Developmental Change in the Integration of Information During Online Sentence Comprehension. Evidence From Eye-Tracking and Event-Related-Potentials

Citation

Permanent link
https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37365539

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Developmental change in the integration of information during online sentence comprehension.

Evidence from Eye-Tracking and Event-Related-Potentials

Abstract

Understanding the moment-to-moment process by which children transform an unfolding stream of sounds into a communicated message can provide much needed insight into the architecture of the language system and the source of developmental change in language comprehension. In this dissertation, we use both eye-tracking and event-related potentials to get insight into how various sources of information constrain interpretation as a sentence unfolds. In Paper 1, we use the visual world eye-tracking paradigm to study comprehension of garden-path sentences. By comparing monolingual and bilingual children in this task, we evaluate the roles that linguistic and cognitive development play in both the selection and revision of syntactic predictions. Using both group comparisons and correlational approaches, we show that improvement in the use of top-down cues during online comprehension reflects increased language experience. In Paper 2, we use the test case of negation to evaluate integration of semantic structure during lexical access. Using the N400 response, we test whether adults and children could use negation to update lexical predictions when negation occurs as a response to a polar question. In the current paradigm, we find that neither adults nor children show online use of negation when the negated proposition is unpredictable. However, we propose that future work should be conducted to see whether incorporation of logical semantics may improve in a more reliable discourse context. Finally, in Paper 3 we use a new method of looking at ERPs
during a natural story-listening task. We find that in a rich, highly predictable discourse, children but not adults show N400 responses that reflect the frequency of the words they hear. However, beyond the effects of frequency or semantic relatedness, both child and adult responses are sensitive to word predictability. Our findings suggest that during naturalistic listening, both adults and children use top-down constraints from the semantic/syntactic/pragmatic context to access upcoming words. However, in children, lower level information continues to play a more central role. We argue that studies using event related potentials, and natural listening tasks more particularly, offer an exciting opportunity to expand the contexts in which online language comprehension has been investigated.
# Table of Contents

Introduction..............................................................................................1  
1.1 Background.......................................................................................2  
1.2 On-going challenges.........................................................................19  
1.3 Present Research...............................................................................21  

Chapter 2: What bilingualism can teach us about online ambiguity resolution and garden-path revision........................................................................26  
2.1 Introduction.......................................................................................27  
2.2 Methods............................................................................................38  
2.3 Results...............................................................................................45  
2.4 Discussion..........................................................................................62  

Chapter 3: When hearing negation, do listeners expect the unexpected? An ERP investigation into online comprehension of negation in response to a yes/no question.........................................................................................69  
3.1 Introduction.......................................................................................70  
3.2 Methods............................................................................................75  
3.3 Results...............................................................................................79  
3.4 Discussion..........................................................................................84  

Chapter 4: The Storytime Paradigm: A naturalistic ERP study of children’s lexical access..................................................................................................92  
4.1 Introduction.......................................................................................93  
4.2 Methods............................................................................................106  
4.3 Results...............................................................................................109  
4.4 Discussion..........................................................................................119  

Chapter 5: Conclusion............................................................................127  
5.1 Conclusion........................................................................................128  

References..............................................................................................135  
Appendix.................................................................................................162
Acknowledgments

I am immensely grateful to the people that have supported me throughout this endeavor. Most importantly, I want to thank my advisor, Jesse Snedeker. Your incredibly patient, thoughtful, enthusiastic, and scientifically rigorous guidance over the years has been invaluable to the development of these projects and, more importantly, to my growth as a scientist, a mentor, and a person. Your wisdom and kindness first inspired my pursuit of this career and will forever serve and an example for me to follow in all aspects of my life. I also want to thank my committee members, Susan Carey, Alfonso Caramazza, and Gina Kuperberg for their expert guidance and boundless enthusiasm. Your thoughts and advice have shaped how I think about these questions and will stay with me as I continue to develop my research program. I look forward to continuing to discuss and develop my ideas through many future conversations.

This work would not have been possible without the support of my friends and labmates. I want to particularly thank Annemarie Kocab for her wise, honest, and helpful advice at every stage of this research project and for her unwavering friendship across the years. I also want to thank Anthony Yacovone, who is always first to offer a helping hand. Your thoughts and advice have helped me develop my research ideas, experimental designs, and statistical abilities and our conversations have also made me a better person and teacher. I am grateful for every member of the Snedeker laboratory with whom I have had the great fortune to cross paths with across the years. The intellectual generosity of everyone in the lab had created a place where my love of science can thrive even in the most difficult times. My work improves with every lab meeting and my drive to keep studying language development strengthens with every conversation. I want to particularly thank the lab managers, Liz Chalmers, Brianne Gallagher, Chantal Hoff, Ellie Kaplan, Sarah Raulston, Tracy Reuter, Margarita Zeitlin, and Briony Waite. You make everything our lab does possible and I thank you for going above and beyond to help me and every other graduate student in the lab. I would also like to thank the many research assistants who have tirelessly devoted countless hours to make this research happen. This project would never have come to fruition without the hard work of Sophie Mandl, Jerry Nelluvelil, Ryan Law, Alexander Robertson, Jojo Wang, Charles Wu, Jake Pechet, Romy Dolgin, Priyanka Alluri, Gabriana Freire, Mae Harrington, Constance Bourguignon, Liliane Emralino, Corinne Bozzini, Sandy Kim, Charlotte Kennish, Tobi Abubakare, Hillary Man, and many others.

I would also like to thank all the participants that so generously donated their time. Participating in a research experiment, and an EEG experiment in particular, takes incredible patience and I want to especially thank the children and their parents for working with us to make testing run smoothly.

This work would also not have been possible without administrative support. I want to especially thank Celia Raia and Allie Pagano for all of their help. I am also grateful for my funding sources, and particularly the Norman Anderson fund for their contribution to this research program.

I would like to thank my family. I am forever grateful for Natalya Bassina and Michael Chechelnitsky whose help and unwavering support put me on this path and allowed me to reach this milestone. None of this would have been possible without you. I am also very fortunate to have an incredible support system, including my grandparents, siblings, in-laws, and friends.

Finally, I would also like to thank my husband, David Levari, and our kids, Noah, Sasha, and Maya. Thank you for always being my biggest cheerleaders and for giving meaning to everything that I do.
Chapter 1:

Introduction
1.1 Background

When we hear a sentence, we understand it almost immediately and seemingly without effort. However, this everyday task is remarkably complex. To understand a sentence a listener must parse a quickly incoming stream of phonemes into words, identify their meanings, infer the semantic and syntactic structure as it unfolds, and recruit information from context, including the preceding conversation, world knowledge, speaker identity, etc. Decades of psycholinguistic literature has provided insight into the cognitive architecture that allows adults to carry out this task (Altmann, 2001; Elman, Hare, & McRae, 2004; McRae & Matsuki, 2013b). However, far less is known about how this ability develops.

By the age of 5, children can carry on a conversation with relative competence. By this age, they can recruit a large working vocabulary (Dail & McGee, 2011), construct and interpret both semantic (Gertner, Fisher, & Eisengart, 2006; C. H. Noble, Rowland, & Pine, 2011; Thothathiri & Snedeker, 2011) and syntactic (Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000; Love, 2007) structures, and even use extra-sentential information such as prosody (Snedeker & Yuan, 2008) to aid comprehension. Despite these linguistic accomplishments, adult-like sentence comprehension emerges slowly across childhood. Unlike adults, children have difficulty using a number of higher-level sources of information to inform sentence comprehension, including referential context (Trueswell, Sekerina, Hill, & Logrip, 1999; Weighall, 2008), world-knowledge (Kidd, Stewart, & Serratrice, 2011; Snedeker, Worek, & Shafto, 2009), and pragmatics (Huang & Snedeker, 2009a, 2009b).

Although prior research has identified several ways in which children’s comprehension differs from adults’, the underlying causes of these differences remains a source of debate. In this dissertation I will discuss two broad theories of developmental change in comprehension. One
possibility is that emerging adult-like comprehension reflects improvements in domain-general executive functioning (EF). Under this theory, children and adults largely share the same language processing architecture and relevant linguistic knowledge but differ in their ability to control which sources of information are used for comprehension and revise interpretations when needed. Alternatively, differences between children and adults may primarily reflect maturation of the language processing system. Across the early-school age years, children's increasing ability to use top-down constraints may reflect improvements in the quality of their linguistic representations and the efficiency with which they can be used.

Our understanding of what may be driving developmental change in sentence comprehension is hindered by several limitations in prior work. The first limitation relates to how cognitive variables are confounded in most developmental research. Specifically, increases in language experience are contemporaneous and often confounded with improvements in children's EF, making it difficult to determine the role each plays in language development. The second limitation of the prior work on the development of sentence comprehension is that most of it has relied on the visual-world-paradigm (VWP). In the VWP insight into comprehension is gained by tracking eye-movements to visually presented referents as a sentence unfolds in time. The use of this paradigm places limits on both the linguistic representations being studied and on the context in which comprehension occurs.

In this dissertation I present three papers that address these challenges. In the remainder of the Introduction, I will first describe the current understanding of the adult language comprehension system and how it differs in early-school-aged children, focusing primarily on syntactic analysis and lexical access. I will then expand on the current theories regarding what may be the driving force of developmental change in this domain. Finally, I will introduce the
three current papers and describe the theoretical and methodological contributions they offer to the field.

1.1.1 Adult language comprehension

How do children and adults differ in how they comprehend sentences? We can first consider how adult listeners accomplish this task. Adult language comprehension can be described as a hierarchy of linguistic representations, from phonological up through pragmatic. Although the exact number, order, and identity of the representations is still highly debated, there is wide consensus that these levels must at least include lower-level phonological processing, lexical access, and higher-level representations of the semantic and syntactic structure (Altmann, 2001). We also know that sentence comprehension is further constrained by numerous factors that go beyond the content of the spoken utterance. For example, sentence comprehension is constrained by the visual referential environment (Knoeferle, Crocker, Scheepers, & Pickering, 2005; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Trueswell et al., 1999), the preceding discourse (Nieuwland & Van Berkum, 2006; Xiang & Kuperberg, 2015), as well as knowledge of the world (Hagoort, Hald, Bastiaansen, & Petersson, 2004; Snedeker et al., 2009) and the speaker (Creel & Bregman, 2011; Grey & van Hell, 2017; Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008). A central topic in psycholinguistics is how these different levels of representation are constructed during language comprehension, and how they constrain one another. Higher levels of analysis are constructed, at least in part on the basis of lower ones. Logically, some degree of perceptual analysis must occur before we can accurately identify a word, and the identity of the word provides critical information for combinatorial semantic and syntactic analyses. However, we often use our higher-level representations to make predictions about upcoming input. Upon hearing “The pirate found the ...” a listener can likely fill in the
word *treasure*, even if the speaker was cut-off prior to its utterance. It remains highly debated how well and how quickly bottom-up sources of information can constrain our higher-level interpretations and the role that top-down prediction may play in how sentences are understood.

1.1.1.1 Incrementality in adult comprehension

Incrementality in the language processing system refers to the propagation of information as it is acquired. We do not wait until we have heard an entire word before we work on identifying it and we do not identify all the words in the sentence before we attempt to construct a syntactic structure. Rather, the system works with partial incoming information to constrain alternatives. The passing of partial information across levels of analysis is perhaps best exemplified by the phenomenon of phono-semantic priming (Marslen-Wilson & Zwitserlood, 1989; Yee & Sedivy, 2006). Yee and Sedivy (2006) showed adults a display containing a target image (*logs*). In addition, the display contained a semantic associate (*key*) of a phonological competitor of the target word (*lock*). Critically, the phonological competitor itself is never shown. Although *logs* and *keys* are neither phonologically nor semantically related, subjects nevertheless show increased looks to the key when hearing a sentence containing the word *logs*. For this pattern to occur, partial phonological information (*lo...*) must activate a set of possible lexical items (a set that includes both *log* and *lock*). What’s more, although the activation of *lock* is only transient, it’s temporary activation must result in activation of close semantic associates, including *keys*. The bottom-up activation from partial input, although ruled out with further information, can be traced across multiple levels of representation.

Studies of how adults identify and access a spoken word have shown that they continuously constrain alternatives as phonological information unfolds (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978). Adults can identify lexical items based on partial input
showing that hearing an entire word is not necessary for lexical access to occur (Grosjean, 1980). In addition, visual eye-tracking studies have shown that listeners will direct their gaze to increasingly smaller sets of visual referents as incoming phonological input rules out alternatives, settling on the proper word once such a word is phonologically disambiguated (Allopenna, Magnuson, & Tanenhaus, 1998).

Similarly, at the syntactic level, listeners will commit to a syntactic structure based on the information available. Consider, for example, the following sentence [1]. In such "garden-path"

[1] The horse raced past the barn fell.

sentences the expected syntactic structure is violated with further input. The temporary confusion caused by the word fell at the end of this sentence shows that listeners do not wait until all the words are spoken to infer how the words are related. Rather, we use the unfolding stream of information to guess the most probable syntactic construction. In the example above, the semantic properties of the verb raced recommend a syntactic structure in which horse is a subject and a subsequent prepositional phrase will reveal a destination. Occasionally, as in garden-path sentences, incomplete information can lead us astray. In fact, the horse was raced past the barn. Adult comprehension of such sentences shows both that listeners use unfolding information to commit to syntactic constructions, but also that they can often revise such commitments when they turn out to be incorrect (Christianson, Williams, Zacks, & Ferreira, 2006; MacDonald, 1994; Trueswell et al., 1999).

1.1.1.2 Interactivity in adult comprehension

Over the past several decades there has been increasing evidence that the adult language system is not only incremental but also highly interactive. While we know that information can flow "bottom-up" from perceptual processing up through increasingly higher levels of analysis, it
has been widely argued that higher-order information can also constrain processing at lower-
levels (Altmann, 2001; Altmann & Kamide, 1999; Federmeier, 2007; Kuperberg & Jaeger, 2016;
McRae & Matsuki, 2013a; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Although
top-down contextual effects have been shown across the processing stream, here I will focus on
evidence for the use of higher-order information on lexical and syntactic stages of processing.

At the lexical level, many studies have shown that context ultimately affects our
interpretation of a word (Khanna & Boland, 2010; Li & Yip, 1998; Zwitserlood, 1989). Much of
the evidence for contextual constraints on lexical access have come from studies of how listeners
handle lexical ambiguity, when the phonological input is either temporarily or globally
consistent with multiple word meanings (e.g. bat). Although adults can ultimately disambiguate
the correct meaning, there is ongoing debate regarding when top-down effects may exert
influence. Does disambiguation occur during early stages of lexical access or does context only
affect a secondary integration or revision stage of comprehension? To get leverage on this issue
numerous studies have been dedicated to identifying the time course of top-down effects, but the
findings to date have been mixed. Using a crossmodal priming paradigm, Zwitserlood (1989)
evaluated activation of both the contextually appropriate and inappropriate meanings of a
homophone as the phonologically ambiguous word unfolded over time. In this study, participants
ultimately showed effects of priming only to a semantic associate of the contextually appropriate
meaning of spoken word. For example, listening to the partial word “gener”, which is consistent
with both general and generous, participants showed facilitated processing of army but not gift.
In support of top-down effects on lexical processing, differential priming based on context was
evident prior to phonological disambiguation. However, the effect of context appeared only after
bottom-up input became available, suggesting that contextual constraints play a role in lexical
item selection but do not constrain the activation of candidate words. Similar findings were shown when lexical items were disambiguated by syntactic construction (Tanenhaus, Leiman, & Seidenberg, 1979). Immediately following a sentence final ambiguous target word, for example *watch*, participants were equally quick to name semantic associates of both the noun and verb interpretation of the word. However, 200ms later, only the syntactically appropriate interpretation showed faster naming latency.

However, subsequent work using the visual-world-paradigm has shown that when sentences contain a strongly constraining verb and a visual context which offers a limited set of referents, context effects appear immediately upon, or even before, phonological input (Brock & Nation, 2014; Dahan & Tanenhaus, 2004). In Dahan and Tanenhaus (2004) participants heard Dutch sentences containing temporarily ambiguous target words, such as *bok* [goat] while looking at a display that contained both the target item and a phonological competitor *bot* [bone]. In a constraining context (e.g. following the verb *klimmen* [to climb]), listeners showed increased looks to a target image of a goat over an image of the phonological competitor upon word onset. Although top-down effects were present early, phonological cues in favor of the competitor over the contextually appropriate target resulted in increased consideration of the competitor, leading the authors to argue that both top-down and bottom-up sources of information are integrated during early lexical activation. More recently, there has been growing evidence that a number of higher-order cues can inform early lexical processes, including syntactic constraints (Strand, Brown, Brown, & Berg, 2018). However, as we will discuss later, eye-tracking studies of top-down effects on lexical disambiguation are largely confounded between predictive activation of specific lexical items and higher-level predictions of the event (Brock & Nation, 2014). In a lexical ambiguity task, Brock and Nation (2014) minimize this possibility by not presenting
target items and their phonological competitors in the same display. Nevertheless, participants show decreased looks to phonologically consistent but contextually irrelevant items even when no contextually relevant images are available to attract their eye-gaze.

Given the ongoing debate regarding when and if top-down effects constrain bottom-up processes during lexical activation, competing theoretical and computational models of lexical access have been proposed. Effects of higher level processing on lower levels of analysis have most often been instantiated both theoretically and computationally through top-down feedback (Kuperberg & Jaeger, 2016; J. L.McClelland & Elman, 1986; McRae & Matsuki, 2013a). Proponents of such feedback have argued that as information is accumulated bottom-up across levels of representation, these in turn can constrain lower-levels of representation. Top-down constraints then result in predictive commitments that shape or even precede the bottom-up input (Kuperberg & Jaeger, 2016; McRae & Matsuki, 2013a). Computationally, models of language comprehension have thus included bi-directional feedback connections that allow for higher-levels of analysis to shape the bottom-up processing stream (e.g. J. L. McClelland & Elman, 1986).

Although the majority of language comprehension models incorporate robust interactivity between bottom-up and top-down feedback, others have argued that top-down feedback is not only unnecessary, but even harmful (e.g. Norris & McQueen, 2008; Norris, McQueen, & Cutler, 2000). Such models argue that effects considered to reflect top-down activation can be achieved solely through highly incremental bottom-up flow of information. For example, both interactive and purely feedforward models of phonological processing can account for faster processing of words over non-words, lexical biases in phonetic categorization tasks, and phoneme restoration effects (Norris et al., 2000). Models of comprehension that rely solely on unidirectional
connections typically employed either separate processing levels for information integration (e.g. Norris et al., 2000), or, more recently through updating of Bayesian priors (Norris & McQueen, 2008). Both types of models argue that no "active" top-down feedback occurs during processing. In Merge (Norris et al., 2000), a model of word identification which incorporates only unidirectional connections between processing levels, lexical effects on word identification occur due to competition between the output of the lexical level and the phonological level, which gets integrated in a separate decision stage that receives input from both processing streams (Norris et al., 2000). In a more recent Bayesian instantiation of unidirectional lexical processing, Shortlist-B (Norris & McQueen, 2008), no active feedback occurs at the time of comprehension, however, the resulting activation pattern is subsequently fed-back down to lower prelexical levels of analysis to shape how the bottom-up input will be used in future comprehension. The lower-level processing remains unaffected while its impact on higher order processes is modulated by the history of those connections. Here effects of semantic context are treated in the same way as effects of word frequency; both are accounted for through patterns that can be acquired over time and shape the prior probabilities of lexical activation (Norris & McQueen, 2008). Thus, although such models do not allow "active" top-down feedback, they often incorporate aspects of top-down processing, either through separate integration processing steps or through the computations used within each processing layer.

As with lexical access, there are parallel effects and controversies at higher levels of processing. There is robust evidence that construction of syntactic structure can be affected by top-down constraints. Take for example sentences such as [4]. In this sentence, there is a

[4] You can tickle the bear with the mirror/paintbrush.
prepositional attachment ambiguity, such that the prepositional phrase is consistent with both a modifier interpretation, in which the preposition modifies the noun (e.g. bear holding a mirror/paintbrush), or with an instrument interpretation, in which the preposition modifies the verb (e.g. tickle using a mirror/paintbrush). Although these sentences are structurally ambiguous, a paintbrush is a much more likely instrument for tickling than a mirror. As Snedeker and Trueswell (2004) show, adults in these sentences interpret the prepositional phrase (with the ...) as an instrument when it concludes with feather and as a modifier when it concludes with mirror, consistent with what they know about the world. Similar studies have shown that adults can use a range of extra-linguistic information to infer syntactic structure, including the referential visual context (Knoeferle et al., 2005; Sedivy et al., 1999; Trueswell et al., 1999).

As in lexical processing, the existence of top-down effects on syntactic disambiguation may reflect either integration of bottom-up cues with higher level constraints or, alternatively, may occur only at subsequent decision or integration stages of processing. While the debate on the architecture that allows effects of higher-level constraints on interpretations at lower levels of analysis continues, any model of adult comprehension must accommodate growing evidence that context informs how a sentence is understood.

1.1.2 Child language comprehension

Like adults, by the time children enter school, they appear to be fluent speakers and competent listeners. Well before the age of five, children master the ability to segment incoming phonological input (Aslin, Saffran, & Newport, 1998; Pelucchi, Hay, & Saffran, 2009), have amassed a large vocabulary (Dail & McGee, 2011), and can construct semantic and syntactic structures of a spoken utterance (Gertner et al., 2006; Hurewitz et al., 2000; Love, 2007; C. H. Noble et al., 2011; Thothathiri & Snedeker, 2011). However, we still know relatively little about
how children make use of these levels of representation in real-time sentence comprehension. The research to date shows that incrementality present from early in life, however, interactivity and the ability to revise develop more slowly across childhood.

1.3.1 Incrementality in child comprehension

Like adults, children seem to process incoming linguistic input as it is acquired and can use the unfolding bottom-up input to constrain likely interpretations. Incremental lexical access has been shown in children as early as 24 months (Swingley, Pinto, & Fernald, 1999). Two-year-olds’ show eye fixation patterns suggestive of activation of both a spoken target word and possible phonological competitors until phonological disambiguation, suggesting that children are narrowing a set of possible lexical items as the relevant phonemes unfold (Fernald, Swingley, & Pinto, 2001; Swingley et al., 1999). For example, Swingley et al. (1999) show that two-year-olds and adults will take longer to identify a target word (doll) in a visual display that includes a phonological onset competitor (dog) then when a display contains a rhyming competitor (ball). Like adults, children can also identify words based on partial input, clearly showing that a complete lexical item does not need to be uttered for lexical access (Fernald et al., 2001).

Incremental processing is also evident across children's syntactic processing. Children have been found to commit to a sentence's syntactic structure before a single syntactic tree can be disambiguated (Snedeker & Trueswell, 2004; Snedeker & Yuan, 2008; Trueswell et al., 1999). For example, children use the semantic bias of a verb to infer syntactic constructions. In a sentence like [5], children hearing the verb put will choose a destination interpretation of the upcoming prepositional phrase (on the napkin) (Trueswell et al., 1999) even when the phrase is also compatible with a modifier interpretation. Although such lexical biases have been most
commonly studied, other bottom-up sources of information, such as prosody (Snedeker & Yuan, 2008), also propagate across levels of representation to make predictive constraints on syntactic analyses.

However, there are age related differences in how bottom-up input is considered. Particularly, children and adults differ in their processing of incoming input when it contradicts earlier commitments. Adults faced with input that is inconsistent with their predicted interpretation, as in the case of garden-path sentences, can revise their original commitments and reinterpret the sentence under a different syntactic structure. However, children faced with the same dilemma show both online looks and offline act-out behaviors that reflect their original predictions, even when those are incompatible with the sentence they ultimately heard (Trueswell et al., 1999; Weighall, 2008).

1.3.2 Interactivity in child comprehension

Despite children's ability to make predictive commitments, adult-like interactivity seems to emerge more slowly (Snedeker, 2013; Snedeker & Huang, 2015). Through age ten, children seem to particularly struggle to use top-down cues to predict upcoming input. Numerous studies using the visual world paradigm have demonstrated children's difficulty using cues from the preceding semantic context (Snedeker & Huang, 2016; Snedeker & Trueswell, 2004) and from their world knowledge (Snedeker et al., 2009) to guide comprehension. As with the preceding section on incremental processing, in the rest of this section I will focus on children's ability to use top-down cues during lexical access and syntactic analysis.

Top-down effects on lexical access have not been studied excessively in children. However, the evidence to date suggests that children do not use top-down sources to guide access to the same degree as adults (Khanna & Boland, 2010; Rabagliati, Pylkkänen, & Marcus,
For example, when hearing a sentence containing a homophone, like [6], adults and children over 12 do not show subsequent facilitation reading the word *grab*. *Grab* is semantically related only to the contextually inappropriate meaning of the homophone, and mature comprehenders rule out this meaning based on the context. However, younger children fail to use context to rule out the contextually inappropriate homophone meaning and show facilitation to *grab* (Khanna & Boland, 2010). Similarly, Rabagliati et al. (2013) found that children can make use of a particular top-down cue, real-world plausibility or the likelihood of the uttered event, to resolve lexical ambiguity. In this study, children were faced with a sentence like [6] which concluded with a homophone. Although semantic association favors the subordinate meaning of

[6] *Jerry was bothered by the shirt’s tag*

the homophone (*knight*), global plausibility of the sentence favors the dominant meaning (*night*). Children select the dominant meaning of the homophone more often in this condition than when global plausibility aligns with semantic association, as in [7], but they do so to a lesser extent than adults and rely more heavily on bottom-up information.

[7] *Elmo watched a funny movie about a castle, and a princess, and a silly dragon. And there was a funny knight*

The evidence that children have difficulty using top-down cues in syntactic analysis is even stronger and more consistent. Specifically, children under the age of 7 consistently fail to use information about the referential context to select the most likely syntactic structure (Kidd & Bavin, 2005a; Snedeker & Trueswell, 2004; Trueswell et al., 1999; Weighall, 2008). For example, Snedeker and Trueswell (2004) played adults and children sentences like example [8].

[8] *Tickle the bear with the mirror*
They varied the kinds of animals in the scene: either there were two bears (e.g., a bear holding a mirror and a bear holding a fan) or there was one bear (holding a mirror) and an irrelevant animal (leopard holding a fan). When there were two bears, adults were more likely to interpret the ambiguous prepositional phrase as a modifier, specifying which bear to tickle. When there was just one bear, they were more likely to interpret the prepositional phrase as indicating an instrument that should be used to do the tickling. Children, in contrast, were uninfluenced by the number of possible referents for the noun and instead responded solely based on the verb (see above). In similar studies using the visual-world-paradigm, children have been shown to struggle with other top-down cues to syntactic structure (Kidd, Stewart, & Serratrice, 2011; Snedeker, Worek, & Shafto, 2009; Huang & Snedeker, 2009a, 2009b).

1.1.3 Theories of Developmental Change

Although differences between adult and child comprehension have been widely reported, the cause of these differences is not yet understood. Two primary theories of developmental change have been proposed. One possibility is that the source of development lies in domain general improvement in executive functioning (e.g. Choi & Trueswell, 2010; Mazuka, Jincho, & Oishi, 2009; Ye & Zhou, 2009). Alternatively, the change may be primarily within the linguistic domain, as the processing architecture becomes more tuned and efficient with increasing language experience. Finally, although these two hypotheses have been most widely discussed, it is possible that both or neither of these factors play a critical role in helping children become adult-like listeners.

1.4.1 Executive Functioning Hypotheses

It has been widely hypothesized that the differences between child and adult language comprehension may reflect differences in domain-general executive functioning (EF). EF refers
to a set of control processes that allow for goal-oriented behavior, including inhibition, selective
attention, cognitive flexibility, and working memory. These skills develop gradually across the
first 20 years of life (P. Anderson, 2002; Diamond, 2001; Mazuka, Jincho, & Oishi, 2009; Ye &
Zhou, 2009). Under this hypothesis, the child and adult language systems are similar, but
children simply lack the EF needed to resolve conflicting cues and inhibit predominant sentence
interpretations.

There has been increasing evidence of recruitment of domain general EF processes in
language comprehension. Resolution of conflicting sentence interpretations has been localized to
the same area of the brain that has been independently associated with domain general conflict
resolution (Novick, Trueswell, & Thompson-Schill, 2005; Novick, Trueswell, & Thompson-
Schill, 2010). When this area is damaged, patients show systematic difficulties with language
comprehension when conflicting sentence representations are in play (Novick et al., 2010). For
example, patients with lesions to this area have difficulties suppressing irrelevant meanings of
homographs and perform similarly to 5-year-olds when faced with garden-path sentences
(Novick et al., 2010).

Although EF may be recruited during language processing, the evidence that EF capacity
relates to variability in sentence comprehension has been largely based on studies of individual
differences. There have been several studies showing that performance on measures of EF are
correlated with garden-path resolution in adults (e.g. Mendelsohn, 2002; Vuong & Martin,
2014). Beyond studies of individual differences, Novick, Hussey, Teubner-Rhodes, Harbison,
and Bunting (2014) tried to directly manipulate executive control in order to determine its causal
relationship with garden-sentence processing. In this study adults were trained on an EF task
aimed specifically at conflict resolution. Those subjects who improved most in their EF training
also showed improvement in sentence revision, suggesting that training in a non-verbal EF task generalizes to linguistic processing.

While the role of EF in sentence comprehension is most commonly associated with revision processes, limitations in EF may also play a part in children’s inability to integrate top-down cues during comprehension. Proponents of this theory argue that children show difficulty with top-down constraints specifically when these conflict with bottom-up sources of information. In support of this theory, when top-down cues are highlighted or competing cues are minimized, children seem to be better able to use top-down sources of information (Meroni & Crain, 2003; Qi, Fisher, & Brown-Schmidt, 2011).

The role of EF in the development of sentence comprehension has been widely proposed, however, the evidence in favor of this theory remains limited. Despite findings of correlations between EF and language tasks in adults, similar findings in children are inconsistent. For example, a recent study by Woodard, Pozzan, and Trueswell (2016) looked at the role of EF on 5 and 6-year-old children's performance on garden-path sentences. They find that garden-path performance is correlated with cognitive flexibility, as measured by the switch cost on an Erickson Flanker task (Woodard et al., 2016). Although they use several other EF measures, they find null effects across other tests and, specifically, on measures of inhibition. Additionally, Huang and Hollister (2019) find that language experience and not EF predict garden-path performance. The few studies investigating a causal relationship between EF and sentence comprehension have been conducted only with adult listeners (e.g. Novick et al., 2014), however, the cause of processing difficulty in adults may differ from the cause of developmental change in the same ability.

1.4.2 Language experience hypotheses
An alternative hypothesis is that children’s improvements in sentence comprehension are a result of maturation of the language system due to continued language experience. Under this hypothesis children have learned the relevant linguistic structures, but repeated activation with use improves the quality of representation and the efficiency of linguistic processing (Dahan, Magnuson, & Tanenhaus, 2001; MacDonald, 1994). Children specifically show difficulty using top-down information to constrain sentence meaning. For top-down constraints to play a role in comprehension, unfolding information needs to work its way to higher level processing before analysis at those levels can be used to constrain lower-levels of representation (e.g. you need to first construct a syntactic structure before that structure can constrain possible lexical items). Faster and more accurate transmission of information across the language system would allow for earlier use of such constraints. Under the current theory, children with less experience with any one language would show slower propagation of information across the language processing hierarchy and would be less likely to integrate such information in time to aid in sentence interpretation.

Change in the efficiency of the language processing system have been well documented across childhood. Across the second year of life children show a marked improvement in their ability to quickly direct their gaze to referents of spoken words, indicating increasing efficiency of lexical comprehension (Fernald, Perfors, & Marchman, 2006). In addition, language processing speed is related to vocabulary growth and grammatical development (Fernald et al., 2006). Borovsky, Elman, and Fernald (2012) presented children and adults with sentences like [9] while looking at a display containing the target object (treasure),

[9] The pirate hides the treasure
an agent-related object (ship), an action-related object (bone), and a distractor. They find that vocabulary, and not age, predicts the speed with which children combined semantic cues from both the subject and verb of the sentence to direct their gaze to the target object. The authors argue that greater language experience, as measured by vocabulary, lead to better ability to integrate available information to predict syntactic structure.

Looking more specifically at syntactic ambiguity resolution and children’s ability to use top-down cues to make predictions, several studies have argued that language experience plays a primary role. For example, S. Anderson, Farmer, Goldstein, Schwade, and Spivey (2011) showed that vocabulary predicted children's ability to use referential context to constrain syntactic analyses. Specifically, children with higher vocabularies showed more accurate act-out behavior when faced with a garden-path sentence like [10] when two possible referents were present in the visual display, a cue that in adults encourages a modifier reading of the prepositional phrase and results in fewer garden-path errors. Although these studies have primarily focused on the role of vocabulary in predicting sentence comprehension, as noted above, vocabulary growth is related to linguistic processing speed and the quantity and quality of language input.

1.2 On-going challenges

Our understanding of children's moment-to-moment sentence comprehension and how it develops is constrained by several challenges. First, the source of developmental change is difficult to isolate because domain-general and domain-specific developments are largely confounded across this age-span. EF is highly correlated with measures of language experience, including vocabulary, as well as other key characteristics such as SES and general intelligence.
Since most of the research looking into this question relies on studies of individual differences, the high correlations across these critical measures poses a challenge for interpretation.

In addition, our current understanding of moment-to-moment comprehension in children has overwhelmingly relied on studies using the visual-world-paradigm (VWP). This paradigm provides time-sensitive information about children’s unfolding sentence interpretation. However, the use of this paradigm places constraints on both the linguistic representations that are targeted for investigation and the context in which sentences are heard.

The design of most VWP studies, relies on presenting a temporarily or globally ambiguous utterance as subjects look at a visual display containing a limited set of alternative interpretations. In relying on linguistic ambiguities, the interpretations of which can be visually discerned, large numbers of studies have been conducted on a limited set of linguistic constructions. In addition, the presence of a referential display can confound the level of representation that was used for disambiguation. For example, in most eye-tracking studies looking at lexical ambiguity, gaze to a target image over a phonological competitor can be interpreted as either a prediction of a specific lexical item or reflecting semantic or event-based commitments that may not extend to word activation. Thus, lexical access remains a highly understudied stage of children's sentence comprehension. While this challenge is particularly evident in studies of lexical ambiguity, similar interpretive concerns arise for studies of higher-level constructions as well.

In addition, the VWP imposes constraints on the context under which comprehension occurs. Many studies are designed such that a sentence can end in only two to four ways and
typically just one of these endings will meet all the constraints of the sentence and discourse. By offering a set of possible interpretations, the use of the VWP can artificially limit or expand the interpretations considered and the degree to which they are activated. What's more, listeners are often faced with a series of isolated sentences or short paragraphs, such that relevant discourse is limited to only the immediately preceding context. Thus, the VWP presents children with a comprehension task that is highly unlike the language they experience in everyday life and the degree to which children rely on various sources of information may differ when language is presented in more naturalistic contexts.

In the remainder of the introduction, I will describe the three papers presented here and how each addresses these limitations. In these, we look at moment-to-moment spoken language comprehension in children and adults in order to identify how information from both bottom-up and top-down processing streams is used for lexical and syntactic processing. By studying how such information streams inform processing at these levels, we work to identify possible internal (Paper 1) and external pressures (Papers 2 and 3) that may affect children’s comprehension ability. Better characterization of children's ability to use top-down constraints during sentence comprehension across a wider set of listening contexts can provide important constraints on theories of developmental change.

1.3 Present Research

In Paper 1, we use the visual-world-paradigm to compare monolingual and bilingual children’s comprehension of garden-path sentences. Garden-path sentences provide insight into two critical steps in sentence comprehension. First, by looking at a temporarily ambiguous grammatical structure we can evaluate the degree to which adults and children are able to commit to a likely sentence interpretation and which information they can bring to bear on this
decision. In the current project we specifically focus on children's ability to use referential context. Second, in a garden-path sentence the syntactic structure typically preferred early in the sentence turns out to be inconsistent with subsequent input, requiring revision. Children, though age eight, show particular difficulty both with the ability to use referential cues to update their original syntactic commitments and with using later arriving information to revise their decisions (Trueswell et al., 1999; Weighall, 2008). Improvements in children's ability to resolving garden-path sentences have most often been attributed to their developing executive functioning skills (Novick, Hussey, Teubner-Rhodes, Harbison, & Bunting, 2014; Woodard, Pozzan, & Trueswell, 2016). However, it is also possible that the difficulty in revision reflects linguistic rather than domain-general limitations (Borovsky, Elman, & Fernald, 2012).

Prior studies investigating the probable causes for children's difficulty with sentence comprehension have largely relied on correlational findings. However, as described earlier, early school age years are associated with contemporaneous improvements in both linguistic and cognitive domains, making it difficult to isolate their respective roles in language development. By comparing monolingual and bilingual children in this task, we can both expand and decorrelate linguistic experience with more general improvements in executive control. We find that language experience rather than executive control best predict performance on garden-path sentences. However, we also suggest that a bilingual language experience may provide an advantage in the ability to use referential context that is not based on differences in EF.

In Paper 2 we evaluate the ability of adults and children to integrate semantic information into their lexical predictions using the test case of negation. Although children as young as two frequently produce the word “no”, though age four or five they still struggle to understand negated sentences (Kim, 1985; Nordmeyer & Frank, 2014; Reuter, Feiman, & Snedeker, 2018).
Prior studies using eye-tracking have suggested that negation is processed by first interpreting the affirmative sentence and then apply the negation offline in a two-step process (Nordmeyer & Frank, 2014). However, more recent work using the same paradigm has shown that given a supportive context, children even as young as two can show online integration of negation (Reuter et al., 2018).

Although adults and children have been shown to successfully interpret negation online, we do not yet know what features of context are necessary to allow incremental use of sentence polarity during sentence comprehension. In Paper 2, we test the hypothesis that a polar question-under-discussion may be sufficient to facilitate the processing of negation. In the current study we turn to ERPs in order to evaluate the incorporation of polarity into online sentence comprehension. Using the N400 response as a measure of lexical access and integration, we can evaluate children’s and adults’ ability to use negation online to update lexical predictions and children’s ability to process negation without a visually constrained set of alternative interpretations. We find that a preceding polar question is not sufficient for online integration of polarity during lexical access. However, future work should be conducted to determine whether this conclusion holds when the overall discourse is more reliable and conducive to predictive processing.

While traditional ERP designs, such as the one used in Paper 2, avoid some of the pitfalls of the VWP, the listener's experience still differs greatly from daily language experience. Like the VWP, ERP studies of language processing are built around controlled contrasts, resulting in repetitive strings of isolated sentences or a series of short disconnected paragraphs. What's more, the task children are asked to perform is often novel and motiveless (e.g., following instructions or remembering sentences). In addition, unlike experiments using the VWP, which are often
short, interactive, and fun, traditional ERP designs are also difficult to run and are often long and tedious. Participants are asked to remain still and attentive often for an hour or more - a challenge for any child.

Due to the limitations associated with both the VWP and traditional ERP designs, we cannot be sure that these studies tap into the skills that are relevant to comprehension in more ecologically valid contexts. Most sentences people hear are integrated into a rich discourse structure rather than a series of disconnected sentences and rarely are the referents of those sentences directly accessible in the visual environment. These differences may significantly impact the comprehension processes that are required for understanding. Lack of a rich discourse can diminish the use of top-down constraints while the presentation of visual referents or the repetitive nature of the paradigm may allow for preencoding, overestimating listeners ability to predict upcoming input.

In Paper 3 we adapt a novel method for studying language comprehension in children using a naturalistic story-listening task. Adult and child participants listened to a story as ERPs time-locked to the onset of every word were recorded. With the current design we were able to record ERPs from hundreds of trials in a task that is short, fun, and ecologically valid. Not only is the task easier to administer than traditional ERP paradigms, but studying language comprehension during a coherent, age-appropriate story or passage may provide a better window into how comprehension occurs in real-world contexts.

Using this method, we evaluate the degree to which children use top-down cues for lexical access. Adults often rely on context to predict upcoming words (Nieuwland & Van Berkum, 2006; Van Berkum, Hagoort, & Brown, 1999; Xiang & Kuperberg, 2015). Much of the evidence for predictive use of top-down constraints comes from studies looking at the N400
response, an event-related potential (ERP) that indexes ease of lexical access. In adults the N400 response reflects a word’s predictability given the semantic, syntactic, and pragmatic constraints of the context (Kutas & Federmeier, 2011). However, we know far less about how children access word meaning or how this is reflected in their neural responses.

To assess the use of both bottom-up and top-down constraints, each content word in the story was coded for frequency, semantic-relatedness (LSA), and cloze probability. Evaluating the degree to which the N400 response patterns with increasingly more interactive models of lexical access can give us insight into the sources of information children and adults are recruiting to access and integrate the words they are hearing. We find that both adults and children make use of top-down constraints during comprehension, however, children rely more on bottom-up cues such as frequency.

Across the three presented papers, we will argue that children's ability to use top-down cues during online sentence comprehension matures with growing language experience. However, it is also affected by the context in which an utterance is heard. ERP paradigms that allow us to study naturalistic language comprehension may allow children to display a greater range of their ability and have the potential to provide important insight into how children use different sources of information in everyday conversations.
Chapter 2:

What bilingualism can teach us about online ambiguity resolution and garden-path revision
2.1 Introduction

By age five most children can be described as capable conversational partners; they understand what is spoken to them and are able to respond in kind. This is a remarkable feat when one considers that understanding a spoken sentence requires one to parse a stream of sounds into words, analyze their meanings, determine the grammatical structure they appear in, and integrate the information into a broader discourse context. Research into the development of online sentence processing has shown that the system that allows a child to accomplish this task is remarkably similar to that of an adult (Trueswell & Gleitman, 2004). As a sentence unfolds, the incoming information is propagated across multiple levels of linguistic analysis; from low level acoustic processing, through identification of lexical items, up to syntactic and semantic analysis, and eventually to higher order constraints from context and pragmatics (Altmann, 2001; McRae, Hare, Elman, & Ferretti, 2005; Snedeker & Trueswell, 2004). The incremental and interactive propagation of information allows both children and adults to constrain upcoming alternatives and thus make predictive commitments regarding upcoming input (Snedeker & Trueswell, 2004). Upon hearing a sentence like [1], we do not need to wait for the final word to know that the sentence will conclude with a noun and most likely a “kite”.

[1] “On a windy day, it is fun to fly a ...”

However, children’s sentence comprehension differs from adults’ in two striking ways. First, children younger than 8 or 10 have difficulty using top-down information to guide sentence interpretation (Hurewitz et al., 2000; Kidd & Bavin, 2005a; Kidd et al., 2011; Snedeker & Trueswell, 2004; Trueswell et al., 1999; Weighall, 2008). For example, unlike adults, children show difficulty using referential context (Trueswell et al., 1999; Weighall, 2008), world-knowledge (Kidd et al., 2011; Snedeker et al., 2009), and pragmatics (Huang & Snedeker, 2009a,
to inform their sentence interpretations. Second, while both children and adults can make predictive commitments, adults are able to revise their predictions if they turn out to be incorrect (Trueswell et al., 1999). Children however often fail to revise their initial predictions and erroneously maintain their original interpretations (Choi & Trueswell, 2010; Trueswell et al., 1999).

These differences between child and adult language comprehension have been widely investigated by looking at the processing of prepositional attachment ambiguities, such as [2].

[2] Put the frog on the napkin in the box.

In such sentences the 1st prepositional phrase, on the napkin, is temporarily ambiguous; it can be interpreted as either the goal of the verb, put (VP-attachment), or as a modifier of noun-phrase, frog (NP-attachment). Trueswell et al. (1999) have shown that when processing such sentences, both adults and children show sensitivity to the lexical bias of the verb and originally interpret the prepositional phrase as the goal of put. Critically, adults do not make this commitment when the context of the sentence contains two possible referents, or frogs. They successfully use referential context to interpret the ambiguous phrase as a modifier of the noun, indicating which referent should be moved. Children however fail to use contextual information, relying on lexical bias alone (Kidd & Bavin, 2005a; Snedeker & Trueswell, 2004; Trueswell et al., 1999). In addition, when adults do commit to a goal interpretation of the prepositional phrase, they are able to revise their prediction once the second prepositional phrase is uttered. Children on the other hand, fail to revise their prediction and ultimately settle on and act out an erroneous interpretation (Kidd et al., 2011; Snedeker & Trueswell, 2004; Trueswell et al., 1999). Children’s difficulty with this task has been shown across numerous studies even when the task is placed within a narrative (Hurewitz et al., 2000), the pragmatic constraints from the referential context
and maximized (by placing two frogs on different colored napkins) and when presentation of the
display is delayed in order to reduce online action planning (Weighall, 2008; but see Meroni &
Crain, 2003).

Although these striking differences between adult and child sentence comprehension
have been widely studied, their cause is not yet understood. Two primary hypotheses have been
put forth. One possibility is that the source of development lies in domain general improvement
in executive functioning. Alternatively, improvements in sentence comprehension may relate to
continued language experience across later childhood. Finally, although these two hypotheses
have been most widely discussed, it is possible that both or neither of these factors play a critical
role in helping children become adult-like listeners.

2.1.1 Executive Functioning Hypotheses

It has been widely hypothesized that the differences between child and adult language
comprehension may reflect differences in domain-general executive functioning (EF) (Choi &
Trueswell, 2010; Mazuka et al., 2009; Ye & Zhou, 2009). Under this hypothesis, the child and
adult language systems are remarkably similar, but children simply lack the cognitive control
needed to resolve conflicting cues and inhibit predominant sentence interpretations.

Cognitive control, or executive function (EF), refers to a set of control processes that
allow for goal-oriented behavior, including inhibition, selective attention, cognitive flexibility,
and working memory. These skills are localized within the dorsolateral prefrontal cortex and
develop gradually across the first 20 years of life (Mazuka et al., 2009; Ye & Zhou, 2009).
Children, for whom the prefrontal cortex is still developing, show deficits in tasks requiring
cognitive control; they often fail to inhibit preeminent responses and face difficulties adapting to
a change in rules, perseverating on a single response pattern (Stahl & Pry, 2005). Childhood is
associated with a marked maturation of the dorsolateral prefrontal cortex, and related cognitive abilities (P. Anderson, 2002; Diamond, 2001; Garon, Bryson, & Smith, 2008). The improvement in executive functioning across this age span has been related to improvements of a variety of concurrently developing skills, such as theory of mind (Carlson, Moses, & Breton, 2002; Hughes, 1998), biological reasoning (Zaitchik, Iqbal, & Carey, 2014), and even Piagetian tests of conservation of number (Poirel et al., 2012).

In the domain of language development, executive function has been most commonly implicated in children’s failure to revise garden path sentences, which require a listener to update their earlier predictions or switch to an alternative interpretation. There is increasing evidence of recruitment of domain general EF processes in such revision tasks. Resolution of conflicting sentence interpretations has been localized to syntactic processes within the left inferior frontal gyrus (LIFG), which is home to Broca’s area (Novick, Trueswell, & Thompson-Schill, 2005; Novick, Trueswell, & Thompson-Schill, 2010). This same area has been independently associated with domain general tasks related to conflict resolution and specifically overriding automatic responses. For example, the LIFG has been shown to be active during EF tasks such as the verbal Stroop task, in which participants must refrain from a dominant response (reading a color word) in favor of an alternative (naming the color of the ink in which it is displayed) (Novick et al., 2005). Additionally, Novick et al. (2010) detail the performance of a patient, IG, with damage to the LIFG, when given the garden-path sentence processing task discussed earlier (“Put the frog [that’s] on the napkin in the box”). Although IG had no difficulty with the unambiguous sentence structure, when faced with a garden-path he showed similar processing to that of 5-year-olds, both in prolonged looks to an incorrect destination and incorrect act-out behaviors on over 60% of trials.
Beyond neurophysiological evidence that EF may be involved in comprehension of garden-path sentences, the role of domain-general resources in accounting for typical variation in garden-path comprehension has been primarily supported by studies of individual differences (Mendelsohn, 2002; Novick et al., 2005; Novick et al., 2010; O'Rourke & Colflesh, 2014; Vuong & Martin, 2014). The evidence to date has found that inhibitory processing, as measured by verbal Stroop tasks, correlates with resolution of syntactic ambiguity (Mendelsohn, 2002; Vuong & Martin, 2014; but see Engelhardt, Nigg, & Ferreira, 2017).

In addition to studies of individual differences, Novick et al. (2014) tried to directly manipulate executive control in order to determine its causal relationship with garden-path sentence processing. In this study adults were trained on an N-back task with lures, an inhibitory control task that specifically tests for conflict resolution, and performed a garden-path sentence comprehension task both before and after training. Those participants that improved most during EF training also showed improvement in sentence revision, suggesting that training in a non-verbal cognitive control task generalizes to linguistic performance. Although this is a step towards suggesting a causal link between EF and language ability, no other measures of language processing were taken, so the results may reflect better learners being better across all tasks.

Although research into individual differences has been primarily conducted with adults, recent work has suggested that the relationships between EF and garden-path resolution may also extend to children. Qi et al. (2011) found that children with better inhibitory control, as measured through a Simon-Says game, made fewer errors when acting out garden-path sentences similar to the ones tested by Trueswell et al. (1999) (e.g. put the frog on the pond into the tent). In a more extensive study, Woodard et al. (2016) gave 5 and 6-year-old children the garden-path “put” task
and an executive functioning battery. They found that performance on ambiguous garden-path sentences was related to cognitive flexibility, as measured by the switch cost on an Erickson Flanker task (Woodard et al., 2016). However, they did not find the same correlation for measures of inhibitory control, which has most often been associated with sentence revision in adults. Woodard et al. (2016) argue that the ability to adapt flexibly to new rules or information, rather than simply inhibiting preeminent responses, underlies children’s ability to recover following garden-path misanalyses.

Finally, proponents of the EF hypothesis have also argued that that pattern of errors seen across garden-path tasks implicates EF as the underlying cause of children’s difficulties. Adapting the Trueswell et al. (1999) paradigm, Choi and Trueswell (2010) studied children’s comprehension of the garden path sentences in Korean. Korean is a head final language, in which the verb comes at the end of the verb phrase. Like the studies conducted in English, children were faced with a temporary ambiguity between a modifier and destination interpretation of a prepositional phrase, but here the intended meaning was clearly disambiguated by the verb at the end of the sentence. Although the verb is a highly reliable and clear cue towards the correct syntactic parse, children were still unable to override their earlier commitments, as seen both in their online eye movements and subsequent errors in act out behaviors. This pattern of results suggests that the difficulty children face is not with computing the final syntactic parse, but specifically with overriding their existing interpretations.

While the EF hypothesis has most often been called upon to explain children’s difficulties with sentence revision, limitations in EF may also play a part in children’s inability to integrate top-down cues during comprehension. As with revision, this hypothesis predicts that children have the relevant linguistic knowledge, but face processing difficulties when top-down
information competes against bottom-up constraints. In favor of this hypothesis, although children show difficulty using context to resolve prepositional phrase ambiguities, by the age of 5 they can reliably produce restrictive modifier clauses and do so specifically when the context requires it (Snedeker & Yuan, 2008). In addition, there is evidence that children can use referential context when attention to other cues is minimized. Particularly, in a study by Meroni and Crain (2003) children were given the “put” task but with two modifications. First, both possible referents, or frogs, were placed on different colored napkins, which was intended to weaken the pragmatic cue that the more salient referent is the intended object and make referential context a more critical source of information. Second, the display was hidden until after the sentence was uttered in an attempt to stop children from making an act-out plan as the sentence unfolded. Under these circumstances children as young as 4 showed sensitivity to referential context. Finally, in a study by Qi et al. (2011) children were given the same task, but the relationship between the subject and modifying object were more closely semantically related (e.g. frog on a pond, rather than frog on a napkin). Subsequently, they too found that children under these circumstances showed more sensitivity to referential context. Qi et al. (2011) argue that replacing the modifier strengthened the competing alternative construction. Although such success on this paradigm is sparse, it suggests that children may be able to use top-down cues correctly when competition from other cues is minimized. As with sentence revision, EF is needed in order for less reliable cues to override the earlier acquired and more robust bottom-up sources of information.

2.1.2 Language Experience Hypothesis

An alternative hypothesis is that children’s improvement in sentence comprehension may reflect increasing language experience. Language comprehension involves the coordination of
many separate pieces of knowledge. In order for top-down cues to successfully constrain incoming input, allowing for prediction, the listener must be able to both formulate robust representations of the relevant linguistic information and be able to use that information in time to inform interpretation. Continued exposure to language may increase the quality of linguistic representations and the efficiency of children’s language processing system, resulting in improved ability to integrate sources of information during online sentence comprehension.

Prior evidence has shown that experience with particular language items or structures improves the quality of their representations (Dahan, Magnuson, & Tanenhaus, 2001; MacDonald, 1994; Savin, 1963). For example, frequent words are acquired earlier (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012), identified faster (Oldfield & Wingfield, 1965), and are better recognized in low signal-to-noise ratios (Savin, 1963). Repeated experience not only effects lexical processing but can impact representations at higher levels of analysis. Like words, frequent syntactic structures are constructed more accurately and are less likely to be over-generalized (Ambridge, Pine, Rowland, & Young, 2008; Gernsbacher & St John, 2013; Wonnacott, Newport, & Tanenhaus, 2008). Low frequency structures, such as use of the passive voice or center embedding, are acquired late (Nippold, 1993) and their comprehension accuracy improves with experience and training (Street & Dąbrowska, 2010).

Not only is there evidence that prior experience plays a role in the representations computed across the processing stream, but language experience has also been associated with efficiency gains across language acquisition (Rabinowitz, Ornstein, Folds-Bennett, & Schneider, 1994). Children exposed to more child-directed speech show processing gains by 24 months, directing their gaze faster to target images upon hearing a spoken word (Weisleder & Fernald, 2013). Across the 2nd year of life, children show marked improvements in the speed with which
they are able to identify spoken referents. These improvements are related to concurrent gains in vocabulary and future acquisition of vocabulary and grammar (Fernald, Perfors, & Marchman, 2006). Increases in language processing efficiency can be seen for higher-level linguistic structures. For example, in a looking-time study investigating children’s and adults’ ability to integrate information from multiple linguistic cues, Borovsky et al. (2012) showed that across ages, greater language experience, as measured by vocabulary, was related to better integration of available information during syntactic analysis.

Finally, recent evidence has raised questions about the role of EF in revision. Huang and Hollister (2019) looked at children’s (ages 3-6) ability to revise a sentence interpretation from an active to a passive syntactic construction. They showed that children’s ability to revise was best modeled by their language ability and not cognitive control. This is consistent with growing evidence that language experience shapes not just global measures of language acquisition, like vocabulary, but online processing and sentence revision (S. Anderson et al., 2011; Huang & Hollister, 2019; Nation, Marshall, & Altmann, 2003).

2.1.3 Current Study

Although both the domain general and domain specific hypotheses have been proposed there is still very limited evidence for either. Most prior studies have relied on investigations of individual differences. However, executive control, language ability, and socio-economic status are strongly interrelated in development (K. G. Noble, Norman, & Farah, 2005), making it very difficult to determine the independent contribution of each.

The current study evaluates these hypotheses by comparing online language comprehension in monolingual and bilingual children on prepositional attachment ambiguities such as those studied by Trueswell et al. (1999) (e.g. *Put the frog [that’s] on the napkin in the*
These populations grow up in very different language learning environments, which may allow us to test how linguistic proficiency and EF may differentially affect the development of ambiguity resolution. First, while bilingual children acquire two languages, they have less experience and consequently show poorer proficiency in either language than monolinguals (Bialystok, Luk, Peets, & Yang, 2010). While the total vocabulary of bilinguals across both of their languages is, on average, equal or greater than the single vocabulary of their monolingual peers (Hoff et al., 2012), bilinguals generally have lower vocabulary scores than monolinguals within each of their languages, with the gap being greater in their second language (Bialystok, 2001; Marchman, Martinez-Sussmann, & Dale, 2004; Volterra & Taeschner, 1978).

While bilingual children may have less language experience, children growing up with two languages have been shown to have an executive functioning advantage over monolinguals, commonly scoring higher on measures of cognitive control processing (Poulin-Dubois, Blaye, Coutya, & Bialystok, 2011). The advantage in executive functioning has been related to the need to suppress one language in order to produce a second one. It is argued that the need to frequently switch between two languages results in a specific inhibitory control advantage in bilingual populations (Bialystok, 2009; Carlson & Meltzoff, 2008). Interestingly, the executive functioning advantage has been shown to differentially affect only some functions typically thought to fall under the umbrella of EF, specifically, bilingual advantage has been shown to affect inhibitory and set-shifting ability while showing no such advantage in working memory (Bialystok & Viswanathan, 2009).

The extent and reliability of this executive control advantage, however, has been questioned (de Bruin, Treccani, & Della Sala, 2014; Duñabeitia et al., 2014; Paap & Greenberg, 2013). One of the goals of the current experimental paradigm is to see whether a bilingual
executive functioning advantage can be replicated with the current population and experimental procedures. In addition, with increasing interest in the variability of the bilingual experience and how it may relate to cognitive benefits (Hwajin, Hartanto, & Yang, 2016; Kaushanskaya & Prior, 2015; Yow & Li, 2015), we will subdivide our bilingual group and evaluate separately children that acquired a second language early and those that acquired it relatively late. Some studies have found that early bilinguals are more likely to show cognitive advantages (Hwajin et al., 2016; Luk, De Sa, & Bialystok, 2011; Yow & Li, 2015), though others have failed to find relevant differences (Pelham & Abrams, 2014).

Given the inconsistency of the EF advantage in bilinguals in prior research, it is possible that no reliable differences may exist in our population sample. We hope that by including both monolingual and bilingual participants we may be able to also evaluate the role of executive functioning by looking at individual differences. Unlike studies of monolingual populations, where language processing is strongly correlated with EF, by including bilingual speakers we may be better able to isolate EF development from English language experience.

By comparing monolingual and bilingual ability to use referential cues, make predictive commitments, and revise erroneous interpretations, we hope to gain insight into the developmental changes that lead to adult-like sentence comprehension. If bilingual children do have lower language proficiency and higher EF, a bilingual advantage in prepositional phrase ambiguity would support the EF hypothesis. On the other hand, monolinguals outperforming bilinguals would support the role of language experience in the development of sentence comprehension. It is also possible that neither improvements in executive functioning nor language ability are related to the development of ambiguity resolution and garden-path revision. Finally, monolingual and bilingual participants may show different developmental trajectories.
and may differentially rely on domain-specific and domain-general capacities. In addition to planned group comparisons between monolingual and bilingual participants, individual difference analyses collapsed across language background can help detangle the effects of domain general and domain-specific improvements and their role in language comprehension. By including children with different language histories, we can obtain a sample with greater linguistic variability and weaker correlations across the key measures of language experience and cognitive control.

2.2 Methods

2.2.1 Participants

81 children between the ages of 5;0 and 7;6 participated in the study. Participants included 34 children that were monolingual English speakers (M<sub>age</sub> = 6.07, SD = 0.57), 31 bilingual children (M<sub>age</sub> = 6.22, SD = 0.56), and 16 children (M<sub>age</sub> = 6.07, SD = 0.82) that were exposed to a second language but did not meet criteria for bilingualism. An additional 41 subjects completed the first session but were not included because they did not conclude both sessions.

Bilingualism was determined by a language use survey administered to the parent. To be classified as bilingual, the child had to be exposed to a non-English language through a primary caretaker at least 75% of the time or be exposed to a non-English language for a minimum of 4 hours a day. Both early sequential and simultaneous bilinguals were accepted. Children whose native language was not English had to be enrolled in a school where English was the medium of instruction, with native English-speaking peers, for a minimum of 2 years in order to participate. Children were classified as monolingual if they spoke only English and had no reported exposure to another language. The bilingual children in our sample spoke a diverse set of non-English
languages, including six speakers of Spanish, four Mandarin, three Japanese, three Portuguese, and two Russian speakers.

In addition, we subdivided our bilingual children into 12 early bilinguals (exposed to two or more languages before the age of 2) and 15 late bilinguals (exposed to a second language after the age of 2). Four children were excluded due to a lack of sufficient information about the age of second language acquisition.

Only the monolingual and bilingual groups were included in the planned group analyses. Children who had exposure to a second language but did not meet the criteria for bilingualism were included only in analyses of individual differences.

2.2.2 Procedure

Children completed a task designed to test their ability to resolve prepositional phrase ambiguities in garden path sentences (Garden-Path Task). In addition, all children completed an executive functioning battery, measures of English language proficiency, and control measures of non-verbal problem-solving ability. Descriptions of each task are included below.

All participants come in for two study sessions. All tests were administered in the same session and test order for all subjects. During the first session participants completed the Executive Functioning battery and standardized measures of English language skills using the Peabody Picture Vocabulary Test (PPVT IV, Dunn & Dunn, 2007) and the Test of Receptive Grammar (TROG-2, Dunn & Dunn, 2007). During the second session, children completed the Garden-Path Game, and took the KBIT Matrices Test (Kaufman & Kaufman, 1990), a measure of non-verbal intelligence. In addition, parents were asked to fill out a self-report measure of socio-economic status. Parents reported their highest achieved level of education, current employment status, and occupation.
2.2.3 Tasks

2.2.3.1 Garden-Path Task

The Garden-Path sentence ambiguity resolution game was adapted from the original work of Trueswell et al. (1999), but was redesigned for presentation on an eye-tracking computer display.

Materials

Children were asked to listen to instruction sentences while looking at a visual display containing relevant items. Based on the original work of Trueswell et al. (1999), critical trials were garden-path sentences that began with the verb “put” and contained a temporary prepositional phrase attachment ambiguity that was resolved with subsequent information (1a).

1a) Put the frog on the napkin in the box

1b) Put the frog that’s on the napkin in the box

In these sentences the first phrase (on the napkin) could be interested as either a destination (VP-attached) or as a modifier (NP-attached). This ambiguity is resolved by the second prepositional phrase (in the box), which requires the first prepositional phrase to be interpreted as a modifier. Control trials (1b) were unambiguous, where only the modifier interpretation was grammatically plausible.

All visual displays were arranged into either a 1-Referent or a 2-Referent context. Each display configuration within a trial consisted of 2 animals; a target animal (Target Animal) that is specified by the instructions and a competitor animal (Other Animal). Target animals always appeared on top of an object, and thus could be accurately described using a modifier phrase (e.g. a frog on a napkin). Half of the critical trails were consistent with a 1-Referent context and contained only one animal that was a possible referent for the object mentioned by the
instructions (e.g. 1 frog). The other half were consistent with a 2-Referent context, where both animals are plausible referents and a modifier phrase is necessary in order to identify the proper Target (e.g. 2 frogs, where one of them appears on a napkin).

Each display also contained 2 possible destinations; a correct goal (Goal) specified by the instructions and a distractor destination (Distractor). The Distractor was always equivalent to the object under the Target animal (e.g. an empty napkin). Thus, each display contained an image consistent with both a modifier and a destination interpretation of the ambiguous prepositional phrase. A sample display is included in Figure 2.1.

![Figure 2.1: Example visual display for the 1-Referent (left) and 2-Referent (right) conditions](image)

**Procedure**

Children were tested individually on a desktop computer with eye movements tracked using a Tobii T-60 remote eye-tracking system. After eye-tracking calibration, each child was told they would see several pictures on the screen and would hear some instructions. They were asked to carry out the instructions by moving the pictures with a mouse. For each trial, participants were shown a set of pictures containing animals and objects. The 4 images appeared on the screen one at a time. Each image was located in a different quadrant of the screen. The
quadrant location in which a particular image appeared was randomized. All images were labeled upon their appearance (‘frog’, ‘napkin’, ‘tunnel’) with prerecorded labels. Following presentation of the stimuli, pre-recorded instructions were played, and then the child was asked to carry out the spoken instruction by dragging the images on the screen. Children could indicate completion of the behavior by clicking on a button in the middle of the screen labeled with “Done!” Participants could only perform the actions after the instruction was over and a mouse pointer appeared on the screen.

Participants were given two practice trials followed by 24 test trials, each of which contained a target and a filler instruction. 12 of the target trials were critical trials and 12 were control. The display remained consistent for both sentences in a trial, with the locations of the images resetting between the two instructions within each trial. The objects were only labeled at the start of each trial, when images were originally displayed on the screen. For half the trials, the target sentence was presented before the filler sentence and for half the trials the target sentence was presented following the filler. Filler instructions also involved the verb “put” but contained only 1 prepositional phrase. These instructions used a variety of ways to indicate the goal destination (e.g. “Put the X in the Y”, “Put the X and Y together”, etc.). Ambiguity was counterbalanced across images, and trials were shown in a blocked, randomized order.

**Action Coding**

Each movement of the images on the screen was recorded as was an image of the final display. Actions were coded as correct if the Target Animal was placed directly on the correct Goal. Any clicks that resulted in no movement were disregarded.
2.2.3.2 Executive Functioning Battery

*Simon Tasks Arrows*

Simon’s tasks represent a classic executive control task where subjects are required to respond to a stimulus while ignoring interference from irrelevant aspects of that stimulus (Davidson, Amso, Anderson, & Diamond, 2006; Simon & Craft, 1970). Specifically, subjects are asked to respond to a stimulus while ignoring the side of the screen that the stimulus appears on. In these tasks a congruent condition represents a trial in which the relevant and irrelevant features of the stimulus correspond to the same response (e.g. pressing a key on the right side of the keyboard when the stimulus appears on the right side of the screen). In incongruent trials, relevant and irrelevant features of the stimuli create a conflict where the side of stimulus presentation does not match the correct response, causing an inhibitory demand (Davidson et al., 2006; Simon & Craft, 1970) (e.g. pressing a key on the right side of the keyboard when the stimulus appears on the left side of the screen.)

In the Simon’s Arrows Task, children were presented with stimuli in the shape of arrows that appeared on either the left or right side of the screen and pointed to a button on the keyboard in front of the computer. They were instructed to press the button that the arrow points to, selecting either the left key (w) or a right key (p) accordingly. On congruent trials the arrow appeared on the same side of the screen as the correct response key and pointed directly down at that key. On incongruent trials, the arrow appeared on the opposite side of the screen from the response key with the arrow pointing at a 45-degree angle toward the key (e.g., an arrow on the right which points to the key on the left).

Children were first shown 2 sets of practice trials with feedback; during the first practice trials they were shown 4 congruent trials and during the second practice round they were shown
2 congruent and 2 incongruent trials. Congruent and Incongruent trials were randomly interspersed throughout the 40 test trials.

**Simon’s Pictures Task**

The Simon’s Pictures Task is very similar to the Arrows Task. Children were instructed to press an orange button on the left side of the keyboard when they saw an orange butterfly and a green button on the right side of the keyboard when they saw a green frog. On the congruent trials, the image appeared on the same side as the correct response button. On the incongruent trials, the picture appeared on the opposite side, near the wrong button. Children were first given 3 sets of practice trials with feedback: in the first set, the images were presented in the middle of the screen (to test whether the child knew the correct response); in the second, they received 4 congruent trials; and in the third, they were given 2 congruent and 2 incongruent trials. This was followed by a block of 40 test trials consisting of Congruent and Incongruent items randomly interspersed.

**Flanker Fish Task**

The Flanker Fish Task is an adaptation of the Attentional Network Test (ANT) for young children (Rueda et al., 2004). Children were shown a display consisting of a row of five fish and were asked to press the arrow that matched the direction that the middle fish was facing (either left or right). On congruent trials, all the fish faced the same direction. On incongruent trials, the middle fish faced the opposite direction from the surrounding fish. Thus, children had to ignore the direction of the four other fish in order to correctly identify the direction of the middle fish. Children were given 2 sets of practice trials with feedback: the first consisted of 4 congruent trials, and the second consisted of 4 incongruent trials. After the practice trials, children were given two sets of test trials. The first set had a blocked design with Congruent and Incongruent
stimuli being presented in alternating blocks with 6 trials in each block (total 36 trials). In the second set, the Congruent and Incongruent stimuli were randomly interspersed (total 24 trials). The stimuli remained on the screen until the child responded.

*Verbal Stroop Task*

Children were shown two blocks of 16 stimuli (8 hearts and 8 flowers). In the first, congruent block, participants were asked to correctly identify each figure (as a heart or a flower). In the second, incongruent block, they were asked to say the opposite word for each picture: saying ‘flower’ when they saw a heart and ‘heart’ when they saw a flower. Their responses were recorded using a standing microphone and coded for response time and accuracy. The first utterance of the recording was used. The stimuli remained on the screen until a response was provided. Children didn’t receive any feedback on this task.

2.3 Results

2.3.1 Control Measures

Table 2.1 reports the mean age, non-verbal intelligence scores, and measures of socio-economic status of the three language groups. Socio-economic status (SES) measured the number of parents employed in the household and the parents’ highest level of education. Parental education level was measured on a scale of 1 – 7, with 1 indicating some high school with no degree and 7 indicating a doctoral degree and was averaged across both parents. Comparisons between monolingual and bilingual participants showed no significant differences in the number of parents employed in the household or the highest level of parental education. However, parental education was marginally higher in the Exposure group (p = 0.06). Early and Late bilinguals did not differ on any control measures.
2.3.2 Language Measures

As expected, Monolinguals outperformed Bilinguals on measures of language experience, showing higher standardized scores on the PPVT-4 ($t_{56.50} = 4.1$, $p < 0.001$) and the TROG-2 test of receptive grammar ($t_{48.29} = 2.6$, $p < 0.05$). Looking at Early and Late bilinguals separately, monolinguals have higher vocabulary than both Early and Late bilinguals ($F_{2,52} = 7.5$, $p < 0.01$). However, Monolinguals and Early bilinguals outperformed Late bilinguals on measures of receptive grammar ($F_{2,53} = 3.5$, $p < 0.05$).

Table 2.1. Participant Demographics

<table>
<thead>
<tr>
<th></th>
<th>Monolingual M (SD)</th>
<th>Bilingual M (SD)</th>
<th>Monolingual vs. Bilingual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>6.07 (0.6)</td>
<td>6.22 (0.6)</td>
<td>n.s.</td>
</tr>
<tr>
<td>Par. Ed †</td>
<td>4.95 (1.4)</td>
<td>4.28 (2.2)</td>
<td>n.s.</td>
</tr>
<tr>
<td>Par. Emp.</td>
<td>0.67 (0.2)</td>
<td>0.65 (0.3)</td>
<td>n.s.</td>
</tr>
<tr>
<td>Non-Verbal IQ (KBIT SS)</td>
<td>111.22 (18.0)</td>
<td>103.26 (13.3)</td>
<td>$\beta = -7.96 (4.3) \dagger$</td>
</tr>
<tr>
<td>Vocabulary (PPVT-4 SS)</td>
<td>125.09 (13.5)</td>
<td>110.61 (13.8)</td>
<td>$\beta = -14.49 (3.5) ***$</td>
</tr>
<tr>
<td>Rec. Grammar (TROG-2 SS)</td>
<td>113.16 (14.1)</td>
<td>101.57 (19.6)</td>
<td>$\beta = -11.59 (4.6) *$</td>
</tr>
</tbody>
</table>

$\dagger$ = $p < 0.1$

* = $p < 0.05$

** = $p < 0.01$

*** = $p < 0.001$

2.3.3 Executive Functioning Measures

Performance on EF tasks targeting inhibition and task-switching was analyzed using R (Venables & Smith, 2012) and lme4 (Bates et al., 2015). We performed a linear-mixed effects analysis for measures of reaction time and mixed-effects logistic regressions for measures of accuracy with Congruency and Language Group as fixed factors and Subject and Trial as random effects, with random intercepts for both factors and random slopes for Congruency by Subject. The results are presented in Table 2.2. The data show no evidence of a bilingual advantage in our sample. In all our EF measures, except for the Flanker Fish Task, subjects were slower and less accurate when facing incongruent trials, indicating sensitivity to EF demand. However, none of
the tasks showed better performance by bilinguals relative to monolinguals. In fact, the significant Congruency x Language interaction in the Simon’s Arrows Task Accuracy reflects better performance by monolinguals than bilinguals, showing a smaller difference in accuracy between congruent and incongruent trials. Early and Late Bilinguals did not show any differences in measures of EF.

Table 2.2: Results of the mixed-effects logistic regression on EF tasks

<table>
<thead>
<tr>
<th>Effect of Congruency</th>
<th>Effect of Language Group</th>
<th>Congruency x Language Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrows Accuracy</td>
<td>β = -1.52, SE = 0.22 ***</td>
<td>β = 0.09, SE = 0.32</td>
</tr>
<tr>
<td>Arrows RT</td>
<td>β = 99.49, SE = 13.73 ***</td>
<td>β = 10.74, SE = 50.43</td>
</tr>
<tr>
<td>Pictures Accuracy</td>
<td>β = -0.69, SE = 0.22 **</td>
<td>β = -0.16, SE = 0.26</td>
</tr>
<tr>
<td>Flanker Accuracy</td>
<td>β = -0.06, SE = 0.18</td>
<td>β = 0.34, SE = 0.28</td>
</tr>
<tr>
<td>Flanker RT</td>
<td>β = 41.15, SE = 16.19 *</td>
<td>β = -33.81, SE = 62.91</td>
</tr>
<tr>
<td>Stroop Accuracy</td>
<td>β = -0.68, SE = 0.17 ***</td>
<td>β = 0.04, SE = 0.24</td>
</tr>
<tr>
<td>Stroop RT</td>
<td>β = 0.36, SE = 0.04 ***</td>
<td>β = -0.04, SE = 0.03</td>
</tr>
</tbody>
</table>

* = p < 0.05  
** = p < 0.01  
*** = p < 0.001

2.3.4 Garden-Path Sentences

2.3.4.1 Gaze Data

*Referential ambiguity:*

In order to evaluate children’s ability to use referential context we compared their ability to identify the correct referent animal (e.g. frog on a napkin) in one referent and two referent contexts. In the 1 Referent Condition, only one animal is consistent with the object in the sentence and we thus expect children to have no difficulties identifying the proper target. However, in the 2 Referent Condition, both animals in the display are consistent with the object of the sentence. In this case, identification of the target animal would suggest a commitment to modifier interpretation of the prepositional phrase.
Two distinct regions were analyzed; (1) the Object region, from the onset of the object (*frog*) through the onset of the preposition (*on the*), offset by 200ms, and (2) the Preposition region, following onset of the preposition (200-1000ms). The dependent measure in these regions is a binary value representing whether subjects looked towards the Target Animal more than the Other Animal (Target/(Target+OtherAnimal) > 0.5).

During the Object region, a mixed-effects logistic regression\(^1\) showed a significant effect of Referential Condition (β = -0.43, SE = 0.14, p < 0.01). In the 1-Referent condition subjects were able to identify the proper referent and directed their gaze to the only animal on the screen consistent with the spoken token. In the 2-Referent condition, subjects looked at both animals. As expected, there was no effect of Ambiguity since the sentence thus far is equivalent between the two Ambiguity conditions.

Comparing monolingual and bilingual participants, there was a moderate Language Group x Ambiguity x Referential Condition (β = 1.01, SE = 0.56, p = 0.07). Gaze patterns did not differ in the 1 Referent Condition, where both groups could easily identify the target object. However, in the 2 Referent Condition, there was a marginal interaction of Language x Ambiguity (β = 0.66, SE = 0.38, p = 0.08); this was due to bilinguals trending towards better identification of the target referent in the Ambiguous 2 Referent Condition although this did not reach significance (β = 0.38, SE = 0.27, p = 0.15), see Figure 2.2, with no difference in looking pattern in the Unambiguous 2 Referent Condition (p = 0.45).

However, looking separately at Early and Late bilinguals, Figure 2.3, showed that the slight advantage in bilinguals in the Ambiguous 2 Referent Conditions was due to increased

\(^1\) The full model specification for this model is given in R code by glmer(LooksToTargetOverObject ~ (Language + Referent + Ambiguity) + (1|CurrentSentence) + (1|Subject), data=Data, family="binomial"). A model with random slopes did not converge.
looks to the Target Object by Early bilinguals, who significantly outperformed both Monolinguals and Late bilinguals, who showed similar fixation patterns.

Figure 2.2: Proportion of Looks to the Target Animal over the Target and Distractor Animal by Monolingual and Bilingual participants timed to the onset of the subject noun

During the region of the preposition, in the 1-Referent condition subjects continued looking at the uniquely identifiable referent. In the 2-Referent condition, subjects were able to
identify the proper referent only in the Unambiguous condition in which the 1st prepositional phrase was interpreted as a modifier. This resulted in a significant effect of Referential Condition \( (\beta = -1.04, \text{SE} = 0.26, p < 0.001) \), Ambiguity \( (\beta = -0.57, \text{SE} = 0.26, p < 0.05) \) and a Referential Context x Ambiguity interaction \( (\beta = 1.23, \text{SE} = 0.49, p < 0.05) \). There were no differences between Language Groups.

*Garden path revision:*

Figure 2.4 plots children’s gaze fixations to the Distractor Object timed to the onset of the modifier (*napkin in the ...*). During this region, looks to the Distractor Object represent a misanalysis of the ambiguous prepositional phrase as a destination and not a modifier. As the second prepositional phrase unfolds, continued looks to the empty napkin reflect a failure to revise this commitment given disambiguating information.

Two distinct regions were analyzed; (1) the Modifier region, from the onset of the object (*frog*) through the onset of the preposition (*on the*) offset by 200ms and (2) the Goal Region, following onset of the goal (200-1000ms). The dependent measure in these regions is a binary value representing whether subjects looked towards the Distractor Napkin. Any looks to the Distractor object were coded as 1, while no looks towards the napkin in the region of interest were coded as 0.
During the Modifier region, participants in the ambiguous condition made more erroneous looks to the empty napkin in Ambiguous relative to Unambiguous conditions, consistent with a garden-path error ($\beta = 0.43$, SE $= 0.19$, p $< 0.05$). This effect was not modulated by Referential Context ($\beta = 0.25$, SE $= 0.37$, p $= 0.48$) and did not differ between Monolinguals and Bilinguals ($\beta = 0.16$, SE $= 0.29$, p $= 0.58$).

During the Goal region, participants continued to look towards the empty napkin in the Ambiguous Condition relative to the Unambiguous Condition ($\beta = 0.33$, SE $= 0.15$, p $< 0.05$), consistent with a lingering misanalysis past access to disambiguating information. In addition, there was a moderate Language Group x Ambiguity x Referential Condition interaction ($\beta = -0.95$, SE $= 0.58$, p $= 0.1$). To investigate this further, we looked at the critical Ambiguous Condition. Bilinguals showed marginally fewer lingering looks to Distractor Object in the Ambiguous 2 Referent Condition ($\beta = -0.59$, SE $= 0.32$, p $= 0.07$).
Looking separately at Early and Late bilinguals, this effect again seemed primarily due to Early bilinguals. Early bilinguals showed numerically fewer looks to the Distractor napkin in the Ambiguous 2 Referent Condition than either the Monolinguals or the Late Bilinguals.

2.3.4.2 Behavior

Figure 2.5 shows the proportion of participants who correctly carried out the instructions by moving the Target object directly to the Goal. Subjects were significantly better in the Unambiguous Condition than the Ambiguous Condition ($\beta = -3.06$, SE = 0.23, $p < 0.001$), with no effect of Referential Context, nor Referential Context x Ambiguous Condition interaction. The behavioral pattern did not differ by Language group with Monolinguals and Bilinguals showing similar patterns and with no further differences between Early and Late bilinguals.

In addition, behavioral patterns were analyzed regarding which animal, Target or Other Animal, was used to carry out the actions. There was a significant effect of Ambiguity Condition ($\beta = 1.64$, SE = 0.27, $p < 0.001$), Referential Context ($\beta = -1.45$, SE = 0.27, $p < 0.001$), and their interaction ($\beta = 1.66$, SE = 0.44, $p < 0.001$). As only one animal is a possible referent in the 1 referent condition, subjects were able to identify the Target regardless of whether the sentence was unambiguous or ambiguous. However, in the 2 Referent Ambiguous Condition subjects were significantly worse at using the proper animal to carry out the actions.
2.3.5 Exploratory Correlations

In addition to preplanned group comparisons, the current study offers an exciting opportunity to look at sentence comprehension through individual differences in language measures and EF across a large and diverse sample of child participants. For the following analyses, we included data from all available subjects regardless of language history to both improve statistical power and to increase the variance of values in our predictors. Thus, we included data from 16 participants who were exposed to a second language but were not included in the bilingual group. A description of this sample is included in Table 2.3 and their scores on EF tasks are presented in Table 2.4. EF scores in the exposure group did not differ from either the monolingual or the bilingual participants on any measure. Looking at overall garden-path task accuracy, this group did significantly better than both monolinguals (β =-1.31, SE = 0.70, p= 0.06) and bilinguals (β =-1.43, SE = 0.71, p < 0.05), but like the other language groups, did significantly worse in Ambiguous conditions (β =-2.33, SE = 0.45, p < 0.001) and showed no effect of Referential Context.

Figure 2.5: Act-out behavioral accuracy across conditions by language group
Table 2.3: Control Measures for the Exposure Group and comparisons to Monolingual and Bilingual participants

<table>
<thead>
<tr>
<th>Measure</th>
<th>Exposure Mean (SD)</th>
<th>Exposure vs. Monolingual</th>
<th>Exposure vs. Bilingual</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td>6.24 (0.41)</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td><strong>Parental Ed</strong></td>
<td>5.5 (1.41)</td>
<td>n.s.</td>
<td>β =-1.22, SE = 0.6, p &lt; 0.05</td>
</tr>
<tr>
<td><strong>Parental Emp</strong></td>
<td>0.85 (0.78)</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td><strong>Matrix K-BIT</strong></td>
<td>116.63 (17.18)</td>
<td>n.s.</td>
<td>β =-13.36, SE = 5.1, p &lt; 0.05</td>
</tr>
<tr>
<td><strong>PPVT-4</strong></td>
<td>117.81 (12.96)</td>
<td>β =7.28, SE = 4.13, p = 0.08</td>
<td>β =-7.21, SE = 4.2, p = 0.09</td>
</tr>
<tr>
<td><strong>TROG-2</strong></td>
<td>106.44 (20.44)</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

The exposure group included participants with higher SES than the bilingual group, as measured by parental education (β =-1.22, SE = 0.6, p < 0.05) and higher non-verbal IQ (β =-13.36, SE = 5.1, p < 0.05). This group had grammar and vocabulary scores that were between those found in the bilingual group and the monolingual group and were not statistically different from either. For subsequent group analyses we will compare the monolingual group to a group that has any reportable access to a second language (collapsed across the bilingual and the exposure group).

In the following analyses of individual differences, we evaluated the role of English language exposure and EF as predictors of accuracy in the garden path task. We included control
factors of age, parental education, and non-verbal IQ. Because parental education had an underlying bimodal distribution, we converted it into a binary measure splitting those children that had parents with education levels up through completion of college and those with higher degrees. Language experience was evaluated using both vocabulary (PPVT-4) standard scores and receptive grammar (TROG-2) standard scores. In addition, two measures of EF were selected. Since the Arrows Task was performed first and showed large effects of EF demand, we included reaction time difference score (incongruent-congruent) as a measure of inhibitory control. We also included a verbal inhibitory control score as measured by the accuracy difference score (congruent-incongruent) on the verbal Stroop task. With both EF measures, a larger difference score was reflective of greater cost associated with increased need of EF resources for incongruent trials. Age correlated with both differences in Arrows Reaction Time (r= -0.42, p < 0.001) and Stroop Accuracy (r= -0.32, p < 0.01), suggesting that these measures showed meaningful improvement across this age span. A full correlation table is available in Table 2.5. In the current sample, neither language measure correlated with either measure of executive control.

Table 2.5: Correlation Table of Factors of Interest

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Parental Ed.</td>
<td>-0.27*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Verb. IQ S.S.</td>
<td>0.17</td>
<td>0.21*</td>
<td>0.27*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary S.S.</td>
<td>0.05</td>
<td>0.29*</td>
<td>0.27*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grammar S.S.</td>
<td>-0.01</td>
<td>0.38***</td>
<td>0.38***</td>
<td>0.74***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arrows RT Cost</td>
<td>-0.4***</td>
<td>0.15*</td>
<td>0.01</td>
<td>0.04</td>
<td>0.21*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroop Acc. Cost</td>
<td>-0.32**</td>
<td>0.1*</td>
<td>-0.08</td>
<td>0.24*</td>
<td>0.1*</td>
<td>0.25*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amb. Accuracy</td>
<td>0.35**</td>
<td>0.05</td>
<td>0.19*</td>
<td>0.06</td>
<td>0.2*</td>
<td>0.01</td>
<td>-0.12*</td>
<td></td>
</tr>
<tr>
<td>Unamb. Accuracy</td>
<td>0.42***</td>
<td>-0.13*</td>
<td>-0.12*</td>
<td>0.05</td>
<td>0.07</td>
<td>-0.09*</td>
<td>-0.17*</td>
<td>0.52***</td>
</tr>
</tbody>
</table>

S.S. = Standard Score
† = p < 0.1
* = p < 0.05
** = p < 0.01
*** = p < 0.001
2.3.5.1 Description of factors of interest

We further evaluated the key factors of language experience and executive control to see how they interrelate. Using model comparison, we ran linear regression analyses to determine which factors predict vocabulary, receptive grammar, Stroop accuracy cost, Arrows reaction time cost, and whether the predictors differ by language group.

For each model we first used forward regression to sequentially add the predictors. We first added the control variables of age, parental education, and non-verbal IQ. For measures of language experience, we then added the two measures of executive control, Arrows reaction time cost, and Stroop accuracy cost. For the two measures of EF, we added the measures of vocabulary and receptive grammar. All factors were retained in the final model to evaluate the independent contribution of each, controlling for the others. Using backwards regression, we then determined the optimal model that describes each factor of interest. Finally, we analyzed whether there was an effect of language group or language group interactions with each predictor. The full models outlining our key factors of interest are included in Table A3 in the Appendix.

**Vocabulary**

In the current sample, adding parental education (F(1) = 3.11, p = 0.08), non-verbal IQ (F(1) = 4.30, p < 0.05), and Stroop Accuracy Cost (F(1) = 5.53, p < 0.05) improved the model fit predicting vocabulary. Backwards regression resulted in a final model in which higher vocabulary scores relative to peers were predicted by better non-verbal IQ (β =0.25, SE = 0.10, p < 0.05). Unlike in earlier work with monolingual participants, in the current sample better EF skills, as indicated by lower costs on Stroop Accuracy, were related to lower vocabulary scores relative to peers (β =40.90, SE = 16.73, p < 0.05). There was a robust main effect of Language
Group ($\beta = -10.57$, $SE = 3.09$, $p < 0.01$), which significantly improved model fit ($F(1) = 11.713$, $p < 0.01$). As expected, monolinguals showing higher vocabulary scores than children exposed to two languages. However, there was no significant interaction between language group and either predictor. If receptive grammar is added to the model ($F(1) = 61.93$, $p < 0.001$), there remains a vocabulary difference between the two language groups but the effect of non-verbal IQ disappears and Stroop Accuracy Cost effects are decreased ($\beta = 22.16$, $SE = 11.66$, $p = 0.06$).

**Receptive Grammar**

For receptive grammar, parental education ($F(1) = 10.20$, $p < 0.01$) and non-verbal IQ ($F(1) = 9.47$, $p < 0.01$) improved model fit. The optimal model, determined by backwards regression, confirmed that better receptive grammar, relative to age-matched peers, was associated with higher parental education ($\beta = 10.26$, $SE = 3.94$, $p < 0.05$) and higher non-verbal IQ ($\beta = 0.35$, $SE = 0.11$, $p < 0.01$), and was not significantly predicted by either measure of EF. Adding language group ($\beta = -7.47$, $SE = 3.76$, $p = 0.05$) into the model marginally increased model fit ($F(1) = 3.93$, $p = 0.05$), there were no significant language group interactions with either predictor, suggesting similar predictive patterns for receptive grammar across children that are monolingual and children that are exposed to a second language. Finally, although the effect of language group disappears, both predictors of receptive grammar remain significant even when vocabulary scores are included in the optimal model ($F(1) = 58.96$, $p < 0.001$).

**Arrows Reaction Time Cost**

A regression model predicting reaction time differences on the Arrows task was improved by the inclusion of Age ($F(1) = 14.612$, $p < 0.01$) and receptive grammar ($F(1) = 5.01$, $p < 0.05$). Backwards regression confirmed an optimal model in which reaction time costs were predicted by Age ($\beta = -129.64$, $SE = 32.52$, $p < 0.001$) and, less robustly, receptive grammar ($\beta =$-
1.80, SE = 0.92, p = 0.05). While cognitive control improved (decrease in reaction time costs) with age, improvements in cognitive control were predicted by a decrease in receptive grammar scores relative to peers. There was no main effect of language group, confirming that we see no EF advantage in our sample of bilingual participants, even when the sample includes a broader range of second language exposure. When language group is included in the model, the effect of receptive grammar was no longer significant (β = 1.65, SE = 0.95, p = 0.08). The predictors of Arrows RT Cost did not differ by language group. Adding the other measure of cognitive control, Stroop Task Accuracy Cost, to the model did not improve the model fit, however, including this factor did not change the overall pattern of the other predictors.

**Stroop Accuracy Cost**

Finally, using forwards regression, a model predicting accuracy costs on the verbal Stroop task was improved by the addition of Age (β = -0.06, SE = 0.02, p < 0.01) and vocabulary (F(1) = 5.49, p < 0.05). These two predictors remained significant in the model obtained through backwards regression with EF (decrease in cost) improving with age (β = -0.05, SE = 0.02, p < 0.01), but showing the opposite relationship with gains in vocabulary (β = 0.002, SE = 0.00, p < 0.05). Differences in accuracy in the Stroop task did not differ by Language group and there were no Language group interactions. Adding Arrows reaction time did not improve model fit nor change the overall pattern of results.

### 2.3.5.2 Predicting Accuracy

To evaluate which factors predict accuracy in garden-path comprehension, we used generalized linear mixed-effects regression (glmer). We first conducted forwards model comparison using generalized linear mixed effects models, with random intercepts for both Subject and Trial. All factors were retained in the model in order to assess the independent
contribution of each factor. Factors were added using a theory driven approach. Our base model contained only the effect of ambiguity condition. Then, we included our control factors in sequential order; Age, Parental Education, and Non-Verbal IQ. Subsequently, we included measures of language experience, first vocabulary and then receptive grammar, followed by measures of EF, first Arrows reaction time costs and then Stroop accuracy costs. To avoid overfitting, we then conducted backward regression to determine an optimal model. Using the optimal model, we evaluated whether these effects interacted with ambiguity condition to determine if any factors specifically predicted garden-path revision above and beyond overall task accuracy. Finally, we evaluated the role of language group. In cases where a predictor interacted with language group, we ran separate analyses on monolingual and children with any second language exposure.

Table 2.6 shows the results of the full model with all predictor values included simultaneously. Only the addition of age ($\chi^2(1) = 13.82, p < 0.001$) and receptive grammar ($\chi^2(1) = 7.58, p < 0.01$) made a significant improvement in model fit beyond the effects of Ambiguity Condition. Backwards regression showed an optimal model in which overall accuracy improved in the Unambiguous relative to the Ambiguous condition ($\beta = -3.12, SE = 0.22, p < 0.001$), with increasing age ($\beta = 0.88, SE = 0.26, p < 0.001$) and standardized receptive grammar ($\beta = 0.57, SE = 0.24, p < 0.05$). Adding language as a factor did not significantly improve model fit, however, the effect of receptive grammar on accuracy was marginally modulated by language group ($\beta = -0.93, p < 0.56, p = 0.09$), which marginally improved model fit ($\chi^2(1) = 2.7, p = 0.1$). In addition, there was a marginal three-way interaction between the effect of receptive grammar, language group, and condition ($\beta = -0.80, SE = 0.49, p = 0.1$). Thus, we conducted separate analyses on the Unambiguous and Ambiguous conditions.
<table>
<thead>
<tr>
<th>Factor</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>2.27</td>
<td>0.47</td>
<td>4.80</td>
<td>n.s.</td>
</tr>
<tr>
<td>Amb. Condition ***</td>
<td>-3.07</td>
<td>0.23</td>
<td>-13.66</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Age ***</td>
<td>0.91</td>
<td>0.30</td>
<td>3.00</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td>Parental Ed.</td>
<td>-0.03</td>
<td>0.55</td>
<td>-0.06</td>
<td>n.s.</td>
</tr>
<tr>
<td>Non-Verbal IQ (KBIT SS)</td>
<td>-0.16</td>
<td>0.28</td>
<td>-0.57</td>
<td>n.s.</td>
</tr>
<tr>
<td>Vocabulary (PPVT-4 SS)</td>
<td>-0.47</td>
<td>0.39</td>
<td>-1.19</td>
<td>n.s.</td>
</tr>
<tr>
<td>Grammar (TROG-2 SS) *</td>
<td>0.98</td>
<td>0.42</td>
<td>2.31</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>RT Cost Arrows</td>
<td>0.32</td>
<td>0.36</td>
<td>0.88</td>
<td>n.s.</td>
</tr>
<tr>
<td>Acc Cost Stroop</td>
<td>-0.15</td>
<td>0.28</td>
<td>-0.54</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

In the Unambiguous condition, only age ($\chi^2(1) = 13.61, p < 0.001$) and receptive grammar ($\chi^2(1) = 4.96, p < 0.05$) improved model fit and showed significant contribution when all other factors were controlled for (Age: $\beta = 1.06, SE = 0.35, p = 0.01$; Receptive Grammar: $\beta = 1.03, SE = 0.46, p < 0.05$). However, the optimal model showed that performance on Unambiguous sentences improved with age ($\beta = 1.06, SE = 0.30, p < 0.001$), and receptive grammar ($\beta = 0.79, SE = 0.29, p < 0.01$), while performance marginally decreased with better non-verbal IQ ($\beta = -0.51, SE = 0.30, p = 0.09$). Neither the main effect of performance on the unambiguous trials nor the effect of age differed by language group.

Looking at Ambiguous trials, age ($\chi^2(1) = 9.97, p < 0.01$) and receptive grammar ($\chi^2(1) = 6.54, p < 0.05$) improved model fit and showed significant contribution when all other factors were controlled for (Age: $\beta = 0.99, SE = 0.38, p = 0.01$; Receptive Grammar: $\beta = 1.12, SE = 0.52, p < 0.05$). The optimal predictive model also confirmed predictive effects of age ($\beta = 0.91, SE = 0.32, p < 0.01$) and receptive grammar ($\beta = 0.74, SE = 0.31, p < 0.05$). Performance did not differ by language group however there was a significant interaction between the role of receptive grammar and language group ($\chi^2(1) = 4.82, p < 0.05$). Receptive grammar significantly predicted Ambiguous accuracy only in the monolinguals ($\beta = 2.26, SE = 0.82, p < 0.01$) but not bilinguals.
Finally, to get a better sense of which factors determine Accuracy in the Ambiguous conditions in both groups, we reran the full analyses separately on monolingual children and children exposed to a second language. In monolingual subjects, a model of Ambiguous Accuracy was improved by the addition of Age ($\chi^2(1) = 3.08, p = 0.08$) and receptive grammar ($\chi^2(1) = 11.83, p < 0.001$) and these were the only significant predictors in the full model. The optimal model obtained through backwards regression confirmed the effects of age ($\beta = 1.07, SE = 0.59, p = 0.07$) and receptive grammar ($\beta = 3.86, SE = 1.20, p < 0.01$). However, backwards regression also showed that better accuracy on ambiguous trials was marginally related to lower vocabulary scores relative to peers ($\beta = -1.38, SE = 0.83, p = 0.1$) and lower non-verbal IQ ($\beta = -0.94, SE = 0.55, p = 0.09$), although neither of these effects were robust. The pattern suggests that those children with specific skills related to grammar over other measures of language experience performed the best on Ambiguous trials. The effect of receptive grammar on Ambiguous accuracy remained even if Unambiguous accuracy was included in the model ($\beta = 3.11, SE = 1.70, p < 0.05$), although the effect of age was no longer significant.

For bilinguals, only age ($\chi^2(1) = 6.82, p < 0.01$) and non-verbal IQ improved model fit ($\chi^2(1) = 3.02, p = 0.08$) and were marginal predictors of Accuracy in the optimal model. Only non-verbal IQ remained a significant predictor when Unambiguous accuracy was included in the model. This suggests that while receptive grammar was not predictive for bilinguals, better IQ may help performance when grammar is insufficient.

To summarize, improvements in receptive grammar play a critical role in garden path performance and ambiguity resolution more particularly. However, this is primarily found in monolingual rather than bilingual participants. Unlike monolingual participants, bilingual participants rely more on IQ than grammar when showing better performance on ambiguous
sentence comprehension. There were no significant effects of EF on any measures of task accuracy, regardless of language history.

2.4 Discussion

Across the early school-age years children show increasingly adult-like sentence comprehension. Specifically, children across this age group improve in their ability to consider top-down cues, such as referential context, and become better able to revise their interpretive commitments when faced with inconsistent information (Kidd & Bavin, 2005a; Trueswell et al., 1999; Weighall, 2008). But what causes this developmental change in comprehension? Here we evaluated two hypotheses. Under the EF hypothesis, improvements in sentence comprehension reflect improvements in domain-general cognitive control. Alternatively, under the language experience hypothesis these changes reflect domain-specific improvements in processing efficiency due to continuing language experience. By looking at comprehension of prepositional phrase ambiguities in a sample of children from diverse language backgrounds, we were able to better isolate the role that language experience and EF play in comprehension.

Both monolingual and bilingual children struggle when faced with an ambiguous garden-path sentence like “put the frog on the napkin in the box”. In fact, consistent with prior work, children correctly interpret “on the napkin” as a modifier only one-third of the time (Kidd & Bavin, 2005a; Trueswell et al., 1999; Weighall, 2008). What’s more, they are unable to use contextual cues to resolve the syntactic ambiguity and revise their garden-path errors (Kidd & Bavin, 2005a; Trueswell et al., 1999; Weighall, 2008). They have similar error rates both when one frog is present, a context that encourages garden-path errors, and when two frogs are on display, a context that, for adults, encourages a modifier interpretation of the ambiguous prepositional phrase (Trueswell et al., 1999).
Although both monolinguals and bilinguals show similar error rates in their final sentence interpretations, online processing measures show subtle differences between the two groups. Children in both language groups are sensitive to verb bias and commit to the prediction that the prepositional phrase following the verb (put) should be a destination. Consistent with this prediction, both groups make looks toward the empty napkin as a potential destination. However, while monolinguals continue looking at the incorrect destination even after an alternative goal is uttered “on the box”, bilingual children show fewer prolonged looks to the empty napkin specifically in the two-referent condition.

In addition, bilingual participants, and specifically early bilinguals, show a preference for the referent that can be modified early in the sentence. Upon hearing the object (frog) bilingual participants in the two-referent condition tend to look towards the frog that can be modified. However, this preference occurs only in the ambiguous condition even though subjects have not yet heard whether a sentence will be ambiguous or unambiguous. There are two possibilities regarding this early effect. First, the sentences the children hear are not spliced, so there may be prosodic differences that can cue a listener into the upcoming syntactic structure. If this is the case, bilinguals show better sensitivity than monolingual to this subtle prosodic information. Alternatively, in the current design, all critical sentences result in a modifier interpretation and occur at least once per trial. It is possible that bilinguals, but not monolinguals, are able to notice and use this consistency to anticipate the syntactic structure.

In summary, gaze patterns of monolinguals are largely consistent with the 5-year-old subjects tested in a similar paradigm by Trueswell et al. (1999). However, the bilingual participants in the current study pattern showed more adult-like gaze patterns with an emerging sensitivity to referential context. Despite improved attention to referential context online,
bilingual children did not show a similar advantage in their offline responses. This is similar to a number of findings showing emerging online sensitivity prior to improvements in behavior (E.-K. Lee & Snedeker, 2016). In the subsequent sections we will discuss how this pattern of findings informs the relative roles that language experience and executive functioning have on children’s improving ability to resolve ambiguous sentences and revise garden path misanalyses.

2.4.1 The role of language experience

According to the language experience hypothesis, increasing language processing efficiency allows subjects to better represent and/or keep online alternative linguistic structures. Consequently, more informational sources are available to inform comprehension as it unfolds. This hypothesis would predict that bilinguals, by virtue of having less experience with English, should show poorer performance on measures of ambiguity resolution (Marchman, Fernald, & Hurtado, 2010). This is not what we find. In fact, the data show the opposite pattern, with gaze results indicating slightly better use of top-down information in the bilingual group despite poorer scores on measures of vocabulary and grammar.

However, in favor of the language hypothesis, analyses of individual differences suggest that receptive grammar is a significant predictor of accuracy in the garden path task and, specifically, in the ability to resolve ambiguous structures. Although overall task performance increased with age, children whose receptive grammar was significantly better than that of their peers showed the best performance on ambiguous trials even when controlling for parental education, non-verbal IQ, and EF ability. However, this finding was primarily driven by our monolingual participants. As we will discuss below, bilinguals may have alternative strategies or capabilities to bear on resolving the ambiguous input.

2.4.2 No Support for the Executive Functioning Hypothesis
Neither our analyses of individual differences nor our group comparisons provided support for the role of EF on garden-path resolution. Accuracy was predicted by receptive grammar and, in bilingual participants, non-verbal IQ. However, neither EF measure tested predicted performance in either group in any measure, despite capturing meaningful improvement in EF across this age group. This is at odds with a recent paper by Woodard et al. (2016) which used similar measures but found that cognitive flexibility, as measured by switch costs on a Flanker task significantly predicted children’s behavior on a similar measure of prepositional attachment ambiguity. To more closely model their analyses, we reran our analyses using a similar measure of cognitive flexibility (comparing trials that switch conditions to those in which the conditions don’t switch, regardless of congruency). However, this analysis did not predict children’s performance with receptive grammar remaining a key predictor. Numerous differences in our experimental designs may account for the discrepant findings. For example, Woodard et al. (2016) ask children to physically act out the relevant instructions while our paradigm is computer-based. EF control may play a larger role when children need to plan out physical movements rather than dragging images across a screen.

Bilingual participants’ online sensitivity to referential context is unlikely to reflect better EF. Contributing to the increasing evidence questioning a bilingual advantage in executive functioning (e.g. de Bruin, Treccani, & Della Sala, 2015; J. B. Morton & Harper, 2007; Paap & Greenberg, 2013), neither our 2nd language exposed group, nor our Early or Late bilingual participants show an executive functioning advantage over their monolingual counterparts. It is possible that bilinguals do indeed have an EF advantage that our tests were not sensitive enough to pick up. However, this is unlikely given that our EF measures showed robust sensitivity to inhibitory demand in all language groups and were able to detect differences in groups split by
age, rather than language history. Alternatively, we may be measuring the wrong EF subskill. While many of our tests focused on inhibitory control, more recent work has argued that cognitive flexibility, rather than inhibitory control, might be advantaged in bilinguals (Prior & MacWhinney, 2010; Wiseheart, Viswanathan, & Bialystok, 2016). However, reanalyzing our tasks to target cognitive flexibility also did not reveal any differences across our language groups.

2.4.3 Bilingual Advantage in Online Processing?

The fact that bilingual children show similar performance on garden-path sentence comprehension and better use of referential context, despite having less experience with English and marginally lower non-verbal intelligence, suggests that bilinguals may be using an alternative strategy. Previous literature has shown that bilinguals, due to the experience of having to monitor two languages and the environments in which they occur show better pragmatic sophistication as well as metalinguistic awareness (Galambos & Goldin-Meadow, 1990; Siegal, Iozzi, & Surian, 2009; Siegal et al., 2010). Better ability at either of these skills would predict a specific advantage at picking out relevant cues in the context to help them navigate sentence comprehension.

The use of referential context requires children to track reference, understand how it may affect speaker intention, and how linguistic structure may reflect that intention. In ambiguous sentences like “put the frog on the napkin in the box”, a child must be able to work through a complicated set of inferences; the speaker wants a specific frog, as indicated by the definite determiner, but there are two frogs in the display, therefore, she must know that I don't know which one she means and will provide more information to distinguish between them. In addition, this pragmatic inference must be so routine that it kicks in before a strong commitment
is made based on the bias of the verb. Better ability to make the necessary pragmatic leaps would improve performance in the current task. While we do not have the ability to directly evaluate pragmatic ability in our participants, the current findings do not provide strong evidence for this possibility. Although bilingual participants show looks consistent with better consideration of referential context this doesn’t seem to affect their behavior; they neither select the correct frog nor carry out the correct action more often than monolinguals.

Alternatively, consideration of the referential context may reflect better metalinguistic awareness. In the current task, greater attention to the language manipulations across trials may cue bilingual participants into features of the display which may inform comprehension. Such considerations could result in more adult-like looking patterns and may even explain early consideration of a frog with an identifying modifier. However, knowing which aspects of the display are important is not helpful unless you know how these inform syntactic analyses. Bilingual participants may show better consideration of the visual context, but with less language experience, may lack the language experience to determine how the relevant cues should be used. While this hypothesis would be consistent with the current findings, the data we have provided don’t offer a way to directly test this proposal.

2.4.4 Conclusion

Our findings provide evidence in support of the hypothesis that increased language experience may benefit online sentence processing. As in previous studies (S. Anderson et al., 2011; Huang & Hollister, 2019; Nation et al., 2003), this support comes primarily from analyses of individual differences and thus warrant further investigation. Contrary to prior work, we do not find evidence that improvement of executive functioning is playing a critical role in helping children use referential context or revise garden-path misanalyses. However, as our groups did
not differ in EF ability, we were unable to directly evaluate this hypothesis. In addition, both monolingual and bilingual children failed to revise their garden-path errors, showing incorrect actions on the majority of trials. It is thus possible that EF is related to garden path revision but that our participants didn’t display sufficient ability. However, bilinguals did show better use of contextual information despite similar EF ability and less English experience to their monolingual counterparts, suggesting that a third alternative may be responsible for the ability to integrate contextual cues.

Future studies could more directly test the role of pragmatics and metalinguistic awareness by providing monolingual and bilingual groups measures targeting these skills. Confirmation of this relationship would suggest that differences across development may reflect neither better bottom-up representation due to increases in language experience, nor better control of information flow due to increases in EF. Rather, this hypothesis would suggest that development of the ability to flexibly use top-down information may be related to better representation at higher levels of processing. Advances in language comprehension may thus reflect better understanding of the role of language use in context. Further research should be done to evaluate pragmatic ability between monolingual and bilingual children and its role in sentence comprehension.
Chapter 3:

When hearing negation, do listeners expect the unexpected? An ERP investigation into online comprehension of negation in response to a yes/no question.
3.1 Introduction

Without the word “no”, it would be incredibly challenging to communicate many of the thoughts we express every day: “I don’t want to go to work today”, “I don’t eat meat”, “I don’t understand this word”, etc. The ability to negate an assertion is critical to communication. Linguistic devices that accomplish this task exist in all the world languages, are highly frequent in conversation (Brysbaert & New, 2009), and are among the first words that children acquire (Dale & Fenson, 1996). In fact, children have been shown to understand the logical function of negation (truth-functional negation) by as early as 27-months (Austin, Theakston, Lieven, & Tomasello, 2014; Feiman, Mody, Sanborn, & Carey, 2017). Although negation is ubiquitous in everyday conversation, numerous studies have shown that negation poses a challenge for comprehension even in adults (Carpenter & Just, 1975; Clark & Chase, 1972; Cornish & Wason, 1970; Just & Carpenter, 1975; Lüdtke & Kaup, 2006). For example, looking at a set of red dots people quickly agree that “the dots are red” but take longer to determine that “the dots aren’t black”, although both statements are true (Just & Carpenter, 1971). Recent work has suggested that the apparent discrepancy between our everyday experience with negation and the processing costs it elicits across experimental paradigms may be due to differences in the context in which negation occurs (Nieuwland & Kuperberg, 2008; Nordmeyer & Frank, 2014, 2015; Tian, Breheny, & Ferguson, 2010). However, which elements of context are necessary to facilitate processing of negation is still poorly understood. In the current study we evaluate whether a preceding polar (yes or no) question might allow listeners to integrate negation into their online sentence interpretation.

3.1.1 Negation can pose a challenge for comprehension
Early work in comprehension of truth-functional negation consistently showed that across a number of measures calculating negation incurred a processing cost. For example, negated sentences are read more slowly (Lüdtke & Kaup, 2006), delay truth-value judgments (Carpenter & Just, 1975; Just & Carpenter, 1975), and are harder to recall (Carpenter & Just, 1976; Cornish & Wason, 1970). This set of findings led to the conclusion that comprehension of negation required two steps; first, the construction of the affirmative proposition, and second, the application of the logical operator (Carpenter & Just, 1975; Clark & Chase, 1972; Kaup, Lüdtke, & Zwaan, 2006; Kaup, Yaxley, Madden, Zwaan, & Lüdtke, 2007). Both offline and online measures of sentence comprehension have provided evidence for the intermediate step in which an affirmative interpretation is constructed. For example, during early stages of offline sentence comprehension, prior to ~1000ms after sentence offset, both affirmative and negated sentences facilitate processing of probes targeting affirmative representations of the uttered proposition. However, given a longer delay, this pattern switches to reflect the truth-value of the given utterance (Hasson & Glucksberg, 2006; Kaup & Zwaan, 2003). In addition, online measures of sentence comprehension using event-related-potentials (ERPs) have shown processing of negated sentences through the affirmative (Fischler, Bloom, Childers, Roucos, & Perry Jr, 1983; Kounios & Holcomb, 1992). When adults hear a negated but ultimately true sentence like “a robin is not a tree” they show a large N400 response, consistent with difficulty accessing and integrating the final word. On the other hand, when hearing a false sentence like “a robin is not a bird”, processing of the final word showed no such difficulty (Fischler et al., 1983). This pattern of findings, in which the N400 response reflects semantic associations across content words and not the truth-value, suggests that the logical operator is not used online to update lexical predictions, but rather, is applied as a secondary step.
3.1.2 The role of pragmatics in comprehension of negation

Recent research has suggested that the challenge posed by negation in prior work may not reflect an inherent difficulty with negation, but may instead reflect the context in which it has been studied (Nieuwland & Kuperberg, 2008; Nordmeyer & Frank, 2015; Tian et al., 2010). Studies that support a two-step process of negation have primarily come from paradigms in which listeners hear a series of disconnected sentences. However, hearing negation out of the blue is typically under-informative and, according to Gricean maxims, pragmatically infelicitous (Grice, 1975). Although "a robin is not a tree" is true, hearing this tells you almost nothing about robins. Evidence for the role of pragmatics in comprehension of negation comes from studies showing that the processing costs associated with negation decrease or disappear when the negated sentence appears in a pragmatically supportive context (Lüdtke & Kaup, 2006; Nordmeyer & Frank, 2015; Tian et al., 2010).

Studies of online comprehension in adults and children have provided evidence that given a supportive context, negation can be used incrementally to update predictions of upcoming referents and lexical items. Nieuwland and Kuperberg (2008) evaluated the role of context on the ability of adults to integrate negation into their online lexical predictions, as measured by the N400 response. In this study, adults are given sentences in which negation is either pragmatically licensed, as in [1a], or unlicensed, as in [1b].

[1a] *With proper equipment, scuba diving isn’t very safe/dangerous…*

[1b] *Bullet-proof vests aren’t very safe/dangerous…*

In line with earlier findings, in the unlicensed context, listeners show large N400 responses to true negated sentences. However, in the pragmatically licensed condition, N400 responses correspond to the truth-value of the sentence, with larger N400s to false sentences in both the
affirmative and negated condition. Adults are able to integrate the logical operator into their lexical predictions as a sentence unfolds.

Pragmatic context seems to not only improve adults' ability to comprehend negated utterances but also enables children to do the same. Eye-tracking studies using the visual world paradigm have suggested that children, like adults, process negation by first interpreting the affirmative (Nordmeyer & Frank, 2014). For example, in an eye-tracking study conducted by Nordmeyer and Frank (2014), four-year-olds looking at a display showing a boy with and without an apple heard either an affirmative [2a] or negated sentence [2b]. Regardless of sentence polarity, they first directed their gaze to the boy with the apple.

[2a] Look at the boy with apples.

[2b] Look at the boy with no apples.

However, children even as young as two can show online use of negation when adequate pragmatic context is provided. For example, in Reuter et al. (2018), two and three-year-old children were presented with a narrative in which a character breaks a bowl and then almost breaks a plate, identifying a clear question-under-discussion (QUD) (what DW did and didn’t break) and highlighting the likely to-be-negated preposition (a plate that almost broke, but didn’t). Within this discourse, three-year-olds hearing a sentence like [3a] correctly looked at the unbroken plate following the verb and did so as quickly and accurately as when they heard the affirmative alternative [3b]. Thus, when both negated and affirmative statements are equally

[3a] DW didn’t break one of the plates.

[3b] DW broke one of the bowls.

pragmatically licensed, children can correctly interpret polarity as they are hearing it (Reuter et al., 2018).
3.1.3 Current study

Prior work has shown, across a number of different paradigms, that a pragmatically supportive context can ease processing of negation. Previously, a pragmatically felicitous context has been established in a number of different ways. For example, online processing of negation is shown when a discourse sets up an event that almost occurred (Reuter et al., 2018), a to-be negated proposition is highlighted as an exception to a rule (De Villiers & Flusberg, 1975), or when informativity of the utterance is maximized (Nordmeyer & Frank, 2015). However, we do not yet know what feature of context is necessary to facilitate processing of negation.

In ordinary discourse we expect each statement to be a response to some implicit question (QUD) that structures the conversation. For example, in a narrative the implicit question is “and then what happened?” In a description the implicit question is “what was the place or person like?” Much of the time these questions are best answered by saying what is the case, rather than what is not (Roberts, 1996). There are, however, times when the implicit question under discussion justifies a negative answer. Perhaps the simplest of these situations is when there is a simple yes/no (or polar question) at the center of the conversation. If Alicia has just got off the phone with potential employer and she knows that her sister is eager to find out whether she has been hired, then it is reasonable for Alicia to tell her “I didn’t get the job.” A polar question introduces both the positive assertion and the negated assertion as alternative possibilities making them both equally felicitous (Groenendijk & Stokhof, 1984; Hamblin, 1973).

This theory has been argued (Gualmini, 2004) to explain why children are better at understanding negative utterances when the possibility of their positive counterpart has been highlighted (De Villiers & Flusberg, 1975). Similar notions of pragmatic felicity are implicit in some of the prior studies of online comprehension of negated utterances (Reuter et al., 2018;
Tian et al., 2010). However, to the best of our knowledge no one has looked at whether directly providing a polar question under discussion can lead to rapid interpretation of negation.

In the current study we evaluate this hypothesis by explicitly presenting a QUD with a polar question about our world (e.g. Do people swim here?). We evaluate whether this allows for online use of negation by seeing whether subsequent predictions of lexical items are updated based on sentence polarity. If polarity can be incrementally incorporated into sentence comprehension, the listener should be able to predict upcoming lexical items in the affirmative case and suspend those same predictions under negation. In addition, by looking at both adults and children we can additionally evaluate whether children can use negation online when there isn’t a negated proposition already implicit in the discourse. Prior work with children has explicitly set up a negated proposition both through the linguistic context and through visual presentation of a limited set of possible referents. In the current study, there are no explicit options provided for either an affirmative (“yes, people swim in …”) or the negated response (“no, people don’t swim in …”). Thus, we can evaluate whether lexical predictions can be updated to reflect sentence polarity without explicit constraints on possible upcoming referents.

3.2 Methods

3.2.1 Participants

20 children, aged 5-7 (mean age = 6.35) and 21 adult (mean age = 17.63) native English speakers participated in the study. An additional 11 adults and 15 children were tested and excluded due to experimental error or excessive artifacts (trial loss of more than 35%). Adult participants were recruited from the Harvard University student body and the greater Boston community and received course credit or cash payment for participation. Child participants were recruited from our database of families in the greater Boston area.
3.2.2 Materials

Each participant was presented with 80 discourses. Each discourse consisted of a yes/no question followed by an affirmative or negated response. Examples are provided in Table 3.1.

Table 3.1: Example Stimuli

<table>
<thead>
<tr>
<th>Question:</th>
<th>List 1</th>
<th>List 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affirmative, Related (True)</td>
<td>Yes, people swim in the pool</td>
<td>Yes, people drive in the street.</td>
</tr>
<tr>
<td>Affirmative, Unrelated (False)</td>
<td>Yes, people swim in the street.</td>
<td>Yes, people drive in the pool.</td>
</tr>
<tr>
<td>Negated, Related (False)</td>
<td>No, people don’t swim in the pool</td>
<td>No, people don’t drive in the street.</td>
</tr>
<tr>
<td>Negated, Unrelated (True)</td>
<td>No, people don’t swim in the street.</td>
<td>No, people don’t drive in the pool.</td>
</tr>
</tbody>
</table>

The final word in each sentence was manipulated such that it was either semantically related or unrelated to the sentence’s subject and verb, making the answer either true or false depending on its polarity. Truth value of the sentence was determined by experimenters. Semantic relatedness of the target word to the preceding words of the sentence was confirmed with latent semantic analysis (LSA) (Landauer, Foltz, & Laham, 1998; Wolfe & Goldman, 2003). Related words (0.32) were significantly more semantically associated than unrelated (0.24) words ($t_{311} = 6.29$, $p < 0.001$). This resulted in four possible answer conditions; Affirmative and Semantically Related (True), Affirmative and Semantically Unrelated (False), Negated and Semantically Related (False); and Negated and Semantically Unrelated (True).

For each discourse in List 1, there is a discourse in List 2 in which a single word was changed in the question such that the target word that was semantically unrelated in List 1 was then semantically related in List 2. This ensured that across the two lists, all relevant features of the particular target word (frequency, length, etc.) were controlled. All answer sentences were spliced such that the same sound file of the target word was used across all conditions. The sound file for the Affirmative and Negated sentence onsets was taken equally often from sentences that were recorded with semantically related and unrelated endings, while the sound
files used across all ending were excerpted from recordings that were taken equally often from Affirmative and Negated sentence onsets. The target word was never mentioned in the question.

3.2.3 Procedure

Following the informed consent procedure, participants were fitted with an EEG cap and seated in front of a computer screen in a quiet testing room. They were told to try to remain still and keep their eyes directed at the screen. Subjects were introduced to two cartoon aliens. It was explained that the aliens have been studying our planet for the past year and were preparing for a test by quizzing each other about a series of images presented on a computer screen. The display included a depiction of a computer screen that was facing away from the participant and towards the aliens, such that the aliens could see the picture, but the participant could not. The purple alien (on the left) asked the question and the green alien (on the right) responded with one of the four possible conditions. Following the answer, the computer provided feedback on the alien’s performance, displaying a cross and a buzz sound if the answer was false and a star and a ding if the answer was true. Every 5 questions, the feedback did not appear, and instead participants provided truth-value judgement regarding the alien’s answer. The full procedure is outlined in Figure 3.1. ERP responses were recorded timed to the target word in the answer sentence and to the onset of the feedback response. In addition, accuracy responses were recorded every 5 trials. To obtain accuracy values, subjects were asked if the answer stated by the alien was correct or wrong. Their verbal responses were recorded.
Figure 3.1: Illustration of the display screen seen by participants and outline of the trial structure. The inter-trial-interval was 750 ms.

3.2.4 Measures

Our key measure of lexical prediction was the N400 response to target words. Mean amplitude of the N400 was calculated in the midline sites (Fz, Cz, Pz, Oz) across three predetermined time windows. First, a standard N400 window was measured from 350-550ms following noun onset. Finally, we also calculated mean amplitude during the P600 time window (550-800ms) at midline sites.

Looking beyond the target word, we also measured the ERP response to the presentation of the feedback slide, in which the sentence truth-value was revealed. Specifically, we measured the mean amplitude during the P300 time window (300-500ms) at midline electrodes. A larger P300 response may indicate surprise or a mismatch between the truth-value presented and the one computed by the participant. Finally, an accuracy score was measured for the subset of trials.

---

2 In addition, we evaluated an early (300-450) and late (400-550) window. This was done because we expected that the peak of the N400 would be earlier for adults than children (Atchley at al. 2006, Holcomb, Coffey, & Neville, 1992) and we wished to have additional windows that might best capture the peak for each of these groups individually. The results for both windows show a similar pattern. Full analyses are available in Appendix (A2).
in which truth-value decisions were requested. Because these were requested only every 5 trials, this measure was primarily intended to confirm the truth-value determined by the experimenter and to ensure that participants were attending to the task.

3.2.5 EEG Recording and Processing

All subjects' EEG data were recorded at 500hz using Brainvision’s Actichamp System with 32 active electrodes placed at International 10-20 System locations and on the left and right mastoids. Impedances were kept below 25kOhm for all relevant electrodes. Passive electrodes connected to the BIP2AUX adapter were attached to the left eye to monitor for vertical eye movements. Offline the EEG signal was resampled to 200 Hz and re-referenced to the average of the left and right mastoids. EEG signals were filtered using an IIR filter with a bandwidth of 0.01–40 Hz. Data were epoched from -200ms to 1000ms and baseline corrected using the pre-stimulus time window (~200 – 0ms). Eye artifacts were removed through independent component analysis. Trials were discarded if they contained artifacts greater than 90 μV. Individual bad channels were interpolated using spherical interpolation.

3.3 Results

To determine whether negation was affecting listeners’ lexical expectations, we compared the N400 response to the semantically related and unrelated target words in affirmative and negated conditions. Additional pre-planned analyses explored the P600 response to the target word, and the P300 response following the presentation of feedback. Finally, we also evaluated accuracy in trials where a behavioral response was recorded. All analyses were conducted in the R programming language, using the lme4 package (Bates et al., 2015). Using model comparison, we evaluated the role of semantic relatedness, polarity, and age group for each dependent variable. All analyses were preregistered on OSF.
3.3.1 Event-Related Potentials

In the current analyses we used linear mixed effects models (lmer) with random effects for both subject and trial. The maximal appropriate random effects structure was used (Barr, Levy, Scheepers, & Tily, 2013; Jaeger, 2008). In cases where a maximal structure failed to converge, random slopes were dropped first from the effects of trial and then from subject. Effects coding was used for all categorical independent variables.

3.3.1.1 N400

We first analyzed whether semantically unrelated words in the affirmative condition produced a larger N400 response relative to words that were semantically related. Looking at the overall N400 response window (350-550ms) across midline sites, there was a marginal overall effect of Relatedness (β = 0.65, SE = 0.40, p=0.10). The effect of Relatedness was significant (β = 1.02, SE=0.48, p < 0.05) at electrode Pz. Consistent with prior research, semantically unrelated endings showed smaller N400 responses than contextually appropriate, semantically related words. This effect did not differ by age group. Although there was a main effect of semantic relatedness, as expected based on prior work with the N400 response, this effect was not robust. As we will discuss later, the weak effect of semantic relatedness in the Affirmative condition may be due to the large number of counterintuitive sentences presented throughout the duration of the experiment.
Next, we compared the N400 response across all trials. Semantically related words showed significantly smaller N400 responses than semantically unrelated words ($\beta = 1.06$, SE = 0.30, $p<0.001$). Adding age group into the model did not improve model fit nor did the effect of semantic relatedness differ between the children and adults. There was no main effect of polarity, nor an age x polarity interaction. However, there was an interaction between Relatedness and Polarity ($\beta = 0.52$, SE = 0.27, $p = 0.05$) that marginally improved model fit ($\chi^2=3.60$, $p=0.06$).

**Figure 3.2:** A) Grand Average ERP waves at midline electrode sites (Fz, Cz, Pz) for Children and Adults and B) Scalp maps showing mean differences between Semantically Unrelated and Related words during N400 and P600 windows. Results show robust N400 responses to semantically unrelated words in the negated condition and a weaker effect for affirmative sentences.

This interaction reflected a larger effect of Relatedness in the Negated ($\beta = 1.71$, SE=0.43, $p<0.001$) rather than Affirmative condition ($\beta = 0.65$, SE=0.40, $p = 0.10$), which as we
noted earlier, was smaller than expected. There was no main effect of age group or a three way interaction across our variables.

3.3.1.2 P600

In addition to looking at the N400 response, we evaluated whether there was a P600 effect to sentences with related words in the negated condition relative to the affirmative. Given no N400 response to semantically related words in the negated condition, a P600 effect in this condition may reflect subsequent analysis based on truth-value. To specifically evaluate whether a P600 effect was present to semantically related words in negated sentences, we compared this condition to semantically related words in the affirmative condition, in which reanalysis would not be required. However, across this time window (550-800ms) we saw no differences between these two conditions.

Looking at the P600 window (550-800ms) at midline sites, we saw that the pattern detected in the N400 response continued through this later time window. There was an overall main effect of relatedness ($\beta = 1.50$, SE=0.34, $p < 0.001$) and a marginal interaction between relatedness and polarity ($\beta = 0.58$, SE=0.32, $p = 0.07$), with a smaller difference in the Affirmative rather than Negated condition.

3.2.1.3 P300

Analyzing the P300 to the feedback response, we saw no main effects of either relatedness or polarity, and neither effect was modulated by Age Group. Children had overall more negative responses than adults ($\beta = 1.44$, SE=0.65, $p < 0.05$). However, there was a significant 3-way interaction between relatedness, polarity, and age group ($\beta = 1.11$, SE=0.29, $p < 0.01$).
In both adults and children, there are no main effects of Relatedness or Polarity. In children, there was a significant Relatedness x Polarity interaction ($\beta = -1.76$, SE=0.53, $p<0.001$). They showed a more positive response in the P300 window specifically to semantically related target words in negated sentences ($\beta = 4.70$, SE = 1.52, $p < 0.01$). This is consistent with a lack of desire to agree that an unusual event doesn’t occur (e.g. surprise that *dogs don’t wear pants* is true). There was a marginal Relatedness x Polarity interaction in adults ($\beta = -1.05$, SE = 0.32, $p = 0.09$). In adults, there was no difference in response to related or unrelated sentences in the affirmative case, but they showed a more positive response to semantically related words in the negated condition ($\beta = -1.80$, SE =1.02, $p = 0.08$). This is consistent with surprise when a semantically related ending is declared as false when under negation (e.g. surprise that *dogs don’t wear collars* is false).

### 3.3.2 Accuracy

Overall both child and adult participants performed very well on this task, suggesting that they agreed with the experimenter's determination of truth-value and could correctly calculate the truth-value. The accuracy results are presented in Table 3.2.

<table>
<thead>
<tr>
<th>Table 3.2. Accuracy by Age Group and Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age Group</strong></td>
</tr>
<tr>
<td>Adult</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Child</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

There was a main effect of Relation ($\beta = -0.67$, SE= 0.22, $p<0.01$) that significantly improved model fit ($\chi^2=8.70$, $p<0.01$), but no Relation by Age Group interaction. Sentences
ending in semantically unrelated endings showed significantly poorer accuracy, meaning that both children and adults have some difficulty responding to statements that do not address a predictable QUD or end with unpredictable words. When Polarity is added to the model we see a marginal increase in model fit ($\chi^2=3.06$, $p=0.08$), with marginally lower accuracy in the negated condition ($\beta = 0.29$, SE=0.17, $p=0.08$), consistent with more difficulty calculating truth value to negated sentences. There was no Relation by Polarity interaction or 3-way-interaction with age.

### 3.4 Discussion

The degree of difficulty listeners face when they hear a negated sentence depends on the context in which it is heard (Nieuwland & Kuperberg, 2008; Nordmeyer & Frank, 2014, 2015; Tian et al., 2010). Although negation out of the blue poses a comprehension challenge (e.g. Fischler et al., 1983), a sufficiently supportive context can allow even young children to use negation to predict upcoming referents (Reuter et al., 2018). But what constitutes a sufficiently supportive context? In the current study we evaluate whether negation can be immediately incorporated into sentence comprehension when a polar question under discussion (QUD) is available. We test this by looking at lexical predictions in affirmative and negated statements when they are presented as answers to an explicit polar question about the world (e.g. Do people swim here?). Surprisingly, under these conditions, N400 responses to target words in negated sentences reflect their semantic association with prior words in the sentence, suggesting that neither children nor adults incorporate negation during online comprehension.

If listeners were able to use sentence polarity to update lexical predictions, one might expect that under negation they should expect the unexpected. Upon hearing "No, people don't swim in the ..." a semantically coherent ending would include any word other than a swimmable body of water. Since the exact target word cannot be predicted, this expectation would result not
in the prediction of the correct (unrelated) target, but rather in a weakening or suspension of prediction based on semantically association. We find exactly the opposite pattern: the effect of relatedness is actually greater in the negated sentences. This pattern cannot be (fully) explained by the alternative hypothesis: if encountering negation has no effect on the processes reflected in the N400, perhaps because negation is not rapidly understood in this context, then we should see a big effect of relatedness (as we do) but no interaction with polarity.

3.4.1 Partial support for two-step theories of negation

The main effect of semantic association in negated sentences is broadly consistent with the two-step theory of negation, in which constructing the affirmative is a necessary first step in comprehension (Carpenter & Just, 1975; Clark & Chase, 1972; Kaup et al., 2006; Kaup et al., 2007). Recent instantiations of the two-step theory of negation rely on perceptual simulation of the referential scene (Kaup et al., 2006; Kaup et al., 2007), arguing that listeners first form a mental representation of the state of affairs under an affirmative interpretation, and then apply the negative operator. In prior work supporting this proposal, activation of the state-of-affairs under an affirmative reading was only tested after sentence offset. The current study may suggest that hearing even the partial sentence "no, people don't swim..." would result in a simulation of people swimming even prior to the onset of the word pool. Decreased N400 responses to the semantically related words would then indicate either prediction based on the simulation or easier semantic integration due to semantic coherence with the simulated event. The truth-value of the sentence would then be calculated in a subsequent stage of processing.

Although simulation of a coherent world in sentences that end in a semantically related target word may play a role in the current study, contrary to our results, the two-step theory of negation would predict similar processing of negated and affirmative sentences. In addition, one
might expect that a secondary step of a reanalysis based on truth-value may show an
electrophysiological signature, such as a semantic P600 (Chow & Phillips, 2013; Van Herten,
Kolk, & Chwilla, 2005). P600 effects in the absence of syntactic errors are typically found in
cases where a semantic anomaly doesn't produce a large N400 response but is detected at a later
stage of processing, leading to a reanalysis. Much like the sentences presented in the current
study, this tends to occur in contexts where there are weak sentential constraints but strong
lexical associations that produce an early illusion of semantic coherence (Chow & Phillips,
2013). We see no evidence of a P600 effect in the current results. However, it is of course
possible that the calculation of truth-value occurs at a later analysis window (after 800ms post
target word onset but before the presentation of feedback). In fact, this would be consistent with
prior work showing that priming facilitation based on truth-value is only visible after 750ms post
sentence offset (Kaup & Zwaan, 2003).

3.4.2 Possible effects of an underspecified QUD

Alternatively, reduced N400 responses to semantically related words may be exaggerated
due to the under-specified QUD in the current design. Although a QUD is explicitly introduced
through a polar question in order to license negation, it refers to an image that is not visible to the
participant and is thus, by design, ambiguous. In this context, it is likely that people will try to
use the polar question to guess the identity of the target image. Since there is no way to make a
guess about a semantically unrelated target, participants are likely to guess and thereby activate,
semantically related alternatives. For example, upon hearing a question about swimming,
listeners may begin guessing possible bodies of water, activating either the target word or its
close semantic associate. Like with the simulation of the affirmative theory above, completing
the under-specified QUD would facilitate processing for both affirmative and negated sentences,
since making a guess about the depicted image is most likely to occur prior to the onset of the answer.

Tian and Breheny (2016) offer a possible explanation for the stronger effect of semantic association under negation. Under this account, negation serves as a cue for retrieving the QUD (Tian & Breheny, 2016). When listening to a sentence out of the blue, negation requires the listener to construct the QUD. For example, upon hearing “the door is not open” people construct the QUD as “whether the door is open”. If negation, and not affirmation, serves as a cue for retrieving the QUD, this would mean that upon hearing a negated answer, subjects may keep active or reactivate the recently filled in QUD (Do people swim in the pool?). However, this is unlikely as the construction of the QUD due to negation is argued to occur specifically when a QUD is not already available. Here, the QUD is explicitly introduced and would thus be equally available for both affirmative and negated sentences.

3.4.3 Effect of semantic relatedness under negation is unlikely to reflect strategic processing

In the current study, half of the sentences are ultimately false. This means that for half of the time a negated sentence concludes in a semantically related word. At the same time, there is no way to guess the target word that would make the sentence true, as this comes from an unconstrained set of options. Thus, counterintuitively, predicting the semantically related word in the negated condition is the most optimal strategy if a listener is trying to predict the word they will be hearing. However, we do not think that this is a likely explanation for the current data pattern. The same optimization of prediction applies to affirmative sentences as well. However, we see decreased reliance on semantic relatedness in the affirmative condition relative to the negated condition.
3.4.4 Interaction between semantic relatedness and sentence polarity may reflect greater processing cost associated with negation

It is possible that the interaction between semantic relatedness and polarity on the N400 may be due to the higher processing load associated with the semantic composition of a negated structure (Clark & Chase, 1972). Numerous studies have shown that negation incurs a processing cost (Carpenter & Just, 1975; Clark & Chase, 1972; Cornish & Wason, 1970; Just & Carpenter, 1975; Lüdtke & Kaup, 2006). If composing negation takes up processing resources, this would predict greater reliance on semantic association specifically in the negated, rather than affirmative condition. In a single trial, there are numerous cues in the current design that may promote top-down activation of semantically related target words including sentential constraints, simulation of possible answers to the polar question, or guesses regarding the image referenced in the QUD. All things equal, this would explain the main effect of semantic relatedness seen in the current results. However, in the broad discourse that unfolds across the trials, the alien only provides a true and coherent response half of the time, establishing a context in which top-down cues to target words are highly unreliable. In discourses where expectations are often violated, adults and children have been shown to rely less on predictive processing (Brothers, Swaab, & Traxler, 2017; Yurovsky, Case, & Frank, 2017). For example, in a self-paced reading task, Brothers et al. (2017) manipulated the proportion of predictable (The volleyball shot barely made it over the net) and unpredictable (The volleyball shot barely made it over the car") sentences included over the course of an experiment. They showed that when most sentences were predictable, reading times were faster for words in a constraining sentence context than a non-constraining one. However, when the discourse contained many unpredictable sentences, this effect disappeared. In the current study, the overall discourse is unpredictable,
since the alien is often wrong. Thus, we might expect decreased predictive processing if subjects are attending to the overall reliability of this cue across the experiment. This is what we see in the affirmative condition. Although there is still a detectable difference between semantically related and unrelated target words in affirmative sentences, this difference is not robust in either children or adults. However, we do not see this same decrease in sentences that are negated. One possibility is that the additional processing load under negation may inhibit consideration of the overall discourse structure. Lack of consideration of the broader unreliable context may result in an over-reliance on low-level cues for lexical access such as semantic association or failure to down-play lexical predictions advanced by less reliable cues.

Increased processing load associated with negation is also supported by adult responses to feedback and poorer performance on truth-value judgements for negated sentences. Although not robust, adults showed larger P300 responses when a semantically related word was presented in a negated condition. A P300 response to the feedback slide suggests that there is a mismatch between their current truth-value calculation and the sentences ultimate truth value, suggesting that subjects occasionally may not have calculated the truth-value in time for the presentation of truth value. Similarly, their explicitly provided judgments confirm that the truth-value of negated sentences may be harder to compute.

3.4.5 Current limitations and future directions

Several methodological constraints limit the conclusions we can draw from the current findings. The main challenge in the current design is that half of the sentences ultimately turn out to be false. As we describe above, the lack of reliability of predictive cues across the discourse may affect how cues are weighed during sentence comprehension, making the comparison between affirmative and negated sentences challenging. In future work we can test whether a
context that is dominated by true statements would result in greater use of negation online. This could be done in two ways. First, a large number of true filler sentences could be added to increase the ratio of true to false statements. This would be difficult in a study with children because they are unable and unwilling to sit through many trials, however, it would be possible to do with adult participants. Alternatively, we could modify the critical sentences so that the unrelated target words create sentences that are only temporarily false. For example, “no people don't swim in the pool” can be made true with an additional clause, for example "if the water is contaminated".

In addition, while there is speculative support in the current study that negation comes with a processing cost, it is possible that the processing cost might be specific to the types of sentences used in the current study. Generics are typically used to describe basic, early acquired, well-known facts about kinds (Leslie, 2008). As such, false generic statements may be particularly surprising or difficult to process. It is possible that we may see decreased reliance on semantic association in non-generic sentences.

Finally, in the current study we see few differences between adult and child comprehension. This is not surprising given that even adults in this task show poor use of negation. However, future work could help determine at least two outstanding questions about children's processing of negation. First, it would be beneficial to determine whether children can use negation to make lexical predictions. Thus far, children have only been shown to succeed in studies using eye-tracking (Reuter et al., 2018). However, eye-tracking studies confound referential and lexical disambiguation. Upon hearing “DW didn’t break a bowl” a look towards the unbroken bowl is consistent with a message-level representation (directing their eyes to an unbroken object or avoiding looking at the object that is clearly broken), or may be consistent
with the use of negation to predictively activate the specific lexical item (directing their eyes to an unbroken bowl due to activation of the lexical item *bowl*). Although children can predict the to-be-negated referent, it is yet unknown whether this information can affect lexical access. By switching to looking at the effects of negation on the N400 to words following negation, we can eliminate visual representations of possible items and look at how negation affects lexical processing. Secondly, earlier work has shown that children can interpret negation online when a very constrained set of possible interpretations is made possible. However, we do not know whether they can do so when possible interpretations are not presented with a visual display. An adaptation of the Nieuwland and Kuperberg (2008) paradigm may allow for insight into these questions.

3.4.6 Conclusions

The robust effect of semantic relatedness under negation in the current study suggests that a polar question under discussion is not sufficient for negation to be used incrementally to update lexical prediction. When sentences are presented in a context where negation is likely, as in a response to a yes or no question, but where the negated proposition itself is unpredictable, neither children nor adults show online updating of lexical predictions based on sentence polarity. However, future work should be conducted to determine whether this conclusion holds when the overall discourse is more reliable and conducive to predictive processing.
Chapter 4:

The Storytime Paradigm: A naturalistic ERP study of children’s lexical access
4.1 Introduction

A central part of language comprehension is recognizing words and retrieving their meanings. When we hear a word out of the blue, with no prior context, we can often identify it bottom-up, using only the sound of the word itself. More typically, however, word recognition occurs in the context of a sentence or a conversation that can guide us to the correct word, before it is fully spoken. For example, we can infer that the statement “He made a peanut butter and jelly ...” is likely to end with “sandwich”. Adults rapidly use top-down cues like these to disambiguate and predict upcoming words (e.g. McRae et al., 2005; McRae & Matsuki, 2013a) but we know very little about how this ability develops. The bottom-up process of mapping speech sounds onto candidate word forms has been extensively studied in infants and children. It develops early, becoming gradually more rapid and efficient during childhood (Fernald et al., 2001; Sekerina & Brooks, 2007; Swingley et al., 1999). In contrast, evidence for children’s use of top-down constraints in word recognition is more limited and more mixed (Henderson, Weighall, Brown, & Gaskell, 2013; Khanna & Boland, 2010). Some studies have found that children make less use top-down information than adults, both in the resolution of lexical ambiguity (Rabagliati et al., 2013) and in the resolution of syntactic ambiguity (Kidd & Bavin, 2005b; Snedeker & Trueswell, 2004; Snedeker et al., 2009). This pattern of findings suggests that there might be a developmental shift from bottom-up to top-down processing as children acquire knowledge about the world and become more adept at coordinating these information streams (see Snedeker & Huang, 2015).

The present study uses EEG to explore the comprehension of spoken words in adults and 5-10-year-old children in an ecologically-valid context. Our task (the Storytime Paradigm) is an adaptation of a new method that has been used to study language comprehension in adults
Participants listen to a story while we collect continuous EEG. We then model the event-related response to each word in the story to draw inferences about the role of top-down constraints during lexical comprehension. In the remainder of this Introduction: 1) we describe prior theories of lexical processing and the findings that led researchers to adopt a top-down interactive model for adults, 2) we discuss the prior work on lexical access in children, noting the absence of strong evidence for the use of top-down lexical prediction, and finally, 3) we describe our method for looking at lexical access during naturalistic listening and introduce three possible hypotheses about children’s use of top-down context.

4.1.1 Theories of Lexical Access in Adults

To understand speech, we must construct a series of representations (phonological, syntactic, and semantic) that link sounds to conversational intentions. This process is incremental: as we hear each speech sound, we update our hypotheses about the word that is being spoken, activate the meanings of those candidate words and begin integrating these potential meanings into the sentence (C.-Y. Lee, Liu, & Tsai, 2012; Yee & Sedivy, 2006). This flow of information from lower to higher levels of representation is called bottom-up processing. Critically, information also flows in the other direction as well: as adults are listening to a sentence, their knowledge of the speaker’s intentions and likely sentence meanings, shapes their hypotheses about upcoming words and sounds, resulting in top-down processing (e.g. Kutas & Federmeier, 2011; Van Berkum et al., 1999). In the present study, we focus on one step in this processing stream: accessing the meaning of a spoken word, a process we will call lexical access. Theories of lexical access in adults have changed over time as we have learned more about the role of context. Below we describe three perspectives that will frame our three hypotheses of how lexical access might develop in children.
Bottom-up models with thresholds based on frequency: All theories of spoken word recognition begin with the premise that, as a word unfolds, we map the sounds we are hearing onto stored lexical representations. The first sounds that we hear are usually compatible with a large number of words (e.g., candy begins in the same way as candle, cup, or cat) but subsequent sounds will typically eliminate most of the alternatives. Listeners make use of this information incrementally, activating a set of candidates on the basis of the first few phonemes and updating this set as the speech unfolds (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978). For example, upon hearing the first syllable of candy, people are more likely to look at candy or candles than an unrelated distractor (like a book). But after the second syllable, listeners will stop looking at the candles shifting their gaze to the candy (Allopenna et al., 1998).

Frequency plays a critical role in lexical access: more frequent words are identified more quickly and on the basis of less perceptual information affecting naming times, lexical decision times, reading, and referent identification (Erker & Guy, 2012). In EEG studies, more frequent words have smaller N400’s suggesting that they take less effort to identify (Halgren & Smith, 1987; Petten, 1993; Rugg, 1990). Because such effects are strong and ubiquitous, frequency is built into the bones of most word recognition models. For example, in many spreading activation models, each word has a set threshold which determines the level of activation required before it will be recognized (Marslen-Wilson, 1990; J. L. McClelland & Rumelhart, 1981; J. Morton, 1969). These thresholds are an inverse function of frequency, such that more common words have lower thresholds and thus require less perceptual input (and less time) to be recognized (Solomon & Postman, 1952).

In many early models, the initial process of lexical access was driven solely by bottom-up activation (Forster, 1979; Marslen-Wilson, 1987; Marslen-Wilson & Tyler, 1980; Norris, 1986).
Effects of context were attributed to later processes of selection, integration, or reconstruction (Forster, 1981; Norris, 1986; West & Stanovich, 1982). This commitment to bottom-up processing was motivated by computational concerns and supported by cross-modal priming studies which found that the unsuitable cohort competitors (Zwitserlood, 1989) and the irrelevant meanings of homophones (Neely, 1977; Swinney, 1979; Tanenhaus et al., 1979) were initially activated, even when the context could have potentially been used to rule them out (e.g., “The congregation rose” initially primes *flower*). In sum, on a bottom-up theory we would expect that initial lexical processes would be sensitive to features of the word (like frequency) but not to features of the context in which the word occurs.

2. Models with lateral semantic priming: The bottom-up theory was challenged by research which found that the speed with which we recognize a word can be influenced by the words that precede it. The classic demonstration of this is semantic priming: we are faster and more accurate in identifying a word (e.g. *cat*) when it is preceded by a related word (e.g. *dog*) than when it is preceded by an unrelated word (e.g. *chair*) (Meyer & Schvaneveldt, 1971; Moss, Ostrin, Tyler, & Marslen-Wilson, 1995; Swinney, Onifer, Prather, & Hirshkowitz, 1979). One way to account for this effect, without positing a fully interactive top-down processing system, is to allow the processing of one word to influence the next. This can be instantiated by associative connections between words (Collins & Loftus, 1975) or by shared semantic features within the lexicon that are primed when activated (Gaskell & Marslen-Wilson, 1997; J. Morton, 1969).

If lateral priming is the primary source of contextual constraint during lexical access, then we should expect that the difficulty of identifying a word in a sentence will depend on how closely it was associated with the set of words that proceed it, regardless of the exact meaning of the sentence or the goals of the conversation (the bag-of-words model). Latent Semantic
Analysis measures these lexical associations by tracking the contexts in which a word occurs and calculating the degree to which two (or more) words share the same contexts (Landauer & Dumais, 1997). This measure predicts the ease of lexical access as measured by lexical decision times and the magnitude of the N400 (Van Petten, 2014). Perhaps the strongest evidence that lateral connections have a privileged role in lexical access comes from studies of cross-modal priming using homophones. As noted above, both meanings of a homophone are usually active immediately after the word is spoken, even when syntactic constraints rule out the irrelevant meaning. However, if one of the words in the context has a strong lexical association with the correct meaning, then the incorrect meaning may not be activated (Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982).

3. Models with top-down prediction: Over the past two decades there has been increasing research into how word meaning is accessed when words occur in a supportive context. A sufficiently constraining sentence or discourse facilitates processing of words that are contextually coherent and can be anticipated based on the preceding information. Predictable words are identified faster (Craig, Kim, Rhyner, & Chirillo, 1993), and are more resistant to noise (Kalikow, Stevens, & Elliott, 1977). Additionally, eye-tracking studies have shown that first-pass reading times are faster when a word is supported by the preceding context (Kliegl, Grabner, Rolfs, & Engbert, 2004; Rayner, Slattery, Drieghe, & Liversedge, 2011) and words in highly constraining contexts are more likely to be skipped over (Rayner, Ashby, Pollatsek, & Reichle, 2004; Rayner, Binder, Ashby, & Pollatsek, 2001). Conversely, words that are inconsistent with a preceding discourse context cause a slowdown in self-paced reading (Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005).
In addition to behavioral evidence that context can facilitate access to word meaning, evidence for top-down constraints on lexical access has come from work looking at the N400 response (Kutas & Federmeier, 2011; Nieuwland & Van Berkum, 2006; Van Berkum et al., 1999). Historically, the N400 response is seen as a marker of semantic incoherence and is detected in cases where a word is difficult to access or integrate into the preceding context. For example, the final noun in a sentence like [1b] will show a larger N400 than [1a], indicating difficulty integrating the unexpected word tree.

[1a] *On a windy day, the boy went to fly a kite.
[1b] *On a windy day, the boy went to fly a tree.

While neural responses to overt semantic violations, like in the example above, can be explained through differences based on lateral semantic association, adult N400 responses have also been shown to be sensitive to higher-level constraints on lexical content from within the sentence (Kuperberg & Jaeger, 2016; Wlotko & Federmeier, 2012) and across a broader discourse (Berkum, Hagoort, & Brown, 1999; Nieuwland & Van Berkum, 2006). The size of the N400 response in adults decreases systematically as words become more predictable (Dambacher, Kliegl, Hofmann, & Jacobs, 2006; Davenport & Coulson, 2011). In fact, prior work in adults has shown that the cloze probability ratings, a measure of how likely a respondent is able to guess the correct upcoming word, are highly correlated to N400 amplitude (r ≥ 0.9) (Kutas & Federmeier, 2011). Beyond the scope of the sentence, the broader discourse can determine which word is easier to access and integrate. For example, Berkum et al. (1999) showed that in a sentence like [2],

[2] Jane told the brother that he was exceptionally quick/slow.
both final words are processed similarly when the sentence is presented in isolation. However, when the preceding context is supportive of one adjective over the other (e.g. the brother completes a task faster than expected) a robust N400 response is seen to the discourse-incoherent word. Nieuwland and Van Berkum (2006) showed that a supportive context can even eliminate N400 responses to verb-object animacy violations, making a sentence like [3a] easier to process

[3a] The peanut was in love

[3b] The peanut was salted

than a sentence like [3b] when presented in a story where peanuts as capable of emotion. In line with this evidence, the top-down model of lexical access suggests that the ease with which we access word meaning probabilistically reflects the constraints of the full discourse structure in which it appears.

4.1.2 Lexical Access in Children

Like adult comprehension, children access words incrementally: as the first sounds of a word are spoken, children begin activating possible lexical items and integrating those candidate words into their interpretation of the sentence (Fernald et al., 2001; Huang & Snedeker, 2011). By as early as 21 months, they can correctly identify a word given only partial phonological input (Fernald et al., 2001). Given a display possible of visual referents, 5-6-year-old children will narrow in on the plausible candidates, directing their gaze to both target items and phonological competitors until a single word can be identified. Children, however, consider competing lexical items for longer than adults suggesting that phonological processing improves with age (Huang & Snedeker, 2011; Sekerina & Brooks, 2007).

The speed with which children can identify words based on bottom-up phonological input is determined by the frequency of the target word and the size of its phonological neighborhood.
Thus, we see robust evidence that children’s behavioral responses reflect statistical properties of the words they are hearing. However, although children’s language processing is sensitive to frequency, its role in spoken word recognition has been debated. Word frequency influences word learning (Bird, Franklin, & Howard, 2001), with children acquiring more frequent words faster. However, in studies of word recognition this has resulted in a confounding effect, where words that are recognized faster and with more accuracy are both more frequent and earlier acquired (Garlock, Walley, & Metsala, 2001).

Like adults’, children’s lexical access is affected by the words that precede it. Numerous studies have provided evidence of spreading semantic activation in children. For example, children show robust effects of semantic priming; they are faster to identify a word after hearing a related word than an unrelated word (e.g. hearing cat after dog vs. chair) (Friedrich & Friederici, 2004; Radeau, 1983; Rämä, Sirri, & Serres, 2013). This contextual facilitation decreases with age (Radeau, 1983) suggesting that children may be more reliant on semantic facilitation than adults.

Far less research has looked specifically at children’s ability to use top-down constraints during lexical access. Khanna and Boland (2010) asked participants to read aloud a target word (e.g. grab) following a sentence that ended with a homophone (e.g. tag) in which one meaning was semantically-related to the target word, while the other was not. For adults and children over 12, hearing the word tag facilitated reading grab only when the preceding sentence constrained for the semantically-related interpretation (e.g. At recess the children played tag). However, younger children (7-9 years) showed facilitation even when the context was consistent with the semantically-unrelated meaning (e.g. Jerry was bothered by the shirt’s tag), suggesting that children struggled to use context to activate a single lexical item. However, when context was
simplified to a single word, even young children showed activation of a unique homophone (read *grab* faster after hearing “*laser tag*” than after hearing “*shirt tag*). It is possible that children can use top-down constraints given lower task demands, however, these results are also consistent with children relying on spreading activation alone. Similarly, Rabagliati et al. (2013), finds that children can make use of a particular top-down cue, world plausibility, to resolve lexical ambiguity, but that they do so to a much lesser extent than adults and rely more heavily on bottom-up information. This leaves open the question of how children prioritize information during natural listening. Does lexical access during natural listening reflect primarily bottom-up lexical cues or top-down predictions from higher levels of processing?

Although there are few studies directly assessing the use of top-down information to access word meaning, broader research of sentence comprehension has shown that children are generally less adept at using top-down cues during comprehension. For example, while four to six-year-olds will readily use bottom-up information such as intonation or lexical information to infer and disambiguate syntactic structure, they often fail to use higher-level cues such as the plausibility of an event or the referential context in which the sentence is used (Kidd & Bavin, 2005a; Kidd et al., 2011; Snedeker & Trueswell, 2004; Snedeker et al., 2009; Trueswell et al., 1999).

4.1.3 N400 as a measure of lexical access

In adults, the N400 has been shown to be sensitive to a wide range of variables that affect the ease of lexical activation (Kutas & Federmeier, 2011). For this reason, it is thought to be a neural marker of the process of word retrieval (Kutas & Federmeier, 2011; Lau, Phillips, & Poeppel, 2008). Consistent with behavioral findings, adult N400 responses are sensitive to word frequency, such that smaller N400s occur for words that are more frequent and thus easier to
access (Dambacher et al., 2006; Rugg, 1990; Van Petten & Kutas, 1990). Adult N400 responses are also sensitive to higher-level constraints on lexical content (Berkum et al., 1999; Kuperberg & Jaeger, 2016; Nieuwland & Van Berkum, 2006; Wlotko & Federmeier, 2012). The effects of frequency and predictability on the N400 response interact. In contexts with low predictability, either when words are presented out of the blue (Rugg, 1990) or early in a sentence (Payne, Lee, & Federmeier, 2015; Van Petten & Kutas, 1990), frequency strongly affects the size of the N400 suggesting that these unpredictable words must be retrieved bottom-up with little help from top-down cues. However, in highly predictable contexts, frequency effects disappear suggesting that the context allows words to be preactivated, minimizing or eliminating the burden of retrieving harder low-frequency words (Kutas & Federmeier, 2011; Payne et al., 2015; Van Petten & Kutas, 1990). When available, adults rely on top-down prediction over bottom-up facilitation during word recognition (Kutas & Federmeier, 2011; Lau, Holcomb, & Kuperberg, 2013).

In children this ERP response has been used to study a narrower range of questions. ERP investigations of children’s comprehension have shown an N400-like response that seems to index lexical retrieval and integration. However, with one exception (Benau, Morris, & Couperus, 2011), these studies have focused on comparing responses to overt semantic violations: either semantically anomalous words in sentences, or a word that mismatches a prior word or picture (e.g., hearing "dog" after seeing or hearing “cup”). In the picture mismatch paradigm, N400 like responses appear as early as 12 months (Friedrich & Friederici, 2010) becoming faster and more reliable over the second year of life (Friedrich & Friederici, 2004, 2005; Mills, Conboy, & Paton, 2005). In the violation paradigm, N400-like responses have been observed in 19-month olds (Friedrich & Friederici, 2005) as well as in toddlers and preschoolers (Silva-Pereyra, Rivera-Gaxiola, & Kuhl, 2005). Finally, children by 24 months show semantic
priming effects on the N400, with decreased responses to spoken words the follow a close
semantic associate (von Koss Torkildsen, Syversen, Simonsen, Moen, & Lindgren, 2007).
Although N400-like responses have been detected in children, the response is typically more
distributed, anterior, larger, and delayed relative to adults (Atchley et al., 2006; Hahne, Eckstein,
& Friederici, 2004). These differences lessen gradually between age 5 and adulthood (Holcomb,
Coffey, & Neville, 1992).

To the best of our knowledge there is no work looking at whether the N400 in children is
sensitive to word frequency or predictability. Although children show ERP responses consistent
with the N400, we do not yet know what constraints on lexical access are being captured by this
response. An N400 to an overt semantic mismatch or violation is consistent with adult-like use of
top-down constraints or bottom-up facilitation based on low-level semantic association. For
example, in a sentence that contains the words bread and toast, the word butter may show a
decreased N400 either because it is predicted based on the sentence and discourse context or it
may be primed, such that accessing semantic features of one word leads to pre-activation of
related words. Identifying the degree to which the N400 response in children reflects the
frequency, semantic relatedness, and top-down predictability of an uttered word we can get
insight into the information that children are able to use to identify word meanings.

4.1.4 Child-friendly approach to ERP studies

The vast majority of our understanding of lexical access and online sentence
comprehension more broadly has come from work using the visual-world paradigm and
traditional ERP designs. While both measures provide a time-sensitive index of lexical
processing, they evaluate language comprehension in restricted contexts that lack ecological
validity. Participants are often asked to interpret a series of isolated sentences or short paragraphs
with no relevant discourse. Recent work using single-trial ERP recordings offer a way to circumvent this challenge. In such experiments, participants read or listen to sentences as ERP responses to every word are recorded. Each word is coded for factors of interest (e.g., frequency or predictability) as they vary naturally across contexts. These factors are then evaluated as continuous predictors of N400 amplitude. Since single-trial ERPs are not restricted to a trial structure, these responses can be recorded as participants perform a natural task, such as listening to a story.

Single-trial ERP studies with adults have generally confirmed findings from more traditional designs; showing that the N400 response is modulated both by low-level features such as word frequency and higher-order processes associated with lexical prediction. Computational models applied to single-trial ERP data have found that the N400 reflects measures of lexical predictability given the preceding context. Several studies have specifically argued that lexical access in adults takes into consideration top-down constraints from the hierarchical syntactic structure (Brennan et al., 2016; Fossum & Levy, 2012) but see (Frank & Bod, 2011; Frank, Otten, Galli, & Vigliocco, 2015). While most of the past research has used isolated sentences presented visually word-by-word, more recently this method has been applied with stories presented auditorily (Brennan & Hale, 2019).

In the current study, we adapt this natural story-listening paradigm for use with children. Adults and children listen to a story as ERPs time-locked to the onset of every word are recorded, allowing us to gather data from hundreds of trials in a task that is short and fun, making it suitable for participants across ages. Most importantly it allows us to study children's language comprehension with greater ecological validity than other temporally sensitive methods. By using natural narratives, we can explore how children use the actual cues provided
by the discourse to guide comprehension in real-time, using a task (listening to a story) that is familiar and relevant to their lives.

In the current paper we evaluate statistical models of word activation that roughly mirror the historical development of our understanding of lexical access. These models, described below, ascribe ease of lexical access on an increasing amount of information; starting from bottom-up statistical properties to top-down prediction. First, we assume a bottom-up *frequency threshold model of lexical access*. Given no other context, the speed with which we can access a word is strongly related to the frequency of that word in our lexicon. In the current experiment, we evaluate the degree to which the N400 response in adults and children can be modeled by frequency alone. We then test a *spreading activation model of lexical access* in which ease of lexical access is affected not only by properties of the word itself but also by its relationship to other words in the context. Using Latent Semantic Analysis (LSA), we evaluate whether, beyond effects of frequency alone, activation of the target word is determined by its relationship to prior words in the discourse, with no consideration of higher-level properties such as the syntactic or pragmatic structure in which the words appear. Finally, to model our current understanding of adult-like word recognition we adapt a *top-down model of lexical access*, in which predictability ratings based on available discourse context predicts lexical access above and beyond models of bottom-up facilitation alone. While there is ample evidence that adults use this type of top-down prediction during sentence comprehension, it remains uncertain whether children are able to do so.

We will evaluate children’s ability to use contextual information during online sentence comprehension by seeing whether increasing use of context predicts the size of the N400 above and beyond simpler models. One possibility is that children’s lexical access is primarily driven
by bottom-up constraints such as word frequency or semantic association, consistent with prior work showing a general difficulty using top-down constraints during online comprehension. On the other hand, it is possible that given a rich discourse and a natural task, both children and adults may be able to recruit top-down constraints to inform their comprehension.

4.2 Methods

4.2.1 Participants

We tested 22 children, aged 5-10 (mean age = 7.65) and 21 adult native English speakers. An additional 9 adult and 21 children were tested and excluded due to excessive artifacts (trial loss of more than 50%) or experimenter error. All participants had normal or corrected-to-normal vision. Adult participants were recruited from the Harvard University student body and received course credit for participation. Child participants were recruited from our database of families in the greater Boston area.

4.2.2 Materials

Participants listened to an excerpt from Ch. 7 of Matilda, by Roald Dahl (Dahl, 2003). The excerpt contains 1594-word tokens of 580 types, out of which 766 tokens of 460 types are content words. The story was recorded at a comfortable reading pace by a female speaker. The onsets of all words in the story were hand-coded. In addition, the onsets and offsets were timed using the Gentle online text to speech aligner (https://lowerquality.com/gentle/) (Ochshorn & Hawkins, 2016). The human and computer coded onset times were highly similar (mean difference = 0.013 sec, 97% of words < 0.1 sec difference).

4.2.3 Procedure

Following the informed consent procedure, participants were fitted with an EEG cap, seated in front of a computer screen in a quiet testing room, and asked to listen to a story. They
were told to try to remain still and keep their eyes directed at the screen. During the story, pictures from the book were shown. The images showed the characters and changed approximately every 100 words. The images were presented in a randomized order and were chosen because they had no strong connection to the events at any given moment in the story. Subjects were told that they would be asked questions about the story at the end of the session and we made good on that promise (5 questions).

4.2.4 EEG Recording and Processing

All subjects' EEG data were recorded at 500hz using Brainvision’s Actichamp System with 32 active electrodes placed at International 10-20 System locations and on the left and right mastoids. Impedances were kept below 25kOhm for all relevant electrodes. Passive electrodes connected to the BIP2AUX adapter were attached to the left eye to monitor for vertical eye movements. Offline the EEG signal was resampled to 200 Hz and re-referenced to the average of the left and right mastoids. EEG signals were filtered using an IIR filter with a bandwidth of 0.01–40 Hz. Data were epoched from -200ms to 1000ms and baseline corrected using the pre-stimulus time window (-200 – 0ms). Eye artifacts were removed through independent component analysis. Trials were discarded if they contained artifacts greater than 90 μV.

4.2.5 Measures

**Frequency:** Frequency was calculated as Global Frequency (per million words) from the SUBTLEXUS corpus (Brysbaert & New, 2009). This is a 51-million-word corpus based on American television and film subtitles. Thus it captures language from aurally presented narratives and conversations that are intended for a wide age range. Frequency was log-transformed for all analyses.
**Semantic-Association:** In order to estimate association between words we used Latent Semantic Analysis (LSA) (Landauer et al., 1998; Wolfe & Goldman, 2003). LSA values were calculated based on the average pairwise cooccurrence values between the target word and the content words that preceded it within each sentence. Preliminary analyses were also concluded using LSA values calculated from the combined words of in the entire preceding sentence context as well as the immediately preceding two words, leading to similar result patterns.

**Cloze:** Discourse cloze probability estimates for each word in the story were gathered from adult participants (N = 423) on Amazon Mechanical Turk with each participant providing responses for one of twelve parts of the excerpt, resulting in ~30 ratings per word. Following informed consent, participants were asked to read an excerpt from the story. This excerpt always began at the beginning of the chapter and included 10 consecutive sentences in which participants were asked to guess words, one by one. The correct word is revealed following each guess and the subject was prompted to guess the next word. Following the critical guessing section, they continue reading the remainder of the excerpt. Thus, participants were asked to guess each word in a sentence based on the entire available discourse up to that point. Discourse cloze was calculated as the percent of correct guesses of the target word. Misspelled words were considered correct if they were phonetically consistent with the target word.

**Concreteness:** Concreteness estimates were gathered from adult participants (N = 235) on Amazon Mechanical Turk. Following informed consent, participants were asked to rate 50 words. Each word was presented with its part of speech. Concreteness was rated on a 7-point Likert scale, ranging from very abstract to very concrete. Subjects were given descriptions of what it means for a word to be concrete (“words refer to things or actions in reality, which you can experience directly through one of the five senses”) or abstract (“words refer to meanings
that cannot be experienced directly but which we know because the meanings can be defined by other words”). Each word received an average of 20 ratings, the mean of which was used as the concreteness value.

**Acoustic Length**: Acoustic Length was calculated as the difference between the onset and offset times of all words in the story, as detected by the Gentle online text to speech aligner ([https://lowerquality.com/gentle/](https://lowerquality.com/gentle/)) (Ochshorn & Hawkins, 2016).

4.2.6 Data Analysis

Our analyses examined the factors that predict the magnitude of the N400 effect. Prior to beginning data analysis, we made the decision to operationalize N400 size as the mean amplitude between 350-550ms averaged across the midline electrode sites (Fz, Cz, Pz, and Oz). We averaged across these electrode sites in hopes that it would increase the stability of the measurement by reducing extraneous noise. Preliminary analyses verified that the size of N400 and the nature of its relationship to our variables of interest did not systematically vary across the midline sites.

Predictors of N400 amplitude were evaluated using linear mixed-effects models via restricted maximum likelihood estimation. All analyses were conducted with the lme4 package (Bates et al., 2015) in R with the maximum appropriate random effects structure, with random intercepts for subjects and items. More maximal models with random slopes failed to converge. All predictors were scaled prior to analysis.

4.3 Results

To evaluate the proposed models of prediction we followed a theory-driven hierarchical modeling approach. First, we constructed a base model (Table 4.1) to control for other properties of the lexical items, or the context, that could influence the size of the N400 response beyond the
factors of interest. Next, we built and tested models that instantiated our three hypotheses about the use of context in lexical access (Table 4.2). These were: 1) The Frequency Model which captures a theory in which words are accessed bottom-up without regard to the context and the ease of lexical access depends largely on word frequency. 2) The Semantic Association or LSA Model in which words are linked by associative relations such that the ease of lexical access depends largely on the degree to which a word is associated with the words immediately before it. 3) The Top-Down Predictive Model in which the comprehenders incrementally construct a representation of the sentence, integrate it into the discourse, and make predictions about the upcoming word. On this hypothesis ease of access should be correlated with the Discourse Cloze Probability of the word.

In each case, our primary question was whether a given model was justified over the simpler, less sophisticated, model that preceded it. We tested this by adding the new factors to the old and comparing the models. For example, to see if there was evidence for the use of frequency, we compared the base model to a model that included all the factors in the base model as well as frequency. This is a conservative approach; it ensures that we can only find an effect of higher-level variables (the discourse cloze) if they predict the N400 response above and beyond what can be predicted by lower-level variables (like LSA). In our initial analyses, we included both the data from the children and the adults to allow for comparison. For each of the three theories, we first introduced the factor of interest (e.g. Frequency) and then explored whether it interacted with age. We followed this up with separate analyses of each age group which used backwards regression to find the best fitting model.

To foreshadow our central findings, we find that both children and adults show robust effects of discourse cloze on the size of the N400, consistent with the top-down prediction
model. Children, however, show larger effects of frequency that persist even when discourse cloze enters the model, suggesting that they may not predict word as consistently as adults and thus may need to engage in more bottom-up processing.

4.3.1 Constructing the Base Model

We constructed a base model to control for extraneous variables that might affect the N400, (see Table 4.1). Our first model included both properties of the target word (such as its concreteness, its acoustic length, the location of the word in the sentence, and the location of the sentence in the story) and properties of the word immediately before and the word immediately after the target (such as Frequency, Semantic Relatedness (LSA), and Discourse Cloze probability). We included this second set of predictors to control for spillover effects of the prior word, effects of the prior word on the baseline, and effects caused by early processing of the subsequent word. In this first model (Base 1 in Table 4.1), we found three reliable effects. First, the location of the sentence in the story is a significant predictor of N400 size, with the N400 response decreasing in size as the discourse progresses (β = 0.44, SE = 0.12, p < 0.001). Second, the size of the N400 to the target word is predicted by the frequency of the immediately preceding word (β = -0.36, SE = 0.16, p < 0.05). N400s were larger, or more negative, when the prior word should have been easier to process (due to its high frequency). This is the pattern that we would expect if processing of the prior word was resulting in greater negativity in the baseline region (-200 ms). Finally, N400 were predicted by the LSA value of that the preceding word to its context (β = 0.003, SE = 0.001, p < 0.05). Here, N400s were smaller when the prior word should have been easier to process. Note that this is the pattern that we would expect if there was a spill-over in processing, with difficulties continuing in the N400 window of the next word. By including these variables in our model, we control for these effects.
Next, we constructed a second baseline model to determine whether the effect of the control factors differed by age group (Base 2 in Table 4.1). We found that the effect of sentence position in the discourse significantly differed by age ($\beta = 0.25$, SE = 0.12, $p < 0.05$) with a robust effect in adults ($\beta = 0.63$, SE = 0.14, $p < 0.001$) and no significant effect in children. In addition, there was a significant interaction of Age and the Discourse Cloze of the word immediately preceding the target word ($\beta = 0.24$, SE = 0.12, $p < 0.05$): this factor had a marginal effect in children ($\beta = 0.40$, SE = 0.24, $p = 0.09$) but not adults. In children words preceded by more predictable words had smaller, or more positive, N400 responses (consistent with a spillover effect, rather than a baseline effect).

Our final baseline model, which served as the point of comparison for the models of prediction, included all of the factors from Base 1, along with the interaction terms that were found to be significant in Base 2 (age group x the position of the sentence in the discourse and age group x discourse cloze probability of the prior word). The full base model is described in Table 4.1.
Table 4.1: Model comparisons determining the control factors to retain in the final Base Model. All main effects and only significant interactions with age were retained.

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed Factors</th>
<th>AIC</th>
<th>BIC</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base 1</td>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lexical Controls:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Concreteness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acoustic Length</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Position in Sentence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Sentence Position in Story</em>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Baseline Controls:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frequency N-1 *</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frequency N+1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LSA N-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LSA N+1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Discourse Cloze N-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Discourse Cloze N+1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base 2</td>
<td>Factors in Model 1 + Concreteness x Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acoustic Length x Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Position in Sentence x Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Sentence Position in Story x Age †</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frequency N-1 x Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frequency N+1 x Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LSA N-1 x Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LSA N+1 x Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Discourse Cloze N-1 x Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Discourse Cloze N+1 x Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base Model</td>
<td>Lexical Controls:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Concreteness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acoustic Length</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Position in Sentence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Sentence Position in Story **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Baseline Controls:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frequency N-1 *</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frequency N+1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LSA N-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LSA N+1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Discourse Cloze N-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Discourse Cloze N+1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Discourse Cloze N-1 x Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Discourse Cloze N+1 x Age</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Factors that significantly improve model fit are highlighted in *italics*

†  = p < 0.1
*  = p < 0.05
** = p < 0.01
*** = p < 0.001
4.3.2 Models of Prediction

To evaluate our three theories of lexical processing, we used a hierarchical modeling approach, in which we first modeled the simplest theory and then added factors to capture the more complex theories. In other words we conducted a forward regression in which the factors were added in the order described in Table 4.2. Once a factor, was entered into the model, we left it in, both because this would allow us to conduct nested model comparisons, and because it provided the most conservative test for accepting a more complex model (does it account for variation in the N400 beyond the simpler model, regardless of whether that simpler model is warranted). The three theories of lexical access that we tested were: the frequency-based bottom-up model, a model of spreading activation as captured by LSA, and finally a model of prediction as captured by discourse cloze probability. For each model we first added the critical factor and then its interaction with age. We then conducted separate models on children and adults to better understand differences across age groups.

Table 4.2: Results of Model Comparisons Predicting Mean Amplitude of the N400 Response in both Adults and Children.

<table>
<thead>
<tr>
<th>Model</th>
<th>Added Factors</th>
<th>AIC</th>
<th>BIC</th>
<th>DF</th>
<th>Model Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base Model</td>
<td>See Table 4.1</td>
<td>109311</td>
<td>109439</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>+ Frequency †</td>
<td>109311</td>
<td>109446</td>
<td>18</td>
<td>$\chi^2(1)=2.72, p = 0.099$†</td>
</tr>
<tr>
<td>Model 2</td>
<td>+ Frequency x Age *</td>
<td>109308</td>
<td>109451</td>
<td>19</td>
<td>$\chi^2(1)=4.96, p = 0.026$ *</td>
</tr>
<tr>
<td>Model 3</td>
<td>+ LSA</td>
<td>109310</td>
<td>109460</td>
<td>20</td>
<td>$\chi^2(1)=0.02, p = 0.887$</td>
</tr>
<tr>
<td>Model 4</td>
<td>+ LSA x Age</td>
<td>109312</td>
<td>109470</td>
<td>21</td>
<td>$\chi^2(1)=0.07, p = 0.788$</td>
</tr>
<tr>
<td>Model 5</td>
<td>+ Discourse Cloze **</td>
<td>109306</td>
<td>109471</td>
<td>22</td>
<td>$\chi^2(1)=7.76, p = 0.005$ **</td>
</tr>
<tr>
<td>Model 6</td>
<td>+ Discourse Cloze x Age</td>
<td>109308</td>
<td>109481</td>
<td>23</td>
<td>$\chi^2(1)=0.14, p = 0.709$</td>
</tr>
</tbody>
</table>

Factors that significantly improve model fit are highlighted in italics
†  = p < 0.1
*  = p < 0.05
** = p < 0.01
*** = p < 0.001
**Frequency Threshold**

We first tested the Frequency Threshold hypothesis by comparing the base model to a model with log frequency as a predictor (Model 1). Frequency marginally predicted N400 size ($\beta = 0.23$, SE = 0.14, $p = 0.1$) and including frequency in the model marginally improved model fit over the base model ($\chi^2(1) = 2.71$, $p = 0.1$). As expected, more frequent words had smaller, more positive, N400 responses. As in the base model, frequency of the word immediately before the target word also significantly predicted the size of the N400 on the target word ($\beta = -0.35$, SE = 0.14, $p < 0.05$). The effect of the word before was in the opposite direction from the effect of the target word, with larger, or more negative, N400 responses when the target word is preceded by a high-frequency word. This is broadly consistent with baseline differences due to N400 variation of the preceding word. More negative going waves in the preceding word would be adjusted more positively in the baseline of the target word.

When age is included as an interaction term, a reliable Age by Frequency interaction emerges ($\beta = -0.25$, SE = 0.11, $p < 0.05$) with a larger frequency effect in children than adults ($\chi^2(1) = 4.96$, $p < 0.05$). As can be seen in Figure 4.1, adults show no significant effect of frequency while children show more negative N400 responses to less frequent words ($\beta = 0.48$, SE = 0.27, $p = 0.07$).
Figure 4.1: Grand average waveforms to low, middle, and high frequency words at electrode Pz and voltage maps of the difference between low and high frequency words in the time window of the N400 response (350 – 550 ms). Patterns shows no significant differences in adults (right) and a significant effect of frequency in children (left).

Semantic Association

To test whether semantic association contributed to lexical prediction, we looked at whether LSA predicted N400 size above and beyond the effects of Frequency. When compared to Model 2, which included both Frequency and a Frequency by Age interaction, we find that LSA did not significantly predict N400 size: the model with LSA (Model 3) did not provide a better fit to the data and the effect for LSA was not reliable. The lack of a reliable effect of LSA did not differ by age (Model 4), with neither adults nor children showing reliable effects. Thus, in a rich discourse context, this measure of lexical co-occurrence does not appear to influence the N400.

Discourse-Level Prediction
In evaluating top-down constraints on lexical access, we compared the model with LSA (Model 4) to a model with Discourse Cloze included as an additional predictor (Model 5). Discourse cloze probability significantly predicted N400 size ($\beta = 0.36, \text{SE} = 0.13, p < 0.01$) and increased model fit ($\chi^2(1) = 7.76, p < 0.01$) over the simpler model. More predictable words show smaller N400 responses. When age is included as an interaction term (Model 5), the main effect of discourse cloze remains significant ($\beta = 0.35, \text{SE} = 0.13, p < 0.01$) with no differences across the two age groups. Looking at each group separately, discourse cloze predicts N400 size in adults ($\beta = 0.36, \text{SE} = 0.15, p < 0.05$) and is only trending in children ($\beta = 0.35, \text{SE} = 0.24, p = 0.14$). Effects of Discourse Cloze are depicted in Figure 4.2.

**Figure 4.2**: Grand average waveforms to low, middle, and high predictability words, as measured by discourse cloze probability, at electrode Pz in adults (right) and children (left) and voltage maps of the difference between low and high predictability words in the time window of the N400 response (350 – 550 ms).
Since LSA did not significantly predict N400 size, we ran an additional model comparison to see whether including Discourse Cloze as a predictor improved model fit beyond the best prior model, which contained Frequency and a Frequency by Age interaction (Model 4.2). Including Discourse Cloze as a predictor significantly improved model fit over the simpler Frequency Model ($\chi^2(1)=7.42$, p < 0.01).

4.3.3 Backwards Regression

In addition to performing a hypothesis-driven forward regression, we also performed a backwards regression separately on the adult and child data. This analysis was performed to find the best model fit for each age group and to account for likely overfitting of a large model to less data. The starting point for this analysis, in each age group, was a model that included all the control variables from the Base 1 model, along with the three factors of interest: frequency, LSA, and discourse cloze. We removed the non-significant factors, one by one, starting with the one with the highest p-value. We stopped when we reached a point at which removing further variables lead to a decrease (p < 0.1) in model fit (an increase in AIC).

The final model for adults included significant effects of discourse cloze ($\beta = 0.28$, SE = 0.11, p < 0.05), position of sentence in the discourse ($\beta = 0.47$, SE = 0.12, p < 0.001), as well as baseline effects Frequency ($\beta = -0.35$, SE = 0.14, p < 0.05) and LSA ($\beta = 0.30$, SE = 0.14, p < 0.05) and Discourse Cloze of the subsequent word ($\beta = -0.26$, SE = 0.12, p < 0.05). There was also a marginal effect of the location of the word in the sentence ($\beta = 0.26$, SE = 0.11, p = 0.06).

The best model for children contained significant effects of Discourse Cloze ($\beta = 0.52$, SE = 0.18, p < 0.01) and target word Frequency ($\beta = 0.37$, SE = 0.18, p < 0.05).
Table 4.3: Best Fit Model by Backwards Regression for Adults

<table>
<thead>
<tr>
<th>Factor</th>
<th>Beta Estimate</th>
<th>Std. Error</th>
<th>T value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position in Sentence</td>
<td>0.218</td>
<td>0.117</td>
<td>1.859</td>
<td>0.063 †</td>
</tr>
<tr>
<td>Frequency N-1</td>
<td>-0.35</td>
<td>0.139</td>
<td>-2.522</td>
<td>0.012 *</td>
</tr>
<tr>
<td>Discourse Cloze N+1</td>
<td>-0.256</td>
<td>0.116</td>
<td>-2.202</td>
<td>0.028 *</td>
</tr>
<tr>
<td>LSA N-1</td>
<td>0.302</td>
<td>0.141</td>
<td>2.139</td>
<td>0.032 *</td>
</tr>
<tr>
<td>Discourse Cloze</td>
<td>0.283</td>
<td>0.115</td>
<td>2.463</td>
<td>0.014 *</td>
</tr>
<tr>
<td>Sentence Position in Story</td>
<td>0.469</td>
<td>0.121</td>
<td>3.884</td>
<td>0.000 ***</td>
</tr>
</tbody>
</table>

† = p < 0.1
* = p < 0.05
** = p < 0.01
*** = p < 0.001

Table 4.4: Best Fit Model by Backwards Regression for Children

<table>
<thead>
<tr>
<th>Factor</th>
<th>Beta Estimate</th>
<th>Std. Error</th>
<th>T value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.370</td>
<td>0.196</td>
<td>1.892</td>
<td>0.059</td>
</tr>
<tr>
<td>Discourse Cloze</td>
<td>0.483</td>
<td>0.197</td>
<td>2.458</td>
<td>0.014 *</td>
</tr>
<tr>
<td>Frequency</td>
<td>0.551</td>
<td>0.198</td>
<td>2.787</td>
<td>0.005 **</td>
</tr>
</tbody>
</table>

* = p < 0.05
** = p < 0.01

Estimates for child models, which contained fewer data points than those of adults, were corroborated with Bayesian models estimated using Markov Chain Monte Carlo (MCMC)

4.4 Discussion

The current study evaluates how children and adults access word meaning when listening to a story. To do so, we looked at what features of the words they hear predict the size of the N400 response, an electrophysiological marker of lexical access. We find that for children, but not adults, lexical access reflects the frequency of the word. However, the N400 response across both age groups reflects adult estimates of cloze probability, suggesting that both children and adults use top-down constraints from discourse to access word meaning.

4.4.1 N400 Responses Reflect Expectation Not Association

The current findings suggest that when listening to a story both children and adults access the words they hear using top-down expectations from the full discourse context, and not semantic association from preceding words. The size of the N400 response across both age
groups similarly reflects the cloze probability ratings of incoming words. Here, cloze probability was calculated from predictions made by adult comprehenders who were provided with all the relevant preceding discourse. While we do not know exactly what information participants were using to make the relevant guesses, there are several features of the data that suggest that they make use of top-down constraints from the semantic, syntactic, and pragmatic context, consistent with adult lexical processing. First, we see increasing predictability as the story progresses, suggesting that subjects are building up a narrative that is being reflected in the accuracy of their predictions. In addition, we had a separate group of subjects perform a similar cloze probability task but they were not provided with the context of the unfolding story, rather, they were asked to guess upcoming words in a series of unrelated sentences. People were better able to correctly guess that upcoming words when they were presented with the entire discourse context, confirming that the narrative informed their predictions. In addition, the robust relationship between the N400 response and cloze probability is broadly consistent with prior work in adults. The current study shows that children’s N400 responses are also guided by similar lexical expectations.

While we see clear evidence of cloze probability influencing the size of the N400 response, we do not see effects of lexical association in either age group. This contrasts with other studies showing modulation of the N400 based on “bag-of-words” models of lexical association (Frank & Willems, 2017; Van Petten, 2014) and LSA values more specifically (Van Petten, 2014). There are several possible reasons for the discrepancy between the current results and prior work. In the current study, lexical association is calculated as the average LSA coocurrence of any particular target word to the of the prior content words in the sentence, consistent with prior work in sentence processing (e.g. Xiang & Kuperberg, 2015). One
possibility is that this operationalization simply fails to capture semantic association and is not the appropriate measure of spreading activation. However, we think this is unlikely for the following reasons. LSA, as a measure of word association, has been shown to predict N400 size better than comparable measures of word cooccurrence and when other factors such as frequency, orthographic neighborhood density, and concreteness are controlled for (Van Petten, 2014). Although LSA may be an adequate model of word association, one might argue that the current operationalization, which uses sentence boundaries as a delimiter, may be more suitable for modeling isolated sentences than an ongoing story. Semantic association may be more local, such that only effects of closest words matter. To rule this out, we recalculated semantic relatedness between the target word and the two immediately preceding words using LSA and showed a similar null pattern of results. Alternatively, the effects of LSA may be more distributed, such that words across a narrative influence activation levels such that the effects for any word are hard to separate from any other word. However, this wouldn’t be a failure at operationalization but rather would suggest that semantic association across a story may not result in robust item-by-item variation.

It is likely that the divergence from prior work may be due to the natural discourse in which the words are presented. Unlike the naturalistic paradigm, earlier work presented series of sentences out of the blue. A single sentence is less constraining than a sentence embedded in a broader discourse, thus, when top-down constraints are weak, the language system may rely less on top-down information and instead uses lexical association. It is also possible that word recognition doesn’t rely on lexical association, but when the only information you have is from a single sentence (constructed by an experimenter) the top-down constraints based on predictability may be very closely captured by the bag of words model. This may be particularly
true of experimental stimuli rather than single sentences taken from real-world contexts because they are designed to be comprehensible and roughly natural with no prior context. Such sentences have to rely heavily on the relationships between the concepts encoded in the content words and if those relations are novel the sentence is likely to be surprising.

4.4.2 No Frequency Effect in Adults

In the current data, the N400 response in adults does not reflect lexical frequency. Furthermore, unlike prior research (Van Petten, 2014), we see no interaction between frequency and any measure of lexical predictability, including cloze probability and the position of the word in the sentence.

Why might adult listeners show no effect of frequency in the current paradigm? It is possible that the frequency of words that can be found in a story suitable for children does not provide the range necessary to see an effect in the N400. However, we do not believe this is likely as the range in the current story is comparable to studies that do detect frequency effects, with minimum frequency at 0.2 words per million and roughly a third of the items with frequencies of fewer than 50 words per million (Van Petten & Kutas, 1990). Effects of frequency on the N400 are primarily detected in comprehension of isolated words (Rugg, 1990) but are inconsistent when words are presented within a context (Fischer-Baum, Dickson, & Federmeier, 2014; Kretzschmar, Schlesewsky, & Staub, 2015). Studies that do show effects of frequency consistently find that frequency interacts with word predictability, such that frequency only matters in low predictability contexts (Van Petten & Kutas, 1990). The story of Matilda, by Dahl, is a highly restricted school-based context with a highly predictable narrative. It is possible that adults are able to constrain lexical access based on discourse enough to minimize any possible effects of frequency.
4.4.3 Why Frequency Plays a Larger Role in Children

A critical finding in the current experiment is that while frequency doesn’t predict the N400 in adults, it does in children. This is consistent with greater reliance on bottom-up information in children. The frequency by age interaction persists even when proper names are removed from the analyses, indicating that the difference is not due to story-specific items, but instead reflects a developmental change in lexical processing more broadly.

We see two non-exclusive explanations for this developmental shift. First, the magnitude of the frequency effect may decline as a direct effect of increasing experience. In many models of language acquisition, connections between meaning and lexical form become stronger with experience, resulting in faster processing. Therefore, more frequent words are processed more frequently. Children have experienced everything fewer times than adults. Thus, all of their connections are likely to be weaker and the increased effort required to activate any one of them may be more noticeable – a bigger part of the effort in lexical access. This hypothesis could be explored by teaching both children and adults novel words of varying frequencies. Second, the increased role of frequency in children may be due to less reliance on top-down information. The current data show that children are able to use top-down constraints to access word meaning, as evidenced by the relationship between the N400 response and discourse cloze probability. However, only adults show decreasing N400 responses as the story progresses, suggesting construction of a broader narrative.

4.4.4 Methodological Relevance

Earlier work in online sentence comprehension has come primarily from work using either the visual world paradigm or traditional ERP designs. However, both paradigms have important limitations. In the VWP, comprehension is assessed by looking at gaze directed at
objects in a visual display as participants listen to a series of sentences. These eye-tracking measures provide fine-grained data about comprehension, but they require that each interpretation under investigation be linked to an object in a visual scene (or its close associate). As a result, VWP studies typically assess comprehension in a highly constrained context. Many studies are designed such that a sentence can only end in two to four ways (given the display) and typically just one of these endings will meet all the constraints of the sentence and discourse. While traditional ERP designs, in which ERPs are compared across manipulated conditions, avoid some of the pitfalls of the visual world paradigm, this paradigm is particularly difficult to use with children. Like the VWP, ERP studies of language processing are built around controlled contrasts, resulting in repetitive strings of isolated sentences or a series of short disconnected paragraphs. However, unlike experiments using the VWP, which are often short, interactive, and fun, traditional ERP designs are often long and tedious. Because the signal-to-noise ratio is lower in ERPs, experiments typically need to have at least 30 stimuli per cell per subject, resulting in hundreds of trials. Furthermore, participants are asked to remain still and attentive often for an hour or more - a challenge for any child. In both of these paradigms, the task is often novel and motiveless (e.g., following instructions or remembering sentences) and participants listen to a series of unrelated sentences, typically with highly repetitive syntactic constructions. When discourse is studied, and in children it rarely is, it is with strings of short vignettes consisting of 2 to 5 sentences.

Due to these limitations we cannot be sure that these studies tap into the skills that are relevant to comprehension in more ecologically valid contexts. Most sentences people hear are integrated into a rich discourse structure rather than a series of disconnected sentences and rarely are the referents of those sentences in the directly accessible visual scene. These differences may
significantly impact the comprehension processes that are required for understanding. Lack of a rich discourse can diminish the use of top-down constraints while the presentation of visual referents or the repetitive nature of the paradigm may allow for preencoding, overestimating listeners’ ability to predict upcoming input.

We want to highlight the potential benefits of the natural listening paradigm for studying language comprehension in children. With the current design we were able to record ERPs from over 700 trials in a task that is short, fun, and ecologically valid. In addition, while the current paradigm relies on correlational analyses, we anticipate that controlled experiments can be grafted on to the natural listening paradigm. Not only is the task easier to administer than traditional ERP paradigms, but studying language comprehension during a coherent, age-appropriate story or passage may provide a better window into how comprehension occurs in real-world contexts.

4.4.5 Children use top-down information to access the meaning of words

Adult listeners make robust use of top-down constraints during word recognition, predictively activating the most likely lexical items based on the semantic, syntactic, and pragmatic context (Berkum et al., 1999; Khanna & Boland, 2010; Li & Yip, 1998; Nieuwland & Van Berkum, 2006; Tanenhaus et al., 1979; Wlotko & Federmeier, 2012; Zwitserlood, 1989). Although there is limited prior work looking at children’s ability to use similar top-down cues, most studies in this domain and in sentence comprehension more broadly, have suggested that children should show specific struggles with top-down prediction (Kidd & Bavin, 2005a; Kidd et al., 2011; Snedeker & Trueswell, 2004; Snedeker et al., 2009; Trueswell et al., 1999). However, the results presented in this paper suggest that both adults and children are able to use top-down constraints during spoken word recognition, as indexed by the N400 response.
Earlier work in sentence comprehension has widely shown that children are generally less adept at using top-down constraints, often failing to use higher-level cues such as the plausibility or referential context (Kidd & Bavin, 2005a; Kidd et al., 2011; Snedeker & Trueswell, 2004; Snedeker et al., 2009; Trueswell et al., 1999). However, most studies in this domain have been conducted using either the visual world eye-tracking paradigm and traditional ERP designs. One important limitation of both of these experimental procedures is that they present listeners with a series of disconnected, highly repetitive sentences or paragraphs, where top-down predictions must be made off of limited sentential information in a disconnected discourse. It is possible that children are more able to harness the relevant discourse context for lexical access when provided with a richer, and longer, discourse structure. There is growing evidence that children can prioritize different sources of information based on the demands of the task (Brothers et al., 2017; Yurovsky, 2017; Khanna & Boland, 2010). In fact, recent work by Yurovsky et al. (2017) provides direct evidence that children as young as five can flexibly adapt the degree to which they rely on bottom-up vs. top-down cues based on their reliability.

In adults, top-down constraints dominate over lower-level information such as frequency. However, in children, frequency continues to play a critical role. The effect of frequency in children is broadly consistent with current theories of the development of sentence comprehension in which children rely more on bottom-up cues than adults.

While the current study shows clear evidence that children are sensitive to top-down constraints, as indexed by adult cloze probability values, we hope that future work can better decipher exactly what sources of information go into the predictions that result in our estimates of cloze probability. In follow-up studies we hope to manipulate the narrative to isolate the use of semantic, syntactic, and pragmatic constraints.
Chapter 5:

Conclusion
5.1 Conclusion

Understanding the moment-to-moment process by which children transform an unfolding stream of sounds into a communicated message can provide much needed insight into the architecture of the language processing system and the source of developmental change in language comprehension. The three papers in this dissertation use both eye-tracking and ERP methodologies to get insight into how various sources of information constrain lexical and syntactic predictions as a sentence unfolds in time. Across diverse language backgrounds (Paper 1) and different contexts (Papers 2 and 3), the work presented here broadly confirms earlier research showing that children rely more on bottom-up information than adults (Snedeker & Huang, 2015; Trueswell & Gleitman, 2004). In Paper 1 we argue that children's ability to make use of top-down constraints increases primarily with language experience, rather than with improvements in executive functioning. In Paper 2 we use the test case of negation to argue that ability to integrate top-down information during online sentence comprehension may depend on the context in which a sentence is heard. Finally, in Paper 3 we use a natural listening paradigm to show that children can make use of top-down constraints when a sentence occurs within a rich discourse structure, but that they rely more on bottom-up information than adults. Using ERPs recorded during an ecologically valid language task suggests that earlier work using the VWP may have underestimated the degree to which children can make use of prior context during online comprehension.

Summary of papers

In Paper 1, we use the visual world paradigm to compare monolingual and bilingual children’s comprehension of garden-path sentences. Garden-path sentences provide insight into two critical components of sentence comprehension. First, by looking at an ambiguous
grammatical structure, we can evaluate the degree to which adults and children are able to commit to a likely sentence interpretation and the degree to which top-down cues such as the referential context can inform this prediction. Second, garden-path sentences allow us to evaluate what happens when initial predictions turn out to be incorrect and revision is required.

By comparing monolingual and bilingual children in this task, we were able to evaluate the roles that linguistic and cognitive development play in both the selection and revision of syntactic predictions. Although EF skills are relatively equal between these two groups (contrary to findings showing an EF advantage in bilingual children, e.g. (Bialystok, Craik, Klein, & Viswanathan, 2004), children with bilingual backgrounds have less experience with English as indicated both by their vocabulary and receptive grammar scores. Using both group comparisons and correlational approaches, we show that the use of top-down cues to make and revise syntactic predictions is better explained by linguistic rather than cognitive development. However, our data also suggest that bilingual children may have a processing advantage that cannot be attributed to differences in cognitive control. We hypothesize that pragmatic challenges faced by a bilingual language learning environment may improve meta-linguistic awareness that may affect sentence comprehension.

An advantage to looking at a broader diversity of language backgrounds when evaluating age-related changes in language comprehension is that it allows for more dissociation between linguistic and cognitive factors. However, this comparison also introduces a number of important confounds. For example, monolingual and bilingual participants often differ in culture, SES, and socio-linguistic environments. What's more, bilingualism is an all-encompassing term that collapses across factors that may play a key role in linguistic development. Bilingual children differ in the languages they speak, the age they were first exposed to a second language, how
often they use each language and switch between the two, etc. All these factors may play a role in what skills they bring to the table when they are faced with an incoming sentence to comprehend. The inherent variability within this sample creates a challenge for drawing concrete conclusions about bilingual abilities and may explain the variability in findings across the field (Luk et al., 2011; Yang, Hartanto, & Yang, 2016).

While we argue that bilingual participants may compensate for their lack of language experience with enhanced attention to context, we have not yet tested this hypothesis as the current study did not include an independent measure of pragmatic or metalinguistic ability. Further work would be needed to determine whether these skills contribute to garden-path comprehension above and beyond the effects of language experience.

Finally, although we argue that executive functioning may not be the driving force of developmental change in garden-path resolution, we still lack a broader understanding of how domain-general capacities may interact with language comprehension and acquisition. While EF training studies are notoriously difficult to conduct on children, manipulating available EF resources may be a critical next step towards a better understanding of its role in language comprehension.

In Paper 2, we evaluated adults’ and children's ability to incorporate negation into their online lexical predictions. Specifically, the study was aimed at determining whether explicit presentation of a polar question-under-discussion may be sufficient for immediate calculation of sentence truth-value or whether integration of negation in lexical predictions required constraints on the to-be-negated preposition. We found that without a predictable conclusion, negated sentences pose a comprehension challenge even for adults, with N400 responses reflecting semantic association and not ultimate truth-value.
Paper 2 used ERPs rather than eye-tracking in order to evaluate lexical prediction. Looking at N400 responses to target words allowed us to test comprehension in the absence of a visually constraining referential environment. However, ERPs introduce a key limitation in the current design. The N400 response relies on detection of semantic incongruity or violations, but the presence of violations in linguistic discourse can change how listeners process upcoming information and the cues they use for predicting upcoming lexical access (Brothers et al., 2017; Yurovsky et al., 2017). The frequent truth-value violations across the discourse context in the current study may have prevented the use of negation for lexical prediction. Before we can draw firm conclusions about listeners’ inability to update lexical predictions based on logical semantics, future work will be necessary to test the use of negation in a context where the speaker remains a reliable communicator.

While the unreliable communicator may be a limitation of the current study, this raises larger questions about how context and reliability of cues impact online sentence comprehension and how the experimental paradigms used may alter the way in which a sentence is understood. Contexts that contain ambiguity, unpredictability, or overt violations may underrepresent listeners’ ability to use top-down cues for predictive processing in more naturalistic contexts.

In Paper 3 we use a new method of looking at ERPs during a natural story-listening task. We find that in a rich, highly predictable discourse, children but not adults show N400 responses that reflect the frequency of the words they hear. However, beyond the effects of frequency or semantic relatedness, both child and adult responses are sensitive to cloze probability. Our findings suggest that during naturalistic listening, both adults and children use top-down constraints from the semantic/syntactic/pragmatic context to access upcoming words. However, in children, lower-level information continues to play a more central role than it does in adults.
These results are broadly consistent with theories in which top-down processing is a core feature of comprehension, but prediction improves with age.

An outstanding theoretical problem highlighted by the current study is the current under-specification of both bottom-up and top-down sources of information. In the current design bottom-up processing is operationalized through word frequency. However, we know that many other bottom-up factors play a role in lexical access, including phonological neighborhood density and age-of-acquisition (e.g. Garlock et al., 2001). In follow-up work we plan to use the current data to better define and control for bottom-up cues. We can do this by evaluating the degree to which N400 responses reflect a wider set of bottom-up predictors. In addition, we can see whether lexical access is better captured by a holistic measure of bottom-up activation, as operationalized by lexical decision times. Additionally, top-down processing in the current design is estimated through adult cloze probability values given the available context. However, a drawback of cloze probability is that we do not know what information within the context adults use to make their predictions. It is possible that combining single-trial neural responses with computational language processing models may provide headway into this critical question.

Although the current methodology provides an exciting opportunity to study comprehension during a task that is both ecologically valid and engaging, a central limitation of the current method is its reliance on correlational analyses. As such, our conclusions are limited by possible confounds and inter-correlations across measures. However, in ongoing and future research we can modify this paradigm to accommodate experimental manipulation. In doing so, we can begin to disambiguate the causes of processing difficulty while retaining the benefits of an ecologically valid and child-friendly procedure.

5.1.1 Methodological advancement
Our current understanding of children's language comprehension relies heavily on research conducted using the Visual World Paradigm (VWP). Although the VWP has provided valuable insight into the development of sentence comprehension, it has several key limitations. Due to its reliance on a visually presented scene, comprehension is typically studied in highly constrained referential environments. In addition, visually depicted targets often confound lexical and referential processes. Finally, the trial structure of the VWP places limits on the context under which comprehension can be tested. The natural listening task presented in the current dissertation presents a promising alternative that can circumvent some of these challenges.

Recording ERPs timed to the onset of every word in a story allows us to gather a considerable amount of data (700-2000 trials) in a short experimental session (10-30 minutes). In measuring a well-studied and well-understood index of lexico-semantic processing (the N400), we can get insight into children's lexical access, a previously understudied step in children's sentence comprehension. In addition, the task is engaging and suitable for a variety of ages and populations. Most importantly, it allows us to study children's language comprehension with greater ecological validity than other temporally sensitive methods. By using natural narratives we can explore how children use the actual cues provided by the discourse to guide comprehension in real-time, using a task (listening to a story) that is familiar and relevant to their lives.

In currently ongoing projects, we plan to replicate and expand the current results. First, by replicating these results with a different story, we hope to ensure that our findings generalize to new materials and new participants. In addition, we are working on expanding this paradigm for use with a broader range of ages and developmental histories. First, by adapting this paradigm for use with preschool children, we can look at language comprehension prior to the
onset of literacy. In a second line of work, we compare reliance on top-down prediction during lexical access in typically developing children and children on the autism spectrum.

5.1.2 Concluding Summary

In this dissertation the set of presented papers argues that children rely more on bottom-up information than adults, but the ability to use top-down cues increases with language experience. In addition, children’s reliance on top-down information depends on the context in which comprehension occurs. Studies using event-related potentials, and natural listening tasks more particularly, offer an exciting opportunity to expand the contexts in which online language comprehension has been investigated.
References:


doi:10.1177/0956797611409589


doi:10.1162/jocn.2006.18.7.1098


Wiseheart, M., Viswanathan, M., & Bialystok, E. (2016). Flexibility in task switching by monolinguals and bilinguals(). *Bilingualism (Cambridge, England), 19*(1), 141-146. doi:10.1017/S1366728914000273


Table A1: Results of Models Describing the Predictors of Measures of Interest in Paper 1

<table>
<thead>
<tr>
<th>Vocabulary Standard Score (PPVT-4)</th>
<th>Receptive Grammar Standard Score (TROG-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor</strong></td>
<td><strong>β Est.</strong></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>74.66</td>
</tr>
<tr>
<td>Age</td>
<td>2.54</td>
</tr>
<tr>
<td>Parental Ed.</td>
<td>4.75</td>
</tr>
<tr>
<td>Non-Verb. IQ (KBIT SS)</td>
<td>0.22</td>
</tr>
<tr>
<td>RT Cost Arrows</td>
<td>0.00</td>
</tr>
<tr>
<td>Acc Cost Stroop</td>
<td>42.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reaction Time on Simon's Arrows Task</th>
<th>Accuracy Cost in Verbal Stroop Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor</strong></td>
<td><strong>β Est.</strong></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>844.38</td>
</tr>
<tr>
<td>Age</td>
<td>-129.9</td>
</tr>
<tr>
<td>Parental Ed.</td>
<td>1.84</td>
</tr>
<tr>
<td>Non-Verb. IQ (KBIT SS)</td>
<td>0.08</td>
</tr>
<tr>
<td>Vocabulary (PPVT-4 SS)</td>
<td>-2.82</td>
</tr>
<tr>
<td>Grammar (TROG-2 SS)</td>
<td>3.41</td>
</tr>
</tbody>
</table>

† = p < 0.1  
* = p < 0.05  
** = p < 0.01  
*** = p < 0.001  
SS = Standard Score
A2:

In addition to looking at the overall N400 window, we analyzed the pattern of results in an early (300-450ms) and a late N400 window (400-550ms) in order to capture possible latency differences in N400 response between adults and children. The pattern across both windows was similar to the overall N400 window. The effect of semantic relatedness was significant across all time windows (Early time window: Β = 0.71, SE = 0.29, p < 0.05; Later time window: Β = 1.25, SE = 0.31, p < 0.001). It was robust for both children (Β = 1.40, SE=0.51, p < 0.01) and adults (Β = -2.52, SE=0.81, p < 0.01) in the later time window. Confirming the pattern found in the overall N400 response window, there was an interaction between Relatedness and Polarity in both the early (Β = 0.56, SE=0.26, p < 0.05) and the late (Β = 0.47, SE=0.28, p = 0.09) windows, There were no main effects of age or age interactions. In summary, both the affirmative and negated conditions, participants showed smaller N400 responses to semantically related words, regardless of ultimate truth-value. This pattern was stronger in the Negated rather than Affirmative condition.
Table A3: Full results for linear mixed effects models predicting N400 amplitude. Model comparisons described in Table 4.2.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>0.23</td>
<td>0.22</td>
<td>0.3</td>
</tr>
<tr>
<td>Log Frequency</td>
<td>0.23</td>
<td>0.14</td>
<td>0.1</td>
</tr>
<tr>
<td>Age Group</td>
<td>-0.02</td>
<td>0.21</td>
<td>0.93</td>
</tr>
<tr>
<td>Log Frequency N-1</td>
<td>-0.35</td>
<td>0.16</td>
<td>0.03</td>
</tr>
<tr>
<td>Log Frequency N+1</td>
<td>-0.02</td>
<td>0.16</td>
<td>0.91</td>
</tr>
<tr>
<td>LSA N-1</td>
<td>0.24</td>
<td>0.17</td>
<td>0.16</td>
</tr>
<tr>
<td>LSA N+1</td>
<td>-0.06</td>
<td>0.2</td>
<td>0.75</td>
</tr>
<tr>
<td>Discourse Cloze N-1</td>
<td>0.17</td>
<td>0.13</td>
<td>0.21</td>
</tr>
<tr>
<td>Discourse Cloze N+1</td>
<td>-0.1</td>
<td>0.12</td>
<td>0.44</td>
</tr>
<tr>
<td>Concreteness</td>
<td>-0.02</td>
<td>0.12</td>
<td>0.86</td>
</tr>
<tr>
<td>Aucustic Length</td>
<td>0.09</td>
<td>0.16</td>
<td>0.57</td>
</tr>
<tr>
<td>Location In Sentence</td>
<td>0.06</td>
<td>0.12</td>
<td>0.63</td>
</tr>
<tr>
<td>Sentence In Discourse</td>
<td>0.38</td>
<td>0.13</td>
<td>0</td>
</tr>
<tr>
<td>Discourse Cloze N-1 x Age Group</td>
<td>-0.27</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>Sentence In Discourse x Age Group</td>
<td>0.21</td>
<td>0.12</td>
<td>0.08</td>
</tr>
<tr>
<td><strong>Model 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>0.24</td>
<td>0.22</td>
<td>0.29</td>
</tr>
<tr>
<td>Log Frequency x Age Group</td>
<td>-0.25</td>
<td>0.11</td>
<td>0.03</td>
</tr>
<tr>
<td>Log Frequency</td>
<td>0.3</td>
<td>0.15</td>
<td>0.04</td>
</tr>
<tr>
<td>Age Group</td>
<td>-0.03</td>
<td>0.22</td>
<td>0.87</td>
</tr>
<tr>
<td>Log Frequency N-1</td>
<td>-0.35</td>
<td>0.16</td>
<td>0.03</td>
</tr>
<tr>
<td>Log Frequency N+1</td>
<td>-0.02</td>
<td>0.16</td>
<td>0.91</td>
</tr>
<tr>
<td>LSA N-1</td>
<td>0.24</td>
<td>0.17</td>
<td>0.16</td>
</tr>
<tr>
<td>LSA N+1</td>
<td>-0.06</td>
<td>0.2</td>
<td>0.75</td>
</tr>
<tr>
<td>Discourse Cloze N-1</td>
<td>0.16</td>
<td>0.13</td>
<td>0.21</td>
</tr>
<tr>
<td>Discourse Cloze N+1</td>
<td>-0.1</td>
<td>0.12</td>
<td>0.44</td>
</tr>
<tr>
<td>Concreteness</td>
<td>-0.02</td>
<td>0.12</td>
<td>0.86</td>
</tr>
<tr>
<td>Aucustic Length</td>
<td>0.09</td>
<td>0.16</td>
<td>0.57</td>
</tr>
<tr>
<td>Location In Sentence</td>
<td>0.06</td>
<td>0.12</td>
<td>0.63</td>
</tr>
<tr>
<td>Sentence In Discourse</td>
<td>0.37</td>
<td>0.13</td>
<td>0</td>
</tr>
<tr>
<td>Discourse Cloze N-1 x Age Group</td>
<td>-0.26</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>Sentence In Discourse x Age Group</td>
<td>0.23</td>
<td>0.12</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Table A3 (Continued):

<table>
<thead>
<tr>
<th>Model 3</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Intercept)</td>
<td>0.24</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>LSA</td>
<td>-0.03</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td><em>Log Frequency x Age Group</em></td>
<td><strong>-0.25</strong></td>
<td><strong>0.11</strong></td>
</tr>
<tr>
<td></td>
<td><em>Log Frequency</em></td>
<td><strong>0.31</strong></td>
<td><strong>0.15</strong></td>
</tr>
<tr>
<td></td>
<td>Age Group</td>
<td>-0.03</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td><em>Log Frequency N-1</em></td>
<td><strong>-0.36</strong></td>
<td><strong>0.16</strong></td>
</tr>
<tr>
<td></td>
<td>Log Frequency N+1</td>
<td>-0.02</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>LSA N-1</td>
<td>0.25</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>LSA N+1</td>
<td>-0.06</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>Discourse Cloze N-1</td>
<td>0.16</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Discourse Cloze N+1</td>
<td>-0.1</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Concreteness</td>
<td>-0.02</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Aucustic Length</td>
<td>0.09</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>Location In Sentence</td>
<td>0.06</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td><em>Sentence In Discourse</em></td>
<td><strong>0.37</strong></td>
<td><strong>0.13</strong></td>
</tr>
<tr>
<td></td>
<td>Discourse Cloze N-1 x Age Group</td>
<td><strong>-0.26</strong></td>
<td><strong>0.11</strong></td>
</tr>
<tr>
<td></td>
<td><em>Sentence In Discourse x Age Group</em></td>
<td><strong>0.23</strong></td>
<td><strong>0.12</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 4</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Intercept)</td>
<td>0.24</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>LSA x Age Group</td>
<td>0.04</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>LSA</td>
<td>-0.03</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td><em>Log Frequency x Age Group</em></td>
<td><strong>-0.27</strong></td>
<td><strong>0.13</strong></td>
</tr>
<tr>
<td></td>
<td><em>Log Frequency</em></td>
<td><strong>0.31</strong></td>
<td><strong>0.16</strong></td>
</tr>
<tr>
<td></td>
<td>Age Group</td>
<td>-0.04</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td><em>Log Frequency N-1</em></td>
<td><strong>-0.36</strong></td>
<td><strong>0.16</strong></td>
</tr>
<tr>
<td></td>
<td>Log Frequency N+1</td>
<td>-0.02</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>LSA N-1</td>
<td>0.25</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>LSA N+1</td>
<td>-0.06</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>Discourse Cloze N-1</td>
<td>0.16</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Discourse Cloze N+1</td>
<td>-0.1</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Concreteness</td>
<td>-0.02</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Aucustic Length</td>
<td>0.09</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>Location In Sentence</td>
<td>0.06</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td><em>Sentence In Discourse</em></td>
<td><strong>0.37</strong></td>
<td><strong>0.13</strong></td>
</tr>
<tr>
<td></td>
<td>Discourse Cloze N-1 x Age Group</td>
<td><strong>-0.26</strong></td>
<td><strong>0.11</strong></td>
</tr>
<tr>
<td></td>
<td><em>Sentence In Discourse x Age Group</em></td>
<td><strong>0.23</strong></td>
<td><strong>0.12</strong></td>
</tr>
</tbody>
</table>
Table A3 (Continued):

<table>
<thead>
<tr>
<th>Model 5</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.24</td>
<td>0.22</td>
<td>0.28</td>
</tr>
<tr>
<td>Discourse Cloze</td>
<td>0.36</td>
<td>0.13</td>
<td>0.01</td>
</tr>
<tr>
<td>LSA x Age Group</td>
<td>0.04</td>
<td>0.14</td>
<td>0.78</td>
</tr>
<tr>
<td>LSA</td>
<td>-0.12</td>
<td>0.18</td>
<td>0.53</td>
</tr>
<tr>
<td>Log Frequency x Age Group</td>
<td>-0.27</td>
<td>0.13</td>
<td>0.03</td>
</tr>
<tr>
<td>Log Frequency</td>
<td>0.26</td>
<td>0.16</td>
<td>0.09</td>
</tr>
<tr>
<td>Age Group</td>
<td>-0.04</td>
<td>0.22</td>
<td>0.86</td>
</tr>
<tr>
<td>Log Frequency N-1</td>
<td>-0.32</td>
<td>0.16</td>
<td>0.05</td>
</tr>
<tr>
<td>Log Frequency N+1</td>
<td>0</td>
<td>0.16</td>
<td>0.99</td>
</tr>
<tr>
<td>LSA N-1</td>
<td>0.28</td>
<td>0.18</td>
<td>0.12</td>
</tr>
<tr>
<td>LSA N+1</td>
<td>-0.1</td>
<td>0.21</td>
<td>0.62</td>
</tr>
<tr>
<td>Discourse Cloze N-1</td>
<td>0.13</td>
<td>0.13</td>
<td>0.31</td>
</tr>
<tr>
<td>Discourse Cloze N+1</td>
<td>-0.11</td>
<td>0.12</td>
<td>0.4</td>
</tr>
<tr>
<td>Concreteness</td>
<td>-0.1</td>
<td>0.12</td>
<td>0.41</td>
</tr>
<tr>
<td>Acoustic Length</td>
<td>0.08</td>
<td>0.16</td>
<td>0.63</td>
</tr>
<tr>
<td>Location In Sentence</td>
<td>0.05</td>
<td>0.12</td>
<td>0.66</td>
</tr>
<tr>
<td>Sentence In Discourse</td>
<td>0.34</td>
<td>0.13</td>
<td>0.01</td>
</tr>
<tr>
<td>Discourse Cloze N-1 x Age Group</td>
<td>-0.26</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>Sentence In Discourse x Age Group</td>
<td>0.23</td>
<td>0.12</td>
<td>0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 6</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.24</td>
<td>0.22</td>
<td>0.28</td>
</tr>
<tr>
<td>Discourse Cloze x Age Group</td>
<td>0.04</td>
<td>0.12</td>
<td>0.71</td>
</tr>
<tr>
<td>Discourse Cloze</td>
<td>0.35</td>
<td>0.13</td>
<td>0.01</td>
</tr>
<tr>
<td>LSA x Age Group</td>
<td>0.03</td>
<td>0.14</td>
<td>0.84</td>
</tr>
<tr>
<td>LSA</td>
<td>-0.11</td>
<td>0.18</td>
<td>0.54</td>
</tr>
<tr>
<td>Log Frequency x Age Group</td>
<td>-0.27</td>
<td>0.13</td>
<td>0.03</td>
</tr>
<tr>
<td>Log Frequency</td>
<td>0.27</td>
<td>0.16</td>
<td>0.09</td>
</tr>
<tr>
<td>Age Group</td>
<td>-0.04</td>
<td>0.22</td>
<td>0.86</td>
</tr>
<tr>
<td>Log Frequency N-1</td>
<td>-0.32</td>
<td>0.16</td>
<td>0.05</td>
</tr>
<tr>
<td>Log Frequency N+1</td>
<td>0</td>
<td>0.16</td>
<td>0.99</td>
</tr>
<tr>
<td>LSA N-1</td>
<td>0.28</td>
<td>0.18</td>
<td>0.12</td>
</tr>
<tr>
<td>LSA N+1</td>
<td>-0.1</td>
<td>0.21</td>
<td>0.62</td>
</tr>
<tr>
<td>Discourse Cloze N-1</td>
<td>0.13</td>
<td>0.13</td>
<td>0.31</td>
</tr>
<tr>
<td>Discourse Cloze N+1</td>
<td>-0.11</td>
<td>0.12</td>
<td>0.4</td>
</tr>
<tr>
<td>Concreteness</td>
<td>-0.1</td>
<td>0.12</td>
<td>0.41</td>
</tr>
<tr>
<td>Acoustic Length</td>
<td>0.08</td>
<td>0.16</td>
<td>0.63</td>
</tr>
</tbody>
</table>
Table A3 (Continued):

<table>
<thead>
<tr>
<th></th>
<th>0.05</th>
<th>0.12</th>
<th>0.66</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Location In Sentence</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sentence In Discourse</strong></td>
<td>0.34</td>
<td>0.13</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Discourse Cloze N-1 x Age Group</strong></td>
<td>-0.27</td>
<td>0.12</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Sentence In Discourse x Age Group</strong></td>
<td>0.23</td>
<td>0.12</td>
<td>0.06</td>
</tr>
</tbody>
</table>