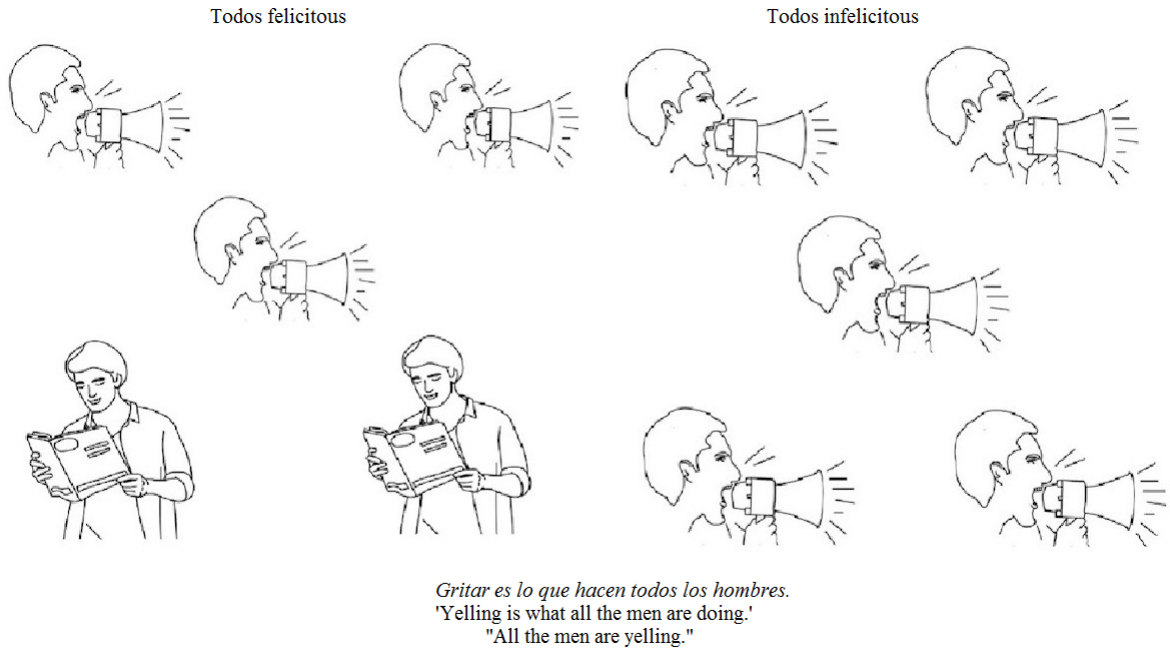


Figure 20: Sample pictures and sentences from Experiment A

Recent ERP work on SIs using picture tasks such as the present study has suggested that putting the quantifier at the front of the sentence (e.g., ***some*** *girls are sitting on sofas*) does not provide enough prior information to make an accurate assessment of the status of the quantifier (see Politzer-Ahles et al. 2013). Fortunate for the present study, in Spanish the infinitive verb, like the gerund form in English, can be nominalized and appear as the subject of a sentence such as *Escalar es peligroso* 'Climbing is dangerous". In an effort to keep the QPs from appearing prematurely, we placed the nominalized verb at the front of the sentences. Participants completed eight practice trials to acquaint themselves with the non-canonical syntax used in the experimental trials. However non-canonical the syntax, however, it is entirely licit. Subsequent to each trial, the participant indicated with a computer mouse whether the sentence described the image well or not, being instructed to use just their natural intuitions of Spanish. Correct and incorrect *all* conditions served as a controls, providing a 3 (Quantifier) x 2 (Type) design.

4.7.4 Materials and procedure – Experiment A

180 sets of black and white pictures arrays were used for the target trials. Fillers were made from unique sets of similar image arrays, thogh not using animate objects. The images were retrieved freely from Google and were adjusted with Paint or Adobe Photoshop when necessary. Fillers were truth value in nature but did not involve quantification. For example, in an image containing five shapes, one of the shapes would be different from the rest. The unique

shape would appear in a particular location on the screen and the accompanying sentence(s) either correctly or incorrectly identified the location of the different shape. Target and filler sentences were matched for length.

Target picture arrays each included five characters performing some activity with contexts that would render the use of the quantifier either felicitous or infelicitous. Target conditions consisted of 30 individual trials (see Table 2). Quantifiers did not appear in conjunction with the same image more than once for any individual trial. For example, infelicitous and felicitous *algunos* conditions shared the same sentence, though an ALL-type picture was used for infelicitous conditions and a SOME-type picture for the felicitous ones. Distinct verbal predicates were used for all of the quantifiers and their relevant individual trials to ensure that no sentence or image would be repeated within the counterbalanced trials. There was an equal number of felicitous and infelicitous trials throughout in order to avoid biasing response patterns.

Condition	**Algunos**	**Unos**	**Todos**	**Fillers**
Felicitious	*n* = 30	*n = 60*	*n* = 30	*n = 90*
Infelicitous	*n* = 30	N/A	*n* = 30	*n =120*
Total	*n* = 60	*n* = 60	*n* = 60	*n =210*

Table 6: Total number of conditions

Seven lists were created for the presentation of the stimuli. The lists were made by counterbalancing the order and type of the stimulus being presented. Each list began with one of the target condition sentences and the rest was randomized. Stimuli were presented in 10 blocks of 36 sentences with a short break between each block. Images were preceded by a 250 ms fixation cross and were presented in the center of the screen for 2000 ms. The images were followed by a 250 ms blank interval before beginning the presentation of the sentences. Individual words were presented for 350 ms followed by a 250 ms blank screen. A randomized blank inter-trial interval between 500-1000 ms was inserted prior to each subsequent trial.

4.7.5 Experiment B – Non-binary free interpretation task

Deriving a pragmatic inference from a scalar term requires some interaction between a speaker and a listener or an interpreter and some utterance. Crucially, the listener or interpreter makes a judgment about the expressed content of an utterance based on cues in the discourse. In order

to ensure that a listener's or interpreter's judgment is not biased, she must not be primed by infelicity already contained within the discourse (e.g., *some elephants have trunks*, or infelicity via pictures); she must also not be primed by instruction (e.g., Barbet & Thierry, 2016; Chemla & Bott, 2014) or asked to rely on reasoning of world knowledge rather than strict interpretation. In this experiment, each trial defines a set of characters and asks the participant to choose all possible quantities associated with the scalar term used in the context (see Figure 4).

Algunos and *unos* each appeared in six conditions with four, five, or six referents respectively in order to determine if set size differentially affected the overall interpretation of each scalar term in Spanish as it does in English when going beyond the subitizing range (i.e., the range within which counting is still possible without significant delay, 1-4) (e.g., Degen & Tannenhaus, 2011, 2015).

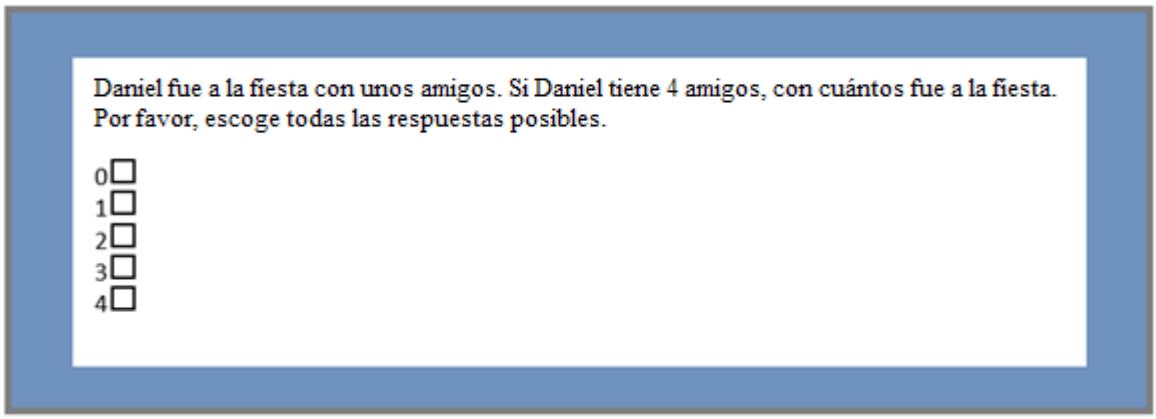


Figure 21: Sample of third experiment Unos condition

In Figure 2, the context is translated as follows: "Daniel went to the store with some friends. If Daniel has four friends, how many did he possibly go to the store with? Please choose all possible answers."

4.7.6 Materials and procedure – Experiment B

Each quantifier appeared three times in each context of four, five or six characters performing an action, totaling nine targets for each. 36 unique fillers were inserted to give a ratio of one target to two fillers. Fillers were made from creating simple arithmetic problems such as "John needs to scale a 100-foot wall. If he climbs at a rate of 10 feet per minute, how long will it take him to reach the summit? Please choose the best answer(s)." In this respect, though we did not ask for judgments of scalar quantifiers per se in the fillers, quantification in a broad sense was the theme of the task in order to maintain a parallel between targets and fillers. This experiment

was completed using SurveyGizmo on a computer with internet in the laboratory where Experiment A was administered. There we no practice trials for this experiment.

4.8 Results

4.8.1 Offline judgments for Experiment A

4.8.1.1 Native controls

Data were analyzed with a binary logistic regression model and pairwise contrasts with Bonferroni correction where appropriate using SPSS, which, in order to get the degrees of freedom to run F-tests on individual variables and their interactions, uses the Satterthwaite Approximation within the model. The dependent variable for the experiment was Acceptance and significance was set at $\leq .05$. We included variables of Condition (partitive versus whole), Word (algunos versus unos), and their interactions. Study 2 (Authors, submitted) showed that the effect of Word was primarily driven by the *algunos/unos* distinction such that *unos* was markedly more acceptable in whole conditions than *algunos*. Furthermore, *algunos* partitive conditions were more acceptable than their whole counterparts. This was taken as evidence that a distinction was not only being made between the two lexical items themselves, but crucially that the (in)felicitousness of algunos in partitive versus whole contexts is reliably affected by the context of the pictures.

One interesting finding from these results was the distinction in the way individuals treated *unos*, which is claimed (theoretically) to be equally acceptable in whole and partitive conditions, a claim confirmed in the previous literature. Our native controls were divided into categories of "logical" and "pragmatic" responders whereby the "pragmatic" responders did not accept *unos* in whole contexts, and the "logical" responders accepted it in both whole and partitive contexts equally. Where there was a clear division between those who were logical and those who were pragmatic, *unos* was accepted by and large in partitive contexts and it was only the whole contexts that differentiated responders.

4.8.1.2 Attriters

Data were also analyzed with a binary logistic regression model and pairwise contrasts with Bonferroni correction where appropriate using SPSS. In order to get the degrees of freedom and run F-tests on individual variables and their interactions, the Satterthwaite Approximation was used. The dependent variable for the experiment was Acceptance and significance was set at $\leq .05$. We included variables of Condition (partitive versus whole), Word (algunos versus unos), AgeofAcquisition, Time in L2 Country and all higher-order interactions of these

variables. Predictor variables were numerically coded within the model, being assigned a unique variable number to distinguish them from the others. The model included random by-subjects intercepts and slopes. There was a significant main effect of Word ($F(2,3834) = 97.39$, $p < .001$, and AoA ($F(6,3834) = 18.07$, $p = .003$. Additionally, there was a significant interaction of Word*Condition*TimeinL2 ($F(25,3834) = 11.25$, $p < .001$, suggesting the longer one remains in the L2 environment, the more their interpretations were distinct from the controls'. No other effects were significant. Pairwise contrasts showed that *algunos* partitive ($M = .16$, $SE = .099$) compared to whole ($M = .182$, $SE = .155$) conditions were treated similarly: contrast = -.075, $t(3,834) = -1.26$, $p > .05$. Furthermore, *unos* partitive ($M = .288$, $SE = .214$) and whole ($M = .219$, $SE = .215$) conditions were also treated similarly: estimate = -.003, $t(3,834) = -.088$, $p > .05$. For the bilinguals, the only contrast that proved significant was that of the *todos* true and false conditions: estimate = .728, $t(3,834) = 19.59$, $p = .001$. Intergroup comparisons revealed significant variables of Word ($F(1,254) = 98$, $p < .001$ and significant interactions of Group*Word ($F(6,254) = 101.42$, $p < .001$ and Group*Word*Condition ($F(9,254) = 824.1$, $p < .001$. Figure 3 provides an overview of the mean acceptance of each group's responses.

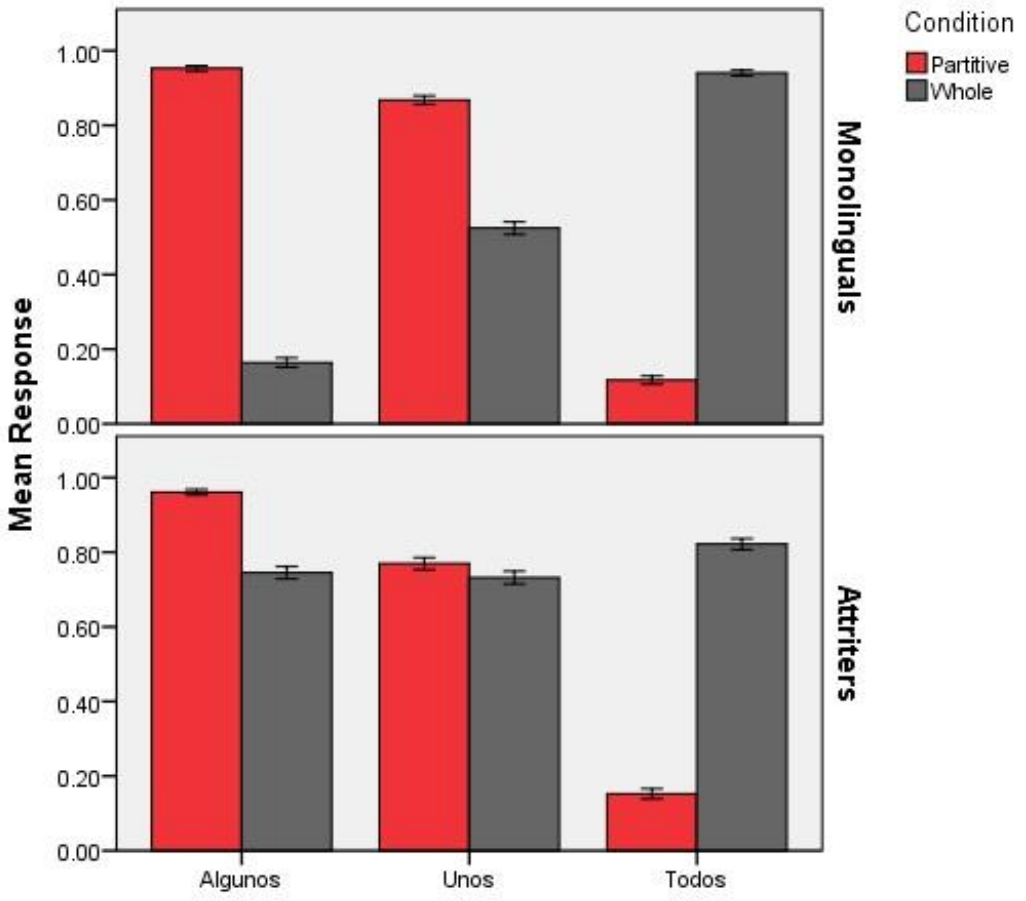


Figure 22: Mean acceptance of each quantifier in context for native controls and attriters

No other variables proved significant. Pairwise contrasts of the Word variable showed that overall *algunos* (*M* = .269, *SE* = .173) is not treated differently than *unos* (*M* = .252, *SE* = .166): estimate = .017, *t*(9,254) = 1.244, *p* = .214. However, *algunos* is treated differently from *todos* (*M* = .491, *SE* = .220): estimate = -.222, *t*(9,254) = -4.544, *p* < .001. Further analysis of the significant interactions revealed that the primary source of difference between the two groups lies in their responses to the *algunos* whole conditions whereby attriters accepted them significantly more than their control counterparts: estimate = -.604, *t*(9,254) = -2.152, *p* = .031.

4.9 Offline judgments for Experiment B

4.9.1 Native controls

Study 2 (Authors, submitted) showed while none of the *algunos* conditions was different from the other, each was significantly different from its counterpart *unos* conditions such that *algunos* was more clearly partitive and *unos* (generally) was either partitive or whole. Unlike previous research in Spanish, especially Miller et al. (2016), our natives make a more reliable distinction between the two quantifiers and, crucially, interpret *algunos* as partitive (see AUTHORS, submitted).

4.9.2 Attriters

For the purposes of making inherently non-binary responses amenable to a binary regression, those responses which constituted a partitive-type judgment were assigned one number (i.e., 1) while those responses that were whole-type were assigned another number (i.e., 2), thus collapsing raw responses into binary responses within the statistical model. The dependent variable for the model was Response Type, this being either partitive or whole. Furthermore, we included variables of Word (*algunos* versus *unos*), Condition (*x* of 4, *x* of 5, *x* of 6), AoA, TimeinL2 and all higher-order interactions of these variables. As with Experiment A, the predictor variables were coded by assigning each a unique number within the model.

The model revealed a main effect of Condition only, ($F(2,408) = 5.923$, $p = .003$, whereby the quantity of referents within the experimental context differentially affects the likelihood of *algunos* being interpreted as whole or not. *Algunos* conditions in which 4 referents were present ($M = .180$, $SE = .221$) were significantly different from those where 6 were present ($M = .707$, $SE = .373$): estimate = -.527, $t(408) = -4.69$, $p < .001$. *Algunos* conditions where 5 referents are present ($M = .293$, $SE = .310$) are significantly different from those where 6 are present: estimate = -.414, $t(408) = -5.1$, $p = .002$. In other words, the greater the reference set, the less likely *algunos* will be interpreted as whole (Degen & Tanenhaus 2011, 2015). This same effect was not seen for *unos* in any context. Figure 4 provides an overview of mean responses within this task for both groups.

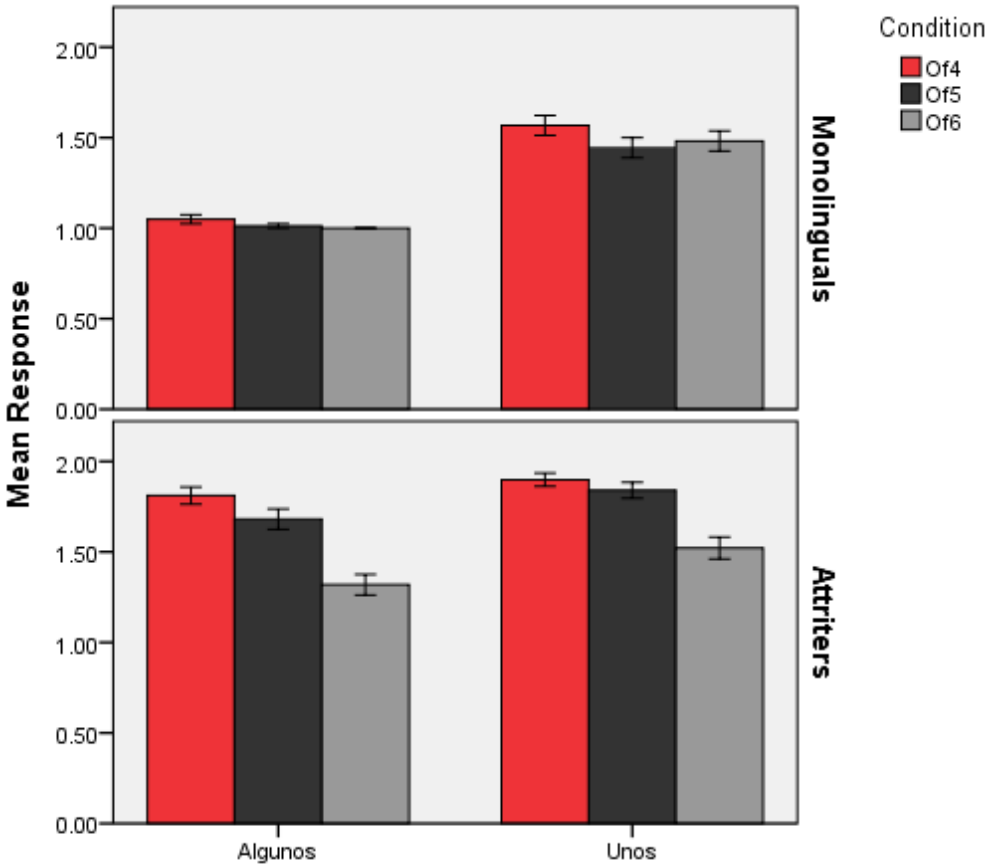


Figure 23: Mean acceptance for responses across both groups

Intergroup comparisons reveal that the L1 controls and the attriters have significantly different performances on this task. The model showed significant effects of Condition ($F(2,888) = 21.533$, $p < .001$, Word ($F(1,888) = 97.501$, $p < .001$, a Group*Word interaction ($F(1,888) = 26.608$, $p < .001$ and finally a Group*Condition*Word interaction ($F(6,888) = 5.05$, $p < .001$. Pairwise contrasts showed that these main effects and interactions arose primarily due to the distinction between *algunos* being used in conditions of 4 and 5 referents wherein the attriters are more likely to associate *algunos* with whole readings. The difference between *algunos* of 4 conditions for controls ($M = .9$, $SE = .126$) and attriters ($M = .180$, $SE = .206$) was significant: estimate = .719, $t(888) = 2.983$, $p = .003$. The difference between *algunos* of 5 conditions for controls ($M = .915$, $SE = .108$) and attriters ($M = .293$, $SE = .289$) was also significant: estimate = .622, $t(888) = 2.016$, $p = .044$. No other conditions proved significant.

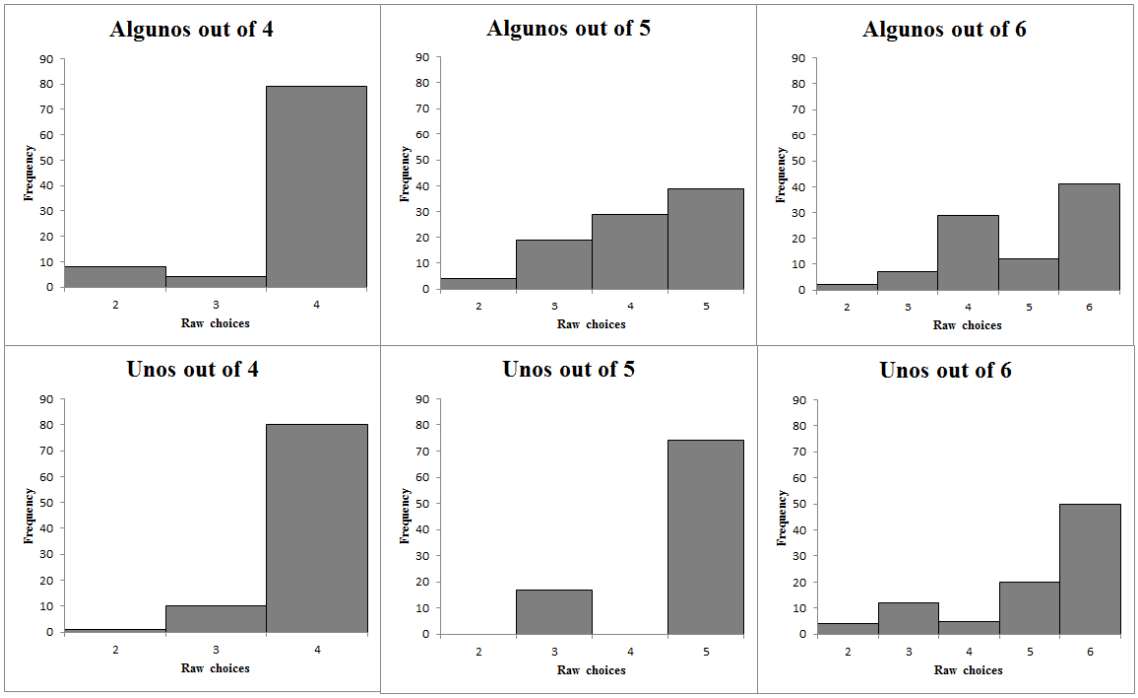


Figure 24: Histograms of individual responses across conditions for bilinguals

Figure 5 provides the frequencies for the individual responses across conditions for biliguals, showing that—compared to the control group—the bilinguals largely accept *algunos* in whole contexts, though this decreases as a function of the number of referents to which the target quantifier refers, and *unos* is largely accepted in whole contexts.

4.10 EEG data

4.10.1 EEG recording and data acquisition

The continuous EEG was recorded from 64 sintered Ag/AgCl electrodes on an Easycap (BrainProducts GmbH, Gilching, Germany). An additional external electrode (IO) was placed on the outer canthus of the right eye to monitor (horizontal) eye movements. EEG was recorded relative to FCz and later re-referenced offline to the average of the mastoid channels. Impedances were maintained below 10 Ω for all channels. Signals were amplified by a BrainAmp MR Plus amplifier (BrainProducts GmbH), applying a bandpass filter of 0.01 to 200 Hz, and digitized at a sampling rate of 1 kHz. Data were later filtered offline using a 30 Hz low-pass filter. Epochs of 1500 ms were selected following the critical phrase and were corrected with a 300 ms pre-stimulus baseline. Raw data inspection was done automatically, rejecting trials due to artefacts of an absolute amplitude difference over 100 V/100 ms, or with activity lower than 0.5V in intervals of at least 100 ms. The minimum number of trials kept within each condition for both controls and attriters was 22 (73%). Overall, artefact rejection

for bilinguals led to the exclusion of approximately 15% of the data prior to grand averaging of the waveforms, inclusive of two bilinguals' recordings due to excessive artefacts.

4.10.2 EEG results

Data were analyzed with a linear mixed effects model and pairwise contrasts with Bonferroni correction where appropriate using SPSS. The dependent variable for the experiment was Mean Voltage. Within the model, we included variables of Condition (partitive versus whole), Word (*algunos* versus *unos*) and the interactions of these variables. Predictor variables were numerically coded within the model.

For the ERP data, following Luck & Gaspelin (2017), we implemented a collapsed localizer analysis in an attempt to remain conservative with our interpretation of results. Because Spanish *algunos* and *unos* have not previously been examined using ERPs, we hypothesized that an N400-like effect may emerge in infelicitous compared to felicitous conditions in accord with previous work on other languages. Though we have no reason to expect major differences in our results, it is possible that a language like Spanish whose quantifiers differ semantically and pragmatically to the other languages tested in the literature would show distinct ERP signatures. Therefore, we explored the data with no concrete commitment to time windows and/or pre-determined regions of interest.

We averaged the waveforms across the conditions that would be used for comparison, later using the timing and scalp distribution from the collapsed waveforms to define the analysis parameters that would be used for the data that had not been collapsed (see Luck & Gaspelin, 2017). Because our working hypothesis was that we would find an N400-like effect for felicitous compared to infelicitous conditions, we first averaged across the relevant conditions. The time window and electrodes showing the greatest negative activity were used to measure the effect in each of the subsequent (relevant) conditions separately, which revealed a bias in centro-parietal electrode sites predominantly in the right hemisphere.

The electrode sites showing the greatest negative effect for *algunos* comparisons, for example, were: Pz, P1, P2, P3, P4, P6, P8, CPz, CP2, CP4, PO2, PO4, PO8, C2, C4. Visual inspection of the waveforms indicated greatest activity for all conditions from 200 to 600 ms post stimulus. Thus, we inspected various sub-intervals within the larger interval (i.e., 200-400, 250-450, 300-500, 350-550, and finally 400-600 ms) in order to determine which time window exhibited the greatest effect. Though each interval visually exhibited an effect, both preliminary collapsing and statistical analyses showed the greatest activity in the 250-450 ms

time window, which is consistent with the N400 latency and which was used for further analysis.

4.10.3 Controls

The model revealed a main effect of Condition. No other effects or interactions were significant. We saw that *algunos* was associated with markedly more negative activity in whole conditions than in partitive conditions and the comparison of whole *algunos* conditions to their *unos* counterparts revealed a distinction such that the latter were less negative than the former. Moreover, *unos* whole conditions compared to their partitive counterparts revealed no differences. The controls also showed a significant effect in the 700-900 ms time window in which a late positivity emerged. While the functional status of this effect is relatively unclear, it was labeled a *Post-N400-Positiivty* following Spychalska et al. (2016) and hypothesized that it could in fact be indicative either of P600-like indices of reanalysis or a "strong semantic violation" (Delong et al. 2014).

4.10.4 Attriters

The model included variables of Condition, AoA, TimeinL2 and all higher order interactions of these variables. Significance was set at $\leq .05$. Overall, the model revealed only a main interaction effect of Condition*AoA ($F(1564) = 3.442$, $p = .002$, indicating that the later (older) one begins L2 acquisition, the more likely she is to process Spanish SIs similarly to adult monolinguals, irrespective of time spent in the L2 environment overall. Moreover, pairwise comparisons showed that *algunos* was not associated with more negative activity in whole conditions (M = -1.12. SE = .624) than in partitive conditions (M = .539, SE = 1.702), $p = .107$. Different from the controls, the attriters exhibited no late positivity for the *algunos* conditions. Additonally, *algunos* partitive conditions compared to *unos* whole conditions revealed a marginal difference, $p = .052$, showing that while *algunos* whole conditions are processed as though there were no infelicity, the *unos* whole conditions do indicate some level of infelicity, which was not reflected in the offline judgments on either task and is not expected according to the literature. There was also a significant difference between the *unos* whole and partitive conditions, $p = .05$, whereas for the controls this condition revealed no difference. See Figures 6-8 for the grand average ERP waveforms and topographical distributions of *algunos* partitive conditions compared to their whole counterparts, *algunos* partitive to *unos* whole conditions, and *unos* partitive to their whole counterparts respectively.

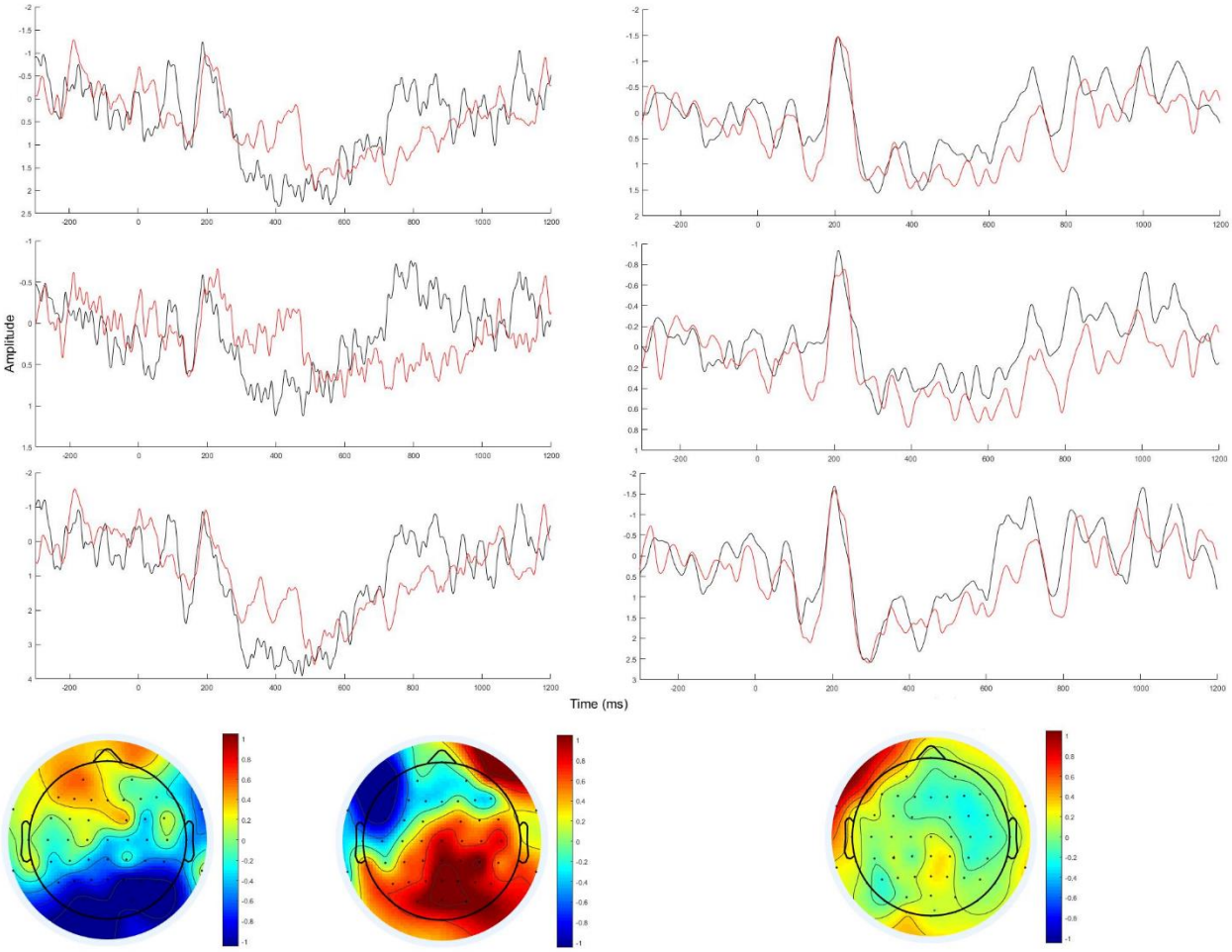


Figure 25: Grand average ERPs at CPz, Pz and Cz for algunos whole compared to algunos partitive among controls (left) and attriters(right). Topographical distribution for same condition made by subtracting felicitous from infelicitous conditions between 200-600 ms (in addition to 700-900 ms for controls).

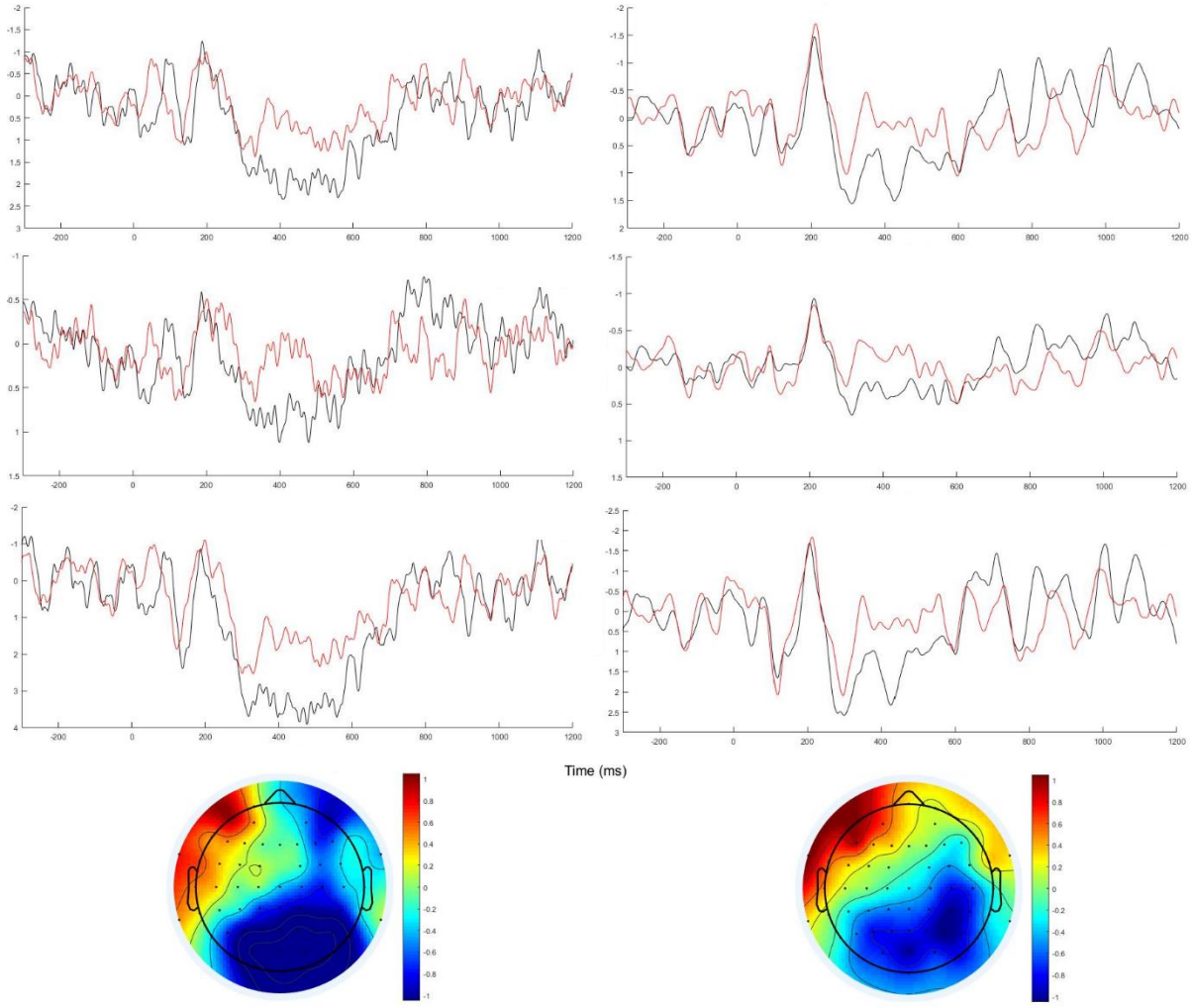


Figure 26: Grand average ERPs at CPz, Pz and Cz for algunos partitive compared to unos whole among controls (left) and attriters (right). Grand average ERPs at CPz for algunos whole compared to unos whole for controls. Topographical distribution for same condition made by subtracting felicitous from infelicitous conditions between 200-600 ms.

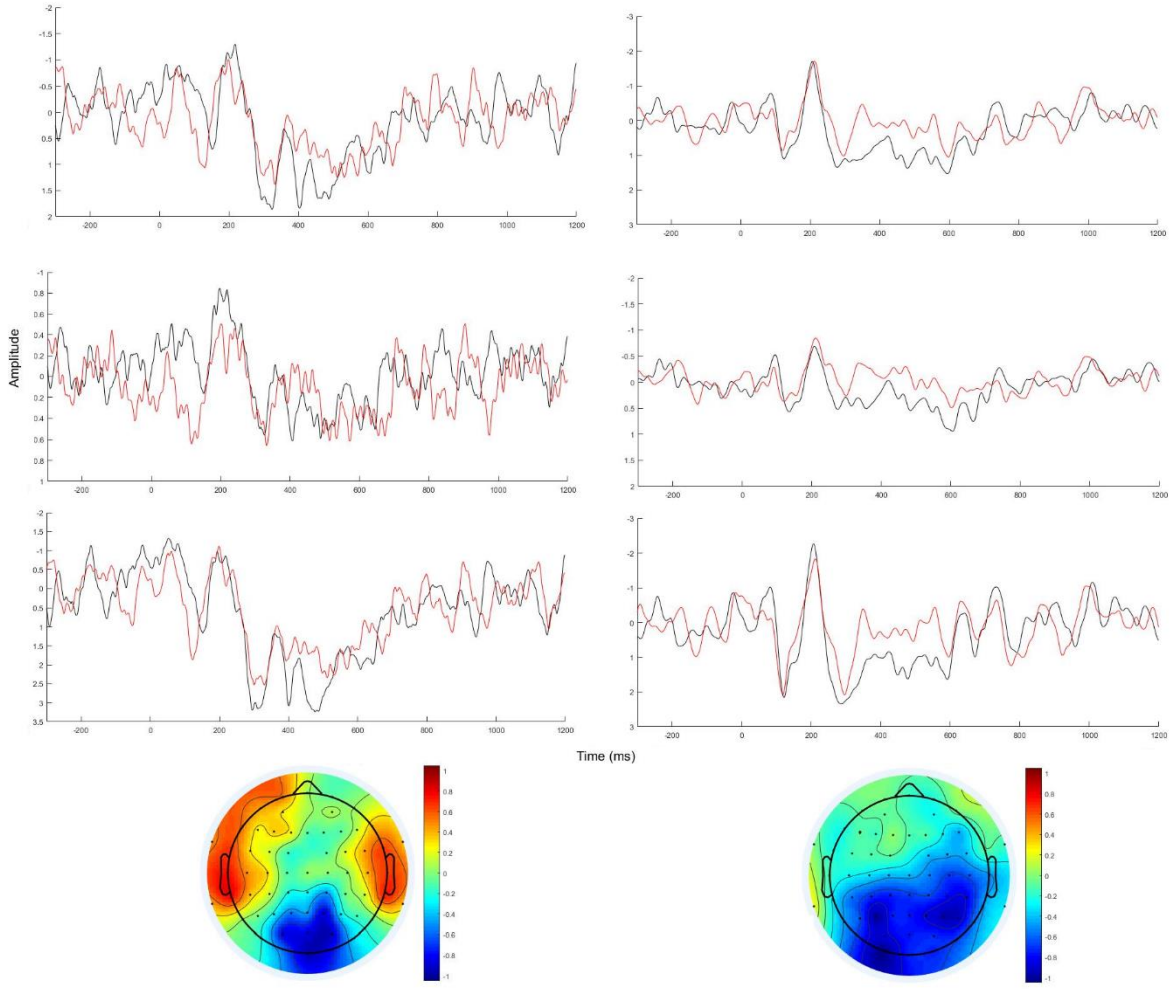


Figure 27: Grand average ERPs at CPz, Pz, and Cz for unos partitive compared to unos whole for controls and attriters. Topographical distribution for same condition made by subtracting felicitous from infelicitous conditions between 200-600 ms.

4.11 Discussion

This study was concerned with examining the extent to which the effects of attrition are a matter of processing mechanisms only or indeed reflect grammatical/representational changes. This question was explored by testing long-term Spanish-English bilinguals' interpretations of scalar quantifiers using both offline truth-value judgments and ERPs. In particular, we examined the above questions in light of the predictions of the Interface Hypothesis (IH), according to which SIs should be a *prima facie* candidate. While the IH predicts that properties that are, strictly speaking, dependent on external interfaces (i.e., involve grammar internal and

external integration) will be more susceptible to effects of bilingualism (attrition being one), research examining properties at external interfaces beyond the domain of anaphoric dependencies with pronouns do not straightforwardly support such a general claim (e.g., Bohnacker, 2010; Bohnacker & Rosén, 2008; Hoot, 2016; Leal, Destruel & Hoot, 2017; Slabakova, Rothman & Kempchinsky, 2012).

While the canonical domain of inquiry for the IH, choice between overt or null subject realization in null subject languages and related anaphoric dependencies, does depend on discourse-related factors such as [+/-]topic shift and [+/-]focus, such factors are predominantly related to the pragmatics of the specific linguistic discourse and not to domain general Pragmatics per se (see Slabakova & Rothman, 2011 for discussion of distinction and argumentation on how the two are often conflated). Thus, one might expect, as can be appreciated by comparing, for example, Belleti et al. (2007) to Rothman (2009) in the case of null vs. overt subject distribution, that there might be less variation of what is and what is not acceptable for native speakers concerning properties that are dependent on linguistic discourse. This is because the discourse constraints themselves (e.g., contrastive focus) effectively delimit an otherwise technically available choice between an overt and null pronoun. Conversely, when discourse constraints do not exist and it is truly pragmatics (not discourse) at play, as is the case with SIs, greater individual differences across native speakers arise precisely because there are no pseudo-systematic constraints on preference, i.e. there is actual preference without affecting grammaticality, truth value or felicitousness. The idea is that *algunos*, for example, becomes infelicitous due to the numerical quantity of referents to which it refers when juxtaposed against an entire set, whereas the overt use of *tú* (2nd person pronoun) in Spanish becomes infelicitous due to the linguistic context in which it is used. In other words, infelicitous uses of overt pronouns in Spanish are such because where one was already used in the linguistic discourse another is not necessary unless changing topic or focus marking. Quantity implicatures via *algunos*, however, do not depend on the already present instantiation of them in previous linguistic discourse. In fact, in spite of the semantic features driving *algunos*' partitive reading, language is irrelevant for its felicitous use. It simply must refer to a subset of items within a larger set. In that respect, it is conceptual. Therefore, of particular relevance to this study was to examine a property that is inherently pertinent to the domain of semantics but also subject to universal pragmatics, namely scalar quantifier interpretations. While these interpretations can be subject to variation caused by, say, prosodic focus, they also vary in large part based on the concept of quantity, a non-language-specific condition. Thus, we predicted

that scalar interpretations would be more prone to instability than subject distribution in the face of prolonged exposure to competing linguistic systems.

English scalar quantifiers derive pragmatic interpretations via one conceptual-to-semantics mapping and Spanish scalar quantifiers do so in another—though related— way due to idiosyncratic lexico-semantic features associated with each lexical item. Because one of the two scalar terms in Spanish derives its pragmatic distribution on the basis of a semantic features and the other has no similar constraint, the way(s) in which they are used and interpreted is distinct from English *some*. Thus, we predicted to see attriting bilinguals in prolonged exposure to English interpret Spanish scalar terms dissimilarly to the L1 controls.

Additionally, it has been suggested that variability at the interfaces, particularly within bilinguals, may be linked to burdened attention allocation that is typical of bilingualism (e.g., Rothman & Slabakova, 2011; Sorace, 2011). Such an effect of bilingualism is said to result from multiple differentially activated linguistic systems, requiring constant inhibition of one grammar while processing the other, and not as a by-product of disparate linguistic representations (see Sorrace & Serratrice, 2009; Wilson, Sorace & Keller, 2010; Sorace, 2011, 2012). Whatever the case for the disparity across groups, this study tested both offline and online performances. Therefore, we might expect to see distinct results from participants who are tested in both ways. In what follows, we contextualize and answer our research questions as they relate to the above discussion.

4.11.1 Research questions 1 and 2

Will SI interpretations of Spanish *algunos* show optionality in attrited bilinguals as compared to controls and, if so, at what level will we find the optionality (i.e., online or behavioral or both)?

Our bilingual attriters show no significant difference between their interpretations of felicitous and infelicitous *algunos* as compared to controls on both the offline and online tasks. As expected, in Experiment A the controls as a group widely rejected *algunos* when used in whole contexts. The bilinguals, however, treated whole and partitive uses equally. In this respect, *algunos* is being treated by the bilingual attriters such that its partitivity is lost.

Experiment B shows much of the same; however, we note some interesting trends. Recall that Experiment B was a free choice interpretation task. Thus, participants were able to choose whatever number or numbers they associated with their interpretation of each scalar quantifier in context. Unlike the controls, for bilingual attriters *algunos* is entirely acceptable in whole conditions. However, we see that as the number of possible referents increases,

acceptability decreases. In other words, contexts in which there are four possible referents are more acceptable than those with five referents, and those with five are more acceptable than those with six. While this variable did not prove significant in the model, it is interesting to note, nonetheless. One open question, then, is whether *algunos* would be acceptable in whole conditions in which there are much greater quantities of referents (i.e., greater than six) and whether at some threshold it would entirely disappear. This effect is in contrast to some recent work suggesting that partitive *some* is preferred, because it is more easily disambiguated, when referring to larger set sizes outside of the subitizing or easily counted range (Degen & Tanenhaus, 2011, 2015; Sun, 2017). In particular, Sun (2017) asked participants to rate whether *some* was preferred in smaller sets of two referents or larger sets of three and found that participants largely preferred the larger set size. However, this could be a reflection of the fact that English (and indeed other languages) have lexical items dedicated to describing sets of two and three objects, namely a *pair* (couple) and a *few*, respectively. Thus, it would be interesting to see what individual preferences would indicate in larger set sizes than those used in Sun (2017), such as Degen and Tanenhaus (2011, 2015). Our data in conjunction with Sun's and others' may reveal that Spanish speakers have conceptualized *algunos*, like English speakers have with *some*, as having an ideal set size of at least four. Given that acceptability of *algunos* in whole contexts decreases incrementally after four, it would be interesting to examine what happens as set size increases and what the threshold quantity is for the acceptability of *some* to refer to *all* or the entirety of a set in Spanish particularly.

4.11.2 Research questions 3 and 4

Will bilingual attriters show a similar degree of response preference for Spanish *unos* as compared to controls and, if so, at what level will we find the preference (i.e., online or behavioral or both)?

A surprising finding arose in the comparision of the *unos* conditions whereby the attriters do show, at the level of the brain, that *unos* whole is produces larger negativities reflective of increased unexpectedness. While this effect did not appear in the offline judgments as it did for the controls, showing such an effect online may be indicative that the potential for attrition in these bilinguals is affecting the offline judgments more than it is affecting their processing. When given more time to reflect and reason with possible answers, it seems that the attriters tend to interpret the quantifiers as having no difference, but online there is a (small) difference between them, particularly in the *unos* whole conditions. Such a finding does not follow neatly from the theoretical proposals claiming the *unos* does not carry an implicature

and can be both partitive and whole in equal measure. Given that both our controls and attriters are similar in their dispreference for *unos* whole conditions, it may stand to reason that there is indeed something missed in the theoretical descriptions. Minimally, these data indicate that there must be room for a strong partitive reading with *unos* with some people and that such an interpretation should be followed up with further research to clarify what may be giving rise to it.

However, in contrast to our control data, the bilingual attriters' offline data reveal no distinction between logical and pragmatic responses on *unos.* Because *algunos* is also treated in much the same way, there is no meaningful way to divide the responses that would be revealing of the pragmatic capacities of our attriters. This is not to say they are incapable of being pragmatic, but that the presence of more than one linguistic system in the brain blurs an adequate description of this ability, particularly as it relates to the derivation of SIs. This begs the question of whether or not SI derivation is, in itself, an appropriate associative metric to determine the pragmatic abilities of individual language users as other researchers have suggested (e.g., Barbet & Thierry, 2016; Nieuwland, Ditman, & Kuperberg, 2010; Zhao et al. 2015) and it motivates future research to take care when recruiting and screening for monolinguals in order to ensure that bilingualism as a confound does not obscure results. As argued by Schmid & Köpke (2017), any theory or argument that attempts to explain language development, acquisition or even attainment must not only describe that which is relevant to monolingualism but also that which is relevant to bilingualism. In a global society that is becoming increasingly multilingual, it may be premature to make cases for specific aspects of language relevant only to the now minority of monolinguals worldwide affecting certain universal properties of cognition, namely the ability to infer, especially if the premises of such arguments do not hold in cases where bilinguals are tested.

As discussed earlier in this paper, the competition brought about by the co-existence of more than one language in the same brain results in certain performance differences between monolinguals and bilinguals on some linguistic tasks (e.g., Bialystok, 2009; 2016). Research has argued that such effects are pertinent predominantly to lexical access or retrieval tasks where there is an apparent difference between monolinguals and bilinguals (Gollan et al., 2005; Sandoval, Gollan, Ferreira & Salmon, 2010). In contrast, other types of performance such as classification show more solidarity, and researchers have taken this to mean that the competition effect constraining bilinguals' performance is relevant only to when bilinguals are charged with accessing lexical representations rather than meaning (Gollan et al., 2005). However, our data suggest that meaning is also susceptible to such a competition effect in that

the bilingual attriters are no longer processing scalar meanings in Spanish as monolinguals do nor are they interpreting them in a similar way. Rather, they are either merging the Spanish form-to-meaning mappings with the single English one such that *algunos* and *unos* are essentially indistinguishable or they are diluting the duality of the forms in Spanish to resemble English *some* (see also Miller et al. 2016). In this sense, the IH is not the only relevant theory describing language attrition. The Feature Reassembly Hypothesis (FRH) (Lardiere, 2009), for example, claims that the level of granularity that formal, language-specific features have across languages can be useful in the explanation of developmental paths and outcomes in a variety of contexts and could be equally, if not more applicable, to attrition as suggested by Schmid & Kopke (2017).

Under the FRH, the task of the learner is to acquire all features (in bundles) associated with a particular functional category, such as Determiner, and map them onto their appropriate lexical forms. In the case where two languages instantiate distinct feature bundles, the learner must then reassemble the features of the L1 to map on to specific lexical forms of the L2. As applied to attrition, then, when a previously monolingual L1 speaker becomes increasingly competent in an L2, this may cause a remapping of the L1 feature bundles to reflect that of the newly acquired L2. Because Spanish *algunos* and *unos* have similar feature bundle configurations to that of English *some* except for the [+partitive] [+D-linked] features of *algunos*, it is plausible that our attriters are simply converging towards an English-like feature configuration with respect to the relevant quantifiers. In other words, *algunos* loses its partitive and/or D-linking feature, leaving both *algunos* and *unos* interpretable like *some* in English at the level of mental representation. An FRH approach to the entirety of our data might be viewed as preferred over an IH one given that they reveal similar performances in both offline and online methodologies within individuals, making it difficult to claim differences exist primarily at the level of processing. This, however, does not explain the lack of interpretation preference for either term. That is, why do the bilinguals accept both terms in both partitive and whole contexts to an equal degree thus exhibiting no pragmatic/logical split offline and yet online there is such a distinction, albeit marginal? This is an open question and one that ultimately would need further testing to adequately answer. However, previous research suggests that linguistic changes that bring about unconventional but not ungrammatical distributional patterns may indeed be more likely in the course of language attrition (Domínguez, 2013). Thus, because this distinction in interpretation between the quantifiers is one of contextual appropriateness and not grammaticality, it may have undergone attrition in a way that other domains do not. Again, the fact that the brain of a bilingual has more competing linguistic

information than a monolingual may lead to variability caused by an over-burdened, finite system of cognitive resources. It would, therefore, be necessary to characterize bilinguals as either pragmatic or logical by some other means of testing.

4.12 Conclusion

Our data suggest that L1 attrition at a truly external interface is not an issue of either processing or competence to the exclusion of one or the other. Both the implicit brain responses measured with ERPs and representation assessed with offline judgments are non-native-like for scalar interpretations, showing evidence of L1 semantic-pragmatic mapping changes after prolonged exposure to a unique L2 with simultaneous (major) decreases to L1 input. However, attrition has not entirely taken effect in our data. That is, the bilinguals do pattern after the monolinguals in one of the conditions. This results is unexpected for both groups given theoretical proposals outlining the nature of Spanish scalar quantifiers.

Although these data do not constitute counterevidence to the Interface Hypothesis per se, we suggested that a more fine-grained approach offered by applying the Feature Reassembly Hypothesis to the domain of L1 attrition might be fruitful for thoroughly explaining the present data and for future studies in L1 attrition more generally.

Chapter 5: General conclusion

The purpose of this dissertation was many-fold. First, there was a need to at least make an attempt to design a reliable and unbiased method not only for testing SIs but also for determining whether an individual is either logical or pragmatic when dealing with scalar interpretations. As discussed at length in each independent study that comprises this dissertation, it is clear that the methodological parameters for testing SIs can have as much to do both with one's interpretation of scalar terms as the so-called underlying cognitive mechanisms for deriving such interpretations. In this vein, it has been shown that a misunderstanding of the methodological design's purpose can lead researchers to misanalyse performance. A clear example of this was detailed in the description of Hartshorne et al. (2017) in articles 1 and 2 where the results of the study showed no meaningful connection between the onset of a scalar quantifier and an expected electrophysiological response. However, as outlined, there were several possible reasons for this supposed null effect, not the least of which being entirely related to the methodological design. Guasti et al. (2005) and Papafragou & Musolino (2003) were among the first studies to show that SI methodology could systematically affect the outcome of the results. These studies provided clear evidence that cases in which individuals did not calculate pragmatic meanings associated with scalar quantifiers were a matter of some artefact of the methodology and that pragmatic interpretations increased as a by-product of instruction, training and providing richer experimental contexts.

In the first study, we also found that task design and task type affected monolingual adult Spanish speaker's interpretation of the scalar terms *algunos* and *unos*. In particular, we found that while the pragmatically enriched interpretation was the most likely one for *algunos* in subset contexts, especially among pragmatic responders, such an interpretation was not impossible with *unos*, contrary to the theoretical description of these terms. Furthermore, we found that a careful examination of individual responses affected the interpretation of the results given that some individuals seemed to be inherently more semantic in their responses and others more pragmatic, a finding that may prove insightful when examining other populations (i.e., children) as argued in detail in article 1. While the purpose of this study was not to challenge or support any psycholinguistic theory as to the nature and/or mechanisms that underlie SI derivation, we provided evidence that subtle properties such as those related to scalar quantifier interpretation require significant introspection/reflection with respect to what previous methodologies have so far revealed. Only then, with principled and necessary

adjustments can one offer a considered, shrewd methodological design to tease apart potentially confounding noise and thus reliably tap individuals' interpretation and/or calculation of SIs more generally.

Second, given the lack of corroboration across studies as to what cognitive mechanisms might be at play (differentially) in those who are so-called logical responders and those who are pragmatic, I applied the same methodological design employed in the first study with the added task associated with electrophysiological data. This was done in an attempt to verify whether offline responses would be corroborated by online, implicit measures taken via ERPs. The results of this ERP study on monolingual Spanish SI processing—especially as informed by previous offline work with the same subjects—showed a clear difference in the way individual participants process scalar terms in Spanish. On the one hand, the experiment was exploratory, aiming to examine how the brain treats Spanish scalar quantifiers. On the other hand, we show that testing whether individuals calculate implicatures, or whether minimally one is pragmatic or logical, should be done on the basis of clear and careful testing parameters such that as much experimental noise as possible is removed from the methodology (Politzer-Ahles et al. 2013). No experimental study will match naturalistic language use with respect to ecological validity; however, future research (ours included) should consider experimental noise that can affect individuals' interpretations of words or language that are highly susceptible to extra-linguistic variables. As argued by Syrett et al. (2017a) for example, because deriving an SI is inherently connected to speaker-hearer interactions, tests that probe their derivation need to allow individuals to derive them in a context that mimics such interactions, crucially without bias from the discourse.

Our novel non-binary free interpretation task is one possible way of doing this. As we showed, not only does the variation with which individuals treat *algunos* in particular decrease, but for the first time in the available data record we see that the theoretical description of *unos* may need adjustment to allow for an associated SI. This experiment was crucial in revealing that Spanish-speakers are indeed either pragmatic or logical as has been found in English (Bott & Noveck, 2004), but this division takes place on the so-called non-SI generating item. We argue that this methodology gave rise to interpretation differences not seen thus far in the Spanish SI literature and that the description of *unos* must take into account the interpretative preference some individuals have for it. This study in conjunction with the first highlight the utility of methodological design and the employment of various methodology types, such as offline and online together.

The general trend in the SI literature has shifted away from theory towards experimentation. Such a shift in focus has led to the development of processing theories whereby online methodologies have probed the pragmatic abilities of individuals who do and do not derive SIs. These theories assume that the mere option of being able to derive an SI is grounds for claiming that individuals should, therefore, derive them. When SIs are not derived, it is assumed that there must be some reason for it, which is that the cognitive mechanisms responsible for SI derivation are deficient. Overall, these assumptions have led researchers to employ tests for autism, systematizing, IQ, working memory, etc., within SI methodologies. While the results are somewhat mixed in that only a select few have been able to correlate scores on the above-mentioned cognitive exams with SI derivation, many others have not been so successful. We argue here that SIs in general may not be the best candidate to make claims about the cognitive functions of individual language users. It is important to highlight that in our data and in those of other studies, SI derivation is a matter of preference. Thus, even so-called pragmatic responders who take *some* or, in our case *unos,* to refer to whole sets of referents are capable of retrieving logical readings; however, they merely prefer the pragmatic reading by majority. The reverse applies to logical responders.

In this respect, we argue that to associate cognitive deficits with the lack of SI derivation is premature and that one must look at cases of more extreme polarization of interpretation. At most, one might hypothesize that an individual who nearly wholly rejects one reading in preference for the other might be more likely to score higher on the ASQ or lower on the SQ-R or have difficulties in working memory. Likewise, it could be the case that the inverse holds in general—if one performs below average on such measures they might be more likely to be logical, but that a mere preference is not necessarily suggestive of such a correlation. Even when such a correlation is borne out in any given data set, it is still unclear what the correlation signifies in terms of the underlying connection (mechanisms) that give rise to this. That is, what is it precisely that should result in the correlation beyond showing one exists? Without understanding the actual causation, it might be premature to generalize anything from it.

For example, one implication drawn from the link of SI preferences to scores on the ASQ is that SIs are good candidates to be used as a diagnostic test for disorders. However, until we know what the correlation actually reflects, jumping too quickly to application might propagate something that future research nullifies. It may be true that there is a correlation—to the extent that the research itself has no confounds—but that does not mean that the wisest eventual interpretation of a preponderance of the facts (including future ones) will conclude that SI calculation informs anything regarding impairment diagnosis. Just like scientists

working on pharmaceuticals have an obligation to understand how drugs precisely interact with the human body before a drug can be considered safe and marketed, we too have an obligation to ensure that recommendations are not precipitous.

Furthermore, the vast majority of the work carried out in this domain has examined monolingual English speakers. Another concern of this dissertation, thus, has centered on the generalizability of the current results in the SI literature. The scant SI work that does exist in other languages and other (i.e., bilingual) populations is revealing on several fronts. First, second language learners have been shown to be either overly pragmatic (e.g., Slabakova, 2010) or not pragmatic enough (e.g., Lieberman, 2009; Miller et al. 2016; Vargas-Tokuda et al. 2009). Additionally, bilingual children tested in Spanish are shown to be sensitive to the subtle distinction between whole and partitive readings of *algunos,* but less so with the universal quantifier *todos* 'all' and not at all with the distinction between *algunos* and *unos* like monolingual adult Spanish speakers seem to be. It is clear that bilinguals in general and developing child monolinguals have different processing strategies for deriving SIs in either the L2 or in a heritage language. However, no literature, to our knowledge, had previously examined the effects of a different type of bilingualism on L1 SI derivation, namely attrition.

The third and final study in this dissertation discusses the effects of L1 attrition on SI derivation among Spanish-English bilinguals. In so doing, we set out to examine the tenets of the Interface Hypothesis applied to L1 attrition and eventually converged on the Feature Reassembly Hypothesis as more explanatory in understanding how to explain the effects attrition apparently had on the mental representation of L1 Spanish scalar quantifiers after significant time in L2 immersion. The data provide evidence that long-term bilinguals may experience changes in both the way they process extra-linguistic information, as well as how they interpret it in their L1. However, because Spanish scalar terms depend on particular semantic features to derive SIs, unlike English *some,* it was worthwhile to interpret the data both from an Interface Hypothesis approach, but also from an approach that captures a more nuanced interplay of feature representations related to specific lexical item's configurations across languages, namely the Feature Reassembly Hypothesis. In this respect, we can see that not only are SIs ripe for attrition, but also when, across language such properties have distinct bundles associated with certain lexical items, the feature bundles can converge towards that of the potentially more dominant L1.

Unlike in articles 1 and 2, the third article did not provide evidence of any logical-pragmatic split between responders. Whether this was a by-product of attritional effects whereby Spanish scalar interpretations (i.e., conceptual-to-semantics mappings) converge

towards those of English or not is open for interpretation. However, we argue herein that bilingualism in general can affect scalar interpretations (among other domains) and, therefore, SIs may not be a reliable or appropriate metric for probing pragmatic abilities of individual language users, especially the bi-/multilinguals who now comprise significantly more than half the world's population. The fact that the bilinguals in this article did not show a pragmatic-logical split does not mean they are (no longer) one or the other, rather a different means of testing those abilities would need to be employed.

In sum, while the available data record for SIs has improved current theories of pragmatics and semantics (e.g., Chierchia, 2017), as well as language processing, there is room for improvement in the way SIs are examined. Furthermore, the bias of English as the language of testing and a general focus on monolingualism have arguably limited the generalizability of the findings. It is clear that other languages that do not work like English with respect to SIs paint a slightly more nuanced picture than the one presented by monolingual English-only studies. We maintain that more data is needed from other languages, other speaker types and that extreme care should be given to methodological design in SI research.

Limitations and suggestions for future research

As is the case with most, if not all, dissertations there is significant room for improvement moving forward to increase the likelihood that what has been claimed herein is accurate and that one can replicate the findings both within Spanish and indeed other languages by considering what I have highlighted. The questions one ends up with after completing a PhD are unlikely to be (solely) the questions with which one started. In a forum such as this, it is useful to reflect on what the limitations are of the present studies which is the first logical step in mapping out how to address these limitations in the course of my emerging research program.

One limitation to consider for the immediate future is the fact that our articles do not offer any concrete data from cognitive metrics from the ASQ and others. In part, this was because this dissertation began nearly five years ago when some of the relevant work showing links between cognitive tests and SI derivation was either brand new or not out yet. Given the questions I focused on, however, and in light of the newness of such correlations before 2012 we decided that it was not worth redesigning the methodology that had already been developed. As discussed in several places of this dissertation, we remain unconvinced—-even in 2017 where a critical mass of studies looking into this correlation have emerged—of what

confirming such a correlation would mean anyway, especially as it relates to the subset of specific questions that comprise the focus of the present investigation.

Moving forward, however, one interesting question is what correlation there is between cognitive tests, particularly those concerning pragmatic abilities, and SI derivation in monolinguals and bilinguals alike, considering the fact that the aforementioned correlation has only be pursued to date in monolingual processing of SIs. Were we to show correlations for monolinguals and bilinguals, for both or neither group we would still be left in a position to ponder in greater depth what such a correlation means. Although still not well understood, it is nonetheless commendable that researchers have begun to test potential direct connections with pragmatic abilities and SIs. However, this fact also makes it stand out even more that no study to date—to my knowledge—has employed tests of logical capacities of individuals and their correlation to SI derivation. Because logic and pragmatics are inherently tied to this concept, it would be worthwhile to test them both moving forward. These could be tests of visuo-spatial logic, mathematical logic, and logic pertinent to language, such as formal and predicate logic, to give a couple of examples. Surely, being at extreme ends of the spectrum of logical or pragmatic abilities can be unfruitful in everyday life, whereas a balance between the two might ameliorate complications arising from language users who are, ourselves, inherently prone to poorly communicating and misunderstanding the messages we receive. Testing individuals' pragmatic and logical abilities together might offer more insight into the workings of human cognition, especially as it relates to language, and indeed uncover in tandem the actual mechanisms underlying the singular correlation that to date has been revealed in some studies.

Another limitation of the dissertation is evidenced in the number of experiments employed to test SIs. While I presented two distinct methodologies, in hindsight there are many more that could have driven home the point of methodological importance. For example, it would be interesting to look at as many variables as possible within task designs that either prompt or suppress an implicature. These variables range from noun drop, prosody, focus, live presentation, set size and more. One could easily employ these offline methods in the same study and run across-study statistics to assess the rate with which SIs are derived or not within specific experimental contexts. A major realization in bringing this dissertation to fruition is that of just how sensitive interpretations are on an individual-by-individual basis. It stands to reason that while some designs may be inherently more reliable concerning their ability to probe specific research questions, others can in fact prejudice the interpretation of the results recorded from them and that such potential bias affects some individuals more than others. In order to more fully evaluate the usefulness and appropriateness of SI task designs, one could

examine differences arising from a multiplicity of methodologies. This then embodies the next logical step in my continuing research program.

Finally, it would be interesting to carry out a meta-analysis of SI studies in general, though specifically of those that have found correlations with ASQ and other tests to SI interpretation. Many of those studies used underinformative occurrences of *some* in world knowledge contexts to determine which participants were logical and which were not and then ran correlation analyses to the cognitive tests of relevance to the individual studies. However, as has been argued, world knowledge judgments are not necessarily an accurate reflection of one's pragmatic abilities per se. Doing an analysis of the studies that have found associations with cognitive tests such as the ASQ and SI derivation may shed light on the specific aspects of the methodology that may have prompted participants to respond the way they did, thus appearing superficially as either logical or pragmatic.

References

Abutalebi, J., Anthony, P., Rosa, D., Green, D. W., Hernandez, M., Scifo, P., … Costa, A. (2012). Bilingualism Tunes the Anterior Cingulate Cortex for Conflict Monitoring, 2076–2086. https://doi.org/10.1093/cercor/bhr287

Ackerman, B. P. (1983). Form and function in children's understanding of ironic utterances. *Journal of Experimental Child Psychology*, *35*(3), 487–508. https://doi.org/10.1016/0022-0965(83)90023-1

Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2014). Morphosyntactic processing in advanced second language (L2) learners: An event-related potential investigation of the effects of L1–L2 similarity and structural distance. *Second Language Research*, *30*(3), 275–306. https://doi.org/10.1177/0267658313515671

Aleman-Bañón, J., Miller, D., & Rothman, J. (2017). Morphological Variability in Second Language Learners : An Examination of Electrophysiological and Production Data. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. http://dx.doi.org/10.1037/xlm0000394

Alonso-Ovalle, L., & Menéndez-Benito, P. (2002). Two types of weak quantifiers: Evidence from Spanish. *Proceedings of the 1st North American Summer School in Logic, Language and Information*, 11-19.

Alonso-Ovalle, L., & Menéndez-Benito, P. (2010). Modal indefinites. *Natural Language Semantics*, *18*(1), 1-31.

Alonso-Ovalle, L., & Menéndez-Benito, P. (2012). Indefinites, dependent plurality, and the viability requirement on scalar alternatives. *Journal of Semantics, 30*(1), 65-102.

Bach, K. (1999). The semantics-pragmatics distinction: What it is and why it matters. In K. Turner (Ed.), *The semantics-pragmatics interface from different points of view* (pp. 65–84). Oxford: Elsevier.

Bach, K. (1999). The myth of conventional implicature. *Linguistics and Philosophy*, *22*(4), 327-366.

Barbet, C., & Thierry, G. (2016). Some alternatives? Event-related potential investigation of literal and pragmatic interpretations of some presented in isolation. *Frontiers in Psychology*, *7*(SEP). https://doi.org/10.3389/fpsyg.2016.01479

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for

confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Belletti, A., Bennati, E., & Sorace, A. (2007). Theoretical and Developmental Issues in the Syntax of Subjects : Evidence from Near-Native Italians. *Natural Language & Linguistic Theory*, *25*(4), 657–689. https://doi.org/10.1007/sll049-007-9026-9

Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1450–1460. https://doi.org/10.1037/a0027850

Bever, T. G., Mehler, J., & Epstein, J. (1968). What children do in spite of what they know. *Science*, *162*, 921–924.

Bialystok, E. (2009). Bilingualism : The good , the bad , and the indifferent. *Bilingualism:, Language and Cognition*, *12*(1), 3–11. https://doi.org/10.1017/S1366728908003477

Bialystok, E. (2016). The signal and the noise. *Linguistic Approaches to Bilingualism*, *6*(5), 517–534. https://doi.org/10.1075/lab.15040.bia

Birdsong, D., & Molis, M. (2001). On the evidence for maturational effects in second language acquisition. *Journal of Memory and Language*, *44*, 235–249.

Blome-Tillmann, M. (2013). Conversational implicatures (and how to spot them). *Philosophy Compass*, *8*(2), 170–185. https://doi.org/10.1111/phc3.12003

Bloom, P., & Wynn, K. (1997). Linguistic cues in the acquisition of number words. *Journal of Child Language*, *24*(3), 511–533. https://doi.org/10.1017/S0305000997003188

Bohnacker, U. (2010). *The clause-initial position in L2 Swedish declaratives: Word order variation and discourse pragmatics*. *Nordic Journal of Linguistics* (Vol. 33). https://doi.org/10.1017/S033258651000017X

Bohnacker, U., & Rosén, C. (2008). The clause-initial position in L2 German Declaratives: Transfer of Information Structure. *Studies in Second Language Acquisition*, *30*(4), 511–538. https://doi.org/10.1017/S0272263108080741

Bonnefon, J. F., Feeney, A., & Villejoubert, G. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition*, *112*(2), 249–258. https://doi.org/10.1016/j.cognition.2009.05.005

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language*, *66*(1), 123–142. https://doi.org/10.1016/j.jml.2011.09.005

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*(3), 437–457. https://doi.org/10.1016/j.jml.2004.05.006

Braine, M. D. S., & Rumain, B. (1981). Development of comprehension of "or": Evidence for a sequence of competencies. *Journal of Experimental Child Psychology*, *31*(1), 46–70. https://doi.org/10.1016/0022-0965(81)90003-5

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition*, *126*(3), 423-440

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434–463. https://doi.org/10.1016/j.cognition.2005.07.003

Büring, D. (1996). A weak theory of strong readings. SALT VI: *Proceedings of the Sixth Conference on Semantics and Linguistic Theory*, (Ed.), T. Galloway and J. Spence, 17-34. Ithaca, N.Y: CLC Publications, Department of Linguistics, Cornell University.

Carston, R. (1990). *Quantity maxims and generalised implicature.* (No. 2).

Carston, R. (1998). Informativeness, relevance and scalar implicature. *Relevance Theory: Applications and Implications*, 179–236.

Cassidy, K. (1998). Three- and four-year-old children's ability to use desire- and belief-based reasoning. *Cognition*, *66*, B1–B11.

Chemla, E., & Bott, L. (2014). Processing inferences at the semantics/pragmatics frontier: Disjunctions and free choice. *Cognition*, *130*(3), 380–396. https://doi.org/10.1016/j.cognition.2013.11.013

Chierchia, G. (2006). Broaden Your Views: Implicatures of Domain Widening and the "Logicality" of Language. *Linguistic Inquiry*, *37*(4), 535–590. https://doi.org/10.1162/ling.2006.37.4.535

Chierchia, G., Fox, D., & Spector, B. (2012). Scalar implicature as a grammatical phenomenon.

In *Semantics: An International Handbook of Natural Language Meaning III.* (pp. 2297–2331). Boston: Walter de Gruyter.

Chierchia, G., & McConnell-Ginet, S. (2000). *Meaning and grammar: An introduction to semantics*. MIT press.

Chomsky, N. (1957). Syntactic structures. Cambridge: MIT Press

Chomsky, N. (1959). On certain formal properties of grammars. *Information and control*, *2*(2), 137-167.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.

Chomsky, N. (1973). Conditions on transformations. In S. Anderson & P. Kiparsky (Eds.), *A Festschrift for Morris Halle* (pp. 232–286). New York: Holt, Rinehart & Winston.

Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding* (Vol. 6). MIT press.

Chomsky, N. (1995). Language and nature. *Mind*, *104*(413), 1-61.

Crain, S. (2012). *The emergence of meaning*. Cambridge: Cambridge University Press.

Crain, S., & Thornton, R. (1998). *Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax an Semantics*. Cambridge: The MIT Press.

Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford: Oxford University Press.

De Bot, K., & Clyne, M. (1994). A 16-year longitudinal study of language attrition in Dutch immigrants in Australia. *Journal of Multilingual & Multicultural Development*, *15*(1), 17-28.

Degen, J., & Tanenhaus, M. K. (2011). Making inferences: the case of scalar implicature processing. In *Proceedings of the Cognitive Science Society 33*(33), 3299-3304.

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, *39*(4), 667-710.

Dekydtspotter, L., & Hathorn, J. C. (2005). Quelque chose … de remarquable in English–French acquisition: mandatory, informationally encapsulated computations in second language interpretation. *Second Language Research, 21*(4), 291–323. https://doi.org/10.1191/0267658305sr249oa

Diesing, Molly. 1992. *Indefinites*. Cambridge, MA: The MIT Press.

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia, 61*, 150-162.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load dual task impact on scalar implicature. *Experimental Psychology, 54*(2), 128–133. https://doi.org/10.1027/1618-3169.54.2.128

De Saussure, F. (1916). Nature of the linguistic sign. *Course in general linguistics*.

Dieussaert, K., Verkerk, S., Gillard, E., & Schaeken, W. (2011). Some effort for some: further evidence that scalar implicatures are effortful. *Quarterly Journal of Experimental Psychology (2006), 64*(12), 2352–67. https://doi.org/10.1080/17470218.2011.588799

Domínguez, L. (2013). *Understanding Interfaces: Second language acquisition and first language attrition of Spanish subject realization and word order variation* (Vol. 55). Amsterdam: John Benjamins Publishing.

Donaldson, B. (2012a). Initial subordinate clauses in Old French : Syntactic variation and the clausal left periphery. *Lingua*, *122*(9), 1021–1046. https://doi.org/10.1016/j.lingua.2012.04.003

Donaldson, B. (2012b). Syntax and Discourse in Near-Native French : Clefts and Focus, (September), 902–930. https://doi.org/10.1111/j.1467-9922.2012.00701.

Doran, R., Baker, R., McNabb, Y., Larson, M., & Ward, G. (2009). On the Non-Unified Nature of Scalar Implicature: An Empirical Investigation. *International Review of Pragmatics*, *1*(2), 211–248. https://doi.org/10.1163/187730909X12538045489854

Doran, R., Ward, G., Larson, M., McNabb, Y., & Baker, R. E. (2012). A novel experimental paradigm for distinguishing between what is said and what is implicated. *Language*, *88*(1), 124–154. https://doi.org/10.1353/lan.2012.0008

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *10*(1), 101. https://doi.org/10.1017/S1366728906002847

Erce, Á. V. (2014). The Role of L2 English Immersion in the Processing of L1 Spanish Sentence Complement/relative Clause Ambiguities (Doctoral dissertation, Pennsylvania State University).

Feeney, A., Scrafton, S., Duckworth, A., & Handley, S. J. (2004). The story of some: everyday

pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, *58*, 121–132. https://doi.org/10.1037/h0085792

Foley, C., & Flynn, S. (2013). The role of the native language. In J. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge Handbook of Second Language Acquisition*, pp. 97-113. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139051729.008

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology-Learning Memory and Cognition*, *22*(5), 1219-1248.

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition, and logical form*. New York: Academic Press.

Gelman, R. (1993). The child's understanding of number. In D. Medin (Ed.), *Learning and motivation* (pp. 61–96). New York: Academic Press.

Gelman, R., & Gallistel, R. (1978). *The child's understanding of number.* Cambridge: Harvard University Press.

Geurts, B. (2009). Scalar implicatures and local pragmatics. *Mind and Language*, *24*(1), 51–79.

Geurts, B., & Pouscoulous, N. (2009). Embedded implicatures?!? *Semantics and Pragmatics*, *2*(4), 1–34. https://doi.org/10.3765/sp.2.4

Gollan, T. H., Montoya, R., Fennema-notestine, C., & Morris, S. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, *33*(7), 1220–1234.

Green, D. (2011). Language control in different contexts : the behavioral ecology of bilingual speakers. *Frontiers in Psychology*, *2*, 1–4. https://doi.org/10.3389/fpsyg.2011.00103

Greeno, J. G., Riley, M. S., & Gelman, R. (1984). Conceptual competence and children's counting. *Cognitive Psychology*, *16*(1), 94–143. https://doi.org/10.1016/0010-0285(84)90005-7

Grice, H. P. (1975). Logic and Conversation. In J. Stainton (Ed.), *Perspectives in the Philosophy of Language: A Concise Anthology* (pp. 41–58). Ontario: Broadview Press.

Grice, H. P. (1989). *Studies in the way of words.* Cambridge: Harvard University Press.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*(1), 42–55. https://doi.org/10.1016/j.cognition.2010.03.014

Guasti, M., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, *20*(5), 667–696. https://doi.org/10.1080/01690960444000250

Gürel, A. (2004). Selectivity in L2-induced L1 attrition: a psycholinguistic account. *Journal of Neurolinguistics*, *17*(1), 53-78.

Gürel, A. (2007). Linguistic determinants of L1 attrition. *Language Attrition: Theoretical perspectives*, 99-120.

Gürel, A. (2011). In search for a unified model of L2 acquisition and L1 attrition: A commentary for the Interface Hypothesis. *Linguistic Approaches to Bilingualism*, *1*(1), 39-42.

Gutiérrez-Rexach, J. (2001). The semantics of Spanish plural existential determiners and the dynamics of judgment types. *Probus*, *13*(1), 113–154.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439-483.

Hagoort, P., Hald, L., Bastiaansen, M., Petersson, K. M., Hagoort, P., Hald, L., & Bastiaansen, M. (2004). Integration of Word Meaning and World Knowledge in Language Comprehension, *304*(5669), 438–441.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive neuroscience*, *11*(2), 194-205.

Hald, L. A., Steenbeek-planting, E. G., & Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension . Evidence from the N400, *6*, 0–8. https://doi.org/10.1016/j.brainres.2007.02.054

Halsey. (2010). Scalar implicatures in L2 Spanish. *Poster presented at GASLA*

Happe, F. G. E. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, *48*(2), 101–119. https://doi.org/10.1016/0010-0277(93)90026-R

Hartshorne, J. K., & Snedeker, J. (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes*, *28*(6), 1–35. https://doi.org/10.1080/01690965.2012.689305

Hartshorne, J., Snedeker, J., & Kim, A. (2014). The neural computation of scalar implicature.

Language, Cognition and Neuroscience, *30*(5), 620–634. https://doi.org/10.1080/23273798.2014.981195

Heyman, T., & Schaeken, W. (2015). Some differences in some: Examining variability in the interpretation of scalars using latent class analysis. *Psychologica Belgica*, *55*(1).

Hirschberg, J. (1991). *A theory of scalar implicature.* New York: Garland.

Hoop, H. (1992). Case configuration and noun phrase interpretation, Doctoral Dissertation, Rijksuniversiteit Groningen.

Hoot, B. (2016). Narrow presentational focus in heritage Spanish and the syntax – discourse interface, *1*(2017), 63–95. https://doi.org/10.1075/lab.14021.hoo

Hopp, H. (2013). Grammatical gender in adult L2 acquisition : Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56. https://doi.org/10.1177/0267658312461803

Hopp, H., & Schmid, M. S. (2013). Perceived foreign accent in first language attrition and second language acquisition : The impact of age of acquisition and bilingualism, *34*, 361–394.

Horn, L. R. (1972). *On the semantic properties of logical operators in English*. Los Angeles.

Horn, L. R. (1992). Pragmatics, implicature, abd presupposition. *International Encyclopedia of Linguistics*, *2*, 260–266.

Horn, L. R. (2004). Implicature. *The Handbook of Pragmatics*, 2–28.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, *58*(3), 376–415. https://doi.org/10.1016/j.cogpsych.2008.09.001

Huang, Y. T., & Snedeker, J. (2011). revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language and Cognitive Processes*, *26*(9), 1161–1172. https://doi.org/10.1080/01690965.2010.508641

Hunt, L., Politzer-Ahles, S., Gibson, L., Minai, U., & Fiorentino, R. (2013). Pragmatic inferences modulate N400 during sentence comprehension: Evidence from picture-sentence verification. *Neuroscience Letters*, *534*(1), 246–251. https://doi.org/10.1016/j.neulet.2012.11.044

Hurewitz, F., Papafragou, A., Gleitman, L., & Gelman, R. (2006). Asymmetries in the acquisition of numbers and quantifiers. *Language Learning and Development*, *2*(2), 77–96. https://doi.org/10.1207/s15473341lld0202_1

Ivanov, I. P. (2012). L2 acquisition of Bulgarian clitic doubling : A test case for the Interface Hypothesis. https://doi.org/10.1177/0267658312452066

Iverson, M. (2012). Advanced language attrition of Spanish in contact with Brazilian Portuguese. PhD dissertation, University of Iowa

Iverson, M., & Miller, D. (2017). Language attrition and maintenance: Two sides of the same coin? *Linguistic Approaches to Bilingualism.* Amsterdam: John Benjamins

Jackendoff, R. (1983). *Semantics and cognition* (Vol. 8). MIT press.

Jackendoff, R. (1991). Parts and boundaries. *Cognition*, *41*(1), 9-45.

Jackendoff, R. (2002). English particle constructions, the lexicon, and the autonomy of syntax. *Verb-particle explorations*, 67-94.

Jackendoff, R. (2006). On conceptual semantics. *Intercultural Pragmatics*, 3(3), pp. 353-358. doi:10.1515/IP.2006.020

Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language. *Cognition, 97*(2), 211-225.

Jakobson, R. (1941). *Kindersprache, Aphasie und allgemeine Lautgesetze.* Uppsala: Almqvist

Johnson, J.S., & Newport, E.L. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition, 39*, 215–258.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*(2), 159-201.

Katsos, N. (2008). The Semantics / Pragmatics Interface from an Experimental Perspective : The Case of Scalar Implicature. *Synthese, 165*(3), 385–401.

Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*(1), 67-81.

Keenan, E. (1973). Logic and Language, *102*(3), 185–194.

Keijzer, M. (2004). First language attrition: A cross-linguistic investigation of Jakobson's

regression hypothesis. *International Journal of Bilingualism*, *8*(3), 389–393. https://doi.org/10.1177/13670069040080030201

Kempson, R. (2003). Pragmatics: Language and communication. *The handbook of linguistics*, 394-424.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205-225.

Kok, A. (1997). Event-related-potential (ERP) reflections of mental resources: A review and synthesis. *Biological Psychology*, *45*(1), 19-56.

Kolk, H.H.J., Chwilla, D.J., van Herten, M., Oor, P.J.W., 2003. Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language 85*, 1–36.

Kratzer, A., & Heim, I. (1998). *Semantics in generative grammar*. Blackwell.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23-49.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*(1), 117-129.

Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. J. (2006). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Language and Cognitive Processes*, *21*(5), 489-530.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language*, *100*(3), 223-237.

Kupisch, T., & Rothman, J. (2016). Interfaces in the acquisition of syntax. In S. Fischer & C. Gabriel (Eds.), *Handbook on Interfaces.* De Gruyter

Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*, *11*(2), 99–116. https://doi.org/10.1016/0301-0511(80)90046-0

Lambert, R. D., & Freed, B. F. (1982). *The Loss of Language Skills*. Rowley: Newbury House Publishers.

Landauer, T. K., Foltz, P. W., Laham, D., Landauer, T. K., Foltz, P. W., An, D. L., … Laham, D. (1998). An introduction to latent semantic analysis An Introduction to Latent Semantic Analysis. *Discourse Processes*, *25*, 2–3. https://doi.org/10.1080/01638539809545028

Lappin, S. (2003). An introduction to formal semantics. In M. Arnoff & J. Rees-Miller (Eds.), *The Handbook of Linguistics,* 369-393. Malden: Blackwell Publishing.

Lappin, S., & Fox, C. (2015). *The handbook of contemporary semantic theory*. John Wiley & Sons.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition, *2*, 173–227.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews.Neuroscience*, *9*(12), 920-33. doi:http://dx.doi.org/10.1038/nrn2532

de Leeuw, E., Schmid, M. S., & Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition*, *13*(1), 33–40. https://doi.org/10.1017/S1366728909990289

Leal, T., Destruel, E., & Hoot, B. (2017). The realization of information focus in monolingual and bilingual native Spanish. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.16009.lea

Levinson, S. C (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature.* Cambridge: The MIT Press.

Lidz, J., & Musolino, J. (2002). Children ' s command of quantification, *84*(2002), 113–154.

Luck, S. J. & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't), *Psychophysiology*, *54*(1), 146–157. https://doi.org/10.1111/psyp.12639

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): assessing language profiles in bilinguals and

multilinguals. *J Speech Lang Hear Res*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Martí, L. (2008). The semantics of plural indefinite noun phrases in Spanish and Portuguese. *Natural Language Semantics*, *16*(1), 1-37.

Martohardjono, G. (1993). Wh-*movement in the acquisition of a second language: A cross-linguistic study of 3 languages with and without overt movement*. PhD dissertation, Cornell University.

Marty, P. P., & Chemla, E. (2013). Scalar implicatures: Working memory and a comparison with only. *Frontiers in Psychology*, *4*, 1–12. https://doi.org/10.3389/fpsyg.2013.00403

Miller, D., Giancaspro, D., Iverson, M., Rothman, J., & Slabakova, R. (2016). Not Just algunos, but indeed unos L2ers can acquire scalar implicatures in L2 Spanish. In *Language Acquisition Beyond Parameters: Studies in honour of Juana M. Liceras* (pp. 125–145). Amsterdam: John Benjamins.

Miller, K., Schmitt, C., Chang, H. H., & Munn, A. (2005). Young children understand some implicatures. In *Proceedings of the 29th Annual Boston University Conference on Language Development*, 389-400.

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, *47*(8), 908-930.

Morris, C. (1938). Foundations of the theory of signs. *International Encyclopedia of Unified Science*, *1*.

Newstead, S. E. (1988). Quantifiers as fuzzy concepts. *Advances in Psychology*, *56*, 51-72.

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, *63*(3), 324–346. https://doi.org/10.1016/j.jml.2010.06.005

Noveck, I. A. (2001). When Children Are More Logical than Adults: Experimental Investigations of Scalar Implicature. *Cognition*, *78*(2), 165–188.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, *85*(2), 203–210. https://doi.org/10.1016/S0093-934X(03)00053-1

Noveck, I. A., & Reboul, A. (2008). Experimental Pragmatics: a Gricean turn in the study of

language. *Trends in Cognitive Sciences*, *12*(11), 425–431. https://doi.org/10.1016/j.tics.2008.07.009

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113.

Ortega, L. (2013). SLA for the 21st Century: Disciplinary Progress, Transdisciplinary Relevance, and the Bi/multilingual Turn. *Language Learning*, *63*(1), 1–24. https://doi.org/10.1111/j.1467-9922.2012.00735.x

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785-806.

Paap, K., Johnson, H., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, *69*, 265–278. https://doi.org/https://doi.org/10.1016/j.cortex.2015.04.014

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics – pragmatics interface. *Cognition*, *86*, 253–282.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism* (Vol. 18). Amsterdam: John Benjamins

Paradis, J. (2007). Early bilingual and multilingual acquisition. *Handbook of Applied Linguistics*, *5*, 15-44.

Partee, Barbara. 1989. Many Quantifiers. *Proceedings of the 5th Eastern States Conference on Linguistics*, (Ed.), J. Powers & K. de Jong, 383-402. Columbus, OH: Ohio State University.

Pesetsky, D. (1987). Binding problems with experiencer verbs. *Linguistic Inquiry*, *18*(1), 126-140.

Phillips, C., & Ehrenhofer, L. (2015). The role of language processing in language acquisition. *Linguistic Approaches to Bilingualism*, *5*(4), 409–453. https://doi.org/10.1075/lab.5.4.01phi

Politzer-Ahles, S., Fiorentino, R., Jiang, X., & Zhou, X. (2013). Distinct neural correlates for pragmatic and semantic meaning processing: An event-related potential investigation of scalar implicature processing using picture-sentence verification. *Brain Research, 1490*,

134–152. https://doi.org/10.1016/j.brainres.2012.10.042

Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. *Language Acquisition*, *14*(4), 347-375. https://doi.org/10.1080/10489220701600457

Ramchand, G., & Reiss, C. (Eds.). (2007). *The Oxford Handbook of Linguistic Interfaces*. Oxford: Oxford University Press.

Reinhart, T. (2006). Interface strategies. *OTS working papers in linguistics*.

Rexach, J. G. (2001). The semantics of Spanish plural existential determiners and the dynamics of judgment types. *Probus*, *13*(1), 113–154.

Roberts, L., Alonso, J. G., Rothman, J., Pliatsikas, C., & Rothman, J. (2016). Evidence from neurolinguistic methodologies : Can it actually inform linguistic / language acquisition theories and translate to evidence-based applications? *Second Language Research, 1-19* https://doi.org/10.1177/0267658316644010

Rothman, J. (2008). Why all counter-evidence to the critical period hypothesis in second language acquisition is not equal or problematic. *Linguistics and Language Compass*, *2*(6), 1063–1088. https://doi.org/10.1111/j.1749-818X.2008.00098.x

Rothman, J. (2009). Pragmatic deficits with syntactic consequences ?: L2 pronominal subjects and the syntax – pragmatics interface, *41*, 951–973. https://doi.org/10.1016/j.pragma.2008.07.007

Rothman, J., & Slabakova, R. (2011). The Mind-Context Divide: On acquisition at the linguistic interfaces. *Lingua*, *121*, 568–576.

Sauerland, U. (2012). The computation of scalar implicatures: Pragmatic, lexical or grammatical? *Linguistics and Language Compass*, *6*(1), 36–49. https://doi.org/10.1002/lnc3.321

Schindele, R., Ludtke, J., & Kaup, B. (2008). Comprehending negation: A study with adults diagnosed with high functioning autism or Asperger's syndrome. *Intercultural Pragmatics*, *5*(4), 421–444.

Schmid, M. S. (2002). *First language attrition, use, and maintenance: The case of German Jews in anglophone countries*. Amsterdam: John Benjamins.

Schmid, M. S. (2009). On L1 attrition and the linguistic system. *Eurosla Yearbook*, *9*(1), 212–244.

Schmid, M. S. (2011). *Language Attrition*. Cambridge: Cambridge University Press.

Schmid, M. S. (2013). First language attrition, *1*, 94–115. https://doi.org/10.1075/lab.3.1.05sch

Schmid, M. S., & Kopke, B. (2004). Language Attrition: The next phase. In M. S. Schmid, B. Kopke, M. Keijzer, & L. Weilemar (Eds.), *First language attrition: Interdisciplinary perspectives on methodological issues* (pp. 1–43). Amsterdam: John Benjamins.

Schmid, M., & Kopke, B. (2007). Bilingualism and attrition. In *Studies in Bilingualism* (pp. 1–7). Amsterdam: John Benjamins.

Schmid, M., Kopke, B., & de Bot, K. (2013). Language attrition as a complex , non-linear development. *International Journal of Bilingualism*, *17*(6), 675–682. https://doi.org/10.1177/1367006912454619

Schmid, M., & Köpke, B. (2017). The relevance of first language attrition to theories of

bilingual development. *Linguistic Approaches to Bilingualism*. Amsterdam: John Benjamins.

Serratrice, L. (2013). Cross-linguistic influence in bilingual development. *Linguistic Approaches to Bilingualism*, *3*(1), 3–25. https://doi.org/10.1075/lab.3.1.01ser

Serratrice, L., & Sorace, A. (2009). Internal and external interfaces in bilingual language development : Beyond structural overlap. *International Journal of Bilingualism, 13(2),* 195-210.

Shook, A., & Marian, V. (2013). Interaction Network for Comprehension of Speech ∗, *16*(2), 304–324. https://doi.org/10.1017/S1366728912000466

Skinner, B. F. (1935). The generic nature of the concepts of stimulus and response. *The Journal of General Psychology*, 12(1), 40-65.

Skinner, B. F. (1938). *The behaviour of organisms: An experimental analysis*. D. Appleton-Century Company Incorporated.

Skinner, B. F. (1963). Operant behavior. *American Psychologist*, *18*(8), 503.

Slabakova, R. (2010). Scalar implicatures in second language acquisition. *Lingua, 120*(10), 2444-2462.

Slabakova, R., Kempchinsky, P., & Rothman, J. (2012). Clitic-doubled left dislocation and focus fronting in L2 Spanish : A case of successful acquisition at the syntax – discourse interface. https://doi.org/10.1177/0267658312447612

Smith, C. L. (1980). Quantifiers and question answering in young children. *Journal of Experimental Child Psychology*, *30*(2), 191–205. https://doi.org/10.1016/0022-0965(80)90057-0

Sorace, A. (2000). Introduction Syntactic optionality in non-native grammars, *2*, 93–102.

Sorace, A. (2011). Pinning down the concept of interface in bilingual development: A reply to peer commentaries. *Linguistic Approaches to Bilingualism*, *2*(2), 209–216. https://doi.org/10.1075/lab.2.2.04sor

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition.* Oxford: Blackwell Publishing.

Spychalska, M., Kontinen, J., & Werning, M. (2016). Investigating scalar implicatures in a truth-value judgement task: evidence from event-related brain potentials. *Language, Cognition and Neuroscience*, *3798*(September), 1–24. https://doi.org/10.1080/23273798.2016.1161806

Syrett, K., Austin, J. R., Sánchez, L., Germak, C., Lingwall, A., Perez-Cortes, S., … Baker, H. (2017). The influence of conversational context and the developing lexicon on the calculation of scalar implicatures: Insights from Spanish-English bilingual children. *Linguistic Approaches to Bilingualism*, 1–35. https://doi.org/10.1075/lab.14019.syr

Syrett, K., Lingwall, A., Perez-Cortes, S., Austin, J. R., Sánchez, L., Baker, H., … Arias-Amaya, A. (2017b). Differences between Spanish monolingual and Spanish-English bilingual children in their calculation of entailment-based scalar implicatures. *Glossa*, *2*(1), 1–19.

Tager-Flusberg, H. (1985). Sentence comprehension in autistic children. *Applied Psycholinguistics*, *2*(5), 5–24.

Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: A Commentary on "Grammatical agreement processing in reading: ERP findings and future directions" by Molinaro et al., 2014. *Cortex*, *66*, 149-

155.

Tanner, D., Grey, S., & van Hell, J. G. (2017). Dissociating retrieval interference and reanalysis in the P600 during sentence comprehension. *Psychophysiology*, *54*(2), 248-259.

Tavano, E. M. (2010). *The balance of scalar implicature.* University of Southern California.

Tsimpli, I., Sorace, A., Heycock, C., & Filiaci, F. (2004). First language attrition and syntactic subjects : A study of Greek and Italian near-native speakers of English, *8*(3), 257–277.

Tsimpli, I., & Sorace, A. (2006). Differentiating interfaces: L2 performance in syntaxsemantics and syntax-discourse phenomena. In *BUCLD Proceedings 30*.

Valenzuela, E. (2006). L2 end state grammars and incomplete acquisition of Spanish CLLD constructions. In R. Slabakova, S. Montrul, & P. Prévost (Eds.), *Inquiries in Linguistic Development: In Honor of Lydia White* (pp. 283–304). Amserdam/Philidephia: John Benjamins.

van Hell, J., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts, *9*(4), 780–789.

Van Herten, M., Chwilla, D. J., & Kolk, H. H. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, *18*(7), 1181-1197.

van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176-190.

Vargas-Tokuda, M., Grinstead, J., & Rexach, J. G. (2009). Context and the scalar implicatures of indefinites in child Spanish. *Hispanic Child Languages*, 93–115.

Villegas, A. (2014). *The role of L2 English immersion in the processing of L1 Spanish sentence complement/relative clause ambiguities.* PhD Dissertation, Pennsylvania State University.

Weissman & Tanner, D. (2017). EuroSLA, 2017. In preparation for CUNY, 2017.

Wilson, F., Sorace, A., & Keller, F. (2009). Antecedent preferences for anaphoric demonstratives in L2 German. In *Proceedings of Boston University Conference on Language Development* (pp. 634–645). Boston.

Wu, Z., & Tan, J. (2009). Scalar implicature in Chinese child language: An experimental study. *Journal of Foreign Languages*, *3*, 69–75.

Wynn, K. (1992). Addition and subtraction by human infants. *Science*, *358*(6389), 749–750.

Yilmaz, G. (2011). Complex embeddings in free speech production among late Turkish-Dutch bilinguals. *Language, Interaction and Acquisition*, *2*(2), 251-275.

Yilmaz, G., & Schmid, M. S. (2012). L1 accessibility among Turkish-Dutch bilinguals. *The Mental Lexicon*, *7*(3), 249-274.

Zhao, M., Liu, T., Chen, G., & Chen, F. (2015). Are scalar implicatures automatically processed and different for each individual? A mismatch negativity (MMN) study. *Brain Research*, (*1599)*, 137-149.