

AN INVESTIGATION OF HOW MULTIPLE SOURCES OF INFORMATION ARE
INTEGRATED DURING ONLINE READING

BY

MALLORY C. STITES

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Kara D. Federmeier, Chair
Associate Professor Kiel Christianson
Professor Gary Dell
Professor Susan Garnsey
Professor Elizabeth A. L. Stine-Morrow

Abstract

The current project investigates the online processing consequences of a conflict between different levels of linguistic representation, specifically focusing on the relationship between the mechanisms supporting word recognition and the information conveyed at the sentence level. The first set of studies asks how readers comprehend compound words containing transposed letters that are swapped between the word's morphemes or that stay within a single morpheme, when these words are encountered in a meaningful sentence context. Will readers weight word-level information more heavily (i.e., letter order, morpheme boundaries), or will the general semantic support of the sentence context allow readers to overcome this type of disruption with relatively small costs? The second set of studies investigates how effectively readers can use syntactic cues to resolve the ambiguity associated with noun/verb homographs, the downstream consequences of their resolution processes, and how these effects differ in older adults. Will readers be able to use constraining syntactic information to select the context-appropriate meaning of the word, especially if the intended meaning occurs less frequently? Furthermore, will the effects observed on the target word itself predict the later availability of the word's context-inappropriate meaning, and does this relationship between early and late effects change with advancing age? The current study addresses these questions through the use two temporally sensitive online measures of language processing, eye-tracking and event-related brain potentials (ERPs) collected during natural reading, as a way to observe how readers integrate and adjudicate between different sources of information as online language comprehension unfolds.

Acknowledgements

The road from starting graduate school to completing the dissertation was a challenging one, and I was lucky to receive help and guidance from so many generous people along the way. I owe many thanks to Dr. Kara Federmeier for her guidance and support throughout my entire graduate career. Kara, I was so blessed to have been welcomed into your lab with open arms, even though I didn't know what an ERP was when I first applied. Without your outstanding training and endless encouragement, I would not be where I am today. I have truly enjoyed our time together, and consider you to not only be the best mentor I ever could have asked for, but also a lifelong friend. Thank you also to Dr. Kiel Christianson and Dr. Elizabeth Stine-Morrow, who allowed me the privilege of collaborating with them from my very first year. I cannot express how lucky I was to have not one, but three amazing mentors during my grad school career. I look forward to many more years of collaborations with both of you. To Dr. Susan Garnsey and Dr. Gary Dell, thank you for serving on my dissertation committee, and for all of your input on my work throughout my time at Illinois. It has been an honor to receive my training from such an elite group of researchers—I can only hope that I will live up to the high standard of excellence you have established.

I want to thank everyone in the Cognition and Brain Lab, both past and present, for providing such a supportive work environment, and for being there for me in ways big and small throughout my time in the lab. In particular, I have been generously helped by: Sarah Laszlo, Chia-lin Lee, Hsu-wen Huang, Edward Wlotko, Charlotte Laguna, Danielle Dickson, and Michelle Leckey. I will miss so much about the lab—especially our "surprise" birthday parties and karaoke nights. I also want to thank the members of The Adult Learning Lab (especially Brennan Payne and Xuefei Gao), as well as the members of the Educational Psychology Psycholinguistics Lab (especially Steven Luke and Cassie Palmer), who all made parts of this work possible.

To my family, especially my Mom and Dad, thank you for always encouraging me to believe in myself. I was the only kindergartener I knew with a college fund, and I am still chasing that dream you instilled in me so long ago. Zoz, thank you for never letting me settle for anything less than my absolute best—I wish you could be here to see this today. This journey was not an easy one, but knowing that my family and friends were cheering me on made it all the more worthwhile.

Finally, a million thanks go to my husband, Matt. It's hard to believe how lucky I was that grad school brought me to you. Thank you for sticking with me over the past five years, for letting me lean on you in the dark moments, and for always pushing me to be a better version of myself. You remind me every day what's really important in life. I love you, and I'm so excited to see what adventures await us.

Table of Contents

Introduction.....	1
Part 1: Transposed Letter Effects in Eye-tracking.....	8
Part 2: Transposed Letter Effects in ERPs.....	32
Part 3: Young Adult Ambiguity Effects in Eye-tracking.....	69
Part 4: Older Adult Ambiguity Effects in Eye-tracking	106
General Discussion	139
References.....	142

INTRODUCTION

At the broadest level, my research seeks to understand how people dynamically bring together varying sources of information to achieve comprehension, especially during skilled reading. In order to understand a text, readers have to actively integrate their prior knowledge with incoming information in real time to create a coherent meaning. Readers must recognize letters, use these letters to recognize the word, activate the semantics of that word, and then integrate it into the ongoing message-level meaning they have been constructing throughout reading the sentence and/or discourse. All of this takes place within roughly half a second after first apprehending the word, and repeats continuously as readers move to new words in the text. Even under the best of circumstances, readers have several ongoing streams of information to synchronize and integrate smoothly to keep comprehension flowing without a hitch. However, this situation is very rarely the case. Far more often than we realize, something in the input conflicts with part of our ongoing representations, prior knowledge, and predictions about the upcoming context, and we have to quickly and flexibly reconcile this conflict in order to create a coherent meaning. My research asks how readers do this—how we integrate context and prior knowledge to achieve comprehension, particularly in the face of conflicting information.

The integration of context, knowledge and predictions during online language comprehension requires the orchestration of processing at a number of different levels—word, sentence, discourse—which differs from how the study of these topics is typically conducted. Language researchers tend to investigate only a single level of processing at any given time, with the aim of characterizing all of the factors that affect that single level of processing. For example, research on word recognition tends to use experimental paradigms in which single words are presented as part of unconnected strings of text. These paradigms often require the

implementation of a secondary task (e.g., lexical decision, speeded naming, category judgment) as a way to measure the impact of different lexical variables on that output measure. This observed difference is then interpreted as the proxy for the effect that word characteristic has on processing in general. However, the secondary tasks required by these types of paradigms introduce additional demands on the reader that are not present during typical comprehension. The extra decision processes that readers must engage likely draw attentional resources away from their ability to fluently recognize the words, and instead encourage them to adopt strategies to achieve the task demands of the experimental design that would not normally be present during reading comprehension. As a result, findings from these studies may overstate the importance of certain lexical characteristics to word recognition, while missing more subtle effects of others that may not directly affect decision-making.

Because of the vastly different experimental paradigms and stimulus materials used to study the different levels of linguistic representation, it is difficult to draw coherent conclusions about the *interaction* of processes across those levels. The current study aims to transcend these barriers in order ask how the mechanisms underlying word recognition are carried out under more naturalistic circumstances, and furthermore, how they interact with higher levels of processing. Our main experimental task—reading sentences for comprehension—in combination with the collection of online measures of processing will allow readers to engage with the text normally, without the need for a secondary task. This will provide the ideal environment in which to study the processing consequences that arise in the face of conflict between multiple levels of representation while using a task that closely matches how reading proceeds outside of the lab.

In the first set of studies, we will explore the effect that misspellings have on readers' ability to effectively comprehend those words. Misspelled words, particularly those containing a pair of transposed letters, have been extensively studied, and the costs associated of transposed letter manipulations on word recognition are well characterized. It is important to note that these conclusions rely heavily on findings from studies in which transposed letter nonwords appear in isolation, or as briefly presented primes preceding correctly spelled words. The subject's task in these studies is often to classify the misspelled word as nonword, or to pronounce it aloud as fast as possible. This stands in stark contrast to a reader's "task" when encountering a misspelled word in the real world, for example in an e-mail, when readers are likely highly motivated to figure out the intended word as quickly as possible. In these situations, readers often have the benefit of a sentence context to aid in their comprehension, which may even cause them to fail to notice misspelled word at all. How does this happen? The first set of experiments will fill in gaps in the literature on how transposed letter manipulations affect meaning comprehension when those words appear in supportive sentence contexts. This will allow us to move beyond the measurement of simple reaction time measures, and instead ask questions about the content of the representations that readers construct when they encounter these types of words. The first part of my dissertation will attempt to answer the question of how readers overcome the conflict introduced when they read a compound word that is misspelled, particularly focusing on the differential effects of letter transpositions that appear either within a single morpheme or that cross the boundary between the word's morphemes. This particular question has the power to investigate the interaction of four different levels of processing—specifically, how manipulations at the *letter* level that disrupt *morpheme* boundaries impact recognition of the *word*, and whether these effects play out differently when targets are embedded in a *sentence* context as opposed

being viewed in isolation. This will give us insights into how the various levels of linguistic representation interact during word recognition, and more generally, come together to achieve comprehension during sentence processing.

Assuming that all of the letters of a word are present and in the right place, readers face an additional, daunting obstacle: a large number of the words in the English language have multiple meanings associated with their particular spelling. The study of lexical ambiguity has also been conducted largely through the presentation of ambiguous words in isolation, or in impoverished prime-target pairings, in which the automatic semantic activation of multiple meanings is examined in the absence of a sentence context. Even those studies that have examined ambiguous words in a sentence context tend to focus their lens mainly on reading time measures on the target word itself. However, most studies of ambiguity resolution stop there, without paying attention to what happens *after* readers have seemingly “completed” ambiguity resolution. This leads to the open question of whether ambiguous words behave like unambiguous words when they appear in constraining contexts—or is there ever evidence for residual activation of the word’s context-inappropriate meaning after readers have moved on in the text?

The second set of studies addresses these questions by examining how readers utilize context to select the correct meaning of an ambiguous word using only syntactic cues, and how their later comprehension is affected by the outcome of their meaning selection. We direct our focus on noun/verb (NV) homographs with two semantically unrelated meanings that fall into distinct grammatical classes. This feature allows us to isolate the separate contributions of sentence meaning (semantics) and structure (syntax) on the disambiguation of these words, two factors that are often confounded in previous research. Their independent manipulation in the

present set of studies will allow us to ask several novel questions about the interaction of lexical and sentential characteristics during reading comprehension. We ask whether the syntactically distinct nature of the homograph's two senses make it possible to select the right (context-appropriate) meaning without the presence biasing semantics? Previous research suggests that encountering NV homographs in syntax-only contexts engenders costs in both ERPs (Lee & Federmeier, 2009; 2011; 2012) and eye-tracking (Stites, Federmeier, & Stine-Morrow, 2013). However, the functional role of these costs has been understudied, particularly in the eye-tracking domain. Specifically, we ask, does the presence of an ambiguity effect predict that the word's context-inappropriate meaning will be unavailable later in the sentence? This study seeks to discover what types of representations readers create after the elicitation of these ambiguity costs in natural reading, and furthermore, how readers use these representations to constrain their later interpretation of the sentence. This study will be uniquely able to investigate the interactive relationships between levels of processing—namely, how syntax aids lexical ambiguity resolution, and in turn, how the outcomes of that resolution affect downstream comprehension.

More complicated still are cases in which readers lack the cognitive resources necessary to help them choose the right meaning, even when given the proper cues—how does this impact their later understanding of the text? Can they eventually recover if they move forward in the text with the wrong meaning? If they are given the opportunity to look back to earlier parts of the sentence, do they choose to do so, and is doing so helpful? To answer these questions, we will examine ambiguity resolution processes in both college-aged students as well as older adults, who have been shown to be less effective at using syntactic cues to select the meaning of an ambiguous word, particularly in the case of noun/verb homographs. Furthermore, we will explore the consequences that their ability (or inability) to choose the correct meaning has for

their downstream processing after they have continued in the text and must use the outcome of their meaning selection process to interpret later parts of the sentence.

I have approached the above research questions by using two different methodologies to study reading—event-related potentials (ERPs) to investigate electrophysiological markers of processing, and eye-tracking to monitor readers' eye movements during natural reading. Both of these methods provide precise information about the timing with which comprehension processes unfold, and I believe their complementary nature has been largely under-utilized. ERPs shine in their ability to provide a detailed picture about the serial intake of words. Their multidimensionality highlights which type of information is available in what time-scale (i.e., semantic, grammatical), making them an especially useful tool for pinpointing when conflict first arises between multiple information streams and which subprocesses of comprehension it affects. Eye-tracking offers readers naturalistic control over the timing and order of their information intake during reading, allowing us to ask how they strategically allocate resources across text when given that control. It is clear that each of these methods has its own strengths and weakness, but both are powerful tools in the study of online language comprehension. Given that my aim is to achieve the fullest understanding of the processes underlying visual word recognition and meaning comprehension in skilled reading, I believe the best way to do so is to employ both methods separately under circumstances that are as well-matched as possible, and then compare outcomes between studies. I have taken this approach in the series of studies presented here, which I believe allows me to demonstrate both the compatibility of findings across measures that can be uncovered when experimental conditions are closely aligned, as well as the complementary way each measure can help illuminate questions left unanswered by the other. The ultimate goal of this project is to provide a fuller understanding of the processes that

produce comprehension in skilled reading, at both the neural and behavioral level, and how these processes interact across multiple levels of representation.

Part 1: Transposed Letter Effects in Eye-tracking

A large number of studies have shown time and again that readers can correctly recognize words with a pair of adjacent letters swapped to incorrect locations, an effect known as the transposed letter similarity effect. A variety of studies have shown that a prime with two adjacent letters transposed provides more facilitation to its base word than does a prime with those same two letters replaced with visually similar letters (Perea & Lupker, 2003). This effect has also been achieved using primes with non-adjacent transposed letters with both one intervening letter (Perea & Lupker, 2004), and to a lesser extent, two intervening letters (Perea, Duñabeitia, & Carrieras, 2009), as well as primes containing multiple pairs of transposed letters (Guerrera & Forster, 2008). Of course, words are not just strings of graphemes, but consist of hierarchically arranged morphemes. More controversial is the question of how the transposed letter effect interacts with morpheme boundaries during visual word recognition. The current study will investigate whether transposed letters that cross morpheme boundaries in compound words cause more disruption to reading times during silent reading than transpositions that are confined to a single morpheme. This will help answer the larger theoretical question of whether compound words are recognized via decomposition into their constituent morphemes, which would predict greater disruption to reading times from the between-morpheme letter transpositions than from the within-morpheme condition. If, on the other hand, compound words are recognized as a single wordform, without being decomposed into its constituents, then we would predict equal disruption from these two misspelling conditions. We are the first to ask this question using an eye-tracking paradigm, which will allow us to expand our current understanding of the transposed letter similarity effect to include how well readers can

comprehend misspellings that compromise a word's morphemic boundaries when presented in a sentence context.

Transposed Letters and Morpheme Boundaries

Over the past 30 years, much research has focused on how morpheme recognition affects whole word recognition (see Diependaele, Grainger, & Sandra, 2012, for a review). The general picture that has emerged is that morpheme recognition plays a role in word recognition, although the exact time course of its influence is still up for debate. Work by Rastle, Davis and New (2004) has suggested that morphemes may play a role in the earliest stages of word recognition. In a masked-priming paradigm, lexical decision times were facilitated by morphologically transparent primes (*cleaner-CLEAN*), as well as pseudo-morphological primes, in which the word ends in a letter string constitutes a real suffix that does not function as such in that particular word (*corner-CORN*). The same priming is not found in non-morphological primes, which ends in a letter string that is not a suffix in the English language (*brothel-BROTH*), despite sharing the same number of overlapping letters between prime and target. From these, and subsequent, findings, it is believed that readers engage in “affix stripping,” a process in which they extract affixes from derived words before they have been fully recognized, allowing the remaining stem to prime the target. Many follow-up studies have found similar results (for a review, see Rastle & Davis, 2008), which have collectively been taken as evidence for prelexical morphological decomposition of affixed words.

Transposed Letter (TL) Similarity Effects from Masked Priming

The transposed letter similarity effect provides an interesting way to better understand the time course of morphological decomposition. If morphological decomposition occurs in the earliest stages of word recognition, then letter transpositions that cross morpheme boundaries

should disrupt this decomposition process and slow comprehension to a greater degree than transpositions that stay within one morpheme. Christianson, Johnson and Rayner (2005) tested this hypothesis using English compound words. They found that naming latencies for compounds were faster when primed by an identical (*sunshine*) or within-morpheme transposition prime (*sunhsine*), relative to those primed by an across-morpheme transposition (*susnhine*) or a single letter replacement prime (*sunsbine*). Christianson et al. (2005) also tested the effect in words with the derivational suffix *-er* (*boaster*), in which across-morpheme transpositions (*boasetr*) did not prime the base words relative to orthographic controls containing a replacement letter (RL) (*boasler*), whereas pseudoderived words like *bluster* benefited equally from a letter transposition prime in the same location (*blusetr*) and an identical prime. In both cases, the benefit of the transposed letter similarity effect was diminished or eliminated when the letters crossed a morpheme boundary. In a study conducted using Basque and Spanish words, Duñabeitia, Perea and Carreiras (2007) also failed to find priming for affixed words when the letter transposition crossed a morpheme boundary in affixed words, although the effect was present for non-affixed words. Finally, in the first study examining cross-morpheme TL effects between stem and inflectional affixes, Luke and Christianson (2013) found that neither stem nor whole-word frequencies interacted with the earliest measures of cross-morpheme TL disruption during the reading of connected text. The reduction or lack of TL priming across morpheme boundaries in these three studies has been taken as evidence for early morphological decomposition in visual word recognition.

However, other research using the masked priming paradigm has complicated this straightforward conclusion, with many studies finding equally-sized TL priming effects within and across morpheme boundaries. For instance, using Basque compound words, Perea and

Carrieras (2006) found significant TL versus RL priming across the morpheme boundary. Rueckl and Rimzhim (2011) used primes containing the derivational suffix *-er*, with transposed or replacement letters on the final letters of the stem (TL: *teah**ce**er-TEACH*, RL: *teak**s**er-TEACH*), or across the stem-affix boundary (TL: *teace**h**r-TEACH*, RL: *teard**i**r-TEACH*). They found significant priming for both within- and across-morpheme TL primes relative to the replacement letter control. Similarly, Beyersmann, McCormick, and Rastle (2013) employed the derivational suffix *-er*, but with stem-internal rather than stem-final transpositions (i.e., *bruner-BURN*), and again found equal within- and between-morpheme TL priming relative to RL controls. Finally, Sánchez-Gutiérrez & Rastle (2013) used affixed Spanish-English cognates, and found equal TL priming from stem-internal and across-morpheme letter transpositions in both sets of words.

In addition to the lexical decision task, Duñabeitia, Kinoshita, and Norris (2011) investigated transposed letter priming effects in a cross-case same-different task using the stimuli from Duñabeitia et al. (2007). In this task, subjects decided whether the initially presented reference letter string matched the subsequent target string presented in a different case, which was primed with transposed or replacement letters within the stem or across the stem-affix boundary. “Same” responses were equally facilitated by within- and across-morpheme transpositions relative to replacement letters. It is important to note that the same-different task might not be sensitive to the same lexical characteristics important for word recognition, so these effects should be interpreted cautiously. Together, though, this group of studies seems to strongly suggest that transposed letter priming is not reduced when the two transposed letters cross morpheme boundaries.

Finally, recent work by Taft and Nillsen (2013) argues that decrements in cross-morpheme TL effects, or the lack thereof, might not constitute decisive evidence either for or

against early morphological decomposition. Letter position coding theories exist in which letter positions are imprecisely coded (e.g., Gómez, Ratcliff, & Perea, 2008), and letter information is activated more strongly when the letter is aligned with its expected slot, but still receives some activation when encountered in an adjacent slot. Under this view, adjacent letters at the end of a stem and the beginning of a suffix, for example, would both still receive some activation. In certain circumstances, this activation might be enough to allow for relatively unhindered word recognition, but in others, it might not be. As such, argue Taft and Nillsen, the presence or absence of cross-morpheme TL effects could be determined by other aspects of the experimental materials that affect letter position coding rather than morphological decomposition per se.

Factors such as the task being performed and the predictability and frequency profile of the target word are two possible influences on the precision with which letters of words are mapped onto positional codes within words. This is especially true when considering visual word recognition within the context of normal reading, which almost none of the visual word recognition studies cited above has done (but cf. Luke & Christianson, 2013). For example, it has recently been shown that the TL similarity effect can be nullified even in monomorphemic words that are highly predictable from context (Luke & Christianson, 2013). Similarly, Luke and Christianson (under review) find that transpositions across stem-affix boundaries in inflected verbs are disruptive to normal reading when the affixed form of the verb is strongly predicted by the syntax, but not if weakly predicted by syntax or semantics. Furthermore, if an affixed verb is strongly predicted, even letter transpositions in unpredicted non-affixed verbs become disruptive, as the parser appears to actively search out the locations of highly probable morphological boundaries.

In sum, previous research has used the TL similarity effect to derive inferences about the time course of morphological decomposition during visual word recognition. The vast majority of this work has measured recognition latencies for individual, decontextualized words in various versions of the masked-priming lexical decision or naming tasks. Only recently have TL similarity effects both within and across morphemes been investigated using more naturalistic reading tasks, including the self-paced masked prime paradigm (SPaM; Luke & Christianson, 2012, 2013a,b) and eye-tracking (Luke & Christianson, under review; White, Johnson, Liversedge, & Rayner, 2008). Valuable insights into the time course of morphological processing and factors that influence it have been gained from the use of these methodologies, which are particularly relevant when multimorphemic words are encountered in connected texts, rather than in word lists.

TL Effects in Natural Reading

As noted above, a major drawback to most previous studies is that they presented words in isolation, removing them from the semantically and syntactically rich sentence contexts in which they typically appear. Further exceptions to this generalization include studies of the transposed letter effect using the boundary change paradigm in eye-tracking. This paradigm takes advantage of the fact that readers extract letter identity information from the next word in the sentence when it is visible in the parafovea (for review, see Schotter, Angele, & Rayner, 2012). Parafoveal information can be thought of as an analogue to a masked prime, in that it is received just before readers fixate the target itself, and it is degraded relative to foveal processing. The preview readers receive of the upcoming word is manipulated, and the critical outcome measure is the preview benefit, or the difference between reading times on the target word when readers initially had a valid preview of it relative to reading times following an

invalid preview. Johnson and colleagues have shown that parafoveal previews containing two internally transposed letters provide a greater preview benefit to target reading times than do previews with visually similar replacement letters (Johnson, Perea, & Rayner, 2007; Johnson & Dunne, 2012). Masserang and Pollatsek (2012) extended this finding by presenting readers with previews containing transposed or replacement letters across a prefix-stem boundary (and in a similar location in non-prefixed words). Their results showed a significant preview benefit for TL versus RL previews regardless of its location relative to morpheme boundaries, providing the first evidence that the TL effect may not be sensitive to morpheme boundaries in natural reading.

Other researchers have simply presented readers with transposed-letter nonwords for full viewing during natural reading, to see how readers cope with overt spelling violations. Unsurprisingly, Rayner, White, Johnson, and Liversedge (2006) found longer sentence reading times for sentences containing TL nonwords, as well as longer average fixation durations, a greater number of fixations and more regressive saccades. In a follow-up study, White et al. (2008) found that word-external transpositions (i.e., involving the word's first two letters, or last two letters) elicited longer reading times than word-internal transpositions, with the most disruption caused by word-initial transpositions (see also Johnson & Eisler, 2012, for more evidence as to the importance of word-external letters). However, in all three of these studies, every word in the sentence five letters or longer contained a letter transposition. It could thus be the case that reading patterns might be altered when readers face an onslaught of misspelled words within a single sentence as opposed to an isolated misspelled word. Furthermore, these studies did not manipulate the placement of transposed letters relative to morpheme boundaries, so the impact of this particular manipulation in natural reading remains an open question.

The Current Study

The current study will attempt to fill several gaps in the literature as to the sensitivity of the transposed letter effects to morpheme boundaries in English compound words during natural reading. While many studies have looked at transposed letter priming in lexical decision or naming tasks, these tasks differ in important ways from natural reading, specifically by requiring an overt response be made to the target word that is not typically part of the reading process. The use of eye-tracking will allow us to understand how readers handle misspelled words under very typical circumstances—while reading a sentence in which the compound word makes sense but is not predictable—which will enable us to draw conclusions about how readers extract meaning from the misspelled compound words in the absence of a secondary task. The only other study to have examined the cost of across-morpheme transpositions in natural reading used the boundary change paradigm (Masserang & Pollatsek, 2012), meaning that subjects never had the opportunity to fully view the misspelled words. It is possible that effects will play out differently when readers are trying to comprehend the transposed letter nonwords rather than just respond to them. This study will also serve as a partial replication of Christianson et al. (2005), in an attempt to better understand under what circumstances transposed letter similarity effects are reduced or not across morpheme boundaries in compound words.

Building on the manipulation used by Christianson et al. (2005), the current study will present compound words with letter transpositions either within one morpheme (i.e., *cupacke* for *cupcake*) or across the morpheme boundary (i.e., *cucpake* for *cupcake*). These nonwords will then be embedded in unconstraining sentences that subjects will read for comprehension while their eye movements are monitored. This will allow us to test for evidence of early morphological decomposition of compound words in a sentence context. Based on previous studies using transposed letter nonwords under full viewing conditions (Johnson & Eisler, 2012;

Rayner et al., 2006; White et al., 2008), we expect to find longer reading times on the transposed letter nonwords than on correctly spelled words. If the extraction of component parts from compounds affects initial visual word recognition, as suggested by the findings from Christianson et al. (2005), we would predict greater disruption to reading times for across- versus within-morpheme transpositions, because these would interfere with the process of recognizing the individual morphemes. On the other hand, if reading times are equally long for the across- and within-morpheme transpositions, this would be evidence that morphological decomposition of English compounds does not affect visual word recognition during normal silent reading of these words in relatively unconstraining contexts.

Two other transposition conditions will also be included to test the severity of the across-morpheme disruption. Following White et al. (2008), we will include word-initial and word-final transpositions, both of which have been shown to cause greater disruption to reading than word-internal transpositions. White et al. (2008) did not, however, manipulate the relationship between transposition location and morpheme boundaries. By including these external transpositions along with the critical within- and between-morpheme transpositions, we can compare the size of the disruption caused by word-external with morpheme-external (but word-internal) transpositions, to see if letters at morpheme boundaries behave more like word-external or word-internal letters. Furthermore, each of our critical sentences will contain only one transposed letter nonword, rather than many misspellings as in White et al. (2008), allowing us to more clearly understand the effect of a single misspelling within a sentence. This is important because previous research has examined the effects of transposed letters on word recognition in isolation with mixed results. As such, by embedded these words in sentence contexts we can get a clearer

picture of the true costs of these types of spelling errors under more naturalistic circumstances, when readers are viewing the words for comprehension.

Method

Participants

Twenty-five undergraduates from the University of Illinois participated for class credit, or payment of \$7.

Apparatus

Eye movements were recorded via an SR Research Ltd. Eyelink 1000/2000 eye tracker, which records the position of the reader's eye once every millisecond (1000 Hz sampling rate), and has a high spatial resolution of 0.01°. Text was displayed in 12-point Courier New font. Participants were seated 69 cm away from a 20-inch monitor. At this distance, approximately 3.5 characters subtended 1° of visual angle. Head movements were minimized with chin and head rests. Although viewing was binocular, eye movements were recorded from the right eye.

Materials and Design

Target words consisted of 49 compound words, ranging in length from seven to nine letters (20 seven-letter words, 22 eight-letter words, and seven nine-letter words), with an average of 7.73 letters. Roughly half of the compounds (24) had two internal consonants (C-C; e.g., *cupcake*), 12 of which had a three-letter initial morpheme (e.g., *airplane*) and 12 of which had a four-letter initial morpheme (e.g., *raincoat*). The remaining 25 compounds had three internal consonants, of which 12 words had an initial morpheme that ended in a single consonant and a final morpheme that started with a two-consonant cluster (C-CC; e.g., *sunshine*). The other

13 words had the opposite pattern, with initial morphemes ending in a two-consonant cluster and final morphemes beginning with a single consonant (CC-C; e.g., *jackpot*). The average whole word frequency of the compound, taken from the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995) was 5.57 appearances per million (range 0-64). The written frequency of the compound word's first morpheme (average 106, range: 0-456) and second morpheme (average 86, range: 0-456) were matched on their written frequency, as confirmed by a paired-sample *t*-test, $t(48) = .94, p = .35$.

The compound words could appear in five different transposition conditions: word-initial transposition, between-morpheme transposition, within-morpheme transposition, word-final transposition, or no transposition (see Table 1). Letter transpositions always involved adjacent letters, and never produced real words. For word-initial transpositions, the first and second letters of the word were inverted (e.g., *ucpcake*), and word-final transpositions involved the last two letters of the word (e.g., *cupcaek*). The between-morpheme transpositions always involved the inversion of the last letter of the first morpheme with the first letter of the second morpheme (e.g., *cucpake*). Within-morpheme transpositions involved either the last two letters of the first morpheme (for CC-C words, and C-C words with four-letter initial morphemes), or the first two letters of the second morpheme (for C-CC words, and C-C words with three-letter initial morphemes). Thus, the within-morpheme transposition appeared equally in the first and second morpheme across the set of compound words, and always involved on either the third/fourth or fourth/fifth letters of the word.

Table 1. The five experimental conditions in which target words could appear. C-C = compound words consisting of initial morphemes ending in one consonant and final morphemes beginning with one consonant; CC-C = compound words consisting of initial morphemes ending with two consonants and final morphemes beginning with one consonant; C-CC = compound words with initial morphemes ending in one consonant and final morphemes beginning with two consonants.

Transposition	C-C	CC-C	C-CC
Identical control	cupcake	jackpot	sunshine
<i>External</i>			
Word-initial	ucpcake	ajckpot	usnshine
Word-final	cupcaek	jackpto	sunshien
<i>Internal</i>			
Between-morpheme	cucpake	jacpkot	susnhine
Within-morpheme	cupacke	jakcpot	sunhsine

The critical compound words were presented in sentence frames for which the target word had a very low cloze probability. To ensure that the target words were not predictable, a norming study was conducted, in which ten subjects (who did not participate in the eye-tracking portion) were presented with the sentence frame up to the critical word and were asked to complete the sentence. For forty-one of the sentences, no participant completed the sentence with the correct compound word. The eight sentences that at least one participant completed with the target compound word were rewritten, and were judged by three native English speakers to not be predictive of the target word. The sentences contained 12 words on average (range: 8-16 words). The critical compound word never appeared as the first or last word of the sentence (average: word 6; range: 3 –10). Every subject read all 49 critical sentences, with 9-10 appearing in each experimental context.

Procedure

After signing informed consent, each participant's eye movements were calibrated using a 9-point calibration screen. In the testing session, each trial involved the following sequence: trials began with a gaze trigger, consisting of a black circle presented in the position of the first

character of the text. Once a stable fixation had been detected on the gaze trigger, the sentence was presented in full. The participant pressed a button on a standard game controller to indicate that s/he had finished reading the sentence. At this point, the sentence disappeared. A question about the content of the sentence appeared after 16 of the 49 critical sentences, which participants answered by pressing the appropriate button on the controller. The question never referred to the critical compound word. After the comprehension question, the next trial began again with the gaze trigger. Sentences were presented in a random order for each participant. In addition to the 49 experimental trials, each list contained 105 other sentences with a variety of structures, and comprehension questions were asked after approximately 34 of these filler sentences as well.

Data Analysis

Within the Eyelink 1000 data analysis package, consecutive fixations shorter than 80 ms and less than 0.5° apart were merged into a single fixation, while other fixations shorter than 80 ms were deleted from analysis. In addition, fixations longer than 800 ms were also deleted from the analysis because these typically indicate track loss. Less than 3.6% of trials were excluded due to track loss or skipping of the critical compound word.

Results

Behavioral Results

Participants performed well on the comprehension questions, with an average accuracy of 85.6%. Because there were no a priori predictions about how misspellings might affect comprehension question accuracy, an omnibus repeated measures ANOVA was conducted on the accuracy data comparing all five conditions against one another. Results showed a marginal

effect of transposition condition, $F(4,96)=2.22$, $p=.07$. This marginal effect likely reflects the fact that comprehension scores ranged from 79% accurate (in the word-initial TL condition) to 93% accurate (in the word-final TL condition), with the other three conditions falling somewhere in the middle (control: 86%, within-morpheme TL: 88%, between-morpheme TL: 92%).

Eye-tracking Results

Reading times on the target compound word were analyzed as a function of the type of transposition the word contained. Two repeated-measures analyses of variance (ANOVAs) were conducted for each measure, considering as random effects both participants (F_1) and items (F_2). We examined four standard reading time measures: first fixation duration (the length of a reader's first fixation on the word), gaze duration (the sum of all fixations on a word before leaving it in either direction), go-past time or regression path duration (a right-bounded measure, including all fixations from the time the reader first lands on the word until they move past it to the right, including regressions back in the text), and total time (the sum of all fixations on a word). We first conducted an omnibus ANOVA to compare all conditions against one another, to test for an overall effect of letter transposition on reading times. We next tested for differences between the external transpositions (collapsing across word-initial and word-final) versus internal transpositions (collapsing across within- and between-morpheme), to see if morpheme-external, but word-internal, transpositions produce shorter reading times than word-external transpositions. Finally, conducted two comparisons between specific conditions: first comparing word-initial and word-final transpositions following White et al. (2008), and then examining the critical comparison of interest: the within- and across-morpheme transpositions.

Overall effect of letter transposition. We first conducted an omnibus one-way ANOVA with the factor of transposition condition, which had five levels (no transposition, word-initial, word-final, within-morpheme, and between-morpheme) to test for an overall effect of letter transposition on reading times (see Table 2 for reading times). For all four reading measures, there was a significant effect of condition (first fixation: $F_1(4,96)=6.84$, $MSE=5773.06$, $p<.001$, $F_2(4,172)=6.52$, $MSE=12394.65$, $p<.001$; gaze duration: $F_1(4,96)=27.25$, $MSE=155214.00$, $p<.001$, $F_2(4,172)=25.15$, $MSE=308853.34$, $p<.001$; go-past time: $F_1(4,96)=19.08$, $MSE=266500.18$, $p<.001$, $F_2(4,172)=19.41$, $MSE=523303.46$, $p<.001$; and total time: $F_1(4,96)=28.16$, $MSE=341162.39$, $p<.001$, $F_2(4,172)=23.11$, $MSE=663027.73$, $p<.001$). Follow-up t-tests comparing each transposition condition to the correctly-spelled control condition revealed that each was significantly longer than control for first fixation ($ts>2.20$, $ps<.05$), gaze duration ($ts>3.53$, $ps<.01$), and total time ($ts>2.16$, $ps<.05$). For go-past time, the within morpheme transposition condition was only marginally longer than control, $t_1(24)=1.76$, $p=.09$, $t_2(49)=1.96$, $p=.06$, while the comparison between every other condition and control did reach significance, $ts>2.85$, $ps<.05$.

External vs. internal transpositions. Next, we examined the effect of transposition location relative to the word boundary by conducting planned comparisons on just the four TL conditions, with the factor of externality of the transposition (external vs. internal). Note that for this analysis, we collapsed across word-initial and word-final transpositions for the “external” condition, and across the between- and within-morpheme transpositions for the “internal” condition. For all four measures, reading times were longer for words with external relative to internal transpositions. This difference only approached significance on first fixation duration, $t_1(24)=1.72$, $p=.10$, $t_2(48)=1.48$, $p=.15$, and it did reach significance for the other three measures

(gaze duration: $t_1(24)=6.49, p<.001, t_2(48)=4.93, p<.001$; go-past time: $t_1(24)=5.16, p<.001, t_2(48)=4.62, p<.001$; total time: $t_1(24)=6.18, p<.001, t_2(48)=6.36, p<.001$). This result replicates White et al. (2008), who did not find significant reading time differences on first fixation duration for external versus internal transpositions, but did find a significant effect on later measures. Thus, readers treated morpheme-external transpositions more like word-internal than word-external transpositions.

Table 2. Mean reading times in ms (and standard deviations) for the compound words.

Transposition Condition	First Fixation	Gaze Duration	Go-Past Time	Total Time
Control	229 (35)	286 (61)	361 (84)	383 (82)
<i>External</i>				
Word-initial	271 (47)	499 (152)	633 (207)	682 (209)
Word-final	251 (45)	389 (118)	473 (163)	542 (195)
<i>Average</i>	<i>261(39)</i>	<i>443 (117)</i>	<i>553 (164)</i>	<i>613 (182)</i>
<i>Internal</i>				
Between	248 (47)	348 (93)	454 (189)	443 (112)
Within	253 (43)	348 (87)	407 (131)	448 (122)
<i>Average</i>	<i>251 (41)</i>	<i>348 (78)</i>	<i>428 (129)</i>	<i>445 (107)</i>

Initial vs. final transpositions. To further explore whether word-initial transpositions were more disruptive than word-final transpositions, as suggested by White et al. (2008), we conducted planned comparisons between just the word-initial and word-final transpositions. Results showed that on all measures considered, readers spent longer reading the word-initial transpositions than the word-final transpositions (first fixation: $t_1(24)=2.20, p<.05, t_2(48)=2.60, p<.05$; gaze duration: $t_1(24)=4.04, p<.001, t_2(48)=4.36, p<.001$; go-past time: $t_1(24)=4.65, p<.001, t_2(48)=4.18, p<.001$; and total time: $t_1(24)=3.85, p<.001, t_2(48)=3.25, p<.001$). These results support White et al.'s (2008) claim that word-initial letters hold greater importance to word recognition during sentence reading than do word-final letters.

Within- and between-morpheme transpositions. Finally, we conducted our planned comparisons to test whether between-morpheme letter transpositions were more disruptive to reading times than within-morpheme transpositions. Across the board, results showed that these two conditions did not differ on any reading measure (first fixation, gaze duration, total time: all $t < 1$, all $p > .52$; go-past time: $t_1(24) = 1.25$, $p = .22$, $t_2(48) = 1.49$, $p = .14$). The only measure on which these conditions differed from each other was second fixation duration, which was 21-ms longer for the between- than within-transposition condition. This difference was marginally significant by subjects, $t_1(24) = 1.94$, $p = .07$, and did reach significance by items, $t_2(48) = 2.60$, $p < .05$. However, there was a substantial amount of data loss for this measure. On 52% of all trials, subjects did not make a second fixation on the target word, and only 40% of subjects made a second fixation in more than half of their ten trials in each condition. Due to the large amount of data loss contributing to this measure, the second fixation duration results should be interpreted with caution. Taking into account all measures, our findings did not support our prediction that between-morpheme letter transpositions would be more disruptive to reading times than within-morpheme transpositions.

Discussion

The current study investigated the consequences of encountering transposed letter nonwords during natural reading, specifically letter transpositions within and across morpheme boundaries in English compound words. We presented readers with compounds with a letter transposition either within one morpheme or across the morpheme boundary, as in Christianson et al. (2005). We also included a comparison condition with transposed letters in the word-initial or word-final position, following White et al. (2008), to determine if word-internal but morpheme-external letter transpositions were as disruptive to reading times as word-external transpositions. Readers encountered the compound words in sentence contexts that were coherent, but unconstraining with respect to the compound word. Findings supported our prediction that transposed letter nonwords would produce longer reading times than correctly spelled words. However, our prediction that compound words with letter transpositions across the morpheme boundary would cause more disruption than those within a single morpheme was not supported by the data. In this way, we failed to replicate Christianson et al. (2005), who found greater facilitation to naming latencies for words preceded by a within- compared to a between-morpheme transposition prime.

This is the first study to show equal disruption from across- and within-morpheme transposed letters in compound words during natural reading. These findings are powerful, and suggest that even under full-viewing natural reading conditions, when readers have control over the amount of time and number of fixations they devote to processing the misspelled words, they do not differentiate between the across- and within-morpheme transpositions. One reading time measure that did show a suggestive difference between these conditions, second fixation duration, indicated that the second time readers fixated the misspelled words, they spent

numerically more time on the between- versus within-morpheme transposition nonwords.

However, this measure should be interpreted cautiously, because readers only made a second fixation on less than half of trials, resulting in substantial data loss for this measure.

The overall lack of reading time differences between these two critical transposition conditions adds to the literature showing that letter transpositions that cross morpheme boundaries are no more disruptive to processing than those that stay within a single morpheme (Diependaele, Morris, Serota, Bertrand, & Grainger, 2012; Duñabeitia, et al., 2011; Masserang & Pollatsek, 2012; Perea & Carrieras, 2006; Rueckl & Rimzhim, 2011; Sánchez-Gutiérrez & Rastle, 2013). Together with previous findings, the current study provides evidence that the identification and comprehension of compound words during natural reading is not differentially affected by the disruption of morpheme boundaries by transposed letters, suggesting that morphological decomposition likely does not occur in the initial stages of word recognition. Accounts of prelexical morphological decomposition would predict greater disruption when the letter transposition crosses a morpheme boundary because this would compromise the reader's ability to successfully extract the constituent morphemes during the initial stages of recognition. That these two conditions did not differ on four of the five reading time measures considered in the present study suggests that readers were not extracting each morpheme separately in order to recognize the whole word. This is not to say that readers never recognized the compound words as containing multiple morphemes, but we have little evidence that readers experienced more difficulty with the across- than between morpheme transpositions at any point during the reading process.

Differences from Previous Work

The current findings diverge from those of Christianson et al. (2005), who also used between- and within-morpheme transpositions in compound words, but whose methodology differed in two important ways. First, Christianson et al. (2005) presented words in isolation rather than in sentence contexts. It has been shown that presenting compound words in neutral sentences can mitigate the processing costs associated with opaque versus transparent compound words (Frisson, Niswander-Kelment, & Pollatsek, 2008; Pollatsek & Hyona, 2005), and that constraining contexts can eliminate even robust effects of morpheme frequency (Juhasz, 2012). It is possible that the weakly constraining sentences in which the compound word appeared in the current study may have allowed readers to generate predictions about the compound, providing enough semantic support for the between-morpheme transpositions to activate their base words better than they did in isolation. Indeed, studies have shown that increasing levels of cloze probability produce graded reductions of both reading times (Rayner & Well, 1996) and N400 amplitudes (Kutas & Hillyard, 1984; Wlotko & Federmeier, 2012) elicited by the expected word. These studies demonstrate that words need not appear in highly constraining sentences to receive contextual support, but that even weakly expected words receive a processing advantage relative to anomalous words (which can essentially be viewed as isolated words, with no pre-activation from their context). With these facts in mind, it is entirely plausible that the larger within- than across-morpheme priming effects found by Christianson et al. (2005) are observable when the words are viewed in isolation, but that the addition of simple sentence contexts in the current study helped readers overcome these relatively small costs when the words were encountered during natural reading, thus providing a clearer understanding of how readers cope with overtly misspelled words under more typical reading conditions.

A second difference between the current study and Christianson et al. (2005) is the task employed. In Christianson et al. (2005), subjects produced the compound words aloud as quickly as possible, whereas subjects' only task in the current study was to read for comprehension. There is evidence that subjects treat letter strings differently when they are preparing to produce them versus reading silently. The dual-route cascaded model of reading aloud (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) posits that a word's phonological units are built up serially from left to right when reading words for production, encouraging more attention to the ordering of letters than when reading for comprehension. Enhanced attention to the serial ordering of letters could have been especially disruptive in Christianson et al. (2005) based on the nature of their stimuli, which contained digraphs (i.e., *sh*, *ch*, *th*, *kn*, *ng*) at the morpheme boundary in 23 of their 44 compound word items. Cross-morpheme letter transpositions inserted an incorrect letter into this unit (i.e., *wishbone* to *wis**h**bone*), as opposed to the within-morpheme letter transposition, in which the two letters of the phoneme unit remained adjacent (i.e., *wishbone* to *wi**h**sbone*). Breaking up the digraph could have severely disrupted the planning of the proper sequence of speech sounds, particularly if readers began planning to produce the sound of each letter separately instead of the single (different) sound they make together. On the other hand, the letters remained adjacent in the within-morpheme transposition condition, possibly making it easier for subjects to recover their proper pronunciation, thus facilitating naming times. Furthermore, there is evidence from neuropsychology showing that some aphasic patients can read words aloud with no comprehension of what they read (Heilman, Tucker & Valenstein, 1976), suggesting that reading aloud and reading for comprehension may be mediated by different (or at least partially non-overlapping) pathways in the brain. As such, the two tasks are

likely subject to different underlying processing mechanisms, which could help explain the different results across the current study and Christianson et al. (2005).

Transparency Effects

The current results add to the literature that transposed letter similarity effects may not be sensitive to morpheme boundaries, at least when they appear in compound words during reading of weakly constraining contexts. However, this topic is still a highly contentious debate. One possible factor that seems to have been overlooked in many studies is the transparency of the affixed words involved. Most of the studies reviewed presently that have found equal TL priming within and across morpheme boundaries have used transparent words, in which the meaning of the whole word is related to the meaning of each morpheme. Diependaele and colleagues (2012) actually manipulated transparency orthogonally to transposition location, and found that TL priming was present across the morpheme boundary in transparent affixed words (*flexible-flex*), but was absent across the same location in pseudo-derived opaque words (*corner-corn*). They explained their results in terms of a dual-route model of morphological processing, suggesting that the morphemes in the transparent words received more top-down activation from the semantically related derived prime than did the opaque words (which were unrelated to their primes), thus facilitating responses to the transparent condition only.

If transposed letter priming is reduced across morpheme boundaries in opaque relative to transparent affixed words, the same could be true of opaque versus transparent compound words. The current materials were not designed to investigate transparency effects, although they covered a range of transparency (transparent: *drugstore*; partially transparent: *catfish*; opaque: *overcast*). In a post-hoc analysis, the authors classified 15 of the 49 items as opaque, and the remaining 34 as transparent, and exploratory analyses were conducted to test for an interaction between transparency and transposition location (between- or within-morpheme) (see reading

times listed in Table 3). Results showed a marginal interaction on first fixation, $F_1(1,24)=3.66$, $MSE=12,313$, $p=.07$, $F_2(1,47)=3.06$, $MSE=4889$, $p=.09$, reflecting the fact that first fixations were shorter to transparent than opaque words in the between-morpheme transposition condition, while the opposite was true for the within-morpheme transpositions. No other measure showed a similar interaction, although go-past times were again shorter for the transparent than opaque words in the between-morpheme transposition condition.

Table 3. Target word reading times in ms (and standard deviations) for between- and within-morpheme transpositions only, split by transparency of compound words.

Transparency	First Fixation		Gaze Duration		Go-Past Time		Total Time	
	Between	Within	Between	Within	Between	Within	Between	Within
Transparent	240 (45)	256 (53)	345 (102)	355 (119)	419 (141)	396 (137)	441 (151)	436 (148)
Opaque	261 (53)	248 (57)	345 (91)	334 (113)	521 (238)	431 (139)	449 (114)	476 (180)

The direction of this first fixation effect, while only suggestive, falls in line with the dual-route model proposed by Diependaele et al. (2012): the morphemes of the transparent compounds may have received more top-down activation from the whole word than the opaque morphemes did, allowing initial processing of the transparent words to be less disrupted by the between-morpheme transposition. Future work could manipulate the transparency of compound words used in conjunction with transposed letters to determine if transparency might be a key factor to determine whether the transposed letter effect is present across morpheme boundaries or not, as suggested by the findings from Diependaele et al. (2012) and our exploratory results.

In general, our study found that compound words with letter transpositions across morpheme boundaries were no more disruptive to processing than those that fell within a single morpheme. Although both word types elicited longer reading times relative to control, they importantly did not differ from each other. These findings thus suggest that English compound

words are not recognized via morphological decomposition when are encountered in a sentence context.

Part 2: Transposed Letter Effects in ERPs

The bulk of the previous work reviewed on the TL similarity effect has used as dependent measures response times and/or reading times. We have until this point largely ignored ERP studies, which examine the electrophysiological markers of the transposed letter similarity effect. The use of ERPs as a tool to measure the type of processing elicited by the TL similarity effect can be extremely fruitful because of the multidimensionality of the data provided by ERPs. They provide information about the polarity, onset, duration, and scalp distribution of effects generated by an input stimulus. ERPs have millisecond level accuracy, so they can be used to observe how processing of a target word unfolds over time. These aspects of ERPs are beneficial to our interpretation of effects because we do not have to rely on factors like prime duration or stimulus onset asynchronies to make claims about how early in processing a certain variable exerts its effect. Because ERPs are elicited obligatorily, participants do not have to engage in a secondary task (i.e., lexical decision, semantic categorization, naming aloud) to elicit a response. Furthermore, secondary tasks can often only measure the end-state of processing (i.e., did the subject correctly categorize the word?), potentially missing processing differences between two stimulus types if they produce the same outcome. The relationship of specific ERP components with different types of processing (i.e., the P2 component with attentional processing, the N400 with semantic processing, and the P600/LPC with later reanalysis or interpretative processing), allows us to ask even more specific questions about transposed letter versus replacement letter nonwords. For example, instead of simply asking which ones produce faster response times, we can instead examine whether transposed letter nonwords allow for less effective contact with semantics than correctly spelled words, which can provide specific insights into the mechanisms underlying transposed letter processing in a way that reaction times cannot.

The current study asks how the presence of transposed letters in compound words affects their comprehension when viewed in a sentence context, and, specifically, whether the placement of a letter transposition at the boundary between the compound word's morphemes disrupts comprehension to a greater degree than a transposition within a single morpheme. We will test this hypothesis by recording subjects' ERPs while they read sentences containing compound words appearing in one of these three spelling conditions, along with noncompound words that also sometimes have letter transpositions. The main theoretical question is whether there is a difference between the within- and between-morpheme letter transpositions (as well as their comparison to correctly spelled compound words). Particularly, we will test if the between-morpheme transpositions elicit larger N400 effects than the within-morpheme transpositions, which would reflect more difficulty in activating the word's semantics, or alternatively if they elicited a larger P600/Late Positive Complex, which would reflect more attentionally-demanding reanalysis or interpretation caused by this misspelling.

To begin, we will review the state of the literature with respect to the types of ERP effects shown to be elicited by various types of nonwords, beginning with misspelled words in general, and then moving to transposed letter nonwords. The first portion of this review will focus on the types of effects that have been garnered from behavioral tasks that have concurrently recorded ERPs during their completion (i.e., lexical decision tasks, semantic categorization, and naming). These studies differ importantly from the design of the current study, especially by presenting words either in isolation or preceded by a prime, but can provide valuable insights into the basic effects that misspellings have on typical word recognition processes. We will next move into a review of the literature regarding the processing of misspelled words appearing in a sentence context, to better understand how the contextual

support of the sentence frame affects comprehension of the misspelled word. Importantly, none of the sentence-reading studies have investigated the effects of transposed letters specifically on reading processes, nor have they explored the relationship between transposed letters and morpheme boundaries, which leaves open the questions to be addressed in the current study.

Behavioral Studies

ERPs to Misspelled Words

McKinnon, Allen, and Osterhout (2003) conducted a lexical decision study using correctly spelled and misspelled words as targets, and found that misspelled words elicited N400s that were much larger (more negative) in amplitude than correctly spelled words within the typical N400 window. Their misspelled words always contained a replacement letter in the word-initial positions, and optionally contained a second replacement letter in either the final or penultimate letter position as well. Importantly, their misspelled words maintained the correct pronunciation of the word (stimuli also known as pseudohomophones), such that readers had to use their orthographic knowledge of the word to make the correct judgment rather than phonology. Additionally, Sauseng, Bergmann, and Wimmer (2004) found in a delayed naming task, in which subjects viewed correctly-spelled or replacement-letter pseudohomophone nonwords, that the replacement letter words also elicited a large negativity relative to the correctly-spelled words from 200-600ms post-stimulus onset. These two studies replicate previous ERP findings that pronounceable pseudowords elicit larger N400s than regular words, although work by Laszlo and Federmeier (2011) has shown that there are in fact no N400 differences between words and pseudowords presented out of context when orthographic neighborhood size is matched across the two stimulus types. However, these results cannot speak

directly to the behavioral studies on the transposed letter effects, because they contained only replacement-letter, but not transposed letter conditions. We would predict that the replacement letter condition would be more difficult than the control condition (as demonstrated by behavioral studies), and these two studies indeed support that prediction. An open question, then, is whether transposed letter nonwords will also elicit larger N400 amplitudes than correctly spelled words.

ERPs to Transposed Letter Nonwords

A study by Grainger, Kiyonaga, & Holcomb (2006) used masked-priming lexical decision task, in which targets were preceded by primes containing either two transposed letters or two replacement letters in the same position. They found an effect on an early negative-going deflection they termed the N250, which was reduced for words preceded by TL versus replacement letter (RL) primes. There was also a nonsignificant reduction in the negativity elicited by the TL relative to RL primes in the N400 window, which was suggestive of easier processing for targets preceded by TL as opposed to RL primes. However, they did not present data from targets that were preceded with their identity prime, so we cannot know for sure whether the N250 would have been reduced to an even greater degree by the presentation of an identical prime. Generally, they concluded that the reduction of the N250 component following the TL primes indicated that subjects considered the TL prime to be a better match to the target than was the RL prime, as the N250 is posited to reflect early orthographic matching between the prime and target.

Putting all three conditions together (correctly spelled, TL, and RL), Carreiras, Vergara, & Perea (2007) conducted a lexical decision study in which the targets were presented under

full-viewing conditions, without the presence of primes. Targets could appear with transposed or replaced letters, involving either two consonants or two vowels, or as correctly spelled words. Their results showed larger N400 amplitudes for all misspelled conditions relative to the correctly-spelled words, with larger N400 amplitudes elicited by the RL than TL nonwords. The replacement letter conditions then also elicited larger late positivities between 500-650 ms post-stimulus onset (which will become important later for sentence-reading effects), and, overall, the vowel-replacement conditions elicited larger (more positive) P600/LPC amplitudes than consonant-replacement conditions. The LPC has been linked to more attentionally-demanding processes, again suggesting that the RL nonwords were harder to process than the TL pseudowords, and furthermore that the type of letter replaced might engage different underlying processing mechanisms. This study is important for a few reasons: it presents ERP data to misspelled words under full-viewing conditions (rather than as primes), which is much closer to approximating how the processes unfold during natural reading; it lays out clearly the differences between TL, RL, and control conditions, which the previously discussed studies have failed to do; and finally, it shows that misspelled nonwords elicit positivities that differ between the TL and RL conditions.

Transposed vowels versus consonants.

The difference between vowel and consonant transpositions observed by Carreiras et al. (2007) has been replicated in a few subsequent studies. Carreiras, Vergara, & Perea (2009) used the misspelled words from their previous study as masked primes in a lexical decision task, again manipulating whether the critical letters on two different dimensions: they could be transposed or replaced, and the manipulation could occur on either consonants or vowels. They found that the vowel-replacement condition produced larger N400s than the vowel-transposition condition,

whereas there was no difference between the N400 amplitudes elicited by the consonant TL and RL prime conditions. This finding suggested greater reliance on vowel order than consonant order in word recognition. Furthermore, Vergara-Martinez, Perea, Marin, & Carreiras (2011) conducted a similar masked-priming lexical decision task, in which primes could contain transposed consonants or vowels, and compared these against a baseline condition in which targets were preceded by identical primes. ERPs elicited by targets showed no differences between the three conditions in the early N400 time window, but in a slightly later time window, from 450-525ms, the TL-vowel primes were significantly more negative than the identity condition. (TL-consonants fell somewhere between the two, but these effects were nonsignificant). It is hard to compare between these studies because they are not perfectly matched on their stimulus or filler types, but the general takeaway message is that the type of TL priming observed may differ importantly based on whether the transposition involved two consonants or two vowels, with a reduction in priming elicited by transposed vowels.

Transposed letter words versus nonwords.

In addition to the differences in effects being due to the type of letter involved in the transposition, ERP work has also shown that the lexical status of the prime can affect the size of the priming effect. Dunabeitia, Molinaro, Laka, Estevez, & Carreiras (2009) conducted a masked priming semantic categorization task, in which subjects were monitoring targets for “profession” names. Targets were preceded by primes that were either a standard TL nonword (barin-BRAIN), replacement letter nonwords (bosin-BRAIN), or a real word, created by the transposition of two letters within the target (casual-CAUSAL) or by the replacement of two letters within the target (carnal-CAUSAL). Results showed a reduced N250 for targets preceded by the TL versus RL primes, replicating Grainger et al. (2006), but showed no difference

between the TL or RL “word” primes. That is, when the prime word had its own representation to activate, it seemingly did not prime its TL neighbor word, perhaps due to lateral inhibition between representations at the word level. This effect also shows that the transposed letter effect is not driven by simple form overlap—even though the word and nonword TL primes shared the same number of letters with the target, only the nonword prime provided facilitation on the N250 component. One caveat to this study is that they only analyzed data within the 200-300 ms time window, meaning that we cannot make any claims about what may have happened in any time window after this, particularly in the classic N400 time window, or the later P600/LPC window, which could have distinguished between conditions by showing more overt noticing of the relationship between the prime and target in one condition over the other.

The interaction between transposed letter effects and frequency.

Finally, one other way in which ERP studies employing the LDT have extended TL effects is by manipulating the frequency of the base word to ask whether the amount of facilitation a TL prime provides to its base word is modulated by its frequency. Vergara-Martinez, Perea, Gomez, & Swaab (2013) used a set of high and low frequency words (>50 per million and <5 per million, respectively) as targets in a lexical decision task. The words were presented in full-viewing conditions, and could appear either in their correct form (i.e., *mother*), with an internal pair of transposed letters (*mohter*) or with a single replacement letter (*mosher*). Their results showed different effects for the two word types, particularly on the N400 component. The high frequency words showed a graded N400 effect, in which the RL nonwords elicited the largest N400 amplitude, followed by the TL nonwords, and finally the correctly-spelled words, suggesting that the TL nonwords were considered as being more word-like than the RL nonwords and more effectively contacted semantics. However, for the low frequency

words, the RL and TL nonwords elicited N400s that were of roughly the same amplitude, both of which were significantly more negative than the correctly-spelled words. Although at some channels the waveforms elicited by the TL nonwords appear to be marginally smaller (more positive) than the RL nonwords, this difference was not significant. These findings thus show that the TL similarity effects differ based on the frequency of the base word. Perhaps the more frequent words are easier to activate, and so the TL pseudowords are able to provide enough evidence so as to partially activate the representation of the base word. However, for the less frequent base words, more evidence would be needed to effectively activate the representation of the base word, and the degraded input of the TL stimulus just could not allow readers to recognize its actual word any better than the replacement-letter stimulus did.

ERP Effects to Misspelled Word in Reading Studies

One feature that all of the above studies have in common is that they do not take full advantage of the power of ERPs to index moment-to-moment word processing in a sentence context, in that they have all required some secondary task (semantic categorization, naming words aloud, lexical decision), placing demands on the subjects and potentially encouraging them to adopt and apply strategies to their processing of the words that would not be present in their everyday reading. Furthermore, for some of the studies involving lexical decision tasks in which the transposed letter words were presented as targets instead of primes, the subjects' job is to reject the TL nonword as such—not a real word. However, this is actually the *opposite* strategy that readers would have when encountering a misspelled word “in the wild” (when actually reading), which would be to consider it as an error that nonetheless represents a real word and attempt to figure out its intended meaning. This involves an entirely different approach

to handling the TL nonwords—not just having to quickly decide how word-like it is compared to others words or nonwords they might see (which, as an attempt at semantic access, would be most likely to elicit larger N400s, as we have seen in previous studies), but to instead attempt to reinterpret the misspelled word in the context of the ongoing sentence context in order to “figure out” its intended meaning.

This type of reinterpretive processing is often associated instead with a late posterior positivity, or P600. Although the P600 is traditionally thought of as a “syntactic anomaly detector,” because of its robust elicitation by tense and number error on verbs (Hagoort, Brown, and Groothusen, 1993), more recent theories regarding the P600 have begun to change this view. In particular, Brouwer and colleagues (2012) propose that the P600 indexes the *updating* of the ongoing sentence representation more generally, rather than the more restricted idea of *reanalysis* following a syntactic anomaly. Under this view, the component is theoretically always present, but its amplitude is modulated by the difficulty of updating the current mental model of the context. This broader take on the P600 component allows their theory to account for the presence of this component under any situation in which integrating the current word with the ongoing context is difficult--particularly one created by a transposed letter nonword, which would likely activate its base word well *enough* to allow for its partial recognition while still posing a challenge to recognition and integration processes.

A few studies have investigated the effects of spelling errors in sentence reading. Although these have not strictly involved transposed letter nonwords, their findings are still relevant to our review. The general finding that emerges is that misspellings elicit a posterior positivity, with roughly the same onset and distribution as positivities elicited by syntactic anomalies, as predicted by Brouwer et al. (2012)’s view. In an early study on the issue, Münte,

Heinze, Matzke, Wieringa, & Johannes (1998) used target words with letter addition misspellings (i.e., *broome* for *broom*) that appeared sentence-medially in moderately constraining contexts. They found a posterior positivity elicited by these words that began quite early (~250 ms after word onset), and was sustained over at least the next two words in the sentence. These words did not elicit a larger N400 than the correctly spelled words did, although because of component overlap, we cannot rule out the possibility that the early onset latency of the positivity may have masked a potential N400 effect. Their syntactic violation condition elicited a small anterior negativity, followed by a shorter-lived posterior positivity that only extended through one subsequent word. Both positivities had a widely distributed scalp distribution that was highly overlapping. In contrast, semantic anomalies in these same sentence contexts elicited large N400 effects relative to correct words. This suggests that readers were largely able to extract the semantics of the misspelled word, or to match it with its correct representation in semantic memory, and then experienced difficulty integrating this degraded representation into the ongoing context.

Expanding on Münte et al. (2003), Vissers, Chwilla, and Kolk (2006) explored the interaction between word predictability and misspelling in a sentence context. They used highly constraining sentences in which target words were the most expected continuation (cloze ~90%). From there, the targets could appear in four conditions: correctly-spelled or pseudohomophone version of the expected continuation (i.e., *In that library the pupils borrow {books/bouks} to take home*), or correctly-spelled or pseudohomophone version of an unexpected continuation (i.e., *The pillows are stuffed with {books/bouks} which makes them feel hard*). Incorrect endings were created by swapping expected endings from other sentence contexts. Their results revealed that the misspellings elicited a large posterior positivity relative to the correctly spelled basewords,

but *only* when the misspelled word was a pseudohomophone of the expected ending. This positivity started around 500 ms post-stimulus onset, suggesting that the misspelled word was close enough so as to not disrupt typical lexico-semantic processing, but that readers still noticed the divergence and elicited the late positivity, which could index either their explicit *noticing* of the word, or potentially their efforts to repair/reinterpret the word. On the other hand, there was no positivity effect for the pseudohomophones when they replaced the unexpected endings.

Laszlo and Federmeier (2009) found very similar outcomes as Vissers, even though their actual research question was quite different. They used high constraint sentences that could appear with a variety of different endings: the most expected word (i.e., *bank*), an unexpected word that was either an orthographic neighbor of the expected word or not (i.e., *bark* or *clam*, respectively), a pseudoword that was also either an orthographic neighbor of the expected word or not (i.e., *pank* or *horm*), or finally an illegal letter string that was again either an orthographic neighbor of the expected word or not (i.e., *bxnk* or *rqck*). Note that their pseudoword condition actually consists of single replacement-letter nonwords, much like the misspelling condition in Vissers et al. (2006). They found that for all item types, sentence completions that were orthographic neighbors of the most expected completion elicited smaller N400s than those that were orthographically unrelated to the most expected ending. This facilitation was significant within words, pseudowords, and illegal strings, despite the overall smaller N400 amplitudes elicited by the illegal strings as compared to the other two stimulus types, which did not differ from each other. Importantly for our current review, the unexpected items also elicited a late positive complex (LPC) over posterior channels. The size of the LPC was graded: it was smallest for words, bigger for pseudowords, and biggest for illegal strings. Within each item type, orthographic neighbors of the expected item elicited larger amplitude LPCs than did non-

neighbors. Their results show that high constraint sentences allow for the specific prediction of not only a word's meaning, but also its orthographic features. Thus, for two words or stimuli that are equally bad matches for the sentence context semantically, an orthographic neighbor of the expected word will show a reduced N400 relative to the non-neighbor, regardless of the lexical status of that stimulus (word, pseudoword, illegal string)

These two studies together have interesting implications for our understanding of how misspellings are interpreted in a sentence context. In highly constraining sentences, the presentation of a stimulus that differs from the most expected ending by a single letter allows readers to notice (perhaps consciously, even) this relationship and work to repair it, thus eliciting the large LPC observed in both Laszlo and Federmeier (2009) and Vissers et al. (2006). However, if the nonword bears no relationship to the expected item, the LPC is much smaller as in Laszlo and Federmeier (2009), or not present at all as in Vissers et al. (2006)—suggesting less engagement of the processes necessary to integrate/interpret a stimulus when it is in very dissimilar from the expected word.

One caveat to these conclusions is that both Laszlo and Federmeier (2009) and Vissers et al. (2006) only investigated the effects of misspellings in highly constraining contexts, in which readers were strongly predicting the target word. While there is much to be learned from very strongly constraining contexts, especially how readers can use previous context to overcome the difficulties generated by misspelled words, understanding how readers deal with misspellings in moderate to weakly constraining contexts is more generalizeable. Most sentences that readers typically encounter provide only moderate or weak constraint, so employing experimental contexts with roughly comparable levels of constraint can provide a more accurate picture of how readers might typically handle these types of errors. Van de Meerendonk, Indefrey, Chwilla,

& Kolk (2011) used high and low constraint sentences, in which the targets were either correctly spelled or a pseudohomophone of the target, again created by replacing a single letter. For example, a high constraint sentence would read *The pillows are stuffed with {feathers/feathurs} which make them feel soft* (cloze =.90), whereas an example low constraint sentence would be *At that spot there sometimes lie {feathers/feathurs} from pheasants and peacocks* (cloze=.10). They also included medium-constraint sentences (cloze=.50) with syntactic violations (*The clean clothes and towels hang/hangs out on the clothesline to dry*). The syntactic violation condition was included as a comparison condition against which to test the specificity of misspelling-related positivity to the more widely-replicated positivity associated with syntactic violations. Their results showed that all targets in the low constraint contexts elicited larger amplitude N400s than did those in the high constraint sentences (regardless of the spelling condition of the target). Subsequent to the N400 effects, all misspelled words elicited a late posterior positivity compared to all correctly spelled words. This effect interacted significantly with constraint—the effect was larger for high-constraint targets, but importantly, the positivity was present and significant for the low-constraint targets as well. This finding replicates Vissers et al. and Laszlo & Federmeier, showing a larger positivity for misspelled targets compared to correctly-spelled ones in high constraint sentence. Furthermore, this study expands on that work by showing that the effect is also present in low constraint sentences. It is not necessary for readers to be generating very strong predictions about the target in order to elicit the positivity to the misspelled words. This shows that they are also elicited in contexts in which readers may have had only very weak predictions about the upcoming words. Unsurprisingly, the syntactic violations also elicited a P600, which did not differ in timing or scalp distribution from the spelling violation.

Additionally, and quite interestingly, the N400 predictability effect was *not* affected by target misspellings. In other words, nonwords created by replacing a single letter of the highly predictable word did not show larger N400 amplitudes than those without a spelling error, and they were generally reduced even compared to correctly spelled words that were less predictable. This replicates Vissers et al. (2006), who also found no N400 differences between the predicted and misspelled words. However, this contrasts sharply with Laszlo and Federmeier (2009), who found that pseudoword neighbors of the expected word, created in the exact same way by swapping out one letter from the correct word, *did* elicit larger N400s than their correctly spelled expected counterpart.

In sum, these studies suggest that spelling errors elicit posterior positivities with largely overlapping timing and scalp distribution as those elicited by syntactic violations, which could suggest that they share at least partially overlapping neural generators—and furthering the idea that the P600 reflects more general processing mechanisms beyond syntactic functions. Note, however, that while the studies reviewed about have covered several topics of interest (especially the type of ERP components elicited by spelling errors during full-viewing natural reading, as well as how these misspellings interact with constraint), that none of them have looked at how misspellings interact with morpheme boundaries. Furthermore, none of them employed the classic transposed letter manipulation, and instead used letter additions or letter replacements. We know that replacement letter nonwords are more disruptive to behavioral tasks than are letter transpositions. This leaves the open question of whether fully viewing letter transpositions during natural reading will elicit the same broad posterior positivity as the single-letter replacement nonwords used in other ERP studies, or whether the effect of TLs will be less

disruptive (and perhaps elicit a smaller P600/LPC effect, potentially along with an N400 effect as well).

Current Study

The current study will examine the processing of transposed letters that cross a morpheme boundary in compound words, to determine if cross-morpheme letter transpositions are more disruptive to processing than within-morpheme letter transpositions, as suggested by Christianson, Johnson, and Rayner (2005). This extends the findings of Christianson et al. (2005) because we ask how transposed letter nonwords are recognized and integrated into a sentence context during natural reading, whereas their study was only able to show that compound words preceded by transposed letter primes were name aloud *slower* when the prime contained a between-morpheme TL than when it contained a within-morpheme TL. The compound words will be presented at a sentence-medial location in contexts that are moderately constraining for the compound word. They will be compared to correctly spelled compound words, as well as to length- and frequency-matched noncompound words, appearing in very similar sentence contexts with letter transpositions in similar word locations. Overall, this design will allow us to ask two questions about the letter transpositions. First, we are interested in how the presence of a letter transposition in the compound word affects a reader's ability to extract meaning from the compound. If the letter transposition makes it more difficult for readers activate the meaning of the compound word, we would predict that the misspellings would elicit larger N400 amplitudes than the correctly spelled condition. Furthermore, if compound words are recognized via morphological decomposition, then the presence of cross-morpheme letter transposition should disrupt this decomposition process and make it more difficult to comprehend the compound word's meaning. If this is the case, then the between-morpheme letter transpositions should elicit

a larger N400 than the within-morpheme transpositions. For these analyses, we will examine mean amplitudes the epoch 300-500 ms post-stimulus onset, in the classic N400 time window, to see if the two compound word misspelling conditions differ from the correctly-spelled condition, and additionally, from each other.

Secondly, we will focus on the epoch 600-900 ms post-stimulus onset, in the typical time window for which the LPC/P600 component is observed. Based on previous studies involving misspellings, we predict to see a late posterior positivity in this time window for the letter transpositions compared to correctly-spelled words. Of critical interest is whether the between-morpheme letter transpositions will elicit larger posterior positivities in this time window than the within-morpheme letter transpositions, which would be evidence that those misspellings elicited more engagement of the processes underlying the late positive complex (i.e. caused more difficult integration/reinterpretation/more conscious ‘noticing’ of the misspelling, potentially because of the disruption they caused to “morphological decomposition.”). Furthermore, we will ask if transposed letters in non-compound words also elicit the later posterior positivity, and if so, whether it differs in timing and/or distribution from that elicited by compound words. Results will help clarify which aspect of processing transposed letters in compound words disrupt, when those words are being processed for meaning in a sentence context.

Methods

Participants

Twenty-one University of Illinois undergraduates participated in the study for course credit (12 males, 8 females; mean age: 19, range: 18-23). All were native English speakers, with no consistent exposure to other languages before the age of five. Subjects did not have a history

of neurological disorder or defect. All were right-handed as assessed by the Edinburgh Inventory (Oldfield, 197x), and 13 reported having left-handed family members.

Materials

Two different types of target words were each embedded in sentences: compound words and non-compound control words. The compound words were on average 7.9 letters long (range: 7-10), with an average whole-word frequency of 4.5 per million (range: 0-64) (from Celex database, cite). The written frequencies of the compound word's first morpheme (mean: 104; range: 0-1093) and second morphemes (mean: 122; range: 0-1147) did not differ significantly, $t(89)=-.62, p=.54$. The compound words had an average neighborhood size of 0.19 (range: 0-2). In their misspelled conditions, the TL-between compounds had an average neighborhood size of 0.02 (range: 0-1), and the TL-within compounds had an average neighborhood size of 0.03 (range: 0-1). The non-compound filler words were on average 8 letters long (range: 7-10), with an average word frequency of 6.29 per million (range: 0-15) and average neighborhood size of 0.85 (range: 0-7). In their misspelled conditions, the filler words had an average neighborhood size of 0.03 (range: 0-2). The only lexical-level factor that differed significantly between the compound and non-compound words was concreteness. This was determined by a norming study, in which 12 undergraduates, drawn from the same subject pool who did not participate in the ERP portion of the study, rated the set of compound words and non-compound target words on the degree to which they felt that the word could be experienced by the senses, with 7 being "highly concrete" and 1 being "less concrete." We also included 140 filler words (70 of which were highly concrete, and 70 of which were highly abstract), producing a total of 280 items for participants to rate. Compound words were on average rated as more concrete (average: 5.4) than

the non-compound target words (average: 4.6), a difference that was found to be significant, $F(1,139)=12.97, p<.001$.

Both types of words appeared in sentences in which they were plausible, and only moderately predictable from the sentence context. In order to control the plausibility and expectancy of the target compound words with respect to their sentence frames, a norming study was conducted with a separate group of 28 undergraduates, also drawn from the same subject pool as the current study and who did not participate in the ERP portion. We started with 122 compound words, each of which was embedded in two different sentences: one plausible and the other implausible. The implausible stimuli were created by swapping the compound words between sentences, such that the same sentence frames and target words would constitute the plausible and implausible conditions, but in different pairings. These 244 sentences were then divided into two lists, with the constraint that each compound word appeared only once per list, and that half of the sentences in each list were plausible. Each subject in the norming study received only one list, meaning that each subject rated 122 sentences (61 plausible, 61 implausible), and that each individual sentence was rated by 14 different subjects. The entire sentence was presented to subjects with the compound word present, which appeared sentence-medially, and was underlined to highlight it as a target word. Subjects were instructed to rate each sentence in two ways. First, they rated the expectancy of the underlined target on a scale from 0-100, with 0 being “not expected at all” and 100 being “completely expected.” Subjects also rated the plausibility of the entire sentence, again on a scale from 0-100, with 0 being “not plausible at all” and 100 being “completely plausible.” From the pool of 122 plausible sentences, we chose 90 to use in the current study, for which the rated expectancy of the target word fell between 40-70 (mean: 54.2; range: 40.3-70.7), and the plausibility was high (average:

96; range: 78.2-100). The final set of 90 sentences was on average 12.9 words long (range: 8-17), and the compound word was never the first or last word in the sentence (average: 5.9; range: 3-11).

The non-compound target words were also embedded sentence-medially in moderately constraining sentences. These sentences were drawn from a previously normed set of materials in which subjects filled in the last word of the sentence in a cloze-norming task (see Federmeier, Wlotko, et al., 2007, for details). The cloze probability of the sentence-final word was moderate, with an average of .46 on a scale of 0-1 (range of .12-.70). Although the expectancy of the target itself was not normed, it was not part of a highly constraining sentence, and, as such, its own cloze probability was likely lower than that of the sentence-final word.

Target words could appear either as correctly spelled or misspelled with a letter transposition. For the compound words, the letter transposition could either occur within one of the word's morphemes or across the morpheme boundary. The between-morpheme transpositions always involved the last letter for the first morpheme and first letter of the second morpheme (i.e., *jackpot* to *jacpkot*). Half of the within-morpheme transpositions involved the last two letters of the first morpheme (i.e., *raincoat* to *ranicoat*), and the other half involved the first two letters of the second morpheme (i.e., *starfish* to *starifsh*). The within-morpheme transpositions were randomly assigned to the first or second morpheme, with the constraint that if one of the word's morphemes contained only three letters, the *other* (longer) morpheme would receive in the letter transposition. Each subject saw all 90 critical compound words, 30 in their correctly spelled form, 30 with a between-morpheme letter transposition, and 30 with a within-morpheme letter transposition (15 in the first morpheme, 15 in the second morpheme). Subjects also saw all 50 of the non-compound filler words, 16-17 of which were correctly spelled, and 33-

34 of which had a letter transposition. As with the compound words, two different transposed letter nonwords were created for each non-compound target. While we chose as many monomorphemic non-compound words as possible, due to the nature of the English language, 32 of the filler words actually had a suffix (typically the inflectional suffixes *-s*, *-ed*, or the derivational suffixes *-er* or *-or*). Importantly, however, the letter transpositions in these words did not cross the morpheme boundary. In total, each subject read 140 sentences, 93-94 of which contained a single misspelled target word (60 misspelled compound words, and 33-34 misspelled noncompound words), and remaining 46-47 contained all correctly spelled words.

Procedure

Participants were seated 100 cm away from a 21" computer monitor in a dim, quiet testing room. They were given verbal instructions before the experiment started to remain as still as possible with their eyes fixated in the center of the screen, and to try to minimize blinks throughout the experiment. Participants were instructed to read for comprehension, and that there would be a memory test on what they read at the end of the experiment. Furthermore, they were told that they might see a few misspellings, but that they should just focus on understanding the sentence to the best of their abilities. Each trial started with the presentation for a string of fixation crosses ("++++") for 500 ms. This was followed by a blank screen for a range of time between 500-1500 ms (jittered to prevent the build-up of slow anticipatory potentials), after which a sentence was presented word-by-word in the center of the screen in normal sentence case. Each word was presented in yellow font on a black background, and was on the screen for 200 ms, followed by a 300 ms blank interval (SOA=500 ms). Sentences were followed by a 2000 ms blank screen, after which time the fixation crosses appeared and the next trial began. As previously stated, participants were encouraged to minimize blinks during sentence presentation,

but no specified “blink interval” was employed. The experiment was divided into four blocks of 35 trials each, each lasting about six minutes. Between each block, participants received a small break.

At the end of the experiment, a paper-and-pencil memory test was conducted to test subjects’ memory for the compound words. A list of 180 correctly spelled compound words was given to each subject, including all 90 target words from the study as well as 90 new compounds. For each word, subjects first indicated whether it had appeared in the study or not. For the words they marked as having seen before, they also indicated whether it was spelled correctly when it appeared in the study or not.

EEG Recording and data analysis

The electroencephalogram (EEG) was recorded from 26 evenly-spaced silver/silver-chloride electrodes attached to an elastic cap. All scalp electrodes were referenced on-line to the left mastoid, and re-referenced off-line to the average of the left and right mastoids. One electrode was placed on the left infraorbital ridge, and referenced to the left mastoid, to monitor for vertical eye movements and blinks. Two electrodes were also placed on the outer canthus of each eye, referenced to each other, to monitor for horizontal eye movements. Electrode impedances for the three electrodes on the face were kept below 10k Ω , and impedances for scalp electrodes were kept below 5k Ω . The continuous EEG was amplified through a bandpass filter of 0.02-100Hz and recorded to hard disk at a sampling rate of 250Hz.

Epochs of EEG data were taken from 100 ms before stimulus onset to 920 ms post-stimulus. Those containing artifacts from amplifier blocking, signal drift, eye movements, eye blinks, or muscle activity were rejected off-line before averaging, using threshold selected for

each participant through visual inspection of the data. Trial loss averaged 8.6%. Artifact-free ERPs were averaged by stimulus type after subtraction of the 100 ms pre-stimulus baseline. Prior to statistical analyses, ERPs were digitally filtered with a low-pass filter of 30Hz.

Results

Behavioral Data

Behavioral results were analyzed from the paper and pencil memory test conducted at the end of the study. Overall, subjects were 67% accurate on the memory test, and specifically recognized 80% of the “old” compound words as having been seen before. Memory performance was assessed with the discriminability index d' , with average scores and standard deviations shown in Table 4. This analysis divided responses into three levels based on the word’s original spelling in the reading portion (i.e., whether it appeared with a within-morpheme transposition, a between-morpheme transposition, or as correctly spelled). Analyses revealed that memory for the two transposed letter conditions was better than that for words that had appeared earlier as correctly spelled. This pattern was supported by a one-way within-subjects Analysis of Variance (ANOVA), which revealed a significant effect of spelling condition, $F(2,40)=6.43$, $p<.001$. Follow-up paired samples tests confirmed that d' scores were higher for both misspelling conditions compared to control (TL-between vs. control, $t(20)=3.34$, $p<.001$; TL-within vs. control, $t(20)=2.82$, $p=.01$), but that they did not differ between each other ($t(20)=.35$, $p=.73$).

We also looked at the second outcome from the memory test, which was the number of words each subject correctly recognized as having been misspelled in the reading portion of the study. We only calculated the percentage of correctly classified words *contingent* upon participants having correctly recognizing the word as “old” (see Table 1 for means). Subjects

were much more accurate in reporting that a target word was spelled correctly than recognizing when it had actually appeared with a misspelling. This outcome was supported by a similar ANOVA as was conducted over the d' results, which showed a main effect of spelling condition, $F(2,40)=9.47, p<.001$. Paired samples t-tests showed that the correctly spelled words had higher accuracy rates than either of the misspelling conditions (TL-between vs. control, $t(20)=-3.47, p<.01$; TL-within vs. control, $t(20)=-3.06, p<.001$), which again did not differ from each other ($t(20)=.26, p=.79$). Overall, these findings show that despite showing much better recognition of words that originally appeared with a misspelling, subjects either did not remember that they were previously misspelled, or there was a large bias to report words as having been spelled correctly previously.

Table 4. Average of d' (and standard deviations), and percentage of correct spelling judgments on the compound words in the memory test.

Spelling Condition	Average d'	% Accurate Spelling Judgments
Between-Morpheme Transposition	1.15 (.46)	40%
Within-Morpheme Transposition	1.12 (.46)	39%
Correctly Spelled	.94 (.41)	73%

ERP Results

We will present ERP results in two different analysis windows, to address two different questions about the current dataset. First, we will examine the epoch 300-500 ms post-stimulus onset, to determine if the transposed letters in compound words elicit larger N400 amplitudes than correctly spelled compound words, and furthermore, if there is a difference between the across-morpheme and within-morpheme letter transpositions. The presence of non-compound words in the analyses will also allow us to test for an N400 effect from letter transpositions *not* appearing in compound words, as well as allowing us to ask whether these two effects differ

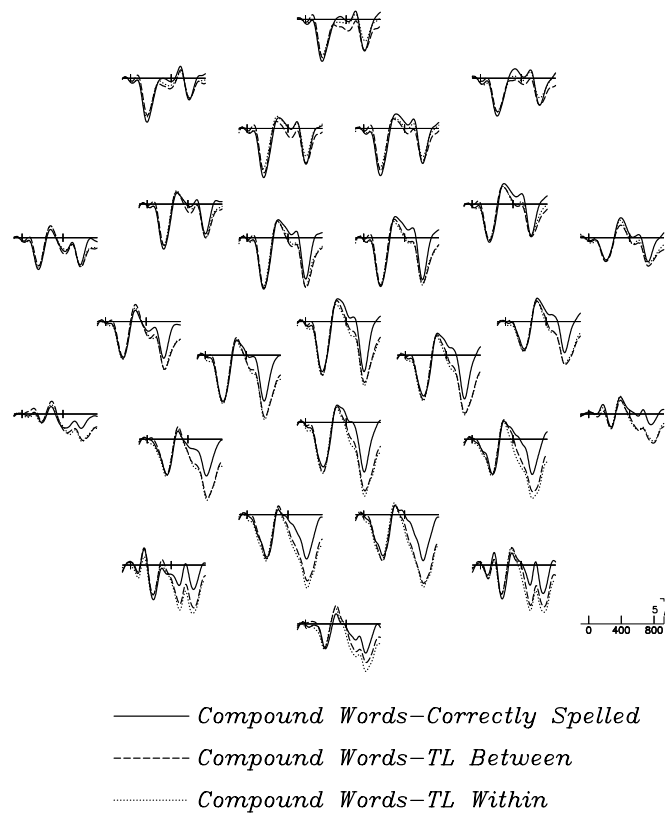
across word types. Secondly, the time window 600-900 ms post-stimulus onset will be examined, to test for a P600 effect elicited by the transposed letter compound words relative to the correctly-spelled compound words, with a special focus on whether this effect differs for the two types of TL conditions. Again, the effect of misspelling on the P600 component will be investigated within the noncompound words, as well as compared across the two word types. The presence of non-compound words provides an important comparison condition in order for us to test whether letter transpositions at morpheme boundaries in compound words behave like other word-internal letter positions (which would provide evidence against morphological decomposition), or whether they are *more* disruptive to processing than similarly placed letter transpositions in non-compound words, which would be evidence for morphological decomposition. All ANOVAs were conducted using a repeated measures design, with the Greenhouse-Geisser correction for the violation of the assumption of sphericity. Electrode will be included as a factor in all ANOVAs, but main effects of electrode will not be reported as they are not of theoretical significance.

The N400 Time Window: 300-500 ms.

Compound words. Mean amplitude of the N400 was measured between 300 and 500 ms over all electrode sites. As can be seen in Figure 1, the N400 amplitudes elicited by the correctly spelled compound words did not differ from those elicited by the transposed letter nonwords. This lack of difference is supported by a one-way within-subjects ANOVA with the factor of spelling (3 levels: TL-between, TL-within, correctly spelled), which showed a non-significant effect of spelling, $F(2,40)=.07$, $p=.93$, and no interaction between the factors of spelling and electrode, $F(50,1000)=.59$, $p=.67$. The effect of spelling is also non-significant even when we

constrain the analysis to just the 15 posterior electrodes that are most commonly shown to exhibit N400 effects $F(2,40)=.10, p=.90$, as was the interaction between spelling condition and electrode, $F(28,560)=.74, p=.62$. Thus, these results suggest that the transposed letter nonwords did not make it more difficult to contact semantics than the correctly spelled words, based on the lack of N400 difference between these conditions.

Figure 1. ERPs to compound words in the correctly spelled condition (solid lines), TL-between condition (dashed lines), and TL-within condition (dotted lines), from target onset to 920 ms post-stimulus onset).

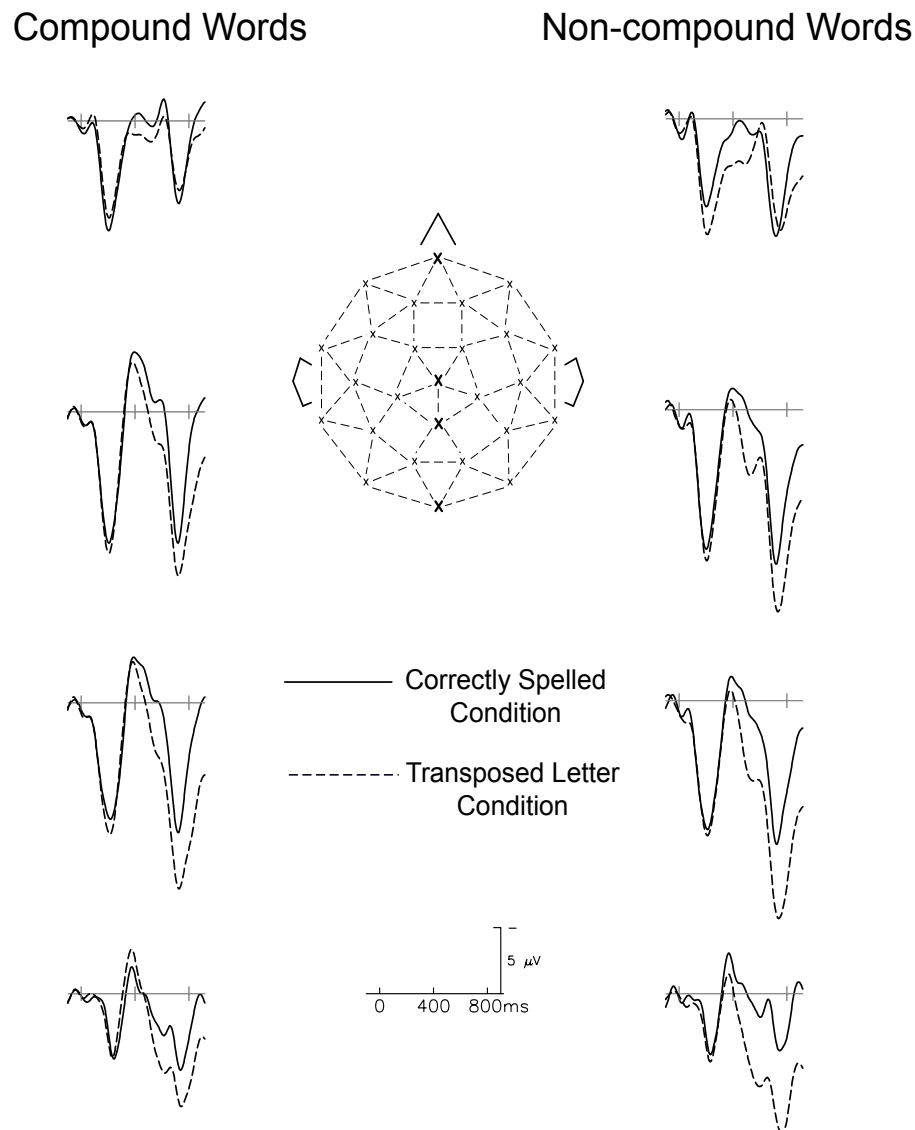


Non-compound words. A similar analysis carried out for the non-compound words also showed that there was no difference in the amplitude of the N400 elicited by the correctly spelled and transposed letter filler words, as can be seen in Figure 2. This was confirmed by an ANOVA conducted over all scalp electrodes, which again shows a non-significant effect of spelling

(correct vs. incorrect), $F(1,20)=.27$, $p=.61$, as well as a non-significant interaction between spelling and electrode, $F(25,500)=1.23$, $p=.31$. Again, as with the compound words, we measured the N400 effect for the difference between the correctly spelled and transposed letter words to see if there is an observable N400 effect over a more restricted range of 15 posterior electrodes most commonly known to show N400 effects (LMCe, RMCe, LDCe, RDCe, MiCe, MiPa, LLTe, RLTe, LDPa, RDPa, LLOc, RLOc, LMOc, RMOc, & MiOc). However, this more restricted range of electrodes also failed to show a difference in the N400 amplitude elicited by the two spelling conditions, $F(1,20)=.60$, $p=.45$, although the interaction between spelling and electrode was significant, $F(14,280)=3.03$, $p<.05$. This interaction is probably due to component overlap—a very large positivity is present for the TL words, which is especially large over posterior and occipital electrodes, which is likely beginning during the latter part of the N400 window.

N400 summary. In sum, the N400 analyses show that transposed letter nonwords elicited N400s that do not differ from their correctly-spelled counterparts in either amplitude or scalp distribution. This finding holds for both compound words as well as noncompound words, suggesting that the misspelled words contacted semantics equally as well as the correctly spelled words.

Figure 2. ERPs to noncompound target words (right) compared to those to compound words (left) in the correctly spelled condition (solid lines) and the TL condition (dashed lines), from target onset until 920 ms post-stimulus onset.



The P600 Time Window: 600-900 ms

Compound words. We will next examine the 600-900 ms epoch, in which we have predicted that we will see larger posterior positivities to the misspelled words as compared to the correctly-spelled words. To begin, we measured mean amplitudes over all 26 scalp electrodes from 600-900 ms after target word presentation. As can be seen in Figure 1, the two letter

transposition conditions elicited a large positivity relative to correctly-spelled compound words, which was largest over posterior channels and does not appear to differ between the TL-within and TL-between conditions. Results showed a main effect of spelling condition, $F(2,40)=6.93$, $p<.01$, as well as a significant interaction between spelling and electrode, $F(50,100)=3.25$, $p<.01$. Next, to explore the distribution of this effect, we conducted a distributional analysis over 16 pre-selected channels to equally represent the effects of hemisphere (left or right), laterality (medial or lateral), and anteriority (anterior to posterior). Results showed a main effect of spelling, $F(2,40)=7.13$, $p<.01$, as well a significant interaction between spelling and anteriority, $F(6,120)=25.99$, $p<.05$. No other effects reached significance (all $F<1.82$, all $p>.10$).

Importantly, this distributional effect shows that the effect of spelling was largest over the posterior channels, as evidenced by the interaction between spelling and anteriority. As such, we again chose the same subset of 15 posterior electrodes used for the N400 analysis and calculated mean amplitudes over just this group of electrodes to focus on the size of the effect in this more circumscribed scalp location. Results revealed a main effect of spelling condition, $F(2,40)=11.45$, $p<.001$, with no interaction with electrode, $F(28,560)=1.49$, $p=.19$. Planned comparisons were conducted to test for differences between the three spelling conditions, in order to test the critical prediction that between-morpheme letter transpositions would produce larger amplitude late positivities than would within-morpheme transpositions. This prediction was not supported by the data. While both the TL-between and TL-within conditions differed significantly from the correctly spelled compound words (between: $F(1,20)=16.56$, $p<.001$, within: $F(1,20)=13.07$, $p<.01$), the two letter transpositions did not differ from one another, $F(1,20)=.42$, $p=.52$. Furthermore, even choosing just a single channel that showed the largest difference between the TL-between and TL-within conditions (MiOc) found the comparison

between conditions at this channel was non-significant, $F(1,20)=1.52, p=.23$. As such, we can conclude that the late positive complex/P600 elicited by the TL-between and TL-within conditions was indistinguishable, suggesting that these two types of misspellings elicit similar processing consequences.

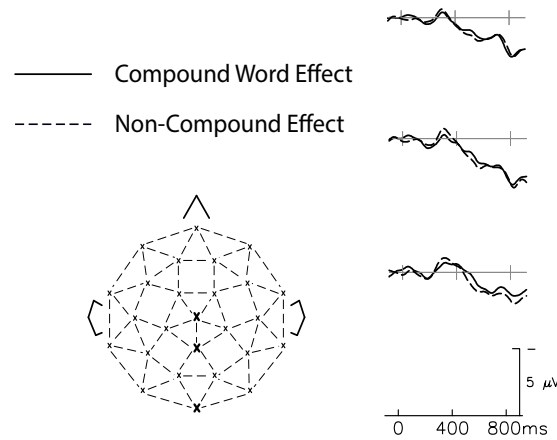
Non-compound words. We also assessed whether transposed letters in the non-compound words elicited late positives relative to correctly-spelled words. We started with an assessment of the effect over the entire head from 600-900 ms, and found a main effect of spelling, $F(1,20)=8.85, p<.01$, which interacted with electrode, $F(25,500)=4.78, p<.01$. The effect appears to be larger over posterior channels than frontal channels, as with the compound words. This observation is confirmed by a distributional analysis, conducted in the same way as described above, which showed a main effect of spelling, $F(1,20)=6.98, p<.05$, indicating larger positivities for the misspelled versus correctly spelled words. As for the compound words, the effect of spelling interacted with anteriority, $F(3,60)=7.77, p<.05$, being larger over the back of the head, as well as with laterality, $F(1,20)=6.23, p<.05$, being larger over more medial sites. For maximal comparability with the compound words, we again conducted follow-up analyses focused on the same set of 15 posterior channels. Results showed a main effect of spelling, $F(1,20)=15.45, p<.001$, as well as an interaction with electrode, $F(14,280)=4.00, p<.01$. The interaction with electrode site, despite the restricted range of electrodes used in the current analysis, likely reflects the difference in effect size between medial and lateral sites, as suggested by the distributional analysis.

Comparison across word types. A direct comparison between the mean amplitude of the compound and non-compound words is not possible, due to factors such as concreteness and cloze probability, which differed between the two types of words. In order to more clearly

highlight the effect of misspelling irrespective of overall amplitude of the waveforms, we will instead focus on the effect size within each wordtype by creating difference waves. Difference waves were created by computing a point-by-point subtraction of the waveforms elicited by the transposed letter condition from those elicited by the correctly spelled condition separately for each word type. In Figure 3, we can see that both difference waves show an obvious positivity (reflecting the difference between correctly spelled and misspelled words within each word type) that is overlapping at the posterior electrode positions, indicating that the difference between the misspelled and correctly spelled conditions is of the same magnitude across the two word types. Analyses conducted over the difference waves confirmed that there is no effect of spelling condition, $F(1,20)=.09$, $p=.76$, nor interaction between spelling condition and electrode, $F(14,280)=1.17$, $p=.33$. Even if we calculate the effect size at the one electrode for which there appears to be a difference between conditions, MiOc, the results indicate that there is no significant difference between these two difference waves, $F(1,20)=1.68$, $p=.21$.

LPC summary. In sum, the transposed letter nonwords elicited a late positivity relative to their correctly spelled counterparts, which was largest over posterior channels between 600-900 ms post-stimulus onset. Critically for our predictions, the TL-between and TL-within compound words elicited LPC components that were of equal magnitude, suggesting that these two spelling violations produced equivalent processing difficulties. Although compound words generally elicited more negative-going waveforms than the non-compound words, likely due to differences in lexical and sentential factors, the size of the *effect* elicited by the two misspelling conditions did not differ relative to their correctly spelled control words.

Figure 3. Difference waves, created by subtracting the waveforms elicited by the transposed letter condition from the correctly spelled condition, for the compound words (black lines) and the noncompound words (red lines), from target onset through 920 ms post-stimulus onset over three posterior channels.



Discussion

The current study investigated the differential effects of transposed letters that appeared either between or within morpheme boundaries of compound words, to test the assumption that compounds are recognized via decomposition into their constituent morphemes. Our results, however, found no evidence for the morphological decomposition of English compound words when presented within moderately constraining sentence contexts. This conclusion is based on the fact that compound words with transposed letters between-morpheme letter transpositions elicited ERP effects that did not differ from those elicited by compound words with within-morpheme transpositions. Specifically, we tested for a difference between these misspelling conditions on two different components, the N400 and P600, which are known to be sensitive to attempted lexico-semantic access and later reanalysis/repair processes, respectively. Overall findings showed that the amplitude of the N400 component elicited by both transposed letter compound word conditions was identical to that elicited by the correctly spelled condition. Furthermore, although the TL-between and TL-within conditions both elicited a late posterior

positivity relative to the correctly spelled compounds, the size of this effect was identical for both misspelling conditions. The implications of each of these findings will be discussed in turn below.

N400 Effects

We first chose to examine N400 amplitude as a measure the efficiency with which the target stimuli were able to make contact with the word's meaning in semantic memory. We predicted that if readers recognize compound words via decomposition into their constituent morphemes, then the between-morpheme transpositions would be more disruptive to processing than the within-morpheme transpositions, resulting in larger N400 amplitudes for the between than within morpheme condition. This prediction was not supported by the data: the TL-between and TL-within compound conditions elicited N400s that were identical to each other in terms of both amplitude, timing, and scalp distribution. Additionally, the N400s elicited by both misspelled compound word conditions were indistinguishable from those elicited by the correctly spelled compound words. This finding suggests that neither the presence of transposed letters nor their placement in the word relative to morpheme boundaries impeded the reader's ability to access the compound word's representation in semantic memory. Results for the *non-compound* words replicated these findings: the transposed letter filler words did not differ on any aspect of the N400 component from the correctly spelled filler words. Together, these findings support our conclusion that letter transpositions in both compound and non-compound words (matched on length and frequency) do not affect lexico-semantic access, as evidenced by the lack of N400 difference between the correctly spelled and transposed letter conditions for both word types.

The current findings can be used to help understand previous inconsistencies in the literature regarding the question of whether the N400 amplitudes elicited by misspelled words presented in a sentence context are larger than their correctly spelled base word (as found by Laszlo & Federmeier, 2009) or equal in amplitude (e.g., Vissers, et al., 2006; van de Meerendonk, et al., 2011). The current findings fell in line with those of Vissers and van de Meerendonk, raising the question of what properties of our stimulus set might cause the divergence of effects between our study and that of Laszlo and Federmeier (2009). First, in Laszlo and Federmeier (2009), the target words and pseudowords appeared in the sentence-final position of highly constraining sentences, whereas target words in the current study appeared sentence-medially in moderately constraining sentences. It could be the case that when readers are not making strong predictions about the target word, that viewing it in its correctly spelled form does not provide much facilitation as compared to viewing it with a transposed letter manipulation. However, the target words in Vissers et al. (2006) and van de Meerendonk, et al. (2011) were also in high-constraint contexts, so sentential constraint cannot fully account for the difference in effects.

Secondly, and possibly more importantly, target word length and neighborhood size differed dramatically between the stimulus sets. Target words used by Laszlo & Federmeier (2009) were on average 3.86 letters long, with an average orthographic neighborhood size of 11.3. For example, the pseudoword *wush* was used to replace the real word *wish*. Note that although *wush* is indeed a misspelled version of *wish*, it also has nine other real-word neighbors, so it could technically be a misspelled version of any of those words as well (e.g., wash, push, rush). It is therefore likely that *wush* would at least partially activate these other words—and that this sweep of activation through semantic memory to its many orthographically similar neighbors

could cause these words to elicit larger N400s than their correctly spelled versions. In contrast, the target words in the current study were long (7.9 letters on average) with very small neighborhood sizes (on average less than one orthographic neighbor per word). For example, the compound word *newspaper* has no orthographic neighbors, nor does its TL-between version, *newpsaper*. It is likely the case, then, that the non-word *newpsaper* would activate its base word the most, with little activation going to any other words, thus producing nN400 amplitudes of equal amplitude as the correctly spelled version. The stimuli used in Vissers et al. (2006) and van de Meerendonk et al. (2011), neither of which found N400 differences between correctly spelled and misspelled targets, were also long (6.2 letters on average) with very few neighbors (3 for words, 1.7 for pseudowords). Thus, we can hypothesize that when transposed letter misspellings of words with a small number of orthographic neighbors appear in a sentence context, the contextual support they receive in combination with few neighbor words produces activation largely for only the base word's representation, eliciting N400 amplitudes that are roughly in size as those for the correctly-spelled word.

P600 Effects

The compound words with transposed letters did elicit a large late posterior positivity compared to correctly spelled words, replicating previous studies that have also found an LPC/P600 effect to misspelled words in a sentence context (e.g., van de Meerendonk, et al., 2011; Vissers, et al., 2006). This component is known to be elicited in circumstances in which there is some sort of “reanalysis” or “repair” process that must take place—either due to the presence of overt grammatical errors (Hagoort, Brown, & Groothusen, 1993), garden path sentences, in which readers have to reinterpret their earlier assignment of a word to its grammatical role (Osterhout & Holcomb, 1992), difficult wh-constructions that are perfectly

grammatical but difficult to parse (Gouvea, Phillips, Kazanina, & Poeppel, 2010), or even from musical sequence violations (Patel, Gibson, Ratner, Besson, & Holcomb, 1998). In light of the varied circumstances that can give rise to this component, it has been suggested that it may generally index the integration of incoming stimuli into one's ongoing mental representation (Brouwer, Fitz, & Hoeks, 2012). This theory provides a unifying explanation for the presence of the P600/late posterior positivity in circumstances *without* over grammatical errors—like spelling errors—by positing that the amplitude of this component is increased any time a reader experiences difficulty integrating a word with its preceding sentence context. We would thus predict that if cross-morpheme letter transpositions were more disruptive to the integration of the compound words with their preceding context, that they would elicit larger late posterior positivities than would within-morpheme transpositions. Critically, however, this prediction was not supported by the data: the size of the positivity effect was identical for both the between-morpheme and within-morpheme conditions.

As a comparison condition, we also included non-compound words in the stimulus set, to test whether transposed letter effects might play out differently in these words. The reasoning was that if compound words are first decomposed into their constituent morphemes, then transpositions located at the word's morpheme boundary may be perceived more like word-external transpositions, which have been shown to cause more disruption to processing than those that occur on word-internal letters (Stites & Christianson, under revision; White, et al., 2008). For example, if *cupcake* is decomposed into *cup* and *cake*, a within-morpheme transposition would disrupt either the final two letters of *cup* (*cpu*) or the first two letter of *cake* (*ack*), which would be highly detrimental to the recognition of those words if they were presented in isolation. We would thus predict that the transposed letters in compound words

would produce greater costs than word-internal transpositions in non-compound words. With the filler words included as a comparison condition, we can test this prediction directly using difference waves. We created the difference waves by subtracting the correctly spelled condition from the transposed letter condition, separately for each word type. This calculation allows us to compare the size, timing, duration, and scalp distribution of the *effect* elicited by the misspelled words, while factoring out overall differences between the compound and non-compound stimuli. We found that the transposed letters were equally disruptive across the two word types: the late positivity elicited by the transposed letter non-compound words was identical to that elicited by the compound words with transposed letters, as evidenced by the difference waves. Thus, morpheme boundaries in compound words are *not* treated similarly to word-external boundaries, but indeed behave in a way that is indistinguishable from other word-internal letters in non-compound words, providing another piece of evidence against the morphological decomposition of compound words during recognition.

Implications of results with respect to the larger literature

Our findings fail to replicate Christianson et al. (2005), who found greater costs to naming times of compound words following masked primes containing between-morpheme letter transpositions relative to within-morpheme transpositions. There are, however, three important differences between their study and the current one. First, as described above, their results were obtained from a masked-priming naming task, in which subjects were presented with a masked prime (containing a TL or not) and their putative task was read the target compound word aloud as fast as possible. The difference observed in effects elicited across tasks suggests that the mechanisms subserving reading aloud may be highly sensitive to letter ordering, particularly when the letters are used in the service of phoneme ordering for speech

production (Coltheart et al). Mechanisms involved in word recognition, on the other hand, may be robust to slight variations in letter ordering, perhaps because of imperfect encoding of letter order as suggested by current models of letter position coding in visual word recognition (e.g., Davis, et al.; Gomez, et al.; Whitney, et al).

A second possible difference between the studies involves the *other* words that were present in the stimulus set. The filler items in Christianson et al. (2005) included many decomposable affixed words, as well as nonwords created by combining two morphemes from other compound words. The presence of a large number of affixed words created in this way may have encouraged readers to pay special attention to the composition of each word, resulting in greater emphasis on *decomposition* processes to aid in fluent production. In contrast, the 90 compound words used in the current study appeared with a wide variety of other stimuli. They were embedded in sentence contexts of anywhere from eight to 17 words, and intermixed with 50 filler sentences, varying in length from seven to 22 words. As such, the overwhelming majority of words in the current study were not compound words; and although there were many inflected and derived words, which are likely recognized via to decomposition processes (Rastle, Davis, & New, 2004), participants were not subject to the same stimulus characteristics as in Christianson et al. (2005) that could have strongly encouraged a strategy of decomposition.

Part 3: Young Adult Ambiguity Effects in Eye-tracking

Ambiguity is a core feature of language and, accordingly, there is now a large body of work, using multiple measures, that has looked at the processing consequences of lexical ambiguity when it is first encountered. Much less clear, however, are the downstream effects of ambiguity and the extent to which these are subject to context-based or individual differences. When are multiple meanings activated, and at what point (if ever) does one of them become suppressed? How effective is suppression when it is used? What representations are readers left with after they have seemingly disambiguated ambiguous words and moved on in the sentence? The literature looking at downstream effects after disambiguation is sparse (Gunter, Wagner & Friederici, 2003; Hagoort & Brown, 1994; Lee & Federmeier, 2012; Miyake, Just & Carpenter, 1994), and so these questions remain largely unanswered. In the present work, we use eye-tracking to conduct a first study of the downstream effects of ambiguity resolution by young adults during natural reading, in order to better understand what happens after the initial ambiguity costs are observed.

The bulk of previous research on the processing of ambiguous words has focused on noun/noun (NN) homographs, whose meanings fall within the same syntactic class (e.g., *calf*). An extensive body of research has explored the individual and joint effects of meaning frequency and context on the processing of NN homographs (for review, see Duffy, Kambe, & Rayner, 2001). One of the most robust findings in this literature is that encountering ambiguity often creates processing burdens in the form of increased reading time for ambiguous words relative to matched controls. For example, longer reading times are found on biased homographs when the preceding context instantiates the word's subordinate, less frequent meaning (the *subordinate bias effect*, Rayner, Pacht, & Duffy, 1994). Reading times are also longer on balanced

homographs with two equally frequent meanings, relative to unambiguous words, when preceded by neutral contexts (Rayner & Frazier, 1989).

Of particular interest to this study is how syntactic cues are used to resolve the cross-class ambiguity associated with noun/verb (NV) homographs, a single written wordform that has both a noun and a verb sense. A line of ERP work has used NV homographs to examine the processing consequences of cross-class ambiguity and to probe how syntactic and semantic cues are each used to resolve that ambiguity. Specifically, several studies by Federmeier and colleagues (Federmeier, Segal, Lombrozo, & Kutas, 2000; Lee & Federmeier, 2006, 2009, 2012) have found that NV homographs elicit a sustained negativity over frontal channels (relative to unambiguous words) when they appear in semantically neutral but syntactically constraining contexts. Importantly, this negativity is only present for NV homographs with two semantically distinct meanings (i.e., *to/the park*), but not when the two senses overlap in meaning (i.e., *to/the vote*) (Lee & Federmeier, 2006). When semantic constraints are available, NV homographs are processed in a qualitatively similar manner as unambiguous words -- that is, without eliciting the frontal negativity. Even in semantically unambiguous contexts, though, the dominant meaning of the ambiguous word remains somewhat active, as evidenced by larger N400 responses to subordinate (but not dominant) instantiations relative to unambiguous controls (Lee & Federmeier, 2009).

Given the prevalence of ambiguity effects for NV homographs in ERP work, and the failure of previous eye-tracking studies to find comparable costs for NV homographs (Folk & Morris, 2003), Stites, Federmeier, and Stine-Morrow (2013) approached this question in natural reading by presenting the stimuli from a previous ERP study (Lee & Federmeier, 2009) in an eye-tracking paradigm. We found effects parallel to those seen with ERP measures, in the form

of increased first fixation durations on the NV homographs in contexts with constraining syntax but lacking coherent semantics, but did not find similarly increased reading times in the semantically rich contexts. Interestingly, the first fixation and frontal negativity effects share several other features beyond their eliciting contexts. First, their onsets are aligned, both coming online around 200-ms after readers first apprehend the target word. Secondly, older adults as a group fail to elicit both the frontal negativity (Lee & Federmeier, 2011) and the first fixation effects. Rather, older adults spend more time *rereading* the NV homographs after leaving them in the first pass, suggesting incomplete initial ambiguity resolution. Finally, both effects showed a similar correlation with individual verbal fluency scores. Lee and Federmeier (2011) found a positive correlation between one's verbal fluency and the size of their frontal effect, and a median split showed that older adults in the high verbal fluency group showed a young-like pattern of effects (cf. Christianson, Williams, Zacks, & Ferreira, 2006). In Stites et al. (2013), both young and older adults with higher verbal fluency showed larger first fixation effects and smaller rereading effects, suggesting a tradeoff between early and late disambiguation that was related to their verbal fluency. Because verbal fluency tests are believed to at least partially reflect frontal lobe efficacy (Henry & Crawford, 2004; Stuss & Levine, 2002), the evidence so far suggests that the first fixation and frontal negativity effects both reflect the recruitment of frontally-mediated meaning selection mechanisms needed to disambiguate NV homographs in the absence of constraining semantics. Taken together, then, the behavioral (Rodd, Gaskell, & Marslen-Wilson, 2002), ERP (Federmeier et al., 2000; Lee & Federmeier, 2006, 2009, 2012), and eye-tracking (Stites et al., 2013) data provide strong support for the idea that in the absence of semantic constraints, additional processes must be recruited to aid ambiguity resolution.

Although the effects of ambiguity at the point it is first encountered have now been fairly extensively studied, much less clear are the consequences of ambiguity for downstream processing. Lee and Federmeier (2012) examined this question by presenting NV homographs in sentences that were preceded by a semantically neutral but syntactically constraining phrase (i.e., *Alan wanted to/the train...*). The target word was followed by a prepositional phrase, containing a noun that was plausible given only one of the homograph's senses. For example, the sentence *Alan wanted to train at the **gym** every day until he lost his excess weight* is plausible, whereas the sentence *Alan wanted the train at the **gym** every day until he lost his excess weight* is not. Thus, the downstream activation of the context-appropriate word can be probed naturally within the sentence context through a plausibility manipulation. This stands in contrast to the more traditional probe technique in which meaning activation is measured via the presentation of a word that is semantically related to one of the homograph's meanings, which Van Petten and Kutas (1987) found may actually produce backward priming of the related meaning rather than simply testing its current activation level. Lee and Federmeier (2012) hypothesized that if the frontal negativity reflects meaning selection, then N400 plausibility effects (i.e. reduced N400 amplitudes for plausible relative to implausible words) should be present at the downstream noun following both ambiguous and unambiguous words, indicating downstream activation of only the context-appropriate meaning. On the other hand, equal facilitation for the plausible and implausible nouns following ambiguous words would indicate that both meanings are still active, suggesting that the frontal negativity reflects meaning maintenance rather than selection.

Lee and Federmeier (2012) replicated the sustained negativity previously found to NV homographs, beginning around 200-ms after target word onset and continuing through the first two words of the following prepositional phrase. Importantly, at the downstream noun, they

found an N400 plausibility effect following both ambiguous and unambiguous words. For young adults, this effect was numerically smaller in nondominant contexts, suggesting residual activation of the dominant-related meaning, despite the unambiguous syntactic cues. Older adults as a group did not elicit the frontal negativity effect, and at the later noun they showed continued activation of the dominant meaning in both dominant- and nondominant-biased contexts. Furthermore, Lee and Federmeier (2012) also found that for both young and older adults, the size of one's frontal negativity effect at the homograph was positively correlated with the size of their plausibility effect at the later noun. This pattern suggests that the frontal negativity reflects suppression of the homograph's context-inappropriate meaning, making it less available downstream and increasing the size of the reader's N400 effect to words plausibly related to it.

The ERP pattern offers strong evidence that the frontal negativity effect reflects meaning selection, rather than meaning maintenance, of NV homographs in the absence of constraining semantics. By using eye-tracking during natural reading, the current study will extend these results and build on the evidence that the first fixation effect likely reflects the same underlying processes as the frontal negativity. The ERP evidence shows that the young adult brain can effectively suppress context-irrelevant meanings of words under conditions wherein words are presented serially and at a relatively slow pace. The use of eye-tracking affords the benefit of observing how readers behave when they determine what to look at, set the pace at which they take in information, and have the opportunity to look back in the text. For example, readers tend to skip function words. If a function word provides the only advance information about a homograph's context-appropriate meaning, will this natural tendency diminish readers' ability to disambiguate these words? Assuming readers do take in the syntactic information, do they take the extra time required to suppress the context-inappropriate meaning of ambiguous words and

when do they first initiate suppression? Finally, how will individuals' tendency to suppress or not affect their later processing of the downstream noun, including their rereading of earlier parts of the sentence?

We will answer these questions by using the stimuli from Lee and Federmeier (2012), giving us the opportunity to compare natural reading patterns with ERP effects previously elicited by the same contexts to look for correspondences and differences between measures. We predict to again see increased first fixation durations to the NV homographs relative to unambiguous words, conceptually replicating Stites et al. (2013). Given that we see the first fixation effect, we will test whether it is elicited with the same timing as the frontal negativity effect, in terms of both its onset and sustained nature. Critically, the current set of materials will also allow us to examine, for the first time, the downstream consequences of ambiguity resolution during natural reading. In particular, we will examine the nature of the plausibility effect at the later noun and its relationship (if any) to effects observed at the prior NV homograph itself.

Previous eye-tracking studies investigating plausibility have used a wide range of plausibility manipulations, from violating selectional or animacy restrictions of verbs (Warren & McConnell, 2007) to pairing unlikely instrument-verb-object triplets (Patson, & Warren, 2011; Matsuki, Chow, Hare, Elman, Scheepers, & McRae, 2011; Rayner, Warren, Juhasz, & Liversedge, 2004), to invoking a fantasy and/or fictional context for literally anomalous words (Filik, 2008; Warren, McConnell, & Rayner, 2008). Perhaps as a consequence, the literature does not provide a clear answer about when and how plausibility effects first come on-line in eyetracking measures. Some studies have reported effects of plausibility on first fixation and gaze durations on the violating word itself (Matsuki et al., 2011; Patson & Warren, 2011

[Experiment 2]; Warren, McConnell, & Rayner, 2008), whereas others have found that plausibility effects emerge only on later reading measures on the violating word or fixations on the next region in the sentence (Filik, 2008; Patson & Warren, 2011 [Experiment 1]; Rayner, et al., 2004; Warren & McConnell, 2007). In contrast, ERP studies that manipulate plausibility find robust N400 plausibility effects elicited by the violating word itself, regardless of the type of violation used (Filik & Leuthold, 2008; Lee & Federmeier, 2012; Nieuwland & Van Berkum, 2006; Paczynski & Kuperberg, 2012) -- but in a timeframe (200-500 ms post-stimulus-onset) that, in natural reading, would span early gaze measures on the violating word and later gaze measures/fixations onto subsequent words. For the purposes of the current study, the pattern of plausibility effects, rather than the timing of their onset, is of import; therefore, we will combine the violating word and its subsequent spillover word into a single region, allowing us to capture the widest range of plausibility effects.

Of critical interest for the present study is whether the plausibility effects observed downstream will differ following ambiguous and unambiguous words. Based on our evidence that the first fixation effect, like the frontal negativity effect, reflects meaning selection, we predict that we will see plausibility effects of roughly equal magnitude following ambiguous and unambiguous words, as Lee & Federmeier (2012) observed in their ERP study. Importantly, to the extent that the first fixation and frontal negativity effects both reflect meaning selection of the NV homograph, we also predict that we will see a relationship between the size of the first fixation and plausibility reading time effects for sentences containing ambiguous words. If we find a positive correlation between these effects in eye-tracking, as was observed in the ERP data, we can feel confident that these two effects likely reflect a similar underlying source and

functional role, namely the recruitment of frontally-mediated selection mechanisms to aid in difficult ambiguity resolution.

Methods

Participants

Twenty-four University of Illinois undergraduate students (19 females and 5 males) with a mean age of 19 years (range: 18-21) participated in the study for course credit. All were native English speakers and were right-handed as assessed by the Edinburgh inventory (Oldfield, 1971)

Materials

The materials and design employed were identical to that in Lee and Federmeier (2012) to allow for the closest comparison between the two studies. Two word types were used: noun/verb homographs, which have both a noun and a verb meaning (i.e. *trip*), and matched unambiguous words. Of the 48 NV homographs used in the study, 33 were considered biased whereas the other 15 were considered balanced (see below for the calculation of meaning bias). Target words were presented in sentence-medial position preceded by a semantically neutral but syntactically constraining phrase (i.e., *John hated to/the trip...*). They were followed by a prepositional phrase, which contained a noun that was either plausible or implausible given the target word (frequency: $M=91.5$, range=0-909; length: $M=5.7$, range=3-11). For sentences containing NV homographs, the noun could be related to either meaning of the homograph. The sentences were made implausible by swapping the syntactic cue preceding the homograph (i.e. *to* and *the*) while leaving the rest of the sentence identical (see Table 5 for example stimuli). This change rendered the noun implausible with the context-appropriate meaning, but still plausible with the context-*inappropriate* meaning of the target word. As such, the same set of words can

serve as both plausible and implausible continuations following the ambiguous words. Stimuli were controlled so that there was little semantic association between the target word and the subsequent noun (according to the University of South Florida free association norms, both forward and backward association strength was 0.01). Similar plausible and implausible sentences were created for the unambiguous words as well, by replacing the plausible noun with an implausible one, of similar length (plausible: $M=6.3$, range=3-13; implausible: $M=6.2$, range=3-13; $t(95)=-0.45$, $p=.65$) and frequency (plausible: $M=106$, range=0-1207; implausible: $M=116$, range=0-1207; $t(95)=0.40$, $p=.69$).

Table 5. Examples of critical used in Experiment 3.

Word Type	Plausibility	Example Sentence
NV Homograph	Plausible	Jared wanted the <i>sink</i> in his <u>bathroom</u> to be repaired.
	Implausible	Jared wanted to <i>sink</i> in his <u>bed</u> to take a long nap.
Unambiguous Words	Plausible	Jared wanted to <i>sink</i> in his <u>bathroom</u> to be repaired.
	Implausible	Jared wanted the <i>sink</i> in his <u>bed</u> to take a long nap.
	Plausible	Joseph started to <i>appear</i> in the <u>show</u> since last season.
	Implausible	Karen expected the <i>member</i> of the <u>band</u> to bring his own instrument.
	Plausible	Joseph started to <i>appear</i> in the <u>mistakes</u> since last season.
	Implausible	Karen expected the <i>member</i> of the <u>mother</u> to bring his own instrument.

Each ambiguous and unambiguous word was repeated once, with a lag of at least 50 sentences and with different continuations. To make the plausibility of the sentence unpredictable, across lists the two appearances of the word were either both plausible, both implausible, or one of each. The order of each pair of sentences was also counterbalanced, creating a total of eight lists. Each participant read 192 sentences, with equal numbers of sentences containing ambiguous and unambiguous words, each with the same number of plausible and implausible continuations. Additionally, ambiguous words were used equally in

their noun and verb sense, and the same number of unambiguous nouns and verbs appeared as well (although the factor of grammatical class will not be examined in this study).

Ambiguous and unambiguous target words were matched on a variety of lexical factors, including log frequency (Kucera & Francis, 1967), word length, and usage-specific concreteness (Lee & Federmeier, 2008; Table 1). The sentential constraint up to the noun of the prepositional phrase, as well as the cloze probability of the noun itself, did not differ across conditions. To ensure the validity of the plausibility manipulation, the plausibility of the downstream nouns was rated as over 75% for the plausible sentences and less than 35% for the implausible sentences (see Lee & Federmeier, 2012 for further details about the norming procedures used to create the stimuli).

In order to compare effects based on meaning dominance, a dominance score was calculated for all of the NV homographs. To do so, total log frequency counts for a homograph's usage in both its noun and verb sense were collected from WordNet 3.0 (Princeton University, 2010; < <http://wordnet.princeton.edu>>) and compared against the word's total usage. The dominance score was created by subtracting the log frequency of a word's noun usage from its verb usage, and dividing this amount by the total frequency count, which produced a bias score ranging from -1 to 1. For a given item, if the dominance score was greater than $|\cdot 3|$, it was considered biased; if its bias score fell between $-.3$ and $\cdot 3$, it was considered balanced. Using this procedure, 33 of the 48 NV homographs were classified as biased, whereas the other 15 were balanced. The average dominance score for the biased words was $\cdot 69$, whereas the balanced words scored an average of $\cdot 13$. For 20 of the biased words, the noun sense was the dominant meaning, whereas for the other 13 biased words, the verb sense was dominant.

Procedure

Subjects were tested individually in a quiet room, seated 71 cm away from the Viewsonic 19-inch CRT monitor. Eye movements were recorded using an SR Research Eyelink 1000 desk-mounted eye-tracker with a sampling rate of 1000 Hz. A chinrest was used to encourage participants to minimize head movements during the experiment. To ensure accurate tracking, a nine-point calibration procedure was used at the start of the experiment, and after every break or movement out of the chinrest. Additionally, a drift-check was performed at the start of every trial, and subjects were recalibrated if the drift-check failed. Recordings were monocular, taken from the right eye.

At the start of the experiment, subjects were presented with written instructions, after which they received seven practice trials to acclimate them with the experimental procedure. Each trial began with the presentation of a drift check target in the upper left corner of the screen. Subjects fixated the target while pressing the “start” button on a hand-held controller. At this time the sentence appeared, which was presented in left-justified, black Courier New font on a light gray background. Three characters subtended one degree of visual angle. Subjects were instructed to read the sentence normally, and press the “advance” button on the controller when they were finished reading the sentence. After each sentence, a probe word appeared in uppercase red font in the center of the screen. Subjects indicated via button-press if the word was presented in the immediately preceding sentence or not. Half of the probe words were new; the old probe words were chosen from syntactic cues, NV-homographs and matched unambiguous words, critical downstream nouns, and the rest of the content words in the sentence.

The experiment was divided into four blocks of 48 trials, each of which took approximately 8 minutes to complete. At the end of each block, subjects were allowed to remove

their heads from the chinrest. A paper-and-pencil sentence recognition test was also administered at this time. The sentence recognition test consisted of 96 sentences total, half of which were new; the old sentences were drawn equally from plausible and implausible sentences containing NV homographs and matched unambiguous words. Participants were told to mark which sentences were present in the immediately preceding block. Because many of the same words were used across the new and old sentences, simple word recognition would not allow good performance on this test. The word and sentence recognition tasks were included to keep task demands as similar as possible to Lee and Federmeier (2012), as well as to encourage subjects to attend closely to every sentence presented, even implausible ones.

Data Analysis

Within the Eyelink 1000 data analysis package, consecutive fixations less than 80 ms in duration and less than .5° apart in visual angle were combined into one fixation. Single fixations that were shorter than 80 ms or longer than 800 ms were then automatically excluded. Fixations shorter than 80 ms are unlikely to represent meaningful cognitive processing (Rayner, 1998), and fixations longer than 800 ms are likely the result of cases in which the tracker temporarily lost the eye, causing inaccurately inflated reading times. Additionally, 1.1% of trials were excluded from the analysis due to track loss or program error.

Results

Behavior

The purpose of the word and sentence recognition tasks was to encourage subjects to pay close attention to the experiment at both the word and sentence level. Overall, subjects answered the word recognition task with 91% accuracy ($d'=2.8$). Performance on the sentence recognition task was also good, at 82% ($d'=2.2$). Both of these values are comparable to how the young adults in Lee and Federmeier (2012) scored on these same tasks: average accuracy of 94% ($d'=3.4$) on the word recognition task and 86% ($d'=2.5$) on the sentence recognition task.

Eye-tracking

Based on the findings of Lee and Federmeier (2012), several areas of interest were examined in the current dataset. First, reading times on the target ambiguous word will be considered, to determine if there are ambiguity effects for the NV homographs relative to the unambiguous words, particularly on first fixation duration as observed by Stites et al. (2013). Next, we will consider reading times on the prepositional phrase following the target words, to look for continuing ambiguity effects as a corollary to the sustained negativity observed by Lee and Federmeier (2012). Reading times on the noun region (including the noun in the prepositional phrase and the word immediately following it) will then be examined to test for a plausibility effect (i.e. longer reading times) on the implausible relative to plausible words, as well as an interaction between plausibility effects and preceding target word ambiguity.

Several standard eye-tracking measures will be employed: first fixation duration, the length of the readers' first fixation on the word/region; gaze duration (a.k.a. first pass time), the sum of all fixations on a word/region the first time it is fixated, before the eyes leave it in either direction; go-past time, a right-bounded measure including all fixations a reader makes on a

word/region, beginning with their first fixation until they move past it to the right (including rereading earlier parts of the sentence); and rereading time, all time spent reading a word/region after leaving it in the first pass. For all regions of interest, two repeated measures Analyses of Variance (ANOVA) were conducted, one averaging over subjects (F_1) and the other averaging over items (F_2). For the by-subjects analysis, both the ambiguity/dominance of the target word and the plausibility of the downstream noun (when applicable) were within-subject factors. For the by-items analysis, the identity of the NV homograph and/or matched unambiguous control will be considered as the item over which to average (considering as separate lexical items the dominant and subordinate instantiations of the biased homographs, and the noun/verb meaning of the balanced homographs), making ambiguity/dominance a between-items factor. Because all lexical items were used both plausibly and implausibly, the factor of plausibility is a within-items factor.

We will present first-pass reading time measures (first fixation, gaze duration, and go-past time) separately from rereading time measures. This will allow us to present the results in a manner that follows naturally from the ERP study, because readers only receive a “first pass” through the sentence when using the RSVP presentation style used in ERP research. Furthermore, the factor of plausibility does not become apparent until readers reach the noun region. As such, plausibility must be included as a factor in analyses of rereading times, but not first pass measures, of the target region.

First-Pass Reading Measures

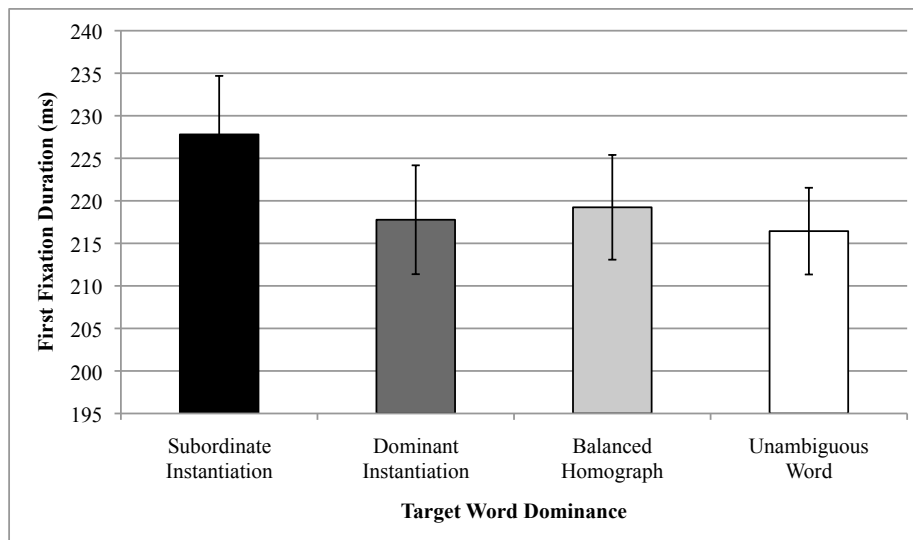
Target Word. Reading times on the target word (listed in Table 6) showed a significant ambiguity effect for first fixation durations, $F_1(1,23)=5.60, p<.05$, $F_2(1,94)=4.59, p<.05$, with longer first fixations for ambiguous relative to unambiguous words. As in Stites et al. (2013), this

effect was driven by cases in which subjects landed directly on the target word without first fixating the preceding function word, $F_1(1,23)=8.70, p<.01$, $F_2(1,94)=6.85, p=.01$. When subjects landed on the function word first, as they did in 42.5% of trials, there was no ambiguity effect on their subsequent first fixation to the target word, F_1 and $F_2<1$. This is likely because the function word was short enough to allow readers to process both the function and target word simultaneously while fixating it, giving readers the opportunity to disambiguate the target word on their previous fixation. An alternative explanation, that there was insufficient power to detect an effect on trials in which subjects had a preview of the target word, seems unlikely given that there were still 1757 first fixations in this condition (of 4125 total first fixations to the target words). There was no ambiguity effect on target word gaze duration, $F_1(1,23)=1.73, p=.20$, $F_2(1,94)=1.18, p=.28$. These findings replicate Stites et al. (2013), who found increased first fixation durations (but not gaze durations) for NV homographs compared to unambiguous words in syntactically constraining contexts that lacked coherent semantics, which were the same contexts that also elicited the sustained frontal negativity to the NV homographs (Lee & Federmeier, 2009, 2012). Additionally, there was an ambiguity effect for go-past time, $F_1(1,23)=7.48, p<.05$, that did not reach significance by items, $F_2(1,94)=1.86, p=.18$, indicating that readers spent more time rereading earlier parts of the sentence before moving past ambiguous words than before moving past unambiguous words.

In addition to overall ambiguity effects, we further explored the first fixation effect as a function of the target word's meaning dominance (dominant, subordinate, balanced, or unambiguous). As can be seen in Figure 4, first fixations on the subordinate instantiations of the NV homographs ($M=228, SD=34$) were substantially longer than the other three conditions, which were very similar to one another (dominant: $M=218, SD=31$; balanced: $M=219, SD=30$;

unambiguous: $M=216$, $SD=25$). Omnibus ANOVAs revealed a main effect of dominance, $F_1(3,69)=5.02$, $p<.01$, $F_2(3,140)=3.42$, $p<.05$, indicating an overall difference between conditions. Follow-up tests were conducted to determine whether the first fixations were especially inflated for the subordinate instantiations compared to the dominant ones, which would be the first piece of evidence for a subordinate bias effect for NV homographs. A paired samples t-test revealed that first fixations were longer in the subordinate than dominant condition, an effect that was significant by subjects, $t_1(23)=3.07$, $p<.01$, but failed to reach significance by items (95% CI 8.09 ± 10). Based on these analyses, we can conclude that the first fixation ambiguity effect was largely driven by increased first fixations to the subordinate instantiations, which is the first time that the subordinate bias effect has been shown for NV homographs.

Figure 4. Target word first fixation duration (by Target Word Dominance)



Prepositional Phrase. To investigate the possibility of sustained ambiguity effects as in the ERP study, we next examined reading times on the prepositional phrase following the target word. Because the prepositional phrase consisted of two short function words, which have an estimated skipping rate of 65% in normal reading (Rayner, 1998), the two-word phrase was

considered as a single region for all analyses. Results showed a significant ambiguity effect on the first fixation to the prepositional phrase region, $F_1(1,23)=5.74, p<.05$, $F_2(1,94)=4.18, p<.05$, with longer reading times following ambiguous than unambiguous words (see Figure 5). There was no ambiguity effect for gaze duration on the region, F_1 and $F_2 < 1$, although there was again a significant ambiguity effect for go-past time by subjects, $F_1(1,23)=7.33, p<.05$, that did not reach significance by items, $F_2(1,94)=2.01, p=.16$. The first fixation ambiguity effect on the prepositional phrase mirrors the sustained nature of the frontal negativity effect in the ERP data, which began on the target word and extended through the prepositional phrase region as well. Given that gaze durations on the prepositional phrase itself were not longer following ambiguous words, and considering that go-past time includes reading times on the prepositional phrase *and* everything to its left before readers move forward in the text, the increased go-past times on the phrase reflect time spent rereading earlier parts of the sentence before moving forward, possibly on the NV homograph or its preceding function word.

The failure of certain by-items statistical tests to reach significance could be due to the way in which bias scores were calculated for each item. As it stands, dominance was calculated using a ratio between the homograph's most common noun and verb meanings. However, if the sum total of the frequencies of all of an item's noun and verb meanings were used to calculate this bias score instead of the frequency of its most common meaning, it would alter the nominal dominance for 12 of the 48 homographs. It is possible that the language processing system weighs item frequencies differently than the way they were calculated in the current study, or could even do so differently for different items (i.e., weighing all frequencies for certain homographs and only the most common frequencies for others), which could produce inconsistencies in the effects observed at the item level.

Figure 5. First fixation durations on target word and prepositional phrase region.

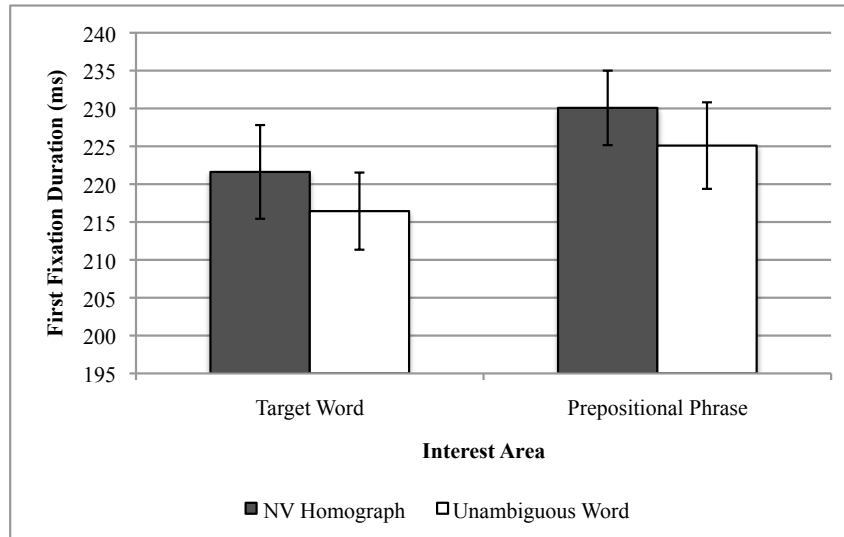


Table 6. First Pass Reading measures (and standard deviations) for the three regions of interest.

Region	Ambiguity	First Fixation		Gaze Duration		Go-Past Time	
Target	Ambg.	222 (30)		251 (40)		310 (56)	
Word	Unamb.	216 (25)		255 (42)		297 (52)	
Prepositional Phrase	Ambg.	230 (24)		323 (63)		390 (93)	
	Unamb.	225 (28)		322 (66)		369 (84)	
Downstream Noun Region	Ambg.	Implaus	Plaus	Implaus	Plaus	Implaus	Plaus
		226 (25)	218 (18)	407 (104)	385 (97)	527 (139)	478 (124)
	Unamb.	228 (28)	221 (23)	429 (104)	399 (95)	551 (155)	473 (106)

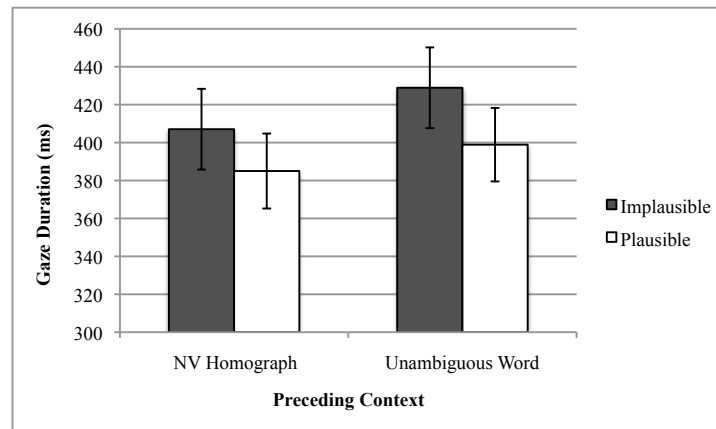
Downstream Noun Region. Next, first-pass reading times were examined for the downstream noun to test for plausibility effects. Recall that the implausible words following the NV homographs were in fact *plausible* with the homograph's context-inappropriate meaning, so longer reading times in this condition would imply successful suppression of the alternate meaning. If, on the other hand, readers were unable to suppress the context-inappropriate meaning of the NV homograph, we would expect reading times on plausible and implausible words to be similarly facilitated following the homographs, indicating continued activation of the

alternate meaning. There were two factors of interest in this region: the ambiguity of the preceding target word and plausibility of the noun itself. Overall, first-pass reading times on the downstream noun region show significantly longer reading times on implausible relative to plausible nouns, with no consistent effects of preceding ambiguity or interaction between the two factors (see Table 6, and Figure 6). First fixations to the target region exhibited a significant effect of plausibility, $F_1(1,23)=11.40, p<.01, F_2(1,94)=10.85, p<.01$, with no effect of ambiguity, $F_1(1,23)=1.35, p=.26, F_2<1$, or interaction, F_1 and $F_2<1$. Gaze durations again showed a main effect of plausibility, $F_1(1,23)=25.02, p<.01, F_2(1,94)=13.95, p<.01$, as well as a main effect of ambiguity that was significant by subjects, $F_1(1,23)=13.94, p<.01$, but not items $F_2(1,94)=1.52, p=0.22$, and no interaction, F_1 and $F_2<1$. Finally, go-past times showed a similar pattern, in which only the effect of plausibility was significant, $F_1(1,23)=26.37, p<.01, F_2(1,94)=25.78, p<.01$, with no effect of ambiguity, $F_1(1,23)=1.60, p=.22, F_2<1$, nor interaction, $F_1(1,23)=1.807, p=.19, F_2(1,94)=1.29, p=.26$.

Readers appear to quickly appreciate the plausibility of the downstream noun as reflected by increased first pass reading times on these words, and this initial plausibility assessment is unaffected by whether it was preceded by an ambiguous word. This is an important point to note, as the implausibility of the downstream noun following ambiguous words was only apparent if readers sufficiently suppressed the context-inappropriate meaning of the NV homograph. The observed pattern of results mirrors the findings from Lee and Federmeier (2012), in which young adults showed N400 facilitation for plausible relative to implausible nouns following both ambiguous and unambiguous words. The lack of an interaction between ambiguity and plausibility across both studies indicates that readers were able to use syntactic cues alone to

suppress the context-inappropriate meaning of the NV homograph well enough to recognize the implausibility of the later nouns in the ambiguous context.

Figure 6. Gaze duration on the downstream noun region, split by preceding ambiguity.



Correlations between ambiguity effects and plausibility effects.

Lee and Federmeier (2012) found in their ERP study that the size of a subject's frontal negativity effect was positively correlated with the size of their downstream plausibility effect. This correlation provided evidence that the frontal negativity reflected meaning selection (rather than maintenance), because a larger initial ambiguity effect made the context-inappropriate meaning of the ambiguous word less available later in the sentence. We thus predicted that if the first fixation ambiguity effect reflects the same underlying processes as the frontal negativity, then the size of a reader's first fixation ambiguity effect would similarly predict the availability of the word's context-inappropriate meaning later in the sentence.

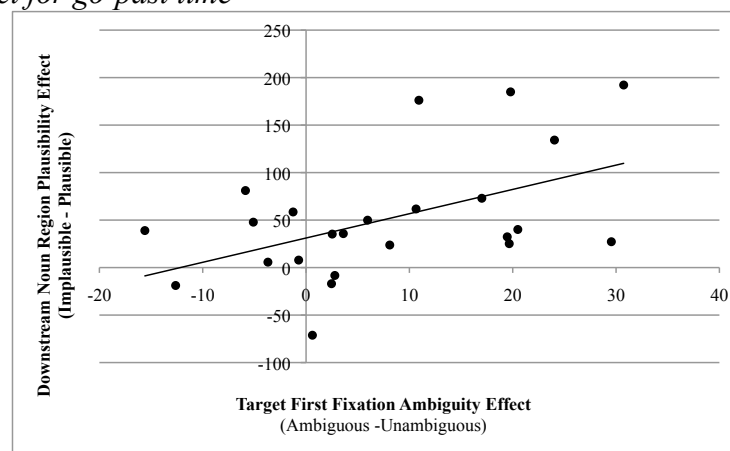
In order to test this prediction, we first calculated an ambiguity effect for each subject by subtracting average first fixation durations to the unambiguous from the ambiguous targets, producing a positive value for longer ambiguous reading times. For this measure, we included only first fixations in which subjects landed directly on the target, because this subset of the data showed clearer ambiguity effects than cases in which they landed on the function word first. We

then calculated three separate plausibility effects for each subject, one for each of the three first-pass reading measures on the downstream noun region following ambiguous words only. Each of these plausibility effects were created by subtracting reading times on the plausible noun regions from the implausible noun regions, again producing a positive number for longer implausible reading times. This procedure was carried out separately for first fixation, gaze duration and go-past times. To test for a relationship between the early and downstream effects, we calculated the correlation between the first fixation ambiguity effect and each of the three downstream plausibility effects separately. Results showed a significant positive correlation between the first fixation ambiguity effect and the go-past time plausibility effect on the downstream noun region ($r=.49$, one-tailed $p=.006$), as can be seen in Figure 7. There was no significant relationship between the ambiguity effect and the other two reading time measures on the noun region (first fixation: $r=.12$, one-tailed $p=.28$; gaze duration: $r=.09$, $p=.32$). Because go-past time includes all reading of the downstream noun region before moving forward in the sentence as well as *rereading* of earlier parts of the sentence, it captures multiple patterns of plausibility effects, including slowing down on the implausible word itself and/or launching a regression to an earlier part of the sentence. These different patterns of plausibility effects are missed in first fixation or gaze duration alone.

Importantly, we wanted to ensure that this correlation was specific to the cases following ambiguous words only rather than reflecting a general tendency of certain readers to slow down for all difficulties equally (i.e., for both ambiguous and implausible words across the board). To do so, we created a plausibility effect for go-past times on the noun region in the *unambiguous* contexts. We then calculated the correlation between this new plausibility effect and the first fixation ambiguity effect, and found no significant relationship between these measures ($r=-.009$,

one-tailed $p=.48$). Finally, to directly test that the size of the correlation between the first fixation ambiguity effect with the go-past plausibility effect following *ambiguous* words is different from its correlation with the same effect following *unambiguous* words, we calculated Steiger's Z (Meng, Rosenthal, & Rubin, 1992) for these two correlations. This test statistic compares the size of two non-independent correlations that share a common variable (i.e., the first fixation ambiguity effect), taking into account the size of the correlation between each of the two unique variables with the shared variable, as well as the correlation between the two unique variables themselves (i.e., go-past time plausibility effect following *ambiguous* vs. *unambiguous* words). The output of this test confirmed that these two correlations are statistically different from each other, $Z=1.98$, $df=21$, $p<.05$. Thus, we can conclude that the first fixation ambiguity effect is selectively correlated with go-past time plausibility effects following *ambiguous* words only, suggesting that the first fixation effect indexed a suppression process that left the context-inappropriate meaning less available downstream. This correlation mirrors that observed in the ERP study, and provides further evidence that the first fixation reading time effect reflects the recruitment of meaning selection mechanisms needed to disambiguate NV homographs in the absence of constraining semantics.

Figure 7. Correlation between target word first fixation ambiguity effect and downstream noun region plausibility effect for go-past time



Later Reading Time Measures

In Stites et al. (2013), we found that certain readers were more likely to reread the NV homograph after leaving it in the first pass, specifically older adults as a group, as well as young adults with low verbal fluency. These rereading patterns were taken as evidence of incomplete ambiguity resolution, resulting in the need to return to the NV homograph after having moved on to later parts of the sentence. In light of these results, we also examined rereading times for the target word in the current study as another way to assess ambiguity effects. Rereading times will be considered for a target word *region*, which we have defined to include both the target word and the class-disambiguating function word immediately preceding it, as readers may choose to return to either word (or both) to gather more information about the homograph's intended sense. Ambiguity and plausibility are both included as factors to investigate whether downstream plausibility had a differential effect on rereading times to the target word region. We calculated rereading time by subtracting gaze duration from total time, excluding all first-pass reading times from the analyses.

Overall rereading times on the target region were longest for the ambiguous words in implausible contexts ($M=158$, $SD=118$), followed by ambiguous words used plausibly ($M=106$, $SD=67$), unambiguous words used implausibly ($M=96$, $SD=75$), and finally, unambiguous words used plausibly, ($M=90$, $SD=73$). Results showed a main effect of ambiguity, $F_1(1,23)=38.12$, $p<.01$, $F_2(1,94)=27.55$, $p<.01$, reflecting the overall longer rereading times for ambiguous than unambiguous words, a main effect of plausibility, $F_1(1,23)=10.49$, $p<.01$, $F_2(1,94)=13.74$, $p<.01$, as well as a significant interaction between these two factors, $F_1(1,23)=6.12$, $p<.05$, $F_2(1,94)=8.13$, $p<.01$. Follow-up t-tests examining the plausibility effects within each ambiguity condition confirm that rereading times were significantly longer for implausible than plausible

ambiguous regions, $t_1(24)=3.05$, $p<.01$, $t_2(47)=4.89$, $p<.01$, but did not differ for the unambiguous regions, $t_1(24)=1.04$, $p=.31$, $t_2(47)<1$. These rereading time measures show that in the implausible sentences containing ambiguous words, readers recognize the NV homograph as the source of the implausibility (because the implausible downstream noun would have been plausible with the *other* meaning of the homograph), and selectively return to the homograph in these sentences. However, the eventual implausibility of the sentence did not affect rereading time for the unambiguous target regions (likely because readers recognized that the implausibility of this noun did not hinge on the target word in these cases).

Based on the subordinate bias effect observed on first-pass reading measures, rereading times on the ambiguous target word region were also examined as a function of meaning dominance (see Table 7). Specifically, we were interested in whether readers continued to experience difficulties with the subordinate instantiations of the NV homographs, or whether the increased reading times during the first pass allowed readers to fully select the subordinate meaning, possibly leading to later reading times that are indistinguishable from the dominant condition. An ANOVA with the factors of plausibility (2: implausible, plausible) and meaning dominance (3: dominant, subordinate, and balanced) showed again that rereading times were generally longer for targets in implausible than plausible contexts, $F_1(1,23)=10.07$, $p<.01$, $F_2(1, 93)=18.42$, $p<.01$. There was also a main effect of dominance condition that was significant by subjects and marginal by items, $F_1(2,46)=3.70$, $p<.05$, $F_2(2, 93)=2.37$, $p=.10$, as well as an interaction between the two, $F_1(2,46)=3.79$, $p<.05$, $F_2(2,93)=3.09$, $p=.05$.

Follow-up t-tests were conducted to test for the significance of the plausibility effect within each of the four dominance conditions. As can be seen in Figure 8, rereading times were significantly longer for implausible than plausible targets in both the dominant and balanced

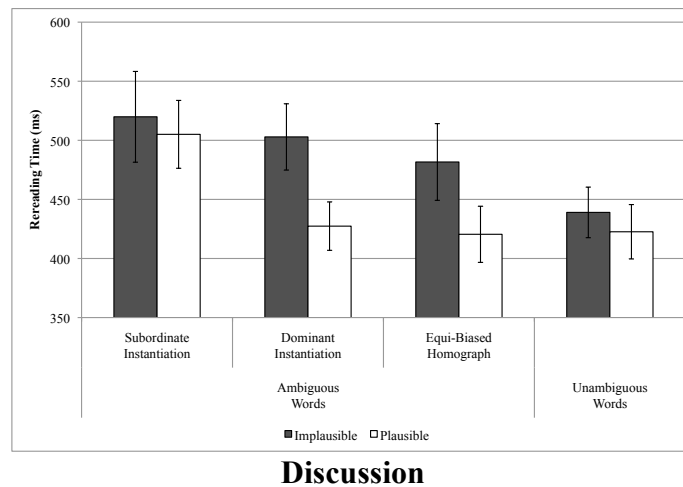
conditions (dominant: $t_1(23)=4.55, p<.01, t_2(32)=4.18, p<.01$; balanced: $t_1(23)=2.61, p<.05, t_2(29)=3.49, p<.01$), while there was no plausibility effect within either the subordinate or unambiguous conditions (subordinate: $t_1(23)=0.98, p=.34, t_2(32)=0.46, p=.65$; unambiguous: $t_1(23)=1.04, p=.31, t_2(47)=0.58, p=.57$). Focusing just on the comparison between the dominant and subordinate conditions, we can see that their implausible reading times only differed by two milliseconds, $t_1(23)=0.14, p=.89$. The striking difference was with respect to the plausible rereading times, which were 56 ms longer for the subordinate than dominant condition, $t_1(23)=2.89, p<.01$ ¹. Even when the target was eventually followed by a noun that was plausible given the context-appropriate meaning of the NV homograph, readers still reread the target word region for just as long as they did when it was followed by an implausible word.

Table 7. Target region rereading times (and standard deviations) by target word dominance

Dominance	Rereading	
	Implausible	Plausible
Dominant	168 (119)	88 (65)
Subordinate	165 (136)	144 (105)
Balanced	149 (87)	87 (68)
Unambiguous	96 (75)	90 (73)

¹ While a similar t-test could not be conducted by-items (because it constitutes a between-items test), we instead calculated Tukey's HSD using the MSW from the by-items ANOVA ($q=3.68, MSW=5695, n=140$), which produced a cutoff value of 51 ms for a significant difference at the $p=.05$ level. We can see that the 2 ms difference between the subordinate implausible and dominant implausible conditions is not significant, whereas the 56 ms difference between the subordinate plausible and dominant plausible is significant.

Figure 8. Target word region rereading time, split by Target Word Dominance.



First-Pass Ambiguity Effects

The current study investigated online lexical ambiguity resolution and its consequences for downstream processing. We expected to replicate our finding that ambiguous words show longer first fixation durations than unambiguous words when only syntactic cues are available for disambiguation (Stites et al., 2013), and indeed we did. First fixation durations to NV homographs were longer than those to unambiguous words when the preceding context contained syntactic cues to disambiguate the homograph and were semantically coherent, but not semantically constraining. These same contexts have previously elicited sustained frontal negativity effects in young adult readers (Lee & Federmeier, 2009, 2012). We furthermore predicted that, if the first fixation and frontal negativity effects share the same neurocognitive source, then they should have a similar timecourse in terms of both their onsets and duration. This prediction was also supported by the data: both effects come online around 200 ms after the NV homographs were first apprehended, and both are sustained through the prepositional phrase region. This study now marks the second to show parallel first fixation and frontal negativity effects to NV homographs that are elicited from the same contexts and with a similar timecourse,

suggesting that these two effects likely reflect the same frontally-mediated selection processes that are recruited in the face of difficult meaning selection.

Additionally, this is the first study to show a subordinate bias effect for NV homographs. Even though the homographs were preceded by a grammatical cue that unambiguously instantiated the subordinate meaning, this syntactic cue alone was not enough to allow selection of the subordinate meaning without eliciting the first fixation effect. This same cue, however, did not elicit the first fixation effect when it instantiated the dominant meaning, suggesting that readers were able to select the more frequent meaning without recruiting the additional meaning selection mechanisms. While the subordinate bias effect has been robustly demonstrated for noun/noun (NN) homographs (Rayner, Pacht, & Duffy, 1994), we are the first to show the effect for NV homographs as well. Furthermore, the balanced ambiguous words failed to show a significant increase in first fixation durations regardless of which meaning was selected by the context (i.e., noun or verb). Previous work has found robust ambiguity effects on balanced NN homographs when preceded by contexts that are semantically neutral with respect to the intended meaning, as in the current study. However, the preceding context in previous studies was truly unhelpful because both of the word's meanings fall within the same syntactic category. Our results thus suggest that when the two meanings of a *balanced* homograph are in different syntactic categories (i.e., nouns and verbs), readers can utilize a syntactic cue alone (*to/the*) to select the context-appropriate meaning. In sum, readers appear able to successfully use a syntactic cue to select the dominant meaning of a biased NV homograph, or either meaning of a balanced one, but are unable to use syntax alone to select the subordinate sense of the word without recruiting the additional processing mechanisms indexed by the first fixation effect.

In general, these results support our hypothesis that these meaning selection mechanisms involve the *active inhibition* of context-inappropriate meanings, and that this inhibition signal slows the eyes on first fixation durations. The fact that the first fixation effect is not elicited by balanced words, or by the dominant meaning of biased words, suggests that readers can perform meaning *selection* of some kind without the need for inhibition. Similarly, readers can select even the subordinate meaning of NV homographs using semantic constraints without eliciting the frontal negativity effect (Lee & Federmeier, 2009). Instead, inhibition seems to be necessary to allow activation of a weak meaning in the face of a competitor, when cues to meaning selection come from outside the semantic system -- i.e., the circumstances in which we observe the first fixation and frontal negativity effects.

First-Pass Plausibility Effects

The core question of the current study involved characterizing plausibility effects at the downstream noun of the prepositional phrase, as a means to test the functional significance of the first fixation effect. The downstream noun served as a “probe word,” and was plausible given only one sense of the target homograph. Reading times to these nouns in the ambiguous (compared to unambiguous) condition thus serve as a test of the extent to which the context-inappropriate meaning of the NV homograph is suppressed. If the first fixation effect reflects meaning *selection*, then its elicitation at the NV homograph should produce longer downstream reading times for implausible compared to plausible nouns. This would indicate that readers selected the context-appropriate meaning, thus rendering the downstream noun associated with the context-inappropriate meaning implausible. If, on the other hand, the first fixation effect reflects meaning *maintenance*, then its elicitation at the NV homograph should produce equally

facilitated reading times on the implausible and plausible nouns, because both meanings would be actively held in memory downstream.

The downstream noun of the prepositional phrase was not directly semantically associated with either meaning of the NV homograph, a feature of the current stimulus set that was important for two reasons. First, this helped avoid reactivation of the homograph's alternate meaning, which could cloud interpretation of facilitation effects observed on the probe (Van Petten & Kutas, 1987). Secondly, it allowed us to improve upon the design of previous eye-tracking studies of plausibility effects, which find no effects of implausibility on first-pass reading measures, but appear to confound target plausibility with its degree of semantic association with the preceding context (Matsuki et al., 2011; Rayner et al., 2004; Warren et al., 2008). The current design allows us to test whether first-pass measures are indeed subject to plausibility effects in the absence of this potential confound.

Although Lee and Federmeier (2012) found evidence linking the frontal negativity with meaning selection in their ERP study, it is possible that the effects could play out differently in an eye-tracking paradigm in which readers have more control over their intake of information from the sentence. Specifically, this control may encourage readers to maintain both meanings of the homograph until more information becomes available, because they know that they can return to earlier parts of the text if they are still confused later. In contrast, it is not possible to move back in the text in the RSVP presentation style necessitated by ERP research. This could instead encourage readers to adopt an early selection strategy, wherein they select a meaning immediately upon encountering the ambiguous word because they know they cannot return to it later. The present study thus allows an important test of the generalizability of the ERP findings to natural reading.

Our results showed that, in eye-tracking (first fixation effects) as in ERPs (frontal negativity effects), ambiguity effects seem to arise from meaning selection processes: first-pass reading times were longer for implausible than plausible downstream noun regions following both NV homographs and unambiguous words. The plausibility effects were just as large following ambiguous as unambiguous words, suggesting that the processes underlying the first fixation effect enabled readers to suppress the context-inappropriate meaning well enough to render it implausible later in the sentence. These findings replicate Lee and Federmeier (2012) in showing that young adults elicit plausibility effects of roughly equal magnitude following ambiguous and unambiguous words. The use of the downstream plausibility manipulation thus allows us to link the observed first fixation effect directly to meaning selection, which was not possible in our previous study (Stites et al., 2013). Furthermore, as will be discussed in more detail below, our results show that *implausibility*, not just *impossibility*, can produce effects on first-pass reading measures.

Relationship Between Ambiguity and Plausibility Effects

If the first fixation effect reflects the recruitment of meaning selection mechanisms, as we have argued it does, then it should share the same relationship with the downstream plausibility effects as was observed in Lee and Federmeier (2012). Lee and Federmeier (2012) found a positive correlation between the size of a subject's frontal negativity effect and the size of their N400 plausibility effect following ambiguous words, suggesting that the context-inappropriate meaning of the NV homograph was less available downstream as a subject's frontal negativity effect at the ambiguous word increased in size. We found the same positive correlation between the size of a subject's first fixation effect on the NV homographs and the size of their downstream plausibility effect. In other words, the bigger one's first fixation effect, the less

available the word's context-inappropriate meaning was at the downstream noun region, which produced larger plausibility effects. This finding strongly suggests that the first fixation effect reflects meaning selection of the ambiguous word. The correlation was largest with the plausibility effect for go-past time on the noun region, which captures all reading times from a reader's first fixation on the noun region until they move past it to the right. As such, it includes initial reading times on both words of the region, as well as any time they spent rereading earlier parts of the sentence (including the earlier ambiguous word) before moving forward. That this measure, rather than first fixation or gaze duration, showed the strongest correlation suggests that readers did not simply slow down on the violating word upon encountering the implausibility, but actively allocated their attention to reading earlier parts of the sentence. Importantly, we demonstrated that the correlation between these effects was specific to cases following ambiguous words and was not due to a reader's general tendency to slow down on difficult sections of a text. The selective relationship between the first fixation ambiguity effect and plausibility effects following ambiguous words provides strong evidence for the involvement of meaning selection processes.

The parallel findings between our current eye-tracking results and the ERP findings using the same stimuli (Lee & Federmeier, 2012) have established a compelling relationship between the first fixation and frontal negativity effects. Our previous work (Stites et al., 2013) showed that both effects are elicited by NV homographs in the presence of syntactic cues without biasing semantics, are reduced or absent in older adults as a group, and are positively correlated with a subject's verbal fluency. The current study adds to our understanding of their relationship by showing again that both are elicited by NV homographs in (different kinds of) semantically neutral contexts, are sustained throughout the following prepositional phrase, and are correlated

with the size of an individual's downstream plausibility effect. Together, the findings from this and our previous study point to the likelihood that these two effects share an underlying function and possibly even the same neural generators. Although the source of the frontal negativity effect cannot be directly assessed using ERPs alone, other imaging work points to the role of the left inferior frontal gyrus (LIFG) in ambiguity resolution (Bilenko et al., 2009; Gennari, MacDonald, Postle, & Sidenberg, 2007; Mason & Just, 2007; Zempleni, Renken, Hoeks, Hoogduin, & Stowe, 2007) and semantic selection more broadly (for reviews, see Novick, Trueswell, & Thompson-Schill, 2010; Thompson-Schill, Bedny, & Goldberg, 2005).

Plausibility Effects in Gaze Measures

The current study also gives us the opportunity to consider the relationship between plausibility effects in eye-tracking and ERPs more generally. Eye-tracking researchers typically examine plausibility effects in terms of which *word* of the sentence first shows increased reading times. This has resulted in what might be a misleading distinction between “immediate” and “delayed” plausibility effects, wherein immediate effects refer to increased first-pass measures on the violating word itself, and delayed effects refer to increased first-pass measures on subsequent words *or* later reading time measures on the violating word. Interpreting plausibility effects in eye-tracking studies becomes especially tricky when we consider that readers have a range of options for how to continue when they reach an implausible word. For example, they could slow down on the violating word, move forward in the text (and possibly slow down there), or immediately regress back to an earlier point of the sentence. The choice of how to proceed may vary from reader to reader, or trial to trial, and, as a result, it is unclear which word in the sentence, or even which reading time measure, provides the “best” representation of

plausibility effects. This may contribute to the inconsistent characterizations of plausibility effects across the eye-tracking literature.

Work using ERPs suggests that it might be more fruitful to focus on the *timepoint* at which plausibility effects are observed rather than looking at reading times on any particular word. ERP work has consistently found that N400 plausibility effects are elicited by the violating word itself (Filik & Leuthold, 2008; Lee & Federmeier, 2012; Nieuwland & Van Berkum, 2006; Paczynski & Kuperberg, 2012), which are always maximal approximately 300-500 ms after readers apprehend the implausible word. We would thus predict that the effects of plausibility would be observed along a similar timecourse in natural reading as well, regardless of where readers' eyes are fixated at that point in time. Our approach in the current study was to apply this prediction from ERP work to our analysis strategy by combining the downstream noun with its subsequent spillover word to make a region. In this way, we can capture readers' processing of the implausible target word as it unfolds over time, even when they have moved their eyes forward in the text. Although this approach will allow us to better align the timelines between eye-tracking and ERP effects, one caveat that remains is that with eye-tracking it is not possible to pinpoint the exact moment that readers first began processing the implausible noun, due to the (potential) availability of preview information (which is not available in typical ERP studies). As such, any reading time measure on the critical word will still likely underestimate the full amount of time readers spent processing that word, because it will be unable to account for any time that readers might have spent gathering information about the word during their previous fixation. Regardless of this shortcoming, when we defined the time window following the apprehension of the downstream noun in this way, we saw effects of implausibility on first pass reading times to the region (which have been rare in eye-tracking studies with *plausibility* violations).

Interestingly, average gaze durations on the noun region fell between 385-429 ms (see Table 6), which line up nicely with the time frame in which N400 plausibility effects are observed, for these materials as well as in other studies.

We believe that our findings represent a more general relationship between the timecourse of ERP and eye-tracking effects -- particularly that effects expressed as modulations of the N400 will often be “smeared” across the reading of several words, due to the extensive amount of saccadic pre-programming hypothesized to underlie skilled reading (Reichle, Pollatsek, & Rayner, 2006). Because readers cannot often recognize that a word is implausible fast enough to affect reading times on the word itself, they will likely continue forward in the text before the semantics of that word can be appreciated in light of the ongoing message-level context. As such, the increased reading times associated with plausibility violations may not show up until the following word (or even two regions later in the sentence, as has been observed in a self-paced reading study [Matsuki et al., 2011])². On the other hand, if a lot of readers choose to look back in the text, either due to individual variation across samples or characteristics of the experimental stimuli, then the plausibility effects might be captured instead in regression out probabilities rather than spillover reading times (as observed by Warren & McConnell, 2007). In the current study, the percentage of first pass regressions out from the noun was low (12-14%, with no effects of plausibility or ambiguity). The vast majority of the time, then, readers moved directly from the noun to the spillover word, where their effects were captured in first-pass noun region reading times.

² Studies employing possibility violations (Warren & McConnell, 2007; Warren, McConnell, & Rayner, 2008), which often report earlier reading time effects, used words that differed greatly from the most expected word. Thus, parafoveal preview information may have allowed readers to begin to recognize its incongruence with the context on the fixation prior to landing on it. This would allow processing of the impossible word to begin while readers were fixated on the previous word and, in line with our proposed timeline of plausibility effects, affect reading times on the target word itself.

We are not the first to suggest that ERP effects, particularly those on the N400, may be spread across multiple fixations in natural reading. Dambacher and Kliegl (2007) suggested a similar relationship between the timing of eye movement and ERP effects. They compared reading times and ERPs elicited by the same group of sentences (recorded at different times) containing words of varying frequencies. Their results showed that low-frequency words elicited larger N400 amplitudes than high frequency words (a typical N400 frequency effect), as well as longer reading times on the *next* word in the sentence. Both sets of findings strongly suggest that the processing of a single word likely continues even after readers are no longer fixating it, even the initial stages of word recognition that are believed to be most affected by word frequency. This processing can influence subsequent fixations at approximately the same timepoint as the effects would be observed on the N400, especially when the word is difficult due to its low frequency or implausibility with the preceding context.

Rereading of Target Word Region

In order to assess whether there were ambiguity effects that were not resolved in the first pass, we examined rereading times on the target word region, calculated as all time spent on the region after gaze duration. We combined the target homograph with the function word preceding it into a region, as either word (or both) would make a reasonable target to return to in order to resolve residual uncertainty as to the intended meaning of the target word. In general, we found that ambiguous target regions that were followed by an implausible noun were reread more than those followed by a plausible sentence continuation. There was, however, no difference in target rereading times for unambiguous words based on the plausibility of their sentence continuation. This pattern of results suggests that in sentences containing ambiguous words, readers recognized that the later implausibility of the noun hinged on the incorrect instantiation of the

homograph and thus allocated more time rereading it. There was no systematic relationship between the implausible word and its unambiguous precursor, and consequently, readers did not return to it.

When we split the ambiguous words into dominant, subordinate, and balanced conditions, we found an intriguing pattern of results (see Figure 5). First, it is striking to note that the implausible reading times did not differ across the three words types. This is especially important for the subordinate instantiations which, when followed by a word plausibly related to the dominant meaning, elicited equally long rereading times as the other two conditions. If the dominant meaning remained active in the subordinate context, we might expect for the subordinate implausible rereading times to be smaller than the other two (because the downstream implausible word would actually be *plausible* with the still-active dominant meaning, thus reducing the need to return to the ambiguous word for reanalysis).

Although implausible reading times were equally long across these three conditions, rereading times on the subordinate plausible stood apart as being substantially longer than the other plausible rereading times. In fact, they did not differ from the subordinate *implausible* rereading times. This interesting finding could have two related, and not mutually exclusive, interpretations. First, it could be taken as evidence for incomplete meaning selection of the subordinate meaning. Residual activation of the dominant meaning could have persisted even though the selection processes indexed by the first fixation effect were at least effective enough to generate the downstream plausibility effects. This lingering activation could have caused readers to incorrectly judge the plausible subordinate-related downstream noun as being implausible (at least some of the time), thus spurring more rereading of the target word region in these cases. The second interpretation of this effect could be that the elicitation of the first

fixation effect at the NV homograph generated an error signal that may have “marked” that word as a place of uncertainty. Readers may have then chosen to return to this region as part of a checking strategy of revisiting the site of a previous uncertainty, even if the ambiguity itself was resolved in the first pass. This second possibility falls in line with a similar proposal by Levy and colleagues (Levy, Bicknell, Slattery, & Rayner, 2009) that readers maintain uncertainty about word identities, especially of short function words, after moving past them in the text, and they can use this uncertainty to selectively return to that word when they later encounter a structural ambiguity whose resolution depends on its interpretation.

Part 4: Older Adult Ambiguity Effects in Eye-tracking

Experiment 4 tackles the question of how older adults use context to constrain their interpretation of ambiguous words in a sentence context, and more importantly, the consequences that this disambiguation has on downstream processing. Specifically, we are interested in the resolution of ambiguity associated with noun/verb homographs, whose meanings fall into distinct grammatical categories. These words provide the unique opportunity to isolate the effects of syntactic information in contexts in which biasing semantic information is unavailable, to better understand if older adults can use the grammatical structure of a sentence to disambiguate the intended sense of a noun/verb homograph, and what representations they are left with when they move forward in the sentence.

Older Adult's Use of Semantic Context in Disambiguation

There is plenty of evidence that older adults can *activate* both meanings of ambiguous words when they are presented out of context. This evidence tends to come largely from semantic priming studies, in which subjects are presented with an ambiguous word, followed by a word related to one of its two meanings. For example, Balota, Watson, Duchek, and Ferraro (1999) found cross-modal semantic priming following the presentation of both ambiguous and unambiguous words that was qualitatively similar across young adults and healthy older adults. Furthermore, following biased homographs, older adults showed equally large priming for words related to both the dominant and subordinate meaning (when compared to unrelated words), suggesting they are able to effectively activate even the word's subordinate meaning. There is also evidence that older adults can quickly use semantic context to select a homograph's meaning (Balota & Duchek, 1991; Hopkins, Kellas, & Paul, 1995; Meyer & Federmeier, 2010; Swaab, Brown, & Hagoort, 1998). In one such study, Balota and Duchek (1991) found that older

adults can use the first word of a three word group to constrain their interpretation of a homograph appearing in the second position (e.g., *kidney-organ-_____*), resulting in facilitated naming latencies on subsequent a word related to the context-appropriate “body part” meaning of the homograph (e.g., *heart*) as compared to the context-inappropriate “instrument” meaning of the homograph (e.g., *piano*). Similar evidence for the use of semantics to choose the context-appropriate meaning of a biased homograph was found when ambiguous words were presented in a simple sentence-context as well (Hopkins, Kellas, & Paul, 1995). In this study, naming latencies to words related to the homograph’s context-appropriate meaning were facilitated relative to those for unrelated words, irrespective of meaning dominance. Together, these studies present behavioral evidence that after encountering an ambiguous word, older adults are faster to respond to words semantically related to the meaning instantiated by the preceding semantic context.

Findings from studies using more temporally-sensitive measures of online processing, such as event-related potentials (ERPs) and eye movements during reading, support the idea that older adults can effectively use semantic context to activate features of ambiguous words. For example, Lee and Federmeier (2011) showed that older adults can use constraining semantics to facilitate their processing of both ambiguous target words presented at the end of congruent sentences (e.g., *My grandpa said he hadn’t played that game since he was a **kid***) compared to cases in which these same words appeared in syntactic prose sentences that had constraining syntactic information but lacked coherent semantics (e.g., *My board said he hadn’t called that volcano since he was a **kid***). Both ambiguous and unambiguous words elicited smaller N400 amplitudes at the end of the congruent relative to syntactic prose sentences, indicating that the build up of message level meaning from the sentence facilitated processing of those words.

Using these identical contexts in an eye-tracking paradigm, Stites et al. (2013) found that first fixations and gaze duration were reduced these same target words in congruent relative to syntactic prose sentences, again suggesting that the older adults effectively used the semantic context to facilitate reading times. Together, these findings add to the evidence from behavioral work to suggest that older adults remain relatively unimpaired at the use of semantic context during online comprehension to engage the processes involved in the initial stages of word recognition indexed by the N400 and first-pass reading times.

Older Adults' Use of Syntactic Information in Disambiguation

This brings us to our next question of whether older adults can use *syntactic* context to select the context-inappropriate meaning of ambiguous words, particularly when those meanings fall into distinct grammatical categories. Behavioral evidence from Dagerman and colleagues suggests that older adults struggle with the use of syntax to constrain their interpretations of NV homographs (Dagerman, MacDonald, & Harm, 2006). They embedded noun/verb homographs, like *fires*, in sentence fragments such as *The union told the reporters that the corporation (warehouse) **fires** could*, and subjects' task was the speeded naming of the target *could*. Naming latencies for young adults showed a cost when the non-dominant verb meaning of *fires* was instantiated (i.e., following *corporation*) relative to its dominant noun meaning (i.e., following *warehouse*), likely due to known difficulties in activating a homograph's less-frequent meaning (Rayner, Pacht, & Duffy, 1994). Older adults, on the other hand, did not exhibit a similar sensitivity to contextual bias in this online measure, although in their offline judgments indicated that they indeed appreciated that the subordinate-biasing context provided a less good fit to the target word. This finding suggests that older adults are unable to use syntactic cues to constrain their *online* interpretation of ambiguous words, despite having the requisite knowledge to do so.

Evidence from both ERPs and our own previous eye-tracking work supports the idea that older adults are less able to use syntax for the selection of a homograph's context-appropriate meaning during online processing. Lee and Federmeier (2009; 2012) have found that young adults elicit a sustained frontal negativity at the onset of noun/verb homographs when there are constraining syntactic cues, but no biasing semantic information to aid in meaning selection. This effect has been posited to reflect the recruitment of frontally-mediated meaning selection mechanisms that are necessary to suppress the context-inappropriate meaning of the ambiguous word under difficult circumstances, for example, when the cues on which to base this suppression come from outside of the semantic system. Older adults, who tend to have less efficient frontal lobe functioning (Raz, et al., 2005), fail to elicit this effect (Lee & Federmeier, 2011; 2012) thus supporting the idea that it may be mediated by frontal sources. Our follow-up eye-tracking work using the same stimulus configuration found that the noun/verb homographs elicit a first fixation ambiguity effect in young adult readers, in terms of longer first fixation times in syntax-only conditions, which is again absent in older adults (Stites, et al., 2013). In both this study and Lee and Federmeier (2011), there was a relationship between the size of a subject's ambiguity effect and their verbal fluency score (a measure of productive vocabulary thought to at least partially reflect frontal lobe efficacy, Benton & Hamsher, 1978), revealing that older adults with higher verbal fluency (and thus more efficient frontal lobe functioning) elicited larger ambiguity effects. These studies thus provide evidence that although older adults as a group struggle with the use of syntactic information to constrain online homograph disambiguation, individuals seem to vary widely in their ability to do so.

Downstream Consequences Following Suppression

More interesting is the question of how older adults' ability (or inability) to use syntactic information to select a homograph's context-appropriate meaning will impact their downstream comprehension. That is, after they encounter an ambiguous word and move past it in the sentence, what types of representations will they be left with, and how can they use these to interpret incoming information? For starters, it is well demonstrated that meaning selection by older adults is rarely perfect, even when it stems from the use of semantic cues. For example, Faust, Balota, and Duchek (1997) embedded homographs and matched unambiguous words appeared at the end of semantically biasing sentences (e.g., *He dug with the spade/shovel*), after which a probe word was presented that was either related to the homograph's context-appropriate meaning (e.g., *garden*), or its context-inappropriate meaning (e.g., *ace*). Older adults were slower to reject words like *ace* than *garden*, which the authors took as evidence of continued activation of the homograph's context-inappropriate meaning despite the semantic cues to the contrary. Young adults run in the exact experimental paradigm also showed these same interference effects (Gernsbacher, Varner, & Faust, 1991), suggesting that all readers may show some level of residual activation for the alternate meaning, although their effects were of a smaller magnitude. Indeed, there is also ERP evidence that even in the face of semantic cues, both young and older adults show residual activation of the context-inappropriate meaning of NV homographs. These words elicit larger N400 amplitudes than unambiguous words in congruent sentence contexts (Lee & Federmeier 2009; 2011), even when matched on cloze probability. Together, then, these findings suggest that even under the best of circumstances, we may still expect to find some level of residual activation of the homograph's alternate meaning.

More importantly for our purposes, then, is whether this residual activation has a detrimental impact on older adult's ability to construct the correct message level meaning of the sentence.

Work from outside of the ambiguity domain suggests that, in general, older adults have a difficult time suppressing any type information that has been shown to be irrelevant. In one study, Connelly, Hasher, and Zacks (1991) found that older adults struggled more than young adults to ignore information presented in a paragraph in a different font style, which they were told indicated that it was irrelevant to the meaning of the passage. In a second series of studies, they found that older adults were also slower to abandon inferences they had formed from a text, even when this inference was clearly overridden by the given material (Hamm and Hasher, 1992; Hartman & Hasher, 1991). For example, Hartman and Hasher (1991) presented subjects with high constraint sentence frames, with the final word delayed. During the delay period, subjects were instructed to guess what the upcoming word would be. The sentence was then completed with either the most expected end, or an unexpected but plausible ending. For example, in the sentence *She ladled the soup into her lap*, the most expected completion *bowl* was “disconfirmed” by the presented ending *lap*. Subjects were later given a sentence completion task, in which they filled in the final word of a moderately constraining frame (e.g., *Scotty licked the bottom of the _____*), where *bowl* would again be the best ending, but a range of other endings would also fit. Results showed that older adults were more likely than young adults to reuse words like *bowl* to complete these sentences, which had been “disconfirmed” by their earlier sentence context, suggesting an inability of older adults to suppress these irrelevant words. These findings could indicate further difficulties that older adults may face after encountering ambiguous words. A similar situation could arise in which they activate both of the word's meanings, but then must use syntactic cues to suppress the irrelevant (“disconfirmed”). Even if

they are able to utilize the syntactic cue online to appreciate which meaning of the homograph is intended by the context, older adults may then face an additional challenge of actually suppressing the word's irrelevant meaning going forward in the sentence, as suggested by these studies of Hasher and colleagues, which could lead to later confusion.

Lee and Federmeier (2012) examined the downstream processing consequences elicited by both young and older adults after encountering NV homographs in sentences that only contained syntactic cues as to the intended meaning, the findings of which will form the basis for the current study. They embedded NV homographs in a sentence-medial position, preceded by semantically neutral but syntactically constraining information, and then followed by a prepositional phrase whose main noun was only plausible given the context-appropriate meaning of the homograph. For example, compare the sentences (1) *Jared wanted **THE sink** in his bathroom to be repaired* and (2) *Jared wanted **TO sink** in his bathroom to be repaired*. In (1), the word *bathroom* is plausible, whereas in (2), it is implausible, a determination that hinges on the context-appropriate meaning of the word *sinks*. Similarly, in (3) *Gary wanted **TO sink** in his bed to take a long nap*, *bed* is plausible, whereas in (4) *Gary wanted **THE sink** in his bed to take a long nap*, *bed* is implausible. The hypothesis was that if readers could select the context-appropriate meaning of the homograph using syntactic cues alone, then this should render sentences (2) and (4) to be implausible, producing N400 plausibility effects on the words *bathroom* and *bed*, respectively. If, however, readers were unable to use the syntactic cues to select the correct meaning, then both meanings of the homograph should remain active regardless of the syntactic cues, facilitating the processing of *bathroom* and *bed* in both of their contexts. Importantly, the downstream probe word (*bathroom*, *bed*) was never semantically related to either meaning of the word *sink*. This manipulation thus avoided the possibility of

backwards priming, in which the presentation of a probe word related to the context-inappropriate meaning may actually *produce* activation of that word, rather than just measure its current activation level (Van Petten & Kutas, 1987).

Lee and Federmeier (2012) showed that young adults elicited the frontal negativity effect at the onset of the NV homographs in the syntactically constraining but semantically neutral contexts, and downstream showed evidence for activation of just the context-appropriate meaning of the ambiguous words. Older adults, on the other hand, *failed* to elicit the frontal negativity effect and, consequently, showed continued activation of the homograph's dominant meaning downstream, regardless of which meaning was instantiated by the context. The pattern exhibited by the older adults suggested that they were unable to elicit the frontally-mediated meaning selection mechanisms needed to choose the meaning of the ambiguous word using syntactic cues alone, and as a result selected a meaning based on dominance. However, when both age groups were considered together, there was a positive correlation between the size of one's frontal negativity effect and the size of their N400 plausibility effect downstream, indicating successful selection of the context-appropriate meaning because the implausible word was plausible with the homograph's context-*inappropriate* meaning. Thus, even though the group pattern did not show an overall frontal negativity effect for the older adults, those who were able to elicit the frontal negativity (and recruit the selection mechanisms they index) showed evidence for successful meaning selection. Overall, then, this study suggested a functional link between the frontal negativity effect and frontally-mediated meaning selection mechanisms.

Our previous eye-tracking work showed that when similar sentences were presented in natural reading, young adults elicited a first fixation effect at the NV homographs, which we

posited was driven by the same underlying mechanisms as the frontal negativity effect, whereas older adults failed to show this effect (Stites et al., 2013). Instead, older adults as a group reread the NV homographs to a greater degree than the young adults. Furthermore, within both age groups, an individual's verbal fluency was positively correlated with the size of the first fixation effect, and negatively correlated with the size of their rereading effect, suggesting a tradeoff between early and late meaning selection strategies that is modulated by an individual's age and language abilities. These findings were unable to draw an explicit link between the first fixation effect and meaning selection, though, as we did not probe the downstream activation of the homograph's context-inappropriate meaning. Experiment 3 of the current project thus sought to test the prediction that the first fixation effect indexes meaning selection. We first replicated our findings that young adults elicit the first fixation effect to NV homographs in syntactically constraining yet semantically neutral contexts. Furthermore, we found a positive correlation between the size of one's ambiguity effect and their downstream plausibility effect, replicating Lee and Federmeier (2012) in suggesting a functional role for the first fixation effect in meaning selection.

The final piece of the puzzle linking the first fixation effect to frontally-mediated meaning selection mechanisms is to test how these effects play out in older adult readers, specifically how their presence (or absence) at the homograph relates to meaning activation downstream. First, we predict that the older adults will not elicit the first fixation effect to the NV homographs in the syntactically coherent yet semantically neutral contexts, because previous studies have shown their inability to elicit both the first fixation effect (Stites et al., 2013) and the frontal negativity effect (Lee and Federmeier, 2011; 2012) under these circumstances. Next, we predict that if the older adults do not show the first fixation effect, then they should show

continued activation of the homograph's dominant meaning when they reach the implausible word downstream, because they will have been unable to recruit the selection mechanisms necessary to choose a meaning using syntax alone. Finally, we predict that there will be a positive correlation between the size of one's first fixation effect and their downstream plausibility effect. Even though we expect for the effect to be absent in older adults as a group, there is likely individual variation in the efficacy of frontal lobe functioning that predicts how successfully they are able to recruit the selection mechanisms, and consequently, elicit the plausibility effects. If these three predictions are borne out, we will be able to confidently claim that the first fixation effect indexes the same frontally-mediated meaning selection mechanisms as the frontal negativity ERP effect, and that these mechanisms are necessary to select the context-appropriate meaning of a NV homograph using syntactic cues in the absence of biasing semantics.

Methods

Participants

Twenty-four older adults (21 women; mean age 69.9; range 63-81) participated in the experiment for cash payment of \$8 per hour. All were also monolingual speakers of English, with no consistent exposure to other languages before age 5 (with the exception of one participant, who reported early language exposure). Twenty-two participants were right-handed, and two reported being left-handed, while 15 reported having first-degree left-handed family members. Participants had no history of neurological or psychiatric disorders or brain damage. We conducted the Montreal Cognitive Assessment to screen participants for cognitive impairments. Participants scored an average of 27 out of 30 possible points (*SD* ; range: 24–30),

which falls within the normal range for this test (suggested cutoff for impairment <23 ; Luis, Keegan, & Mullan, 2009).

Materials/Procedure

The materials and procedure were the same as those used for Experiment 3.

Data Analysis

Data analysis procedures were the same as those used for Experiment 3. 4.3% of trials were excluded from the analysis due to track loss or program error.

Results

Behavioral Measures

The purpose of the word and sentence recognition tasks was to encourage subjects to pay close attention to the experiment at both the word and sentence level. Overall, the older adults answered the word recognition task with 88% accuracy ($d'=2.6$). They also performed well on the sentence recognition task, at 81% ($d'=2.1$). Both of these values are comparable to how the older adults in Lee and Federmeier (2012) scored on these same tasks: average accuracy of 86% ($d'=2.5$) on the word recognition task and 84% ($d'=2.3$) on the sentence recognition task, as well as very close to the performance of the young adults in Experiment 3.

Eye-Tracking Measures

Based on the outcomes of Experiment 3, as well as the findings from Lee and Federmeier (2012), the following regions of interest will be examined in the current experiment. First, reading times on the target ambiguous word will be considered, where we previously observed first fixation effects for young but not older adults, where we predict not to observe first fixation

ambiguity effects. Next, reading times will also be measured on the prepositional phrase region following the target, where the young adults in Experiment 3 exhibited continued first fixation costs following ambiguous words. Finally, reading times on the downstream noun region (i.e., the noun and its subsequent spillover word) will then be examined to test for a plausibility effect (i.e. longer reading times) on the implausible relative to plausible words, as well as an interaction between plausibility effects and preceding target word ambiguity.

Several standard eye-tracking measures will be employed: first fixation duration, the length of the readers' first fixation on the word/region; gaze duration (a.k.a. first pass time), the sum of all fixations on a word/region the first time it is fixated, before the eyes leave it in either direction; go-past time, a right-bounded measure including all fixations a reader makes on a word/region, beginning with their first fixation until they move past it to the right (including rereading earlier parts of the sentence); and rereading time, all time spent reading a word/region after leaving it in the first pass. For all regions of interest, two repeated measures Analyses of Variance (ANOVA) were conducted, one averaging over subjects (F_1) and the other averaging over items (F_2). For the by-subjects analysis, both the ambiguity/dominance of the target word and the plausibility of the downstream noun (when applicable) were within-subject factors. For the by-items analysis, the identity of the NV homograph and/or matched unambiguous control will be considered as the item over which to average (considering as separate lexical items the dominant and subordinate instantiations of the biased homographs, and the noun/verb meaning of the balanced homographs), making ambiguity/dominance a between-items factor. Because all lexical items were used both plausibly and implausibly, the factor of plausibility is a within-items factor.

As in Experiment 3, we will present first-pass reading time measures (first fixation, gaze duration, and go-past time) separately from rereading time measures. This will allow us to present the results in a manner that follows naturally from the ERP study, because readers only receive a “first pass” through the sentence when using the RSVP presentation style used in ERP research. Furthermore, the factor of plausibility does not become apparent until readers reach the downstream noun region. As such, plausibility must be included as a factor in analyses of rereading times, but not first pass measures, of the target region.

First-Pass Reading Measures

Target word. Older adults’ reading times on the target word did not differ between ambiguous and unambiguous words on first fixation duration, F_1 and $F_2 < 1$ (see Table 8 for reading times). This finding replicates Stites et al. (2013) in showing that older adults fail to elicit first fixation ambiguity effects to NV homographs embedded in syntactically constraining but semantically neutral sentences, the same situations in which young adults do exhibit these effects (cf, Stites & Federmeier, in prep). The other first pass measures also failed to show ambiguity effects for the NV homographs. Gaze durations were marginally *shorter* for ambiguous relative to unambiguous words in the subject analyses, $F_1(1,23)=3.83$, $p=.06$, but the difference was not significant by items, $F_2(1,94)=1.99$, $p=.16$. Go-past times showed no effect of ambiguity, $F_1(1,23)=1.34$, $p=.26$, $F_2 < 1$. Older adults thus showed no ambiguity costs on their first pass through the target word.

Based on the subordinate bias effect observed for young adults on first fixation duration (i.e. longer first fixation times for subordinate versus dominant instantiations), we also examined reading times on the homographs as a function of target word dominance, to test that possibility that the subordinate word reading times were in fact longer than those for the dominant but were

masked by averaging the two conditions together. However, this possibility was not supported. First fixations across the four dominance conditions were nearly equivalent (dominant: $M=253$, $SD=33$; subordinate: $M=255$, $SD=39$; balanced: $M=255$, $SD=44$; unambiguous: $M=254$, $SD=39$), and the analysis yielded a non-significant effect of target dominance, F_1 and $F_2 < 1$. As such, we have no evidence that older adults exhibited the first fixation ambiguity effects on the target homographs

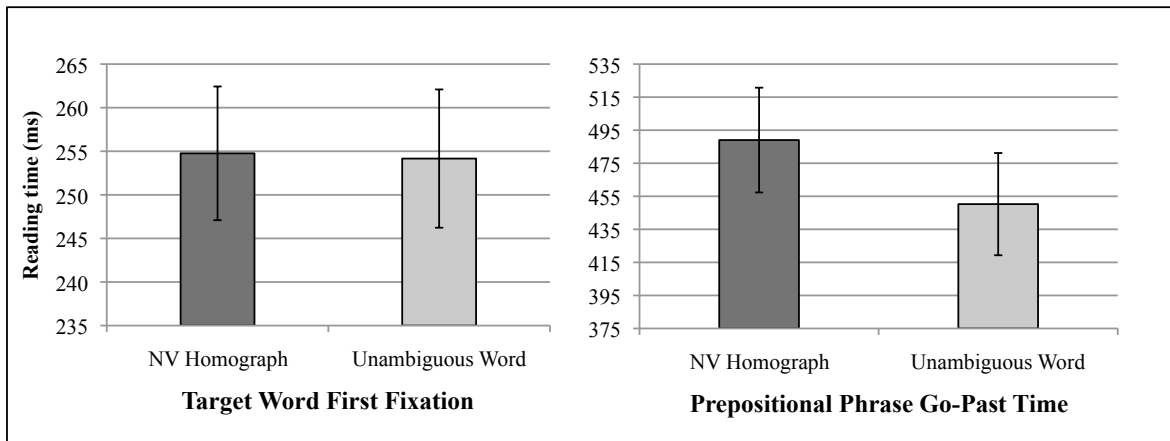
Table 8. Reading times (and standard deviations) on areas of interest (Experiment 4).

Region	Amb	First Fixation		Gaze Duration		Go-Past Time		Selective Go-Past Time	
Target Word	AA	255 (38)		298 (65)		401 (113)		327 (84)	
	UW	254 (39)		308 (78)		392 (106)		335 (92)	
Prep. Phrase	AA	239 (39)		337 (83)		489 (155)		397 (107)	
	UW	237 (38)		346 (92)		450 (152)		390 (113)	
Later Noun Region	AA	Impl	Plaus	Impl	Plaus	Impl	Plaus	Impl	Plaus
		264 (34)	257 (35)	498 (159)	457 (131)	796 (249)	664 (205)	627 (201)	552 (165)
	UW	266 (34)	263 (42)	528 (162)	474 (131)	818 (276)	626 (186)	666 (211)	543 (154)

Prepositional phrase. Reading times on the prepositional phrase directly following the target words were also examined, because both Lee and Federmeier (2012) and Stites and Federmeier (in prep) found sustained effects of ambiguity on this region in young adults, in terms of sustained frontal negativity and increased first fixation durations, respectively. The first two words of the prepositional phrase were considered as a region, because many readers skipped one or the other of these short function words (skip percentage for first word of prepositional phrase was 37%, and 43% for the second word). Results showed no ambiguity effects on first fixation duration in this region, F_1 and $F_2 < 1$, nor gaze duration, $F_1(1,23)=2.74$,

$p=.11$, $F_2<1$. Interestingly, go-past times did show a substantial ambiguity effect—they were 39 ms longer following ambiguous than unambiguous words (see Figure 9). This difference in go-past times was significant by subjects, $F_1(1,23)=14.61$, $p=.001$, and marginally significant by items, $F_2(1,94)=3.09$, $p=.08$. Because go-past times include reading times on both the prepositional phrase region *and* all preceding regions in the text before moving forward, and because the effect was not significant for gaze duration on the prepositional phrase itself, this effect is thus likely driven by rereading of the ambiguous word and/or function word before moving forward in the text.

Figure 9. First fixation times for the target word (left) and go-past times on the prepositional phrase (right) for the older adults (Experiment 4).



Because older adults elicited ambiguity effects on the go-past time of the prepositional phrase region, it was also of interest then to see if this effect was sensitive to the dominance of the preceding target word. Go-past times were longest for the prepositional phrase following the subordinate instantiations ($M=526$, $SD=177$), followed by the dominant instantiations ($M=487$, $SD=166$), balanced words ($M=452$, $SD=151$), and finally the unambiguous words ($M=450$, $SD=152$). An ANOVA comparing these four conditions revealed a significant effect of dominance, $F_1(3,69)=8.712$, $p<.001$, $F_2(3,140)=3.43$, $p<.05$. Follow-up t-tests were conducted to

test for a subordinate bias effect. These tests confirmed that go-past times following the subordinate condition were significantly longer than the unambiguous words and the balanced words by both subjects and items (unambiguous words: $t_1(23)=4.66, p<.001$; by-items 95% CI of the difference: 80.7 ± 80 ; balanced words, $t_1(23)=3.43, p<.01$, by-items 95% CI of the difference: 81.4 ± 71.8) The difference between subordinate and dominant go-past was significant the by subjects, $t(23)=4.66, p<.001$, but did not reach significance by items (95% CI of the difference: 51.6 ± 78). By breaking the prepositional phrase region down by the dominance of the preceding target word, we can show that the longer go-past times following the ambiguous words were driven by an increase following the subordinate instantiations, and to a lesser extent, following the dominant instantiations as well. This is important, as it shows not only that the older adults noticed the ambiguity, but they were able to modulate their reading times over the immediately following sentence region to accommodate the additional processing difficulties created by the subordinate words.

Downstream noun region. Next, reading times were considered on the downstream noun region of the prepositional phrase, for which there were factors of both the ambiguity of the preceding target word as well as the plausibility of the noun itself. Previous eye-tracking studies have found that plausibility violations can cause increased reading times on either the offending word itself, or on the spillover word directly following it. However, this debate is tangential to the main theoretical question addressed in this study, which is whether older adults can elicit plausibility effects *at all* when those effects depend on successful ambiguity resolution earlier in the sentence. Furthermore, plausibility effects have not been well-characterized in older adults, so we do not have a clear prediction as to where they are most likely to show processing costs. As a result, we combined the violating downstream noun with its subsequent word to create a

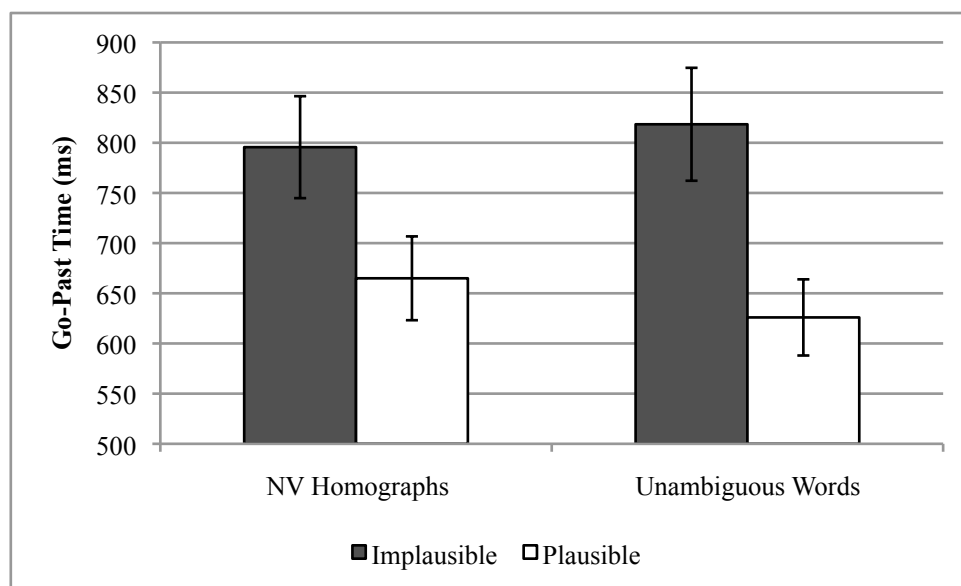
region (as we also did for young adults), which we will analyze for the presence of plausibility effects.

In general, reading times were longer for implausible than plausible regions on all measures, but the two factors only showed signs of an interaction on go-past times (see Table 8 for reading times). First fixations showed a significant effect of plausibility by subjects, $F_1(1,23)=4.14, p=.05$, that was marginally significant by items, $F_2(1,94)=2.92, p=.09$, indicating longer reading times for implausible than plausible noun regions. There was no effect of ambiguity, $F_1(1,23)=1.15, p=.30, F_2(1,94)=1.14, p=.29$, nor interaction, F_1 and $F_2 < 1$. Gaze durations also showed longer reading times for implausible than plausible regions, as well as slightly longer reading times on regions following *unambiguous* than ambiguous words. These findings were supported by a main effect of plausibility, $F_1(1,23)=23.10, p<.01, F_2(1,94)=21.64, p<.01$, a main effect of word type that was significant by subjects, $F_1(1,23)=5.99, p<.05, F_2(1,94)=1.43, p=.24$, and no interaction, F_1 and $F_2 < 1$. Go-past times showed a plausibility effect within both word types that, importantly, was reduced following ambiguous words (see Figure 10). This pattern was supported by a main effect of plausibility, $F_1(1,23)=56.40, p<.01, F_2(1,94)=86.13$, no effect of word type, F_1 and $F_2 < 1$, and a marginal interaction between the two, $F_1(1,23)=3.38, p=.08, F_2(1,94)=2.50, p=.12$. Follow-up t-tests showed that the plausibility effect was indeed significant following both the ambiguous words, $t_1(23)=5.34, p<.01, t_2(47)=6.19, p<.01$, and the unambiguous words, $t_1(23)=6.28, p<.01, t_2(47)=6.94, p<.01$, although the magnitude of the effect was smaller following ambiguous words.

The interactions present on the go-past time indicates that older adults exhibited marginally smaller plausibility effects for the downstream noun regions following ambiguous words. Lee and Federmeier (2012) also observed smaller plausibility effects following

ambiguous words in their study, which was largely driven by continued facilitation of the dominant meaning in a subordinate-biased context. In the current results, the smaller plausibility effects following ambiguous words appear to be driven by both a reduction in implausible reading times compared to the unambiguous words, as well as increased plausible reading times. The current results were calculated collapsing across all ambiguous words, so the strongest conclusion we can draw is to say that the word's *alternate* meaning was likely still active, regardless of whether the dominant or subordinate meaning was instantiated.

Figure 10. Go-past time on the downstream noun region of the prepositional phrase for the older adults (Experiment 4).

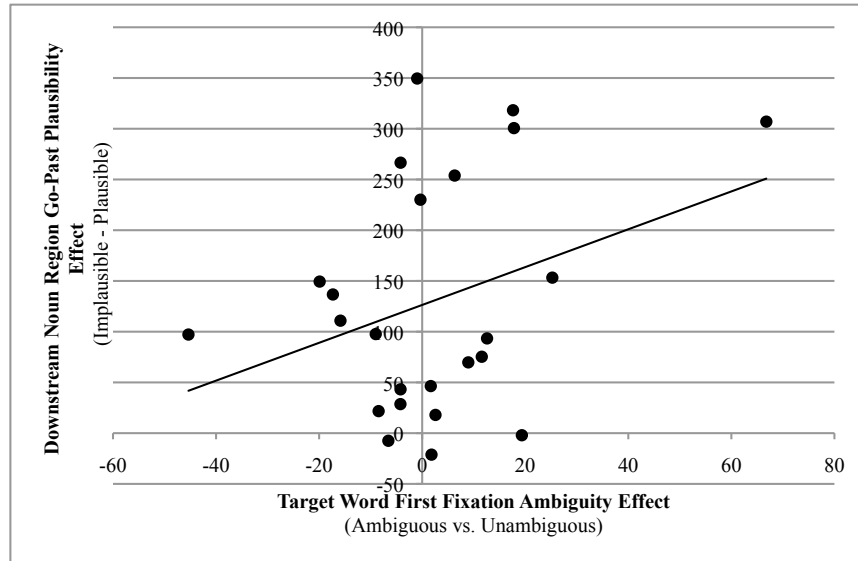


Correlations Between Ambiguity and Plausibility Effects

If the functional significance of the first fixation and frontal negativity effects truly is meaning selection, then there should be a measurable relationship between the size of one's first fixation effect and the availability of the word's alternate meaning downstream. Accordingly, Lee and Federmeier (2012) found a positive correlation between the size of participants' frontal negativity effect at the ambiguous word and the size of their N400 plausibility effect at the

downstream noun. Using the same materials, Stites and Federmeier (in prep) also found a positive correlation between the magnitude of young adult readers' first fixation ambiguity effect and the magnitude of their go-past time plausibility effect on the noun region. Together, these findings suggest that the initial ambiguity effects (both first fixation and frontal negativity) reflect suppression of the context-inappropriate meaning of the NV homograph, thus increasing the magnitude of readers' disruption when they later encountered a word plausibly related to the homograph's alternate meaning. To determine if a similar relationship exists between early and late effects for the older adults, we calculated a first fixation ambiguity effect by subtracting first fixations to the unambiguous from the ambiguous target words (when readers did not first fixate the function word, in line with the young adult data), producing a positive value for longer ambiguous word reading times. We also calculated four different plausibility effect scores, one for each first-pass reading measure on the noun region (first fixation, gaze duration, and go-past time). These were created in a similar fashion as the ambiguity effect score: reading times on plausible regions were subtracted from those on implausible regions, producing a positive value for longer implausible reading times. Analyses showed that there was no significant correlation between first fixation ambiguity effects and plausibility effects for downstream noun region first fixation ($r = -.01$, one-tailed $p = .48$) or gaze duration ($r = .04$, one-tailed $p = .43$). This correlation was, however, marginally significant for go-past time ($r = .33$, one-tailed $p = .06$). Although it failed to fully reach significance, it indicated a trend in the predicted direction as can be seen in Figure 11. (When the older adults in Lee & Federmeier were considered as a group, their overall correlation also failed to reach significance [$r = .15$, one-tailed $p = .2$]). Given the lack of a first fixation in the older adults as a group, it is perhaps not surprising that it does not serve as a strong predictor of downstream plausibility effects. .

Figure 11. Correlation between ambiguity and plausibility effect for the older adults.



Rereading Time: Target Word Region

Rereading times were calculated on the target word region, which consisted of both the target homograph as well as its preceding function word (*to/the*). This measure was of particular interest because in our previous study, older adults' group tendency was to reread the ambiguous words (as opposed to slowing down on them in the first pass). Investigating rereading effects in the current study will help provide important insights as to whether rereading the ambiguous words is a general ambiguity resolution strategy of older adults, or if the stimuli in the previous study (i.e., syntactic prose sentences that lacked coherent semantics) encouraged an atypical strategy. We chose to include rereading times on both the target and function word because either (or both) would be a reasonable place to return to in order to resolve residual uncertainty as to the homograph's intended meaning, particularly because the function word was the only unambiguous cue as to the homograph's context-appropriate meaning. Rereading times were calculated by subtracting gaze duration from total time, producing a measure that included all reading times on the region after readers left it for the first time.

Rereading times on the target word region were generally longer for all implausible relative to plausible contexts, and for ambiguous relative to unambiguous words, with the longest rereading times for ambiguous words used implausibly (ambiguous implausible: $M=522$, $SD=265$; ambiguous plausible: $M=350$, $SD=186$; unambiguous implausible: $M=404$, $SD=234$; unambiguous plausible: $M=336$, $SD=184$). Results showed main effects of both ambiguity, $F_1(1,23)=33.95$, $p<.01$, $F_2(1,94)=18.71$, $p<.01$, and plausibility, $F_1(1,23)=40.30$, $p<.01$, $F_2(1,94)=60.24$, $p<.01$, as well as a significant interaction, $F_1(1,23)=15.92$, $p<.01$, $F_2(1,94)=12.60$, $p<.01$. Follow-up t-tests indicated significant plausibility effects both within the ambiguous target regions ($t_1(23)=8.58$, $p<.01$, $t_2(47)=8.19$, $p<.01$) and the unambiguous ones ($t_1(23)=2.67$, $p=.01$, $t_2(47)=2.91$, $p<.01$). The interaction was seemingly driven by the much larger plausibility effect for the ambiguous words. This finding suggests that the older adults, like the young adults, noticed that the implausibility of the nouns following NV homographs invoked its context-*inappropriate* meaning, and as such, they allocated the most time to rereading the area of the text containing the homograph.

Table 9. Target word region rereading times (and standard deviations) by target word dominance for Experiment 4 (older adults)

Dominance	Plausibility	
	Implausible	Plausible
Dominant	532 (273)	339 (194)
Subordinate	534 (280)	389 (221)
Balanced	491 (270)	320 (190)
Unambiguous	404 (234)	336 (184)

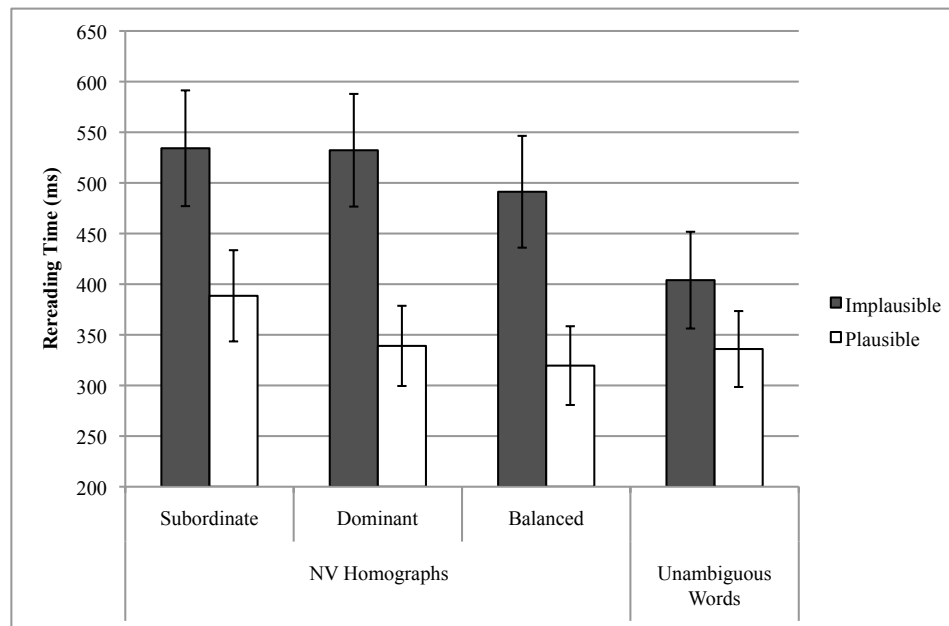
Rereading times were also examined on the target word region as a function of the word's dominance. This measure is of particular interest to us because of the intriguing effects found on young adult target region rereading times, in which the subordinate-plausible condition was inflated relative to the other plausible conditions, and did not differ significantly from the

subordinate-implausible condition. Examining these effects in older adults will help us to better understand if the effect in the young adults was caused by residual activation of the dominant meaning in the subordinate context -- in which case, the older adults would be expected to show a similar effect. Our second hypothesis for the locus of the rereading effect was that the first fixation effect generated an “uncertainty marker” that caused the young adults to return to a location of previous uncertainty, even if the ambiguity was already resolved, in which case we would not expect the older adults to exhibit the effect given the lack of first fixation ambiguity effects.

Target region rereading times broken down by dominance can be found in Table 9 (and Figure 12). Overall, rereading times were longer for implausible than plausible target word regions, but this difference was much smaller within the unambiguous words than any of the three homograph conditions. Results showed a main effect of plausibility, $F_1(1,23)=66.65$, $p<.01$, $F_2(1,140)=80.75$, $p<.01$, a main effect of dominance, $F_1(3,69)=9.79$, $p<.01$, $F_2(3,140)=6.51$, $p<.01$, as well as a significant interaction between the two, $F_1(3,69)=3.83$, $p<.05$, $F_2(3,140)=4.83$, $p<.01$. Follow-up t-tests confirmed that the plausibility effect was significant within each word type (dominant: $t_1(23)=6.25$, $p<.01$, $t_2(32)=5.22$, $p<.01$; subordinate: $t_1(23)=4.55$, $p<.01$, $t_2(32)=4.04$, $p<.01$; balanced: $t_1(23)=4.55$, $p<.01$, $t_2(29)=5.67$, $p<.01$; unambiguous: $t_1(23)=4.55$, $p<.01$, $t_2(47)=2.91$, $p<.01$). This differed from the young adult data, for which the subordinate-plausible reading times did not differ significantly from the subordinate-implausible reading times. However, for the older adults, the subordinate-plausible condition was still 50 ms longer than the dominant-plausible reading times. The difference between these same two conditions in the young adults was 56 ms, making this increase roughly equal across the two groups. However, the vastly increased rereading times in general for the

older adults implausible conditions ensured that the overall plausibility effect was still significant.

Figure 12. Target region rereading times for older adults, broken down by target dominance.



Discussion

First Pass Ambiguity Effects

This study investigated the online processing costs of encountering NV homographs in semantically neutral but syntactically constraining sentence contexts, and the downstream consequences after readers have had the opportunity to disambiguate the NV homographs. Our results showed that, as predicted, older adults did not elicit the first fixation ambiguity effects associated with encountering NV homographs when only syntactic cues were available. This replicated our previous research, showing that while younger adults elicit first fixation ambiguity effects in these circumstances, older adults as a group do not show a similar effect (Stites et al., 2013). This finding also conceptually replicates previous ERP work using similar stimulus materials, showing that older adults as a group did not elicit processing costs (i.e., the frontal

negativity effect) associated with NV homographs in syntactically neutral contexts whereas young adult readers did (Lee & Federmeier, 2009; 2011; 2012). The first fixation and frontal negativity effects are hypothesized to index frontally-mediated selection mechanisms that must be brought online for meaning selection of NV homographs in the face of constraining syntax but neutral semantics. Older adults are known to have diminished frontal lobe functioning, and so the fact that they do not elicit either of these effects adds more power to the claim that these processes likely are mediated by frontal sources.

Interestingly, older adults did still show some costs associated with the processing of the NV homographs. Go-past times were longer on the first two words of the prepositional phrase directly following the ambiguous compared to unambiguous words. Even though the older adults did not as a group slow down on their initial fixation on the homographs, in the region spanning the next the two words following the ambiguous words they were able to alter their reading patterns to slow down and/or launch regressions back to earlier parts of the sentence to address the ambiguity they just encountered. Furthermore, this effect was modulated by meaning dominance, such that the longest go-past times were following the subordinate instantiations, creating a subordinate bias effect in this region. Because the ambiguity effect was on go-past time but not gaze duration for the prepositional phrase, it indicates that the effect was driven by rereading of *earlier* parts of the text. This possibility is intriguing, especially because the ambiguity effect we observed for older adults in our previous study was in the form of increased *rereading* of the ambiguous relative to unambiguous words. Although we cannot pinpoint exactly which areas of the text were being revisited during go-past times, the majority of the sentence preceding the prepositional phrase consisted of the target ambiguous word and its preceding function word, which instantiated the context-appropriate meaning of the homograph.

One way to assess how much of the go-past time was actually spent on regions of the sentence prior to the prepositional phrase (i.e., likely the target and/or function words) is by subtracting selective go-past time for the phrase from overall go-past time for the phrase, leaving only those reading times on earlier parts of the sentence. We can also calculate this same measure for the young adults as a form of baseline against which to compare the older adults. In doing so, we find that older adults spent 92 ms, on average, on the region before the prepositional phrase for ambiguous words (or 19% of their overall go-past time), compared to 60 ms (or 13% of go-past time) for the unambiguous words. In contrast, the young adults spent only 32 ms (8% of go-past time) on this region for ambiguous words, and 21 ms (6% of go-past time) for the unambiguous words. Thus, older adults spent a much larger percentage of their go-past times reading earlier parts of the sentence, which likely indexes fixations to the target and/or function words, before moving forward in the text.

Furthermore, we can also assess the probability of launching a regression back from the prepositional phrase as another way to compare reading patterns across the two groups. Looking at this probability can help ensure that the increase in average reading times on this region for the older adults is not driven by a few over-inflated cases but rather reflects a general tendency across the group. Older adults launched a regression from the prepositional phrase in 22% of ambiguous trials and 16% of unambiguous trials, compared to the young adults, who did so on only 11% of ambiguous and 7% of unambiguous trials. The older adults were thus twice as likely as the young adults to launch a regressive saccade from the prepositional phrase region. These descriptive measures together portray two different reading patterns on the prepositional phrase region across the two age groups. The young adult readers elicited first fixation ambiguity effects on both the target word and the prepositional phrase, as well as increased go-past times on the

prepositional phrase, while generally continuing their forward movement through the text. The older adults, on the other hand, were much more likely make a regression back from the prepositional phrase region, and to spend proportionally more time in the region preceding the prepositional phrase compared to the young adults.

This additional time spent on rereading earlier parts of the sentence, particularly on the homograph and/or function word, before moving past the prepositional phrase, may have helped the older adults “make up” some of the suppression that they failed to do earlier (based on the lack of the first fixation effect on the NV homographs). In our previous study on ambiguity resolution in older adults (Stites et al., 2013), their ambiguity effects manifested as more rereading of the ambiguous words relative to unambiguous words. The increased go-past times for ambiguous relative to unambiguous words, particularly those on the pre-prepositional phrase region, effectively replicate this finding. Together with our previous study, the current results suggest that older adults may need more time to recognize the ambiguity and bring online their selection mechanisms—thus causing their ambiguity resolution processes to manifest in increased rereading times rather than earlier onsetting effects. All of these rereading times captured in the go-past ambiguity effect take place before readers reach the downstream noun in the prepositional phrase, so any work they did to disambiguate should affect their plausibility effects downstream. Even though older adults exhibit much bigger rereading ambiguity effects than do the young adults, they still do not seem to achieve the same level of meaning selection, as evidenced by the reduction in plausibility effects at the downstream noun region following ambiguous words (to be discussed below).

Plausibility Effects

The central question addressed by this study targeted the downstream consequences of ambiguity resolution processes in older adults. The stimuli were designed to probe the outcome of meaning selection by manipulating the noun in a downstream prepositional phrase to only be plausible given the preceding homograph's context-appropriate meaning. Importantly, the implausible continuations were actually plausible with the homograph's context-*inappropriate* meaning. As such, a large plausibility effect at this region would indicate successful meaning selection, whereas facilitated reading times for the implausible words would indicate continued facilitation of the context-inappropriate meaning. Given that the older adults did not elicit the first fixation ambiguity effect, we originally predicted that they would show continued activation of the context-inappropriate meaning downstream, and that as a result their plausibility effects at the later noun region would be greatly reduced or absent. However, older adults elicited robust plausibility effects following ambiguous and unambiguous words. This result indicates largely successful meaning selection: for each of the first pass measures on the downstream noun region, readers exhibited longer reading times on implausible than plausible noun regions, which were only slightly smaller than those following unambiguous words for the go-past time measure. Even in the absence of the first fixation effect, readers were able to eventually engage processes that enabled them to select the correct meaning well enough for them to recognize the implausible downstream noun region as such.

The reduction in plausibility effects following ambiguous words, coupled with the lack of first fixation effects, adds another piece of evidence to the claim that the first fixation effect reflects meaning selection mechanisms: in the absence of the effect, older adults were less able to fully suppress the context-appropriate meaning of the homograph, thus reducing (but not

eliminating) the plausibility effects observed following ambiguous words. It is important to note, however, that the older adults *did* exhibit increased go-past times on the prepositional phrase region directly following the target homographs, indicating that they spent more time rereading earlier parts of the sentence before moving forward. This effect was especially large for the regions following subordinate instantiations. The increase of go-past times on the prepositional phrase would have allowed the older adults time to reread the homographs before reaching the later noun region (as previously discussed), which could explain why they were still able to exhibit plausibility effects in the absence of the first fixation effect. Instead, the smaller magnitude of the plausibility effects following ambiguous words suggests that even with the increased go-past times on the prepositional phrase, the older adults were still unable to fully select the context-appropriate meaning of the homograph, thus facilitating reading times on the implausible words (as well as increasing reading times on the plausible words).

Relationship Between Ambiguity and Plausibility Effects

To further develop the explicit link between the first fixation effect and meaning selection, we tested the prediction that bigger first fixation effects at the NV homograph should produce bigger plausibility effects at the downstream noun region. This analysis uncovered a marginally significant positive correlation between the two, indicating that the bigger one's first fixation effect was, the more effectively they were able to suppress the context-inappropriate meaning of the NV homograph, thus producing larger plausibility effects downstream. Although this effect is only marginal, it is important to be able to highlight the individual variation in the expression of the first fixation effect. Furthermore, its predictive relationship with downstream plausibility effects provides a functional link between the effect and its consequences for ambiguity resolution. One reason that the effect did not fully reach significance could be that it

was not the only ambiguity resolution strategy available to the older adults, as it was in the ERPs. In this study, the older adults were able to control the speed and order with which they progressed throughout the sentence, and as a group they chose to slow down and/or regress back in the text from the prepositional phrase region following ambiguous words to a greater extent than they did following unambiguous words. Nevertheless, the extra rereading of the target word before reaching the later noun region was not enough to completely override the importance of their initial first fixation effect on their eventual ambiguity resolution. This correlation highlights the importance of the processes underlying the first fixation effect in the selection the ambiguous word's meaning. Despite slowing down more following ambiguous than unambiguous words, this extra reading time cannot make up for failing to elicit the frontally-mediated selection mechanisms upon first encountering the NV homograph.

Rereading Effects on Target Word Region

Lastly, rereading times on the target word region were examined to test for lingering effects of ambiguity after readers left them in the first pass. In the ambiguous sentences, total times on the target word region were differentially inflated in the implausible context compared to the implausible-unambiguous target word regions. This effect shows that older adult readers did not treat the implausibility equally across the two sentence types. They recognized when the implausibility of the downstream noun was caused by the incorrect instantiation of the NV homograph, and preferentially looked back to the target and/or function words, possibly to double-check that they interpreted it correctly. On the other hand, when implausible downstream nouns followed unambiguous target words, readers did not spend as much time rereading the target word region (although the plausibility effect was still significant within unambiguous words). This could be due to a general rereading strategy over the entirety of the sentence when

it was implausible, but in this case, the rereading was not focused specifically on the target word region as it could not offer them assistance in interpreting the implausible sentence. This tendency to preferentially return to the ambiguous words in the implausible sentences was also found when young adults read these sentences (Stites & Federmeier, in prep), suggesting that both young and older adult readers adjusted their strategic allocation of attention to different parts of the sentence depending on the type of implausibility they encountered.

One difference between the young and older adult data was in the rereading of the subordinate-plausible target word. We found an interesting pattern in the young adults, in which the subordinate-plausible condition stood out, with reading times as long as those in the implausible condition. This suggested either 1) incomplete ambiguity resolution/inability to fully suppress the homograph's dominant meaning, thus rendering the plausible continuations following the subordinate instantiations "less good"/"less plausible" and causing rereading of the target word region, or 2) a double-checking process—perhaps driven by the generation of an uncertainty marker left by the first fixation at that point in the text, which readers may be driven to return to even if the ambiguity has been resolved. The older adults provide an interesting test case for these two predictions---if the increased subordinate-plausible reading times were due to incomplete suppression of the dominant meaning, then we would expect to see increased subordinate-plausible rereading times for the older adults as well (compared to the other plausible conditions), because they have been shown to experience particular difficulty in suppressing dominant meanings. However, if the increased subordinate-plausible reading times are just about double-checking on residual *uncertainty* (i.e., recognizing that something was ambiguous earlier, even if the ambiguity has been resolved), potentially driven by the first fixation effect, then we would not expect the older adults to show this effect because as a group

they did not elicit the first fixation ambiguity effect. Interestingly, the older adults seem to show a little bit of both patterns. They do reread the subordinate-plausible condition more than the dominant-plausible (and in fact, display almost the same numeric effect as do the young adults), suggesting that they may have deemed the subordinate-plausible a little *less* plausible than the dominant or balanced conditions, potentially because of residual activation of the dominant meaning. The fact that they at least partially show this increase in subordinate-plausible reading times suggests that the effect in the young adults was not driven entirely by an uncertainty marker generated by the first fixation effect. However, older adults' rereading responses were still dominated by the overall effects of plausibility, in which all implausible conditions were reread for far more time than the plausible conditions for all NV homographs, regardless of dominance. Particularly striking was the lack of difference between the implausible reading times across the dominant and subordinate contexts—older adults *must* have been able to select the context-appropriate meaning, otherwise they would not have elicited plausibility effects to the dominant-related words used implausibly (i.e., in the subordinate-biased contexts) because the dominant meaning would have continued to remain active. The rereading results thus suggest that the older adults were generally successful at suppressing the context-inappropriate meaning, although with the suggestion for slightly more difficult suppression of the dominant meaning.

The Importance of Control

Overall, the findings from this study seem to suggest the critical role that having *control* over their rate of information intake plays in older adults' ability to select the meaning of an ambiguous word, over and above than the *amount* of time they have to read the homographs. Our index of successful meaning selection is the size of one's downstream plausibility effect, which can only be elicited if readers have suppressed the context-inappropriate meaning of the word.

While readers in the current study showed robust plausibility effects following both ambiguous and unambiguous words (which were only slightly reduced on some measures following ambiguity), the older adults in the ERP study using the same stimuli showed significantly reduced plausibility effects, particularly following subordinate instantiations (Lee & Federmeier, 2012). The presence of an effect in the current study, and lack thereof in the ERP study, may at first seem counterintuitive when we consider that aspects of the stimulus presentation style in ERPs may actually be beneficial for older adults' meaning selection. First, the presentation rate in the ERP study was slower than natural reading (one word every 500 ms). Additionally, readers are presented with every word, including the function words directly preceding the NV homographs that critically cue its context-inappropriate meaning but are typically skipped in natural reading. This combination of factors could potentially aid in older adults' ambiguity resolution, by giving them more time to recruit their meaning selection mechanisms. However, they were still unable to suppress the dominant meaning of the NV homographs, as evidenced by their downstream N400 plausibility effects. There was a hint that older adults may have recognized the ambiguity of the NV homograph, though—they elicited a small negativity to the ambiguous words from 250-500 ms post-stimulus onset, which was restricted relative to that elicited by the young adults in terms of both amplitude and duration. In the current study, although there were no first fixation effects at the homograph, it is possible that the same signal which produced the small negativity in the ERP study was also generated in the current study when the older adults encountered the ambiguous word. However, due to a combination of older adults' inefficiency in recruiting selection mechanisms along with pre-programmed motor routines for reading (Reichle, Pollatsek, & Rayner, 2006) the older adults were only able to utilize this cue after they had already moved forward in the text. As such, older adults used the

control allowed to them by the eye-tracking methodology to move back in the text and reread the target word region before reaching the implausible words, which they chose to do for ambiguous words more than unambiguous words, and especially so for subordinate instantiations. This strategy gave older adults more time to process the ambiguous than unambiguous words, resulting in more effective ambiguity resolution relative to the older adults in the ERP study. However, older adults in the eye-tracking study still moved through the sentence at a faster pace than in the ERP study—landing time on the noun in the prepositional phrase ranged from 1935-2125 ms on average in the eye-tracking study, but was held constant at 3500 ms after sentence-onset in the ERP study. The freedom that the eye-tracking methodology allowed readers to go back in the text and reread the homograph before moving forward enabled the older adults to achieve much better ambiguity resolution in the eye-tracking study as opposed to the ERP study, despite the 1400 extra milliseconds that the older adults in the ERP study received before being presented with the downstream noun. The word-by-word presentation fashion used in ERPs unfortunately does not allow subjects to control the timing or order of stimulus presentation. In this way, the strategies and strengths of older adult readers are missed because they do not conform to the timescale put forth by ERP data collection. Part of the value of parallel studies such as this one is to allow us to compare outcomes and strategies across measures, to illuminate the processes underlying ambiguity resolution, and language comprehension more generally, regardless of the tools used to study it.

General Discussion

The purpose of this project was to investigate the consequences for online processing that occur when a conflict arises between different levels of linguistic representation. This question was carried out by studying the interaction between the mechanisms supporting word recognition and the semantic and syntactic information conveyed at the sentence level. The first set of studies tested the prediction that between-morpheme letter transpositions would be more disruptive than within-morpheme transpositions in compound words encountered during natural reading. Our results did not support this prediction. Instead, Experiment 1 showed equally inflated reading times for these two types of letter transpositions relative to correctly spelled words in the eye-tracking domain, and Experiment 2 found that they elicited equally large late posterior positivities compared to correctly spelled words when we compared ERP effects across the word conditions. The ERP evidence also showed that the misspelled words, both compound and non-compound, elicited N400 amplitudes that were equal to their correctly spelled counterparts, further suggesting that the presence of transposed letters does not impede the word's ability to effectively contact semantics, regardless of their placement relative to morpheme boundaries.

The current study thus provides evidence against the role of early morphological decomposition in English compound word recognition during natural reading. A number of studies suggest that morpheme boundaries can be salient in visual word recognition in some tasks or contexts (Christianson et al., 2005; Duñabieta et al., 2007; Luke & Christianson, 2011, 2013, under review; Rastle et al., 2004), but appear not to be so salient in others (Duñabieta et al., 2007; Masserang & Pollatsek, 2012; Rueckl & Rimzhim, 2011; Sánchez-Gutiérrez & Rastle, 2013). To date, most of these studies have been conducted using single-word presentation and have focused on inflectional or derivational morphology. Given the conflicting results, it is likely

that differences in task and/or materials, along with other potential factors (e.g., frequency, Luke & Christianson, 2011; extreme predictability, Luke & Christianson, 2012), can affect results in ways that are not yet completely understood. Compound words in English are semantically complex and orthographically inconsistent – with some compounds spaced and some unspaced – and their processing is influenced by sentence context (Juhasz, 2012). The present study suggests, however, that the morphemic boundaries in English compounds are not salient in visual word recognition in relatively unconstrained contexts during normal silent reading.

The second set of studies investigated readers' ability to use syntactic cues to resolve the ambiguity of noun/verb homographs, and the downstream consequences of these resolution processes. Results showed that young adults readers elicited first fixation ambiguity effects on the NV homographs in these contexts, but that older adults instead elicited longer go-past times on the region immediately *following* the NV homographs. In both groups of readers, these ambiguity effects were driven almost entirely by cases in which the context instantiated the subordinate meaning of the homograph, suggesting that the recruitment of additional resources was not necessary for meaning selection in general, but instead that the frontally-mediated selection mechanisms were especially necessary to suppress a more frequent, dominant meaning when the cues for doing so came from outside of the semantic network. Furthermore, in both age groups, the size of their first fixation ambiguity effect was correlated with the size of their downstream plausibility effect (although this effect was only marginally significant in the older adults). Together, these findings suggest that the size of one's first fixation ambiguity effect indexed how successfully they were able to suppress the context-inappropriate meaning of the NV homograph, providing evidence for the functional link between the first fixation ambiguity effect and frontally-mediated meaning selection mechanisms.

In general, then, to answer the question of how readers handle conflict that arises between multiple streams of information, the findings from the current research suggests that readers weight the information coming from their semantic system more heavily than they do for information coming from structural cues. The first set of studies showed that even relatively unconstraining sentence contexts allowed for the compound words with transposed letters to be used to successfully make contact with semantics, despite the disruption of the word's morphemic structure. The second set of studies showed that readers exhibit reading time costs when they have to use syntactic cues to select the context-appropriate meaning of a noun/verb homograph, especially in cases when the sentence structure must be used to override the automatic activation of the word's more frequent dominant meaning. Overall, these findings reflect a language processing system that weights semantic information most heavily when a conflict arises between levels of representation, whereas structural cues such as a word's morphological boundaries or the purely syntactic information contained in a sentence have a more limited capacity to affect comprehension, particularly the initial stages of word recognition.

References

- Andrews, S., Miller, B., & Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology*, 16(1-2), 285-311.
- Baayen, R. H., Piepenbrock, R. & Gulikers, L., The CELEX Lexical Database (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA, 1995.
- Balota, D. A., & Duchek, J. M. (1991). Semantic priming effects, lexical repetition effects, and contextual disambiguation effects in healthy aged individuals and individuals with senile dementia of the Alzheimer type. *Brain and Language*, 40, 181-201.
- Balota, D. A., Watson, J. M., Duchek, J. M., & Ferraro, F. R. (1999). Cross-modal semantic and homograph priming in healthy young, healthy old, and in Alzheimer's disease individuals. *Journal of the International Neuropsychological Society*, 5, 626-640.
- Benton, A. L., & Hamsher, K. (1978). *Multilingual aphasia examination manual*. Iowa City, IA: University of Iowa.
- Beyersmann, E., McCormick, S. F., & Rastle, K. (2013). Letter transpositions within morphemes and across morpheme boundaries. *The Quarterly Journal of Experimental Psychology*, 37–41. doi: 10.1080/17470218.2013.782326
- Bilenko, N. Y., Grindrod, C. M., Myers, E. B., & Blumstein, S. E. (2009). Neural correlates of semantic competition during processing of ambiguous words. *Journal of Cognitive Neuroscience*, 21(5), 960–975. doi:10.1162/jocn.2009.21073.Neural
- Brouwer, H., Fitz, H. Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, 1446, 127-143.
- Carreiras, M., Vergara, M., & Perea, M. (2007). ERP correlates of transposed-letter similarity effects: are consonants processed differently from vowels? *Neuroscience letters*, 419(3), 219–24. doi:10.1016/j.neulet.2007.04.053
- Carreiras, M., Vergara, M., & Perea, M. (2009). ERP correlates of transposed-letter priming effects: the role of vowels versus consonants. *Psychophysiology*, 46(1), 34–42. doi:10.1111/j.1469-8986.2008.00725.x
- Christianson, K., Johnson, R. L., Rayner, K. (2005). Letter transpositions within and across morphemes. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 31(6), 1327-1339.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–56.

- Connelly, S. L., Hasher, L., & Zacks, R. T. (1991). Age and reading: The impact of distraction. *Psychology and Aging*, 6(4), 533-541.
- Dagerman, K. S., MacDonald, M. C., & Harm, M. W. (2006). Aging and the use of context in ambiguity resolution: Complex changes from simple slowing. *Cognitive Science*, 30, 311-345.
- Dambacher, M., & Kliegl, R. (2007). Synchronizing timelines: Relations between fixation durations and N400 amplitudes during sentence reading. *Brain Research*, 1155, 147-62. doi:10.1016/j.brainres.2007.04.027
- Diependaele, K., Grainger, J., & Sandra, D. (2012). Derivational morphology in skilled reading: An empirical overview. In M. J. Spivey, K. McRae, & M. F. Joanisse (Eds.), *The Cambridge Handbook of Psycholinguistics* (311-332). New York: Cambridge University Press.
- Diependaele, K., Morris, J., Serota, R. M., Bertrand, D., & Grainger, J. (2012). Breaking boundaries: Letter transpositions and morphological processing. *Language and Cognitive Processes*, 1-16.
- Duñabeitia, J. A., Kinoshita, S., & Norris, D. (2011). Is morpho-orthographic decomposition purely orthographic? Evidence from masked priming in the same – different task. *Brain and Language*, 37-41.
- Duñabeitia, J. A., Molinaro, N., Laka, I., Estévez, A., & Carreiras, M. (2009). N250 effects for letter transpositions depend on lexicality: “casual” or “causal”? *NeuroReport*, 20(4), 381-387. doi:10.1097/WNR.0b013e3283249b1c
- Duñabieta, J. A., Perea, M., Carreiras, M. (2007). Do transposed-letter similarity effects occur at a morpheme level? Evidence for morpho-orthographic decomposition. *Cognition* 105, 691-703.
- Faust, M. E., Balota, D. A., Duchek, J. M., Gernsbacher, M. A., & Smith, S. (1997). Inhibitory control during sentence comprehension in individuals with dementia of the Alzheimer type. *Brain and Language*, 57, 225-253.
- Federmeier, K. D., Segal, J. B., Lombrozo, T., & Kutas, M. (2000). Brain responses to nouns, verbs and class-ambiguous words in context. *Brain : A Journal of Neurology*, 123 Pt 12, 2552-66. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11099456>
- Filik, R. (2008). Contextual override of pragmatic anomalies: evidence from eye movements. *Cognition*, 106(2), 1038-46. doi:10.1016/j.cognition.2007.04.006

- Filik, R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: evidence from the N400. *Psychophysiology*, 45(4), 554–8. doi:10.1111/j.1469-8986.2008.00656.x
- Folk, J. R., & Morris, R. K. (2003). Effects of syntactic category assignment on lexical ambiguity resolution in reading: an eye movement analysis. *Memory & Cognition*, 31(1), 87–99. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12699146>
- Frazier, L., & Rayner, K. (1987). Resolution of syntactic category ambiguities : Eye movements parsing lexically ambiguous sentences. *Journal of Memory and Language*, 26, 505–526.
- Frisson, S., Niswander-Kelment, E., & Pollatsek, A. (2008). The role of semantic transparency in the processing of English compound words. *British Journal of Psychology*, 99, 87–107.
- Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(3), 430–445.
- Gomez, P., Perea, M., & Ratcliff, R. (2008). The overlap model: A model of letter position coding. *Psychological Review* 115(3), 577–601.
- Gouvea, A. C., Phillips, C., Kazanina, N., Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, 25(2), 149–188.
- Grainger, J., Kiyonaga, K., & Holcomb, P. J. (2006). The time course of orthographic and phonological code activation. *Psychological Science*, 17(12), 1021–1027.
- Guerrera, C. & Forster, K. (2008). Masked form priming with extreme transposition. *Language and Cognitive Processes* 23(1), 117–142.
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an eRP measure of syntactic processing. *Language and Cognitive Processes*, 8(4), 439–483.
- Hamm, V. P., & Hasher, L. (1992). Age and the availability of inferences. *Psychology and Aging*, 7(1), 54–64.
- Hartman, M., & Hasher, L. (1991). Aging and suppression: Memory for previously relevant information. *Psychology and Aging*, 6(4), 587–594.
- Heilman, K. M., Tucker, D. M., & Valenstein, E. (1976). A case of transcortical aphasia with intact naming. *Brain*, 99, 415–426.
- Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance in patients with traumatic brain injury. *Neuropsychology*, 18(4), 621–8. doi:10.1037/0894-4105.18.4.621

- Hopkins, K. A., Kellas, G., & Paul, S. T. (1995). Score of word meaning activation during sentence processing by young and older adults. *Experimental Aging Research*, 21, 123-142.
- Johnson, R. L., & Eisler, M. E. (2012). The importance of the first and last letter in words during sentence reading. *Acta Psychologica*, 141(3), 336–51. doi:10.1016/j.actpsy.2012.09.013
- Juhasz, B. J. (2012). Sentence context modifies compound word recognition: Evidence from eye movements. *Journal of Cognitive Psychology*, 24(7), 855-870.
- Kutas, M. & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Brain*, 307, 161-163.
- Kutas, M. & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621-647.
- Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, 61(3), 326–338. doi:10.1016/j.jml.2009.06.004
- Laszlo, S., & Federmeier, K. D. (2010). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology*, 48, 176–186. doi:10.1111/j.1469-8986.2010.01058.x
- Lee, C.-L., & Federmeier, K. D. (2006). To mind the mind: an event-related potential study of word class and semantic ambiguity. *Brain Research*, 1081(1), 191–202. doi:10.1016/j.brainres.2006.01.058
- Lee, C.-L., & Federmeier, K. D. (2009). Wave-ering: An ERP study of syntactic and semantic context effects on ambiguity resolution for noun/verb homographs. *Journal of Memory and Language*, 61(4), 538–555. doi:10.1016/j.jml.2009.08.003
- Lee, C.-L., & Federmeier, K. D. (2011). Differential age effects on lexical ambiguity resolution mechanisms. *Psychophysiology*, 48(7), 960–72. doi:10.1111/j.1469-8986.2010.01158.x
- Lee, C.-L., & Federmeier, K. D. (2012). Ambiguity's Aftermath: How age differences in resolving lexical ambiguity affect subsequent comprehension. *Neuropsychologia*, 50(5), 869-879.
- Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences of the United States of America*, 106(50), 21086-21090.

- Luis, C. A., Keegan, A. P., & Mullan, M. (2009). Cross validation of the Montreal Cognitive Assessment in community dwelling older adults residing in the southeastern US. *International Journal of Geriatric Psychiatry*, 24, 197–201. doi:10.1002/gps.2101
- Luke, S. G., & Christianson, K. (2011). Stem and whole-word frequency effects in the processing of inflected verbs in and out of a sentence context. *Language and Cognitive Processes*, 26(8), 1173-1192.
- Luke, S. G., & Christianson, K. (2012). Semantic predictability eliminates the transposed-letter effect. *Memory & Cognition*, 40(4), 628-641.
- Luke, S. G. & Christianson, K. (2013) The influence of frequency across the time course of morphological processing: Evidence from the transposed-letter effect. *Journal of Cognitive Psychology*, DOI: 10.1080/20445911.2013.832682.
- Luke, S. G., & Christianson, K. (under review). Predicting inflectional morphology from context.
- Masserang, K. M. & Pollatsek, A. (2012). Transposed letter effects in prefixed words: Implications for morphological decomposition. *European Journal of Cognitive Psychology*, 24(4), 476-495.
- Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(4), 913–34. doi:10.1037/a0022964
- Meyer, A., & Federmeier, K. D. (2010). Event-related potentials reveal the effects of aging on meaning selection and revision. *Psychophysiology*, 47, 673-686.
- McKinnon, R., Allen, M., & Osterhout, L. (2003). Morphological decomposition involving non-productive morphemes: ERP evidence. *Neuroreport*, 14(6), 883–6. doi:10.1097/01.wnr.0000070192.28954.7e
- Müntz, T. F., Heinze, H. J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: no evidence for specificity of the syntactic positive shift. *Neuropsychologia*, 36(3), 217–26.
- Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18(7), 1098–111. doi:10.1162/jocn.2006.18.7.1098
- Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the Cognitive Control Connection. *Language and Linguistics Compass*, 4(10), 906-924.

- Osterhout, L., & Holcomb, P. J. (1992). Event-related potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785-806.
- Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language*, 67(4), 426–448. doi:10.1016/j.jml.2012.07.003
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, 10(6), 717-733.
- Patson, N. D., & Warren, T. (2011). Eye movements when reading implausible sentences: Investigating potential structural influences on semantic integration. *Quarterly Journal of Experimental Psychology*, 63(8), 1516–1532. doi:10.1080/17470210903380822
- Perea, M. & Carreiras, M. (2006). Do transposed-letter effects occur across lexeme boundaries? *Psychonomic Bulletin & Review*, 13(3), 418–22.
- Perea, M., Duñabeitia, J.A., Carreiras, M. (2009). Transposed-letter priming effects for close versus distant transpositions. *Experimental Psychology*, 55(6), 384-393.
- Perea, M. & Lupker, S.J. (2003). Does judge activate COURT? Transposed-letter similarity effects in masked associative priming. *Memory & Cognition*, (31)6, 829-841.
- Perea, M. & Lupker, S.J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language*, 51, 231-246.
- Pollatsek, A. & Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes*, 21(1-2), 261-290.
- Princeton University "About WordNet." (2010). WordNet. Princeton University. <<http://wordnet.princeton.edu>>
- Rastle, K. & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, 23(7/8), 942-971.
- Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, 11(6), 1090-1098.
- Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 779–90. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2528602>

- Rayner, K., Pacht, J. M., & Duffy, S. A. (1994). Effects of prior encounter and global discourse on the processing of lexically ambiguous words: Evidence from eye fixations. *Journal of Memory and Language*, 33, 527–544.
- Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1290–301. doi:10.1037/0278-7393.30.6.1290
- Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, 3(4), 504–509. doi:10.3758/BF03214555
- Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., Dahle, C., Gerstorf, D., & Acker, J. D. (2005). Regional brain changes in aging healthy adults: General trends, individual differences, and modifier. *Cerebral Cortex*, 15(11), 1676–1689.
- Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E–Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research*, 7(1), 4–22. doi:10.1016/j.cogsys.2005.07.002
- Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, 46(2), 245–266. doi:10.1006/jmla.2001.2810
- Rueckl, J. G. & Rimzhim, A. (2011). On the interaction of letter transpositions and morphemic boundaries. *Language and Cognitive Processes*, 26(4/5/6), 482–508.
- Sánchez-Gutiérrez, C. & Rastle, K. (2013). Letter transpositions within and across morphemic boundaries: Is there a cross-language difference? *Psychonomic Bulletin & Review*. doi:10.3758/s13423-013-0425-0
- Sauseng, P., Bergmann, J., & Wimmer, H. (2004). When does the brain register deviances from standard word spellings? An ERP study. *Cognitive brain research*, 20(3), 529–532. doi:10.1016/j.cogbrainres.2004.04.008
- Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics*, 74(5), 5–35.
- Stites, M. C., Federmeier, K. D., & Stine-Morrow, E. A. L. (2013, May 20). Cross-age comparisons reveal multiple strategies for lexical ambiguity resolution during natural reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. doi: 10.1037/a0032860

- Stuss, D. T., & Levine, B. (2002). Adult clinical neuropsychology: lessons from studies of the frontal lobes. *Annual Review of Psychology*, 53, 401–33. doi:10.1146/annurev.psych.53.100901.135220
- Swaab, T., Brown, C., & Hagoort, P. (2003). Understanding words in sentence contexts: The time course of ambiguity resolution. *Brain and Language*, 86, 326–343.
- Taft, M. & Nilsen, C. (2013). Morphological decomposition and the transposed-letter (TL) position effect. *Language and Cognitive Processes*, 28, 917–938.
- Thompson-Schill, S. L., Bedney, M., & Goldberg, R. F. (2005). The frontal lobes and the regulation of mental activity. *Current Opinion in Neurobiology*, 15, 219–224.
- Van de Meerendonk, N., Indefrey, P., Chwilla, D. J., & Kolk, H. H. J. (2011). Monitoring in language perception: Electrophysiological and hemodynamic responses to spelling violations. *NeuroImage*, 54(3), 2350–63. doi:10.1016/j.neuroimage.2010.10.022
- Vergara-Martínez, M., Perea, M., Gómez, P., & Swaab, T. Y. (2013). ERP correlates of letter identity and letter position are modulated by lexical frequency. *Brain and Language*, 125(1), 11–27. doi:10.1016/j.bandl.2012.12.009
- Vergara-Martínez, M., Perea, M., Marín, A., & Carreiras, M. (2011). The processing of consonants and vowels during letter identity and letter position assignment in visual-word recognition: An ERP study. *Brain and Language*, 118(3), 105–17. doi:10.1016/j.bandl.2010.09.006
- Vissers, C. T. W. M., Chwilla, D. J., & Kolk, H. H. J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, 1106(1), 150–63. doi:10.1016/j.brainres.2006.05.012
- Warren, T., & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review*, 14(4), 770–5.
- Warren, T., McConnell, K., & Rayner, K. (2008). Effects of context on eye movements when reading about possible and impossible events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 1001–10. doi:10.1037/0278-7393.34.4.1001
- Wlotko, E. W. & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage*, 62(1), 356–66.