Impact of Animacy and Aspectual Information on Semantic and Syntactic Processing: Evidence from Event-Related Potentials

A dissertation submitted by

Martin Paczynski

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Psychology

COMMITTEE:

Gina R. Kuperberg, M.D., Ph.D. (chair)

Phillip J. Holcomb, Ph.D.

Ray Jackendoff, Ph.D.

Tufts University

Tali Ditman-Brunyé, Ph.D.

Massachusetts General Hospital

TUFTS UNIVERSITY

May 2012

Abstract

Part I examined the role of animacy in verb-argument processing. Experiment 1 found distinct patterns of neural activation in the processing of verb arguments that violated the animacy restrictions of the preceding verb compared to violations of real-world knowledge. Experiment 2 found that animacy affected the processing of subject noun arguments as well as the subsequent verb. Importantly, this processing was independent of thematic role assignment. Together, these results underscore animacy's critical role in semantic and syntactic processing. Part II examine the time-course of aspectual interpretation. Experiment 3 found that aspectual information is computed as soon as it becomes licensed. Additionally it demonstrated that processing costs associated with iterative coercion are independent of iterativity itself as well as general shifts in aspectual interpretation. Experiment 4 found that iterative interpretation of punctual verbs in the progressive is independent of neural processing associated with aspectual coercion. Additionally, the results indicated that high and low proficiency language users differ qualitatively in aspectual interpretation requiring enriched composition but quantitatively when aspectual interpretation is complex but compositionally transparent. Together the two studies suggest that online event interpretation relies on a complex interplay of multiple neural substrates.

KEYWORDS: semantic, syntactic, verb argument, animacy, lexical aspect, grammatical aspect, aspectual coercion, ERP

Acknowledgements

As with all large undertakings, this dissertation is the work of numerous individuals whose contributions were critical for the completion of the project. First and foremost, I would like to thank my advisor and mentor, Gina Kuperberg, who provided me with this wonderful opportunity to study here at Tufts and who spent countless hours, morning and night, helping me develop ideas and structure my prose, and always insisting on the highest level of intellectual and professional rigor. I would also like to thank my committee, Phil Holcomb, Ray Jackendoff, and Tali Ditman-Brunyé who provided invaluable feedback, critique, and support throughout the conception and life of this project. Additionally, I am grateful to the Neurocognition Lab members, past and present, who made coming to work a pleasure, rather than a burden. In particular, I would like to thank Neil Cohn for his role in keeping me sane and helping me rechannel excess energy into martial arts rather than nail biting.

As no scientist is just a brain in a jar, I would like to thank my friends and family for their part in this undertaking. A big thank you to my parents, Hanna Paczyńska and the late Bohdan Paczyński, for getting me started on the road to scientific inquiry at an early age and tirelessly encouraging me to follow my passion, rather than settle for anything less. I would like to thank my sister, Agnieszka Paczyńska, for being the trailblazer who showed me how much dogged perseverance is necessary to succeed. I would like to thank my many friends for providing social support and sense of community even when work tried to swallow me whole, especially Jessica Wandrei, Shane Landrum, and the Fitos/Bartlett family, Michel, Imre, Sabine, and Simone. A special thanks to my now ex-wife, Marta Schoonmaker Rivera Monclova, without whom I would never have joined the Tufts community and who made sure that I would never starve nor become intellectually, socially, or emotionally complacent. Finally, I would like to extend my most heartfelt gratitude to my partner, Rachel Mello, for always being there, all the way up until this not-so-bitter end, whether it was/is providing comfort and support, being an active sounding board for ideas, or occasionally giving me a desperately needed kick in the pants to

get me over life’s and work’s hurdles. Thank you, this dissertation could not have reached completion without you.

TABLE OF CONTENTS

Impact of Animacy and Aspectual Information on Semantic and Syntactic Processing: Evidence from Event-Related Potentials

Part I: Effects of Noun Phrase Animacy on Online Verb Argument Processing

Distinct Effects of Semantic Relatedness on Violations of Real-World Event/State Knowledge and Animacy Selection Restrictions During Online Processing: Evidence from Event-Related Potentials

Introduction

In order to rapidly and proficiently understand sentences, comprehenders use different types of stored semantic information. These include: 1) semantic relatedness between individual groups of words (e.g. *music*, *bass* and *guitarist* are semantically related to each other by sharing a common schema), 2) more structured real-world knowledge about the likelihood of a person carrying out a specific action or participating in a particular state in a given context (e.g. knowing that a *bass* is more likely to be *strummed* by a *guitarist* than by a *drummer*), and 3) animacy selection restrictions of a verb on its argument(s) (e.g. the verb *strum* requires that that its Agent be animate, like a *guitarist*, not inanimate, like *drum*). Each of these different types of stored semantic knowledge can influence online sentence and discourse comprehension, as indicated by findings from both behavioral studies as well as electrophysiological studies that focus on the N400—an event-related potential (ERP) thought to reflect lexico-semantic processing of words in relation to their semantic context (Kutas & Federmeier, 2011). What is currently less clear is how these different types of stored semantic knowledge interact with one another and with incoming material as meaning is built incrementally during word-by-word processing. In this ERP study, we explore these interactions by determining whether, when and how semantic relatedness between individual words influences the processing of Agent arguments that violate either our real-world knowledge about events and states, or the animacy selection restrictions of their preceding verbs.

Below we first review behavioral and ERP evidence that each of these three different types of stored semantic knowledge can influence online sentence and discourse processing. We then consider previous studies that have addressed the question of whether semantic relatedness between individual

content words can override the effects of either real-world event/state knowledge or verb-based selection restrictions, before describing how the present study was designed to address this question.

Semantic relationships between individual words: semantic relatedness

From the earliest descriptions of the 'semantic priming effect' (Meyer 1971; Becker 1979), it has been known that multiple types of semantic relationships can influence lexico-semantic processing of upcoming words. Electrophysiologically, semantic priming manifests as an attenuation of the N400 to target words that follow a semantically related prime (Bentin, 1985; Rugg, 1984). N400 attenuation is seen to targets that are semantically related to a prime along a variety of dimensions, including category membership (e.g. *tulip-ROSE*) (Grose-Fifer & Deacon, 2004), semantic features (e.g. *wig-MOP*) (Deacon, Grose-Fifer, Yang, Stanick, Hewitt, & Dynowska, 2004), or through an indirectly related mediator (e.g. *lion*-[tiger]-*STRIPES*) (Chwilla, Kolk, & Mulder, 2000; Kreher, Holcomb, & Kuperberg, 2006; Silva-Pereyra, Harmony, Villanueva, Fernandez, Rodriguez, Galan, et al., 1999). The N400 is also attenuated to target words that are related to their preceding word(s) through broader script-based or schema knowledge (e.g. scalpel-SURGEON, Deacon et al., 2004; director-bribe-DISMISSAL, Chwilla & Kolk. 2005).

Semantic priming is not only seen in paradigms presenting pairs or triplets of words. Both behavioral (Duffy, Henderson, & Morris, 1989; Foss, 1982; Morris, 1994) and ERP (Van Petten, 1993; Van Petten, Weckerly, McIsaac, and Kutas, 1997; Coulson, Federmeier, Van Petten, and Kutas, 2005; Camblin, Gordon and Swaab, 2007; Otten and Van Berkum, 2007) studies suggest that words can be primed by semantically related words within sentences. Importantly, semantic priming effects between individual words in sentences can be detected even when the sentence lacks overall semantic coherence (e.g., "When the moon is rusted it is available to buy many stars…", Van Petten, 1993). This suggests that semantic relatedness between individual words can influence processing, regardless of a sentence's message-level meaning.

Real-World Event/State Knowledge

It is also well established that real-world knowledge about events or states is used rapidly during online language processing. Participants are faster to detect (Marslen-Wilson, Brown & Tyler, 1988) and read (Rayner, Warren, Juhasz, and Liversedge, 2004; Warren and McConnell, 2007; Camblin, Gordon and Swaab, 2007) words that are congruous than incongruous with the likelihood of a particular event or state occurring in the world. A similar facilitation in processing has been observed neurophysiologically. Words which are congruous with real-world knowledge consistently been shown to attenuate the N400, compared to words which violate expectations about actions, events or states based on real-world knowledge (Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; Hagoort, Hald, Bastiaansen, Marcel, & Petersson, 2004; Ferretti, Kutas and McRae, 2007; Filik & Leuthold, 2008; Bicknell, Elman, Hare, McRae, & Kutas, 2010; van de Meerendonk, Kolk, Vissers, and Chwilla, 2010).

One might argue that the effects of real-world knowledge on sentence processing result from semantic priming between its individual content words. For example, in the sentence, "Every morning at breakfast the boys would eat/*plant…," the smaller N400 seen to *eat* (versus *plant*) might be driven by the relatedness between the group of words, *morning*, *breakfast*, *boy* and *eat*, as these share a common schema. However, there is at least some evidence to suggest that influences of real-world knowledge on language processing are not always reducible to semantic relatedness between individual content words that share schema membership. For example, we have shown that the N400 is modulated by degree of match with real-world knowledge, even when semantic relatedness between individual words is matched across conditions (Kuperberg, Paczynski, & Ditman, 2010; Nieuwland & Kuperberg, 2008). Such findings suggest that as comprehenders incrementally build a message-level interpretation of a sentence or discourse, they rely, at least to some extent, on structured real-world event/state knowledge representations to generate more specific expectations about the semantic features of an upcoming word.

When the semantic features of an incoming word match such expectations, lexico-semantic processing, reflected by the N400, is attenuated (Ferretti, Kutas, & McRae, 2007; Kutas & Federmeier, 2011) [1].

Selection Restrictions on Verb Arguments

Another way in which stored semantic information can influence lexico-semantic processing of upcoming words is through a verb's selection restrictions—the semantic constraints of a verb on its argument(s). By the far the most common type of selection restriction explored in psycholinguistics is the animacy constraints placed on verb arguments. For example, in "The farmer plowed the *laborer", *laborer* is anomalous because the verb selects for an inanimate rather than an animate direct object. Again, selection restrictions can be dissociated from semantic relatedness between individual words. For example, in "The pillow *slept," the word *pillow* violates the animacy selection restriction of the verb *sleep,* despite the two words being highly semantically related.

By definition, violations of selection restrictions also violate real-world knowledge. However, in contrast to the types of real-world knowledge event/state representations discussed above, selection restrictions may be distinctly encoded within the lexicon (Chomsky, 1965; Katz, 1963; but see Jackendoff, 2002). Consistent with this distinction, some behavioral studies have indicated a processing difference between violations of real-world knowledge and violations of selection restrictions (Marslen-Wilson, Brown, & Tyler, 1988; Warren & McConnel, 2007). These differences are not only quantitative, but qualitative as well. For example, in an eye-tracking study, Warren and McConnell (2007) reported longer regression time durations on words that violated real-world knowledge such as "The man used a blow-dryer to dry (versus a strainer to drain) ?the thin spaghetti yesterday evening.". In contrast, they reported longer first fixation durations to selection restriction violations, such as "The man used a photo

[1] It is debated exactly how and when such expectations are generated. They may arise through passive resonance mechanisms (e.g. Myers & O'Brien, 1998) or from more active prediction mechanisms (e.g. Lau, Holcomb & Kuperberg, Under Review). It is also unclear how often the semantic features of upcoming lexical items are pre-activated *before* any bottom-up information becomes available (e.g. DeLong et al., 2005; Van Berkum et al., 2005), or whether facilitation arises through an interaction with bottom-up perceptual information (e.g. Marslen-Wilson, 1987). In this study, we are less concerned with mechanism by which lexico-semantic facilitation occurs. Rather, we focus on the interactions between different types of stored knowledge to influence lexico-semantic processing of upcoming words.

to blackmail (versus a strainer to drain) *the thin spaghetti yesterday evening." The authors suggested that these types of selection restrictions might have a privileged status during processing, either because they are accessed earlier than more general real-world knowledge, or because the processing of coarse-grained constraints, such as animacy, is prioritized over more finer-grained semantic representations, as proposed by Sanford and Garrod (1998). They also noted that the selection restriction violations, but not the real-world knowledge violations, were associated with additional downstream effects past the critical noun-phrase region (see also Rayner et al., 2004).

In ERP studies, the effects of selection restriction violations can once again manifest on the N400. Words that violate animacy selection restrictions evoke a larger N400 than non-violated words, regardless of whether such violations occur on the verb itself, e.g. "The honey was *murdered." (Rösler, Pütz, Friederici & Hahne, 1993; see also Friederici & Frisch, 2000; Hahne & Friederici, 2002; Bornkessel-Schlesewsky, Kretzschmar et al. 2010) or on one of its arguments, e.g. "The businessman knew whether the secretary called the *article at home." (Garnsey, Tanenhaus et al. 1989; see also Ainsworth-Darnell, Shulman et al. 1998; Friederici & Frisch, 2000; Nieuwland & Van Berkum 2005; Li, Shu et al. 2006; Paczynski & Kuperberg, 2011).

The N400 effect on selection restriction violations has sometimes been interpreted as reflecting the implausibility of the proposition that is generated once full semantic-syntactic integration of the word into its context has occurred (Friederici & Frisch, 2000; Garnsey, Tanenhaus, & Chapman, 1989). However, N400 amplitude does not necessarily pattern with ratings of propositional plausibility (e.g. Kuperberg et al., 2003; Kuperberg, Choi et al., 2010; van de Meerendonk et al., 2010; Paczynski & Kuperberg, 2011). Instead, both our group (Kuperberg, 2007; Kuperberg et al., 2010; Paczynski & Kuperberg, 2011), as well as others (Kutas & Hillyard, 1984; Kutas & Federmeier, 2011) have argued that the amplitude of the N400 indexes the degree of match (or mismatch) between the expected and actual semantic features of an incoming word. Evidence for this position comes from studies that have

examined the effects of semantic relatedness on the modulation of the N400 evoked by violations of real-world knowledge and verb selection restrictions. We discuss these findings in the following section.

Interactions between semantic relatedness and violations of real-world knowledge and selection restrictions

As reviewed above, there is evidence that various types of stored semantic information—semantic relatedness between individual words, knowledge about the likelihood of people or things taking part in familiar real-world events or states, and animacy selection restrictions of a verb—can each modulate lexico-semantic processing of an upcoming word, as reflected by the N400. Of course, during everyday communication, all these types of information co-occur: words used to describe likely events will tend to be (or become) semantically related. As noted above, violations of selection restrictions are necessarily violations of real-world knowledge. Therefore, during normal language processing, expectations based on one type of stored knowledge will be consistent with expectations based on another. For example, in the sentence, "The pianist played his music while the bass was strummed by the guitarist…," the words *pianist, played, music, bass, strummed* and *guitarist* are semantically related through common schema membership. The sentence also conforms to our more specific world knowledge that a *bass* is likely to be strummed by a *guitarist*. Finally, the word *guitarist* is animate and conforms to the animacy selection restrictions of the verb *strum*. However, as discussed above, these different types of stored information are not all simply reducible to one another. Words can be semantically related through schema membership, but still violate real-world knowledge or selection restriction constraints. And an argument that violates real-world knowledge in a particular context does not necessarily violate the selection restrictions of a given verb. This raises the question of what happens when an incoming word matches expectations based on one type of semantic knowledge, but not another: which types of information take precedence in influencing the semantic processing of upcoming words?

Can expectations based on semantic relatedness override expectations based on real-world knowledge?

There has been some work addressing this question by pitting semantic relatedness between individual words against more specific expectations based on real-world knowledge. In an important study by Federmeier and Kutas (1999), the authors showed that the N400, evoked by a word which was incongruous with real-world knowledge, was attenuated when the word shared semantic features with the expected, congruent word (see also Kutas & Hillyard, 1984). A similar attenuation of the N400 was reported by Camblin, Gordon, and Swaab (2007) to real-world knowledge violations that were semantically related with the preceding content words (see also Ditman, Holcomb, & Kuperberg, 2007).

In some cases, the N400 to a real-world knowledge violation can be completely attenuated by semantic relatedness—a phenomenon that has been termed a neural 'semantic illusion' (Hoeks et al., 2004; see also Nieuwland & Van Berkum, 2005 and Sanford, Leuthold, Bohan, & Sanford, 2011) [2]. For example, Kolk et al. (2003) found no N400 effect to critical verbs that violated real-world knowledge expectations when both of their arguments were strongly related to both the verb and to each other, e.g. "The cat that from the mice *fled…" (versus "The mice that from the cat fled…"). Similarly, Sanford et al. (2011) failed to see an N400 effect to words that violated real-word knowledge when these were semantically related to the preceding content words through common schema membership, e.g. "Child abuse cases are being reported much more frequently these days. In a recent trial, a 10-year [care order/*sentence] was given to the victim…". Such findings indicate that, under some circumstances,

[2] In the behavioral literature, the term semantic illusion is used to describe phenomena such as 'The Moses Illusion' (Erickson, 1981), e.g. *How many animals of each kind did Moses bring on to the Ark?*. People often fail to notice the violation of real-world knowledge in such sentences, responding with the answer, *two* when, in fact, the correct answer is *zero* (see also Barton and Sanford, 1993 for similar results). It has been proposed that such behavioral illusions occur when semantic relatedness overrides the overall implausibility of propositional meaning, leading to an erroneous final sentence-level interpretation. In contrast, a neural semantic illusion is defined by an attenuation of the N400 effect produced by a violation during online processing. It need not necessarily lead to a full behavioral semantic illusion on the final interpretation of the sentence. Also, note that the term 'neural semantic illusion' was first used to describe attenuations of the N400 effect produced by certain types of selection restriction violations (Hoeks et al., 2004, Nieuwland and Van Berkum, 2005). However, as we discussed in the next section, it is not clear whether the attenuation of the N400 to these types of selection restriction violations is actually driven by semantic relatedness.

semantic relatedness between individual words can override expectations based on real-world knowledge.

Can expectations based on semantic relatedness override expectations based on selection restrictions?

It is less clear whether semantic relatedness between individual words can also override expectations based on a verb's selection restrictions to influence lexico-semantic processing. Initial evidence seemed to indicate that it could. For example, Kuperberg et al. (2003) showed that selection restriction violations occurring on verbs, which were semantically related to their preceding content words, failed to evoke an N400 effect, e.g. "Every morning at breakfast the boys/*eggs would eat…," instead evoking a P600 effect as discussed further below (see also Hoeks et al., 2004; Kim & Osterhout, 2005). However, later studies suggested that other factors were at play in reducing the N400 evoked by these types of violations. For example, we found no N400 effect to selection restriction violations on verbs, even when they were not semantically related to their preceding content words, e.g. "Every morning at breakfast the eggs would *plant…" (Kuperberg et al., 2006; see also Stroud & Philips, In Press; but see Kim and Osterhout, 2005, Experiment 2).

Most relevant to the present study, we recently showed that inanimate nouns that violated the selection restrictions of their preceding verbs did produce a robust N400 effect, even when they *were* semantically related to preceding content words, e.g. "At the homestead the farmer penalized the *meadow/farmer…" (Paczynski & Kuperberg, 2011).[3] This suggested that, at least when a selection restriction violation falls on an inanimate argument, the N400 effect produced may be relatively impervious to semantic relatedness between individual words, i.e. close semantic relatedness between individual words may fail to override broader verb-based expectations for inanimate arguments. To date,

[3] In contrast, Paczynski and Kuperberg (2011, Experiment 1) showed that selection restriction violations falling on *animate* arguments did not produce an N400 effect (e.g. "At the homestead the farmer plowed the *laborer/meadow…), despite their being rated equally implausible as the selection restriction violations on inanimate direct objects. We suggested that this was because encountering an animate argument in the direct object position of an active sentence violated *verb-independent* expectations about animacy. This is considered further in the Discussion.

however, no study has directly tested this hypothesis by manipulating semantic relatedness on inanimate arguments that violate the selection restrictions of their preceding verb.

The P600 and propositional implausibility

We have argued that the amplitude of the N400 reflects the degree to which expectations, based on the interaction between context and different types of stored semantic knowledge, match the semantic features of an upcoming word. These expectations may be based on different types of semantic relationships. However, in all cases, they reflect a *direct influence* of stored semantic information on lexico-semantic processing of upcoming words, what we have termed 'semantic memory-based' analysis (Kuperberg, 2007). We distinguished this from the processes by which a word is fully integrated into its context by combining semantic and syntactic information to construct a full propositional representation of meaning, which we refer to as full 'combinatorial analyses' (Kuperberg, 2007).

It has been known for some time that disrupting combinatorial processing by violating syntactic expectations and/or constraints can trigger a posteriorly-distributed late positivity effect, known as the P600. This waveform is thought to reflect a continued analysis or reanalysis associated with full semantic-syntactic integration of a word into the context in which it appears (Osterhout & Holcomb, 1992; Hagoort, Brown & Groothusen, 1993). There is also evidence that, under some circumstances, a P600 effect can be evoked by certain semantic violations, termed a 'semantic P600' (Kuperberg et al. 2003, Kolk et al., 2003; Hoeks et al., 2004). There has been much debate about what exactly triggers the P600 effect to semantic violations. Although several frameworks have been proposed to explain this phenomenon, they can be divided into two broad categories: those that emphasize the P600 as reflecting conflict between semantic and syntactic processes and those that emphasize the P600 as related to overall propositional plausibility.

The first category of accounts proposes that the semantic P600 is triggered by a mismatch between a semantic and syntactic interpretation. Kim and Osterhout (2005) proposed that a strong

'semantic attraction' between a verb and its argument(s), e.g. "The hearty meals were *devouring…" would cause the parser to arrive at an incorrectly interpretation, e.g. "The hearty meals were devoured…," which conflicts with the full syntactic interpretation of the phrase. More recently, Hagoort, Baggio and Willems (2009) proposed a more general version of this account in which they suggest that strong semantic cues, encompassing strong semantic relatedness between words, are sufficient to bias the initial interpretation towards one which conflicts with full syntactic interpretation, triggering a P600.

The second set of accounts emphasizes the overall propositional *implausibility* of a full combinatorial analysis as a critical factor that influences whether or not a P600 effect will be evoked (Kuperberg, 2007; van de Meerendonk et al., 2009; Bornkessel-Schlesewsky & Schlesewsky, 2008). On these accounts, severely implausible/impossible propositions are thought to evoke a P600, regardless of the level of semantic relatedness between the critical word and its preceding content words. For example, our own group found that selection restrictions violations on verbs evoked a similar P600 effect, regardless of whether the critical verb was semantically related or not to preceding content words in the context, e.g. "For breakfast the eggs would *eat/*plant…" (Kuperberg et al., 2007). There are some differences between these three accounts, including the emphasis they each place on conflicting intermediate representations in triggering the P600 effect, and we return to these distinctions in the Discussion section. For now we note that, by emphasizing propositional implausibility, all three accounts hold that task plays an important role. For example, in a recent study we demonstrated that, during passive reading, a P600 effect was observed to animacy selection restriction violations, but this was attenuated compared to when participants were asked to make explicit plausibility judgments (Wang et al., 2010). We refer the interested reader to Kuperberg (2007) for a more in-depth discussion of the interplay between propositional plausibility, semantic relatedness and task.

The present study

In the present study, we explored the effects of semantic relatedness, real-world event/state knowledge and animacy selection restrictions on the processing of Agent nouns in passive English sentences (see Table 1).

We operationally defined semantic relatedness through Semantic Similarity Values (SSVs) generated by Latent Semantic Analysis (LSA). LSA uses a large training corpus to create representations of words and relationships between them within a multidimensional semantic space (see Methods). In addition to utilizing the co-occurrence of words within the training document, LSA generates inferences regarding semantic relationships, such that words can have a high SSV even without co-occurring within a document, so long as the semantic contexts in which they appear are similar (Landauer, Foltz, & Laham, 1998). LSA has been shown to reliably model and predict human performance in various linguistics tasks, including word categorization (Laham, 1997; see also Landauer, McNamara, Dennis, & Kintsch, 2007, for additional discussion of the relationships between LSA driven analysis and human performance). Importantly for our purposes, LSA is insensitive to word order/syntax. For example, "the chef cooked the pasta" and "the pasta was cooked by the chef" share propositional meaning but not syntactic structure. On the other hand, "the chef cooked the pasta" and "the pasta cooked the chef" share syntactic structure but not propositional meaning. However, all three sentences are equally related to the word *kitchen*—a judgment accurately modeled by LSA.

We distinguish 'semantic relatedness' between groups of content words that share schema membership (as indexed by LSA) from more structured real-world event/state knowledge. In this study, we construe real-world knowledge representations as reflecting people's knowledge about familiar, repeatable events and states. For example, given an introductory context like "At the estate sale, prices are announced by the…" our knowledge about the world allows us to understand that it is more likely that an *auctioneer* would make such an announcement, rather than a *bidder*, despite both being highly semantically related to the preceding content words and both being equally able to *announce* something.

Finally, similarly to our previous studies, selection restriction violations were on the animacy of the Agent argument of verbs—specifically, the requirement that these Agents be animate, rather than inanimate. Thus, all critical words that violated these restrictions were inanimate. Unlike violations of the real-world knowledge, selection restriction violations were independent of the preceding sentential context, relying only on the relationship between the verb and its argument. Moreover, rather than yielding implausible-yet-possible propositions, selection restriction violations yielded implausible-and-impossible propositions.

We fully crossed semantic relatedness with type of violation (real-world event/state knowledge[4] versus verb-based animacy selection restriction), giving rise to five conditions: 1) plausible control, 2) semantically related violations of real-world event/state knowledge, 3) semantically unrelated violations of real-world event/state knowledge, 4) semantically related violations of animacy selection restrictions, and 5) semantically unrelated violations of animacy selection restrictions. See Table 1 for example sentences in each of the five conditions.

Our first question concerned the impact of semantic relatedness on lexico-semantic processing of Agent nouns that violated real-world knowledge expectations established by the preceding context, as reflected by the N400. Based on several studies reporting that semantic relatedness can partially (Camblin et al., 2007; Ditman et al., 2007; Federmeier & Kutas, 1999) or even completely (Kolk et al., 2003; Sanford et al., 2011) attenuate the N400 effect evoked by words that mismatch real-world knowledge expectations, we predicted that the N400 effect evoked by violations of real-world knowledge would be attenuated by strong semantic relatedness. We were interested in the open question of whether strong semantic relatedness between content words was sufficient to completely attenuate

[4] In previous studies, following Marslen-Wilson, Brown and Tyler (1988) and Tyler (1992) we used the term ‘pragmatic violations’ to describe violations of this type of real-world knowledge (Kuperberg, Caplan, Sitnikova, Eddy, & Holcomb, 2006; Kuperberg, 2007; Kuperberg, Caplan, et al., 2006; Kuperberg, Kreher, et al., 2006; Kuperberg et al., 2007; Kuperberg et al., 2003). However, we now use the term ‘real-world knowledge violations’ to avoid confusion with the use of the term ‘pragmatics’ to describe the relationship between sentence meaning and a speaker’s meaning, or to refer to a variety of phenomena at the level of discourse.

this effect, despite a lack a longer discourse context (as in Sanford et al., 2011) or strong 'semantic attraction' between the verb and the critical noun argument (as in Kolk et al., 2003).

Our second set of questions concerned the impact of semantic relatedness on lexico-semantic processing of Agent arguments that violated animacy selection restrictions of their preceding verbs, again as indexed by the N400. As discussed, previous studies have yielded mixed results as to whether semantic relatedness can attenuate the N400 effect evoked by selection restriction violations. However, in many of these studies critical arguments not only violated selection restrictions of the verb, but *also* mismatched broader verb-independent animacy expectations, specifically that animate arguments appear first and inanimate arguments appear second (see Paczynski & Kuperberg, 2011 for a more in depth discussion). In the present study, we used passive constructions, such that selection restriction violating inanimate Agentive arguments appeared *after* the verb, rather than before it, as opposed to our earlier studies (e.g. Kuperberg et al., 2003, 2006, 2007). We asked whether animacy selection restriction violations on inanimate post-verbal Agentive arguments would evoke an N400 effect, similar to our findings on post-verbal inanimate direct object arguments (Paczynski & Kuperberg, 2011), or whether they would fail to evoke an N400 effect, similar to when inanimate Agentive arguments appear before the verb (Kuperberg et al., 2003, 2006, 2007). If these post-verbal animacy selection restriction violations did evoke an N400 effect, we asked whether it would be modulated by semantic relatedness between the critical noun and the preceding content words. If selection restriction violations are processed like more general violations of real-world event/state knowledge, this would predict an attenuation of the N400 by close semantic relatedness (see above). On the other hand, if animacy selection restriction violations are functionally distinct and impervious to semantic relatedness, this would predict no N400 modulation by semantic relatedness.

Our third question concerned the semantic P600. While not the primary focus, our study offered an opportunity to explore two questions regarding the semantic P600. First, whether real-world knowledge violations occurring on nouns which are strongly semantic related to their preceding contexts

would evoke a semantic P600 within single sentences, as Sanford et al. (2011) found in a longer discourse context, or whether this effect would be seen only on violations of selection restrictions. Second, whether semantic relatedness would modulate the semantic P600 evoked by selection restriction violations. The two broad categories of frameworks highlighted above make different predictions here. According to the frameworks proposed by Kim & Osterhout (2005) and Hagoort et al. (2009), a semantic P600 effect should only be observed when semantic cues are stronger than syntactic cues. This would only be the case for related (not unrelated) inanimate selection restriction violating nouns. On the other hand, the proposals of Kuperberg (2007), Kolk and colleagues (van den Meerendonk, 2009), and Bornkessel-Schlesewsky and Schlesewsky (2010) would all predict a P600 effect for both types of selection restriction violations, as in both cases the overall propositional meaning would be implausible-and-impossible. This would be particularly likely given that participants were biased towards the detection of propositional implausibility/impossibility by making explicit acceptability judgments about the sentences.

Methods

Construction and ratings of materials

Five types of sentences were constructed (see Table 1 for explanation and examples of each type of sentence; see http://www.nmr.mgh.harvard.edu/kuperberglab/materials.htm. for the full list of stimuli). We selected 120 verbs that required animate Agents (e.g. *strummed*). For each verb, we created a fairly constraining introductory context (e.g. "The pianist played his music while the bass was strummed by…"). Plausible Control sentences were created by adding an animate noun—the critical word—that was semantically related to the content words in the preceding context to serve as a plausible Agent of the verb (e.g. *guitarist*). Related Real-World Knowledge Violation sentences were created in a similar fashion, with the exception that the semantically related animate noun could not serve as a plausible Agent within the context (e.g. *drummer*). No animate critical nouns were repeated and critical

nouns in the Control and Real-World Knowledge Violation sentences did not differ significantly on either length ($t(239)=0.13$, $p = 0.89$) or frequency ($t(239)=0.76$, $p=0.45$), see Table 2. Related Animacy Selection Restriction Violations were created by selecting an inanimate noun that was related to the preceding sentential context (e.g. *drum*). No inanimate critical nouns were repeated. Compared with animate critical nouns in the Control and Related Real-World Knowledge Violation sentences, inanimate critical nouns were, on average, one letter shorter ($ts > 4.82$, $ps < 0.00001$), and more frequent ($ts > 2.74$, $ps < 0.01$), see Table 2.

To create the Unrelated Real-World Knowledge Violation and Unrelated Animacy Selection Restriction Violation sentences, scenarios that were not semantically related to each other were paired up. Unrelated Real-World Knowledge Violation sentences were created by substituting animate critical nouns from the Control sentences (50% of scenarios) or Related Real-World Knowledge Violation sentences (50% of scenarios) of the paired scenario (e.g. "The pianist played his music while the bass was strummed by the gravedigger…"). Unrelated Animacy Selection Restriction Violations were created by substituting the critical noun from the Related Animacy Selection Restriction Violation sentences of the paired scenario (e.g. "The pianist played his music while the bass was strummed by the coffin…").

To confirm that the semantic relatedness between the critical nouns and their preceding content words in the related sentence types was indeed closer than in the unrelated sentence types, we determined their Semantic Similarity Values (SSVs) using LSA (Landauer and Dumais 1997; Landauer et al. 1998; available on the internet at *http://lsa.colorado.edu*). As noted in the Introduction, LSA uses a large training corpus to develop a multidimensional representation in which each word is represented by

a single vector. The Semantic Similarity Value (SSV) between two words (or texts) is computed by finding the cosine of the two vectors representing the words (or texts).[5]

We calculated SSV for each sentence by averaging the SSVs, based on term-by-term pair-wise comparisons, between the critical noun and the content words that preceded it using the tasaALL space corresponding to a 1st year college student reading level, using all 300 factors. Mean SSV values and standard deviations for the five sentence types are shown in Table 2. A 5-way ANOVA revealed a significant effect of Sentence Type, $F(4,476) = 51.51$, $p < 0.0001$. Planned pair-wise comparisons were carried out between the Control sentences and each of the Violation sentences. There was no difference in semantic relatedness between the Control sentences and either the Related Real-World Knowledge Violation sentences or the Related Animacy Selection Restriction Violation sentences. As expected, SSV values in the Control sentences were significantly higher than in the Unrelated Real-World Knowledge Violation sentences ($t(119)=9.759$, $p < 0.0001$) and the Unrelated Animacy Selection Restriction Violation sentences ($t(119)=9.877$, $p<0.0001$). A 2 (Relatedness) x 2 (Violation Type) ANOVA revealed a main effect of Relatedness ($F(1,119)=124.789$, $p < 0.00001$), owing to the Unrelated sentences types having significantly lower SSVs than the Related sentence types. There was no main effect of Violation Type and no Relatedness by Violation Type interaction.

The experimental sentences were then assigned to ten lists such that each scenario appeared twice within each list in two out of the five conditions. An equal number of each possible combination of condition pairs appeared within each list and, across all lists, each scenario appeared in each of the five conditions the same number of times.

To each list, 144 plausible filler sentence were then added so that participants would have an equal likelihood of encountering a plausible or implausible sentence. These filler sentences had the same construction as the experimental sentences but used verbs that did not impose animacy selection

[5] One problem with LSA is that it is insensitive to polysemy (e.g. chair in ‘department chair’ and ‘easy chair’ are treated as instances of the same token). For most words, however, a single meaning dominates usage. Thus, the impact of this confound is relatively minimal in a large analysis, such as in the current experiment in which several hundred stimuli were generated.

restrictions on the critical nouns. In 96 of these fillers, the critical noun was inanimate (e.g. “After the injury his leg was supported by the pillow to reduce swelling.”) and in 48 fillers, the critical noun was animate (e.g. “At the circus the kids were entertained by the clown who was extremely funny.”).

Fillers and experimental sentences were then pesudorandomized in each list. Because each scenario appeared twice in each list, in two different conditions, constraints were imposed during randomization. First, no two sentence types of the same scenario occurred within forty sentences of each other. This was done to minimize potential repetition priming effects. It also reduced the potential for participants being able to remember the plausibility of the first presentation of the scenario and use it to predict the plausibility of the second presentation. Second, for a given scenario, the Control sentence was never presented before the Related Violation sentence. This was done in order to prevent the prior presentation of a congruous critical word from interfering with the processing of a semantically related violated critical word the second time a scenario was presented.

To summarize, in each list, there were 240 experimental sentences (48 sentences in each of the five sentence types) and 144 filler sentences. In total, each list consisted of 192 plausible sentences (96 with animate and 96 with inanimate critical nouns) and 192 implausible sentences (96 with animate and 96 with inanimate critical nouns).

Because at the point of the critical word, passive sentences are ambiguous as to whether the *by*-phrase is Agentive, as intended for our critical manipulation, or Locative (e.g. “… the bass was strummed by the drummer/drum…” can potentially be interpreted as “… the bass was strummed next to the drummer/drum…”), we conducted a rating study. All sentences (experimental and fillers) were presented up to the point of the critical noun to 20 Tufts student volunteers who did not participate in the ERP study. Three periods after the critical nouns were used to indicate that the sentences could continue after this point. Each of the ten lists was presented to two participants. Participants were told that they were seeing ‘beginnings of sentences’ and were asked to give ratings from 1 through 7, with 1 indicating that the sentence described something that would be very unlikely to occur in the real world and 7

indicating that the sentence described something that would be very likely to occur in the real world. Several examples were given but participants were told to go with their first instincts and that there were no right or wrong answers.

As can been seen in Table 2, results of our plausibility rating study clearly indicate that the critical nouns in our stimuli were interpreted as Agentive, rather than Locative. An overall ANOVA indicated a significant main effect of Sentence Type on both subjects ($F(4,76)=655.71$, $p < 0.0001$) and items ($F(4,476)=830.36$, $p < 0.0001$) analyses. Planned pair-wise comparisons indicated that, at the point of the critical noun, each of the Violation sentences was rated as significantly less plausible than the Control sentences (subject analyses, $ts > 28.66$, $ps < 0.0001$; items analyses, $ts > 36.27$, $ps < 0.0001$). A 2x2 ANOVA crossing Relatedness and Violation Type revealed main effects of Relatedness ($F1(1,19)=25.43$, $p < 0.0001$, $F2(1,119)=16.38$, $p < 0.0001$) and Violation Type ($F1(1,19)=65.57$, $p < 0.0001$, $F2(1,119)=101.61$, $p < 0.0001$). The effects were due to the Unrelated Violation sentences being rated as slightly more implausible than the Related Violation sentences, and Animacy Selection Restriction Violation sentences being rated as slightly more implausible than Real-World Knowledge Violation sentences. Additionally, the Relatedness by Violation interaction was significant ($F1(1,19)=7.01$, $p < 0.05$, $F2(1,119)=4.24$, $p < 0.05$). Follow-up pair-wise comparisons indicated a significant difference in plausibility between the Related and Unrelated Animacy Selection Restriction Violation sentences in both the subjects analysis ($t(19)=7.34$, $p < 0.0001$) and the items analysis ($t(119)=9.56$, $p < 0.0001$), while the difference in plausibility between Related and Unrelated Real-World Knowledge Violation sentences was smaller, reaching significance on the subjects analysis ($t(19)=2.44$, $p < 0.05$) but only approaching significance on the items analysis ($t(119)=1.83$, $p=0.07$).

Event-related potentials

ERP recording

Twenty-nine tin electrodes were held in place on the scalp by an elastic cap (Electro-Cap International, Inc., Eaton, OH), see Figure 1 for montage. Electrodes were placed below the left eye and

at the outer canthus of the right eye to monitor vertical and horizontal eye movements, and also over the left mastoid (reference) and right mastoid (recorded actively to monitor for differential mastoid activity). All EEG electrode impedances were maintained below 5 kΩ (impedance for eye electrodes was less than 10 kΩ).

The EEG signal was amplified by an Isolated Bioelectric Amplifier System Model HandW-32/BA (SA Instrumentation Co., San Diego, CA) with a bandpass of 0.01 to 40 Hz and was continuously sampled at 200 Hz by an analogue-to-digital converter. The stimuli and participants' behavioral responses were simultaneously monitored by a digitizing computer.

ERP Procedure

Twenty participants (12 female; mean age 19.75 (2.75)) were recruited by advertisement and were paid to participate. All were right-handed native speakers of English, who had not learned any other language before the age of five, and who had normal or corrected-to-normal vision.

Each participant was given 15 practice trials at the start of the experiment and was then assigned to one of the ten experimental lists (i.e. each list was viewed by two different participants). Participants sat in a comfortable chair in a dimly lit room, separate from the experimenter and computers. Sentences were presented word-by-word on a computer monitor located 47 inches in front of participants. Text was centered and displayed in white on a black background. Text subtended approximately 1° visual angle vertically and 1-3° visual angle horizontally. Each trial (one sentence) began with the presentation of a fixation point at the center of the screen for 450ms, followed by a 100ms blank screen, followed by the first word of the sentence. Each word appeared on the screen for 450ms with an interstimulus interval (ISI) of 100ms separating words. The final word of each sentence appeared with a period. A 750ms blank-screen interval followed the final word in each sentence, followed by a "?". This cue remained on the screen until the participant made his/her response at which point the next trial started. Participants' task was to decide whether or not each sentence made sense by pressing one of two buttons on a response box with their left or right thumb (counterbalanced across participants). They were told that

sentences may not make sense in different ways and that if a sentence seemed at all odd or unlikely, they should indicate that it did not make sense. They were instructed to wait until the "?" cue before responding. This delayed response was designed to reduce any contamination of the ERP waveform by response sensitive components such as the P300 (Donchin & Coles, 1988).

Electrophysiological Recording

Twenty-nine tin electrodes were held in place on the scalp by an elastic cap (Electro-Cap International, Inc., Eaton, OH), see Figure 1. Electrodes were also placed below the left eye and at the outer canthus of the right eye to monitor vertical and horizontal eye movements, and on the left and right mastoids. Impedance was kept below 2.5 kΩ for all scalp and mastoid electrode sites and below 10 kΩ for the two eye channels. The EEG signal was amplified by an Isolated Bioelectric Amplifier System Model HandW-32/BA (SA Instrumentation Co., San Diego, CA) with a bandpass of 0.01 to 40 Hz and was continuously sampled at 200 Hz by an analogue-to-digital converter. The stimuli and behavioral responses were simultaneously monitored with a digitizing computer.

Data Analysis

ERPs were formed by off-line averaging of artifact-free trials, and time-locked to the onset of critical words in each sentence. Each trial was baselined using average voltage between a -50 pre-stimulus to 50ms post stimulus onset. Three time-windows of interest were chosen a priori: the 300-500msec, 500-700msec and 700-900msec. The first time window encompasses the N400. The second and third time windows encompassed the early and late P600. We also carried out analyses of the 100-200msec, and 200-300msec time windows to ensure later effects were not driven by early effects, possibly due to artifact. The modulation of average ERPs within each of these time windows was examined using analyses of variance (ANOVAs) for repeated measures at each of four electrode columns (see Figure 1). The Midline column had five levels of electrode sites along the anterior-posterior distribution (AP Distribution) of the scalp (FPz, Fz, Cz, Pz, Oz), the Medial column had three levels for AP Distribution and two levels for Hemisphere (FC1/FC2, C3/C4, CP1/CP2), the Lateral

column had four levels for AP Distribution and two levels for Hemisphere (F3/ F4, FC5/FC6, CP5/CP6, P3/P4), and the Peripheral column had five levels of AP Distribution and two levels of Hemisphere (FP1/FP2, F7,F8, T3/T4, T5/T6, O1/O2). At each electrode column, we carried out a 3 (Adverbial Phrase) x 2 (Verb) ANOVAs, with AP Distribution and Hemisphere (for the medial, lateral, and peripheral columns) as additional within-subject factors. These were followed up with planned simple effects ANOVAs, contrasting the effects of Adverbial Phrase for each verb type. Pair-wise comparisons were then carried out for each sentence type with one another. Significance was set at alpha equal to 0.05. A Geisser-Greenhouse correction was applied to all repeated measures with more than one degree of freedom, for which original degrees of freedom and corrected probability levels are reported.

Linearly interpolated voltage maps showing the scalp distribution of differences in ERPs elicited by critical nouns were produced using EEGLAB v4.512 for MatLab software.

Results

Participant Responses

We will refer to the agreement between participants' acceptability judgments during the ERP experiment and our prior categorizations of each sentence type as *degree of match*. Overall, participants' judgments matched our prior categorizations 90% of the time (see Table 3 for percent match broken down by sentence type). An ANOVA revealed significant differences in the degree of match across the five sentence types ($F(4,76) = 15.33$, $p < 0.001$). Post-hoc t-tests indicated that the degree of match was less for the Related Real-World Knowledge Violation sentences than all other sentence types (all $ts > 3.06$, $ps < 0.01$), but more for the Unrelated Animacy Selection Restriction Violation sentences than all other sentence types (all $ts > 3.36$, $ps < 0.01$). Degree of match did not differ significantly between the Control, Unrelated Real-World Knowledge Violation and Related Animacy Selection Restriction Violation sentence types.

ERP data

Approximately 5% of the trials were rejected for artifact (Control: 5.0% (3.1); Related Real-World Knowledge Violations: 4.4% (2.9); Unrelated Real-World Knowledge Violations: 5.8% (2.6); Related Animacy Selection Restriction Violations: 5.9% (3.0); Unrelated Animacy Selection Restriction Violations: 5.5% (2.9)). An overall ANOVA indicated there was no significant effect of sentence type on rejection rates, $F(4,76) < 1$, $p > 0.87$. ERP analyses only included trials in which participants' acceptability judgments matched our prior categorizations of the five sentence types.

ERPs on critical nouns

Voltage maps and grand-average ERPs elicited by the critical nouns at selected electrode sites are presented in Figure 2 (Control versus Real-World Knowledge Violations) and Figure 3 (Control versus Animacy Selection Restriction Violations).

Early Time Windows

Within the first 100ms post stimulus onset there were no significant main effects or interactions involving Sentence Type at any electrode column (see Table 4).

Visual inspection of the waveforms indicated a modulation of ERPs prior to the N400 time window (see Figure 2). We therefore carried out additional analysis within the 150-250ms time window, which captured this early effect. Omnibus ANOVAs that included all five sentence types revealed main effects of Sentence Type, but no further interactions involving Sentence Type, at all electrode columns (see Table 4).

Follow-up pair-wise ANOVAs, comparing each Violation sentence type with the Control sentences, revealed a smaller early positivity to critical nouns in the Unrelated Real-World Knowledge Violation sentences than the Control sentences at all electrode columns (Table 5). This was qualified by an interaction with AP Distribution at the midline and medial columns, reflecting an anterior distribution of the effect. Critical words in the Unrelated Animacy Selection Restriction Violation sentences likewise

evoked a small attenuation of this early positivity, though the effect was only significant at the midline electrode column. No other pair-wise contrasts were significant at any electrode column.

Additional 2 (Violation Type) x 2 (Relatedness) ANOVAs revealed a significant main effect of Relatedness as well as significant Violation Type by Relatedness interaction at all electrode columns (Table 6). This interaction was driven by the smaller positivity/larger negativity to the Unrelated than the Related Real-World Knowledge Violations, with no effect of Relatedness on the Animacy Selection Restriction Violations.

The N400 (300-500ms)

Omnibus ANOVAs including all five sentence types revealed main effects of Sentence Type at all electrode columns as well as Sentence Type by AP Distribution interactions at the midline and peripheral columns (see Table 4), reflecting larger N400 effects at posterior than anterior scalp locations (see Figures 2 and 3). There were no differences in the hemisphere distribution of N400 modulation by critical nouns across the five sentence types, as indicated by the lack of Sentence Type by Hemisphere or Sentence Type by Hemisphere by AP Distribution interactions.

Pair-wise ANOVAs contrasting the N400 to critical nouns in Control sentences and Unrelated Real-World Knowledge Violations (Table 5B), Related Animacy Selection Restriction Violations (Table 5C), and Unrelated Animacy Selection Restriction Violations (Table 5D) all revealed significant main effects of Sentence Type at all electrode columns and Sentence Type by AP Distribution interactions at some electrode columns, indicating significant N400 effects to these violations, particularly over cento-parietal sites. However, the comparison between critical nouns in the Control sentences and those in the Related Real-World Knowledge Violation sentences did not show robust N400 modulation, as reflected by the absence of a main effect of Sentence Type at any electrode column, with only the midline column showing a significant Sentence Type by AP Distribution interaction (Table 5A).

2 (Violation Type) x 2 (Relatedness) ANOVAs revealed significant main effects of both Violation Type and Relatedness at most electrode columns and significant Violation Type by Relatedness interactions at all electrode columns (Table 6). No interactions involving Hemisphere and/or AP Distribution reached significance. The Violation Type by Relatedness interactions arose because of a significantly smaller N400 to the Related than the Unrelated Real World Violations ($Fs > 13.34$, $ps < 0.01$), but no difference in the N400 evoked by the Related and Unrelated Animacy Selection Restriction Violations ($Fs < 1$, $ps > 0.84$).

The P600 (700-900ms)

Omnibus ANOVAs showed highly significant main effects of Sentence Type at all electrode columns and significant Sentence Type by AP Distribution interactions at the midline, medial and lateral electrode columns (see Table 4).

Pair-wise ANOVAs comparing critical nouns in Control sentences and those in the Related and Unrelated Real-World Knowledge Violation sentences showed no main effects of Sentence Type or Sentence Type by AP Distribution interactions at any electrode columns (see Table 5). On the other hand, ANOVAs comparing the Control sentences with the Related and Unrelated Animacy Selection Restriction Violation sentences revealed a clear posteriorly-distributed P600 effect, as reflected by main effects of Sentence Type at all columns and Sentence Type by AP Distribution interactions at several electrode columns (see Table 5).

2 (Violation Type) x 2 (Relatedness) ANOVAs confirmed highly significant main effects of Violation Type at all electrode columns (Table 6), as well as interactions between Violation Type and AP Distribution at all columns ($Fs > 5.11$, $ps < 0.05$) except the peripheral column ($F(4,76)=2.16$, $p=0.24$). This was due to a larger posteriorly-distributed P600 to both types of Animacy Selection Restriction Violations than to both types of Real-World Knowledge Violations. There were, however, no significant main effects of Relatedness and no interactions involving Violation Type and Relatedness (Table 6).

ERPs on sentence-final words

Grand-average ERPs elicited by sentence-final words at select midline electrode sites are shown in Figure 4. A negativity starting at approximately 300ms and persisting until 500ms is apparent on sentence-final words following all four types of Violation sentences compared to Control sentences, i.e. an N400 effect. Omnibus ANOVAs within this epoch comparing all five sentence types confirmed highly significant main effects of Sentence Type (all $Fs > 6.92$, $ps < 0.001$) and Sentence Type by AP Distribution interactions (all $Fs > 3.88$, $ps < 0.01$) at all electrode columns. Follow-up simple effects ANOVAs confirmed more negative N400s on sentence-final words in all four Violation sentences than in Control sentences, with significant main effects of Sentence Type (all $Fs > 6.01$, $ps < 0.05$) and significant Sentence Type by AP Distribution interactions at all electrode columns (all $Fs > 6.63$, $ps < 0.05$). 2 x 2 ANOVAs examining the effects of Violation Type and Relatedness on N400 amplitude revealed no significant main effects or interactions between these two variables at any electrode column (all $Fs < 1$, all $ps > 0.4$).

Discussion

We used ERPs to investigate the online use of three types semantic information: 1) semantic relatedness between content words, 2) knowledge about who is likely to take part in familiar real-world events or states, and 3) animacy selection restrictions on verb arguments. We examined how and when these types of information interact in passive English sentences by contrasting plausible post-verbal Agentive arguments with arguments that either violated real-world event/state knowledge expectations or animacy selection restrictions of their preceding verbs. Consistent with our pre-rating plausibility studies, we found that violations of either real-world event/state knowledge or animacy selection restrictions were generally classified as unacceptable, suggesting that the critical nouns were indeed interpreted as (implausible) Agents, rather than (plausible) Locatives (e.g. "…the bass was strummed [by the/next to] the drummer/drum…"). Both types of violations evoked robust N400 effects, relative to

non-violated arguments. However, the two types of violations differed with regards to how they were modulated by semantic relatedness. While the N400 effect to the real-world knowledge violations was almost completely attenuated when the critical noun was semantically related to preceding words in the context, semantic relatedness failed to modulate the N400 evoked by selection restriction violations. Additionally, we found that selection restriction violations, but not real-world knowledge violations, evoked a robust P600 effect, regardless of semantic relatedness. On the sentence-final word, all four types of violations produced an N400 effect, which was not modulated by either violation type or by semantic relatedness. Below we will discuss our findings in greater detail before considering their general implications and open questions.

The N400

Effects of semantic relatedness on violations of real-world event/state knowledge

Our finding that the semantically unrelated violations of real-world knowledge evoked a significant N400 effect (relative to non-violated nouns) is consistent with previous work from our group (Kuperberg et al., 2003; Kuperberg et al., 2006, Kuperberg et al., 2007) as well as others (Camblin et al., 2007; Federmeier & Kutas, 1999; Hagoort et al., 2004). As noted in the Introduction, the N400 effect evoked by semantic violations has sometimes been interpreted as reflecting the *implausibility* of the proposition formed by full semantic-syntactic integration of a critical word into its context. Although this type of explanation can account for the N400 effect evoked by unrelated real-world knowledge violations, it does not easily account for the near-complete attenuation of the N400 effect evoked by *related* real-world knowledge violations. Related real-world knowledge violations were rated as significantly more implausible than the control sentences (a difference of 3.9 on a seven-point scale), while the N400 amplitude difference between these two conditions was almost non-existent. On the other hand, the difference in plausibility between the related and unrelated real-world knowledge violations was very small (0.2 on a seven-point scale), yet the difference in N400 amplitude was

significant and substantial. Because we only used trials that matched our prior classifications of sentence types in the ERP analysis, the difference in plausibility between the control and related real-world knowledge violation sentence types was likely exaggerated. Additionally, the magnitude of the N400 effect produced by sentence-final words was the same following related and unrelated real-world violations.

Rather than being driven by sentence-level plausibility that is assessed once a critical word has been fully semantically-syntactically integrated with its preceding context, the N400 evoked by a word more likely reflects the match between its semantic features and expectations based on the interaction between its context and semantic information stored at various grains of representation within semantic memory (Kuperberg, 2007; Kutas & Federmeier, 2011). As discussed in the Introduction, some attenuation of the N400 to real-world knowledge violations by semantic relatedness has been reported before (Camblin et al., 2007; Ditman et al., 2007; Federmeier & Kutas, 1999), and this attenuation can sometimes be complete (Kolk et al., 2003; Sanford et al., 2011) or near-complete, as in the present study—termed a neural semantic illusion. Federmeier and Kutas (1999) suggested that semantic context activates not only the expected critical word but also specific semantic features related to it. Kolk and colleagues (Kolk, 2003; van Herten et al., 2006) suggested that the attenuation of the N400 reflected a plausible heuristically-determined meaning based on semantic relatedness between the critical verb and its preceding arguments, a proposal similar to that of Kim and Osterhout (2005) who discuss the notion of 'semantic attraction'—the likelihood of particular arguments playing an alternative thematic role around a verb.

Neither of these accounts, however, can fully explain the N400 attenuation to the related real-world knowledge violations in the present study. In many of our related real-world knowledge violation sentences, the Agent shared few semantic features with the expected critical word and could not plausibly take an alternative thematic role. For example, in "The wreckage of the sunken ship was salvaged by the victims…," *victims* shares few semantic features with the expected critical word, *divers*.

Victims are also unlikely to either *salvage* or *be salvaged*. We therefore suggest that semantic relationships between individual words in the context activated general schemas or scripts (Sanford, 1998; Sanford et al., 2011; Schank & Abelson, 1977). We propose that such schema activation overrode more specific predictions of semantic features of a likely animate Agent based on real-world knowledge about events or states, thus leading to an attenuation of the N400 effect evoked by related real-world violations (see Sanford et al., 2011 for a similar interpretation).

There has been much debate about how quickly semantic memory-based expectations for a given word can influence processing. There is evidence that in some situations, predictions about the properties of a critical word can impact processing very quickly—sometimes before that word is encountered (see DeLong, Urbach, & Kutas, 2005; Federmeier, 2007; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005) or during the earliest stages of its perceptual processing (e.g. Dikker et al., 2009; Federmeier, Mai, and Kutas, 2005). In the present study, we found evidence for such an early influence of context: the waveform evoked by the unrelated real-world violations started to diverge (becoming less positive/more negative) from that evoked by critical words in the control and the related real-world violation sentences at around 150ms. Speculatively, this may reflect an early detection of a mismatch between schema-based predictions and the semantic features of the incoming critical word. It is unclear whether this early divergence should be conceived as the early part of the N400 itself or whether it reflected a separate component. In favor of the latter idea, the early effect was more anteriorly distributed than the N400 effect and, as discussed below, did not pattern with N400 modulation to the animacy violations.

Finally, it is important to note that, although a near-complete attenuation of the N400 to related real-world knowledge violations can, in some cases be associated, with behavioral ‘semantic illusions’ (Sanford et al., 2011), in the present study, the neural illusion was only a temporary phenomenon during processing (see Footnote 1). By the end of the sentences, participants had registered the implausibility of the related real-world knowledge violations, as indicated by both their acceptability judgments, and a

clear N400 effect on the sentence-final words, which did not differ significantly from that evoked on sentence-final words in any of the other three violation conditions.

Effects of semantic relatedness on violations of animacy selection restrictions

Inanimate nouns that violated the selection restrictions of their preceding verb evoked a robust N400 effect, consistent with numerous previous studies (Ainsworth-Darnell, Shulman, & Boland, 1998; Bornkessel-Schlesewsky et al., 2010; Friederici, Pfeifer, & Hahne, 1993; Garnsey et al., 1989; Li, Shu, Liu, & Li, 2006; Nieuwland & Van Berkum, 2005; Rösler et al., 1993). Unlike violations of real-world knowledge, however, the N400 effect to these violations was not modulated by semantic relatedness between the critical noun and the preceding content words.

The failure of semantic relatedness to impact N400 amplitude to the animacy selection restriction violations in the current study cannot be attributed to our relatedness manipulation being ineffective. As noted in the Methods, the difference in SSVs between the related and unrelated selection restriction violations was highly significant and the same as for the difference in SSVs between the related and unrelated real-world knowledge violations which, as discussed above, led to marked N400 modulation.

We suggest that the difference in sensitivity to semantic relatedness in the real-world event/state knowledge violations and the animacy selection restriction violations relates to how the two types of violations are treated initially by the parser. In the case of the real-world event/state knowledge violations, the verb's argument matched the animacy selection restrictions imposed by the preceding verb: while neither a *drummer* nor *gravedigger* are likely to *strum* a *bass*, both are animate and therefore both meet broad semantic category requirements of the verb *strum*. Having matched such broad animacy requirements, more fine-grained semantic matching, on the basis of real-world event/state knowledge and schema-based semantic relatedness between content words, could continue. In contrast, selection restriction violations on inanimate arguments failed to match coarse-grained animacy constraints of the verb. We propose that, having failed to match at this coarser level, the parser did not pursue more fine-grained semantic feature matching, thus preventing the N400 from being attenuated. Consistent with the

idea that coarse-grained semantic feature matching occurs quite early, we found a difference in how semantic-relatedness impacted processing within the 150-250ms time window. As noted above, we found that this early waveform was modulated by semantic relatedness for real-world knowledge violations. On the other hand, no such modulation was observed for selection restriction violations.

The current findings therefore suggest that violations of animacy selection restrictions are treated as functionally distinct from violations of real-world event/state knowledge during online sentence processing. On the one hand, both types of violations impact online processing within the same N400 time frame. However, the processing of real-world knowledge violations, but not animacy selection restriction violations, is affected by semantic relatedness. This functional distinction between the two types of violations is consistent with previous eye-tracking results by Warren and McConnell (2007).

It is important to note that our interpretation does not contradict previous proposals arguing that verbs encode additional event-specific information which can be used to facilitate verb argument processing, as shown by McRae, Ferretti, and Amyote (1997), McRae, Ferretti, and Hatherell (2001), Matsuki et al. (2011), and others. Indeed, it is likely that both coarse-grained as well as more fine-grained semantic information associated with a verb can be utilized by the parser to facilitate online processing. What our data rather suggest is that broad animacy selection restrictions on verb arguments may have a relatively privileged status during verb-argument processing and that after encountering mismatches at this level, the parser may not pursue more fine-grained matching between the context and finer-grained semantic features of an incoming word.

Our current findings are partially consistent with our earlier findings for animacy selection restriction violations on verbs (Kuperberg et al., 2007). Recall that we found that there was no difference in the N400 evoked by *eat* versus *plant* in sentences like “For breakfast the eggs would only *eat/*plant…” despite the former but not the latter verb being related to preceding content words in the context. Together, these findings suggest that the processing of animacy selection restriction violations is impervious to the effects of semantic relatedness. However, while in our earlier study, the amplitude

of the N400 evoked by selection restriction violations was similar to that evoked by verbs in control sentences (no N400 effect), in the current study, the N400 evoked by animacy selection restriction violations was significantly larger (an N400 effect).

What could account for this difference in N400 modulation between our earlier and current experiment? Results from another study we recently conducted shed at least some light on this question (Paczynski & Kuperberg, 2011). Similar to Kuperberg et al. (2007), we examined the effects of violating selection restrictions in active English sentences. Similar to the current study, our critical manipulation was on the post-verbal argument. Rather than manipulating semantic relatedness, however, we manipulated the animacy of this post-verbal argument. We found that selection restrictions falling on inanimate arguments (e.g. "...penalized the *meadow…) evoked an N400 effect, while those falling on animate arguments (e.g. "…plowed the *laborer…) did not. We argued that the difference between the two conditions was based on the degree of match between verb-independent animacy expectations (for inanimate arguments to appear later in the sentence). These expectations were met in the former case but not in the latter. Similarly, in Kuperberg et al. (2007), in which selection restrictions became apparent on the verb, the subject argument was inanimate, thus violating both verb-dependent animacy expectations (selection restrictions) and verb-independent animacy expectations (the animate first principle) and no N400 effect was observed. In the current experiment, inanimate critical nouns violated verb-dependent animacy expectations but not verb-independent animacy expectations and an N400 was observed.

Taken together, these two studies suggest a complex interplay between differ types of semantic knowledge. On the one hand, violations of both verb-dependent and verb-independent animacy expecations results appears to result in the language parser not pursing further semantic processing within the N400 time-window. On the other hand, violating only verb-dependent animacy constraints (i.e. selection restrictions) appears to results in further processing within the N400 time-window that is not modulated by semantic relatedness, suggesting that the parser does not pursue more fine-grained semantic feature matching.

The P600: Effects of severe implausibility

As outlined in the Introduction, there has been some debate about what exactly triggers a P600 to semantic violations. The initial set of studies describing this effect (Kuperberg et al., 2003; Kolk, 2003; Hoeks, Stowe, & Doedens, 2004; Kim & Osterhout, 2005) highlighted the fact that the P600 effect was present when the N400 effect was absent, and the two components were often interpreted as functionally related to one other. For example, Kim and Osterhout's (2005) 'semantic attraction' hypothesis suggests that, when a selection restriction violating verb argument can plausibly occupy an alterative thematic role about the verb, the N400 is attenuated and instead a P600 is triggered. More recently, Hagoort, Baggio and Willems (2009) proposed a more general version of this theory, suggesting that linguistic errors trigger an N400 when syntactic cues are strong but semantic cues are weak, while a P600 is triggered if semantic cues are strong but syntactic cues are weak. Within this model, violations involving 'semantic attraction' between a verb and its argument(s) constitute a subset of circumstance under which a P600 is evoked.

The present findings are inconsistent with these semantic attraction/semantic relatedness types of accounts. We found that selection restriction violations evoked a P600 effect regardless of whether the critical noun was semantically related or unrelated to the preceding verb or other words in the context. This is in line with several previous studies that have also reported clear P600 effects to unrelated selection restriction violations (e.g. Kuperberg et al. 2007; Stroud & Philips, In Press, Kuperberg et al. 2010, but see Kim & Osterhout, 2005, Experiment 2). Indeed, in a post-hoc analysis (data not reported here), we found that the P600 evoked by related selection restriction violating critical nouns was not modulated by whether or not the critical noun could be a plausible Theme for the preceding verb (see also Kuperberg et al., 2006, and Paczynski & Kuperberg, 2011).

Our results are more consistent with frameworks that emphasize the overall *implausibility/impossibility* of the proposition, resulting from syntactically and semantically integrating a critical word with its preceding context, as being a trigger for the P600 effect (Kuperberg, 2007; van

de Meerendonk, 2009; Bornkessel-Schlesewsky & Schlesewsky, 2008). According to all these accounts, a P600 effect, reflecting continued analysis or reanalysis, is most likely to be evoked by selection restriction violations, as opposed to real-world knowledge violations, because the former result in highly implausible/impossible propositions, but the latter result in implausible-but-possible propositions. Consistent with this idea, eye-movement studies have shown that violations resulting in implausible-impossible, but not implausible-possible, propositions are associated with downstream effects (Rayner, Warren, Juhasz, and Liversedge, 2004; Warren and McConnell, 2007).

One might wonder whether the observed effects on the P600 were related to the structural ambiguity of our materials, with inanimate critical nouns more likely to be interpreted as Locations rather than Agents. This seems unlikely. First, as noted in the Methods, our norming study indicated that sentences with inanimate critical nouns were rated as more implausible than those with animate critical nouns. If participants had interpreted the inanimate nouns as Locations, this would predict an opposite pattern, with animacy violations being interpreted as at least semi-plausible locations. Second, Locative *by* phrases are relational (an action occurs in proximity of a contextually relevant landmark). Since animate nouns can serve as equally plausible Locations in this context, it is unclear why a P600 effect would only be evoked by inanimate rather than animate critical nouns. However, even if we assume that continued analysis/reanalysis is more likely to occur for inanimate than animate nouns, this does not alter our interpretation of the P600. As noted in the Introduction, the syntactic P600 is not only evoked by syntactic anomalies, but also by syntactically ambiguous structures when the initial syntactic analysis yields an 'impossible' initial interpretation (e.g. "The banker persuaded to sell…" (Holcomb & Osterhout, 1992). In such cases, it may index successful reanalysis processes. Similarly, in the present study, we suggest that the semantic P600 is triggered by an *initial* interpretation of a word resulting in a highly implausible proposition, and that it reflects continued combinatorial analysis or reanalysis in an *attempt* to reassign thematic roles and recover a coherent discourse meaning, regardless of whether this attempt is successful (for discussion, see Kuperberg et al. 2006).

Relevant to this interpretation, the P600 evoked by nouns in the present study began at around 600ms and lasted for several hundred milliseconds. This is similar to the latency of the P600 seen on semantically violating noun-phrase arguments in our previous studies (e.g. Paczynski & Kuperberg, 2011; Kuperberg, Choi, Cohn, Paczynski & Jackendoff, 2010), and is around 100ms later than the P600 evoked by selection restriction violations on verbs (e.g. Kuperberg et al., 2003, 2006, 2007; Kim & Osterhout, 2005, Hoeks et al. 2004)[6]. We have speculated that differences in the timing of the semantic P600 on nouns and verbs reflect the likelihood that a semantic violation is recoverable. On nouns, comprehenders may delay reanalyzing in case a subsequent word disambiguates a highly implausible interpretation (see Paczynski & Kuperberg, 2011 for discussion), whereas selection restriction violations falling on verbs will hardly ever be recoverable as this is where thematic roles are (usually) unambiguously assigned.

Open questions

In the present study, we studied interactions between three types of semantic knowledge during sentence comprehension: semantic relatedness between individual words, real-world event/state knowledge, and selection restrictions of verbs on their arguments. It is important to recognize that each of these types of stored knowledge is divisible into multiple subcomponents. We operationalized 'semantic relatedness' between content words quite broadly to encompass different types of semantic relationships based on close association (e.g. black-white; Kiss, 1973; Nelson, 1998), category membership (e.g. carnation-orchid), semantic features (e.g. ball-planet), as well as co-occurrence due to shared situational or schema membership, whether concrete (e.g. victim-jail) or abstract (e.g. treason-justice). Although all of these modulate the N400 similarly in isolation, it is an open question whether

[6] Indeed, the P600 produced by verbs that violate the selection restrictions of their preceding argument(s) can sometimes begin within the N400 time window. This can lead to difficulties in interpretation. The N400 and P600 both have a posterior scalp distribution and have opposite polarities. Thus, when the P600 starts within the N400 time window, it can mask the appearance of an N400 on the surface of the scalp. Because of this, it has sometimes been unclear whether the absence of an N400 effect to a given manipulation is an artifact of this component overlap, or whether it reflects a true absence of neural modulation within this time window (see Kuperberg et al. 2007 for discussion).

each interacts similarly with representations of both real-world event/state knowledge and/or animacy selection restrictions during sentence processing.

We operationalized real-world knowledge as capturing what we know about the likelihood of people taking part in events or states. Future studies should explore how different types of real-world knowledge affect online processing, distinguishing not only between events and states, but also between knowledge we have about objects and locations.

The current study examined only animacy-based selection restrictions. An important question for future studies will be to determine whether a relative impermeability to semantic relatedness is specific to selection restrictions on argument animacy, or whether it generalizes to other types of restriction. For example, would a similar pattern of results be observed for selection restrictions based on argument concreteness (e.g. The pirates buried the treasure/*mutiny), when animacy is held constant? What about more finer-grained semantic features (e.g. The man drank the *sandwich)? These questions have theoretical implications. Early versions of generative grammar proposed a mental lexicon with verb-argument selection restriction constraints separate and independent from real-world knowledge (Chomsky, 1965; Katz, 1963). Although more recently it has been argued that the selection restrictions and real-world knowledge associated with a given verb are difficult to disentangle (e.g. Jackendoff, 2002; Matsuki, Chow, Hare, Elman, Scheepers, & McRae, 2011), there is evidence that animacy may be a privileged semantic feature. In some languages, animacy is formally encoded within syntactic structure (Craig, 1977; Hale, 1972; Minkoff, 2000; Van Valin, 1997) as well as word morphology (Aristar, 1997; Malchukov, 2008; Wiese, 2003). Even in languages where it does not formally constrain syntactic structure, such as English, animacy information can influence noun ordering (Rosenbach, 2008; Snider & Zaenen 2006). Thus it will be important for future studies to determine whether animacy is a privileged, and perhaps unique, semantic feature when it comes to the processing of verb selection restrictions.

The third set of questions left unresolved by the current study is under what circumstances semantic violations elicit a P600 effect. As noted in the Introduction, our study was only designed to address the question of whether or not semantic relatedness between a critical word, on which a semantic violation occurred, and words in the preceding sentential context was necessary in order to evoke a P600 effect. As noted above, our study adds to a growing literature indicating that semantic relatedness is not a necessary criterion for a P600 to be evoked, at least in the case of animacy selection restriction violations (e.g. Kuperberg et al, 2007; Kuperberg et al., 2010; Paczynski & Kuperberg, 2011; Stroud & Philips, In Press), and it favors accounts that emphasize the role of propositional implausibility as an important factor for the production of a semantic P600 effect (Kuperberg, 2007; van de Meerendonk et al., 2009; Bornkessel-Schlesewsky & Schlesewsky, 2008). The present study, however, was not designed to distinguish between these three accounts, which do differ, particularly with respect to the emphasis they place on conflict between alternative representations. Thus, future studies will need to explore the more nuanced difference in predictions made by these three models.

Finally, there is the question of what our findings mean for more naturalistic language comprehension. As is common in many psycholinguistic studies, we explored how the language system behaves when pushed against limits. For example, syntactic processing of complex argument structures has been explored by examining object relative clauses in which both arguments are animate (e.g. The lawyer that the banker irritated…; cf. Traxler, 2002), despite such constructions rarely, if ever, appearing in natural discourse. In the current study, we used visual word-by-word presentation along with a plausibility task in order to determine, amongst other inquiries, whether animacy selection restriction violations are processed similarly to violations of real-world knowledge. Our results clearly demonstrate that the two types of violations engender different patterns of processing, at least within the experimental paradigm used in the current study. Thus our results suggest that that animacy selection restrictions are not simply a “more severe” type of real-world knowledge violation, but rather that the two types of knowledge can be treated distinctly during processing. However, it remains an open

question whether, and to what extent, such functional distinctness impacts more natural language comprehension, both auditory comprehension and more passive reading, which varies in its demands, depending on the comprehender's attention and motivation.

Conclusions

In the present study, we have drawn a distinction between semantic memory-based processes, which modulate the N400, and the implausibility/impossibility of the proposition formed by semantic-syntactic integration, which modulates the P600. Semantic memory-based mechanisms refer to the generation of expectations through the interaction between context (prior to the critical word) and stored semantic relationships of various types about the semantic features of an incoming critical word. We showed that strong semantic relatedness between content words can override expectations based on real-world event/state knowledge, but not necessarily a verb's selection restrictions on argument animacy. Combinatorial mechanisms refer to the integration of a critical word with its preceding context using both syntactic and semantic constraints to produce a propositional interpretation. We showed that when the resulting proposition is implausible-impossible, but not implausible-possible, it will lead to additional analysis/reanalysis, reflected by a P600 effect, at least when readers explicitly judge plausibility. Finally, we demonstrated that this continued combinatorial analysis is not necessarily modulated by semantic relatedness, arguing against the idea that the semantic P600 is triggered purely by semantic relatedness or attraction. Taken together, our findings suggest a complex interplay between different types of semantic information that can influence early and later stages of online word-by-word sentence comprehension.

References

Ainsworth-Darnell, K., Shulman, H. G., & Boland, J. E. (1998). Dissociating brain responses to syntactic and semantic anomalies: evidence from event-related potentials. *Journal of Memory and Language, 130*, 112-130.

Aissen, J. (2003). Differential object marking: iconicity vs. economy. *Natural Language and Linguistic Theory, 21*, 435 – 483.

Aristar, A. R. (1997). Marking and hierarchy types, and the grammaticalization of case-markers. *Studies in Language, 21*, 313 – 368.

Barton, S. B., & Sanford, A. J. (1993). A case-study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition, 21*, 477–487.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology, 60*, 343–355.

Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language, 63*, 489-505.

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., Schlesewsky, M. (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language, 117*(3), 133-152.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2006). The extended Argument Dependency Model: a neurocognitive approach to sentence comprehension across languages. *Psychological Review, 113*, 787-821.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews, 59*, 55-73.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977-990.

Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language, 56*, 103-128.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*: MIT Press.

Chwilla, D. J., Kolk, H.H.J., Mulder, G. (2000). Mediated priming in the lexical decision task: evidence from event-related potentials and reaction time. *Journal of Memory and Language, 42*, 314-341.

Chwilla, D. J., & Kolk, H. H. (2005). Accessing world knowledge: Evidence from N400 and reaction time priming. Cognitive Brain Research, *25(3)*, 589-606.

Coulson, S., Federmeier, K. D., & Kutas, M. (2005). Right hemisphere sensitivity to word- and sentence-level context: evidence from event-related brain potentials. *Cognition, 31*, 129 -147.

Craig, C. (1977). *The Structure of Jaceltec*. Austin/London: Texas Press.

Deacon, D., Grose-Fifer, J., Yang, C.-M., Stanick, V., Hewitt, S., & Dynowska, A. (2004). Evidence for a new conceptualization of semantic representation in the left and right cerebral hemispheres. *Cortex, 40*(3), 467-478.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117-1121.

Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2007). The contributions of lexico-semantic and discourse information to the resolution of ambiguous categorical anaphors. *Language and Cognitive Processes, 22*, 793-827.

Dikker, S., Rabagliati, H., & Pylkkänen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition, 110(3)*, 293-321.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Science, 11*, 355-372.

Duffy, S., Henderson, J. M., & Morris, R. K. (1989). Semantic facilitation of lexical access during sentence processing. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 15*, 791-801.

Erickson, T. D., & Matteson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Behaviour, 20*, 540–551.

Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology, 44*(4), 491-505.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: long-term memory structure and sentence processing,. *Journal of Memory and Language, 41*, 469-495.

Federmeier, K.D., Mai, H., & Kutas, M. (2005). Both sides get the point: Hemispheric sensitivities to sentential constraint. *Memory and Cognition, 33(5),* 871-886.

Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *Cognition, 33*, 182-196.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language, 44*, 516- 547.

Filik, R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: evidence from the N400. *Psychophysiology, 45*, 554-558.

Foss, D. J. (1982). A discourse on semantic priming. *Cognitive Psychology, 14*, 590-607.

Friederici, A. D., & Frisch, S. (2000). Verb argument structure processing: the role of verb-specific and argument-specific information,. *Journal of Memory and Language, 43*, 476-507.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1*, 183-192.

Garnsey, S. M., Tanenhaus, M. K., & Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research, 18,* 51-60.

Grose-Fifer, J., & Deacon, D. (2004). Priming by natural category membership in the left and right cerebral hemispheres. *Neuropsychologia, 42*, 1948-1960.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. In: Garnsey, S.M. (Ed.), Language and Cognitive Processes. Special Issue: Event-Related Brain Potentials in the Study of Language, vol. 8(4). Lawrence Erlbaum Associates, Hove, pp. 439–483.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Science, 9*(9), 416-423.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic Unification. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (4 ed., Vol. 7, pp. 819-836). Cambridge, MA: MIT Press.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*, 438-441.

Hale, K. (1972). A note on subject-object inversion in Navajo. In B. Kachru, R. Lees, Y. Malkiel, A. Pietrangeli & S. Saporta (Eds.), *Issues in Linguistics: Papers in Honor of Henry and Renée Kahane* (pp. 300–309). Urbana: University of Illinois Press.

Hoeks, J. C. J., Stowe, L., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive Brain Research, 19*, 59-73.

Jackendoff, R. (2002). Foundations of Language: Brain, Meaning, Grammar, Evolution. Oxford: Oxford University Press.

Katz, J. J. & Fodor, J.A. (1963). The structure of a semantic theory. *Language, 39*, 170-210.

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

Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In R. W. B. A. J. Aitkin, & N. Hamilton-Smith (Ed.), *The Computer and Literary Studies* (pp. 153-165). Edinburgh: Edinburgh University Press.

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

Kreher, D. A., Holcomb, P. J., & Kuperberg, G. R. (2006). An electrophysiological investigation of indirect semantic priming. *Psychophysiology, 43,* 550-563.

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

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

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

Kuperberg, GR., Choi, A., Cohn, N., Paczynski M., & Jackendoff, R. (2010) Electrophysiological correlates of Complement Coercion. *Journal of Cognitive Neuroscience, 22(12),* 2685-2701.

Kuperberg, G. R., Paczynski, M., & Ditman, T. (2010). Establishing causal coherence across sentences: an ERP study. *Journal of Cognitive Neuroscience, 23*(5), 1230-1246.

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

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectation and semantic association. *Nature, 307* (5947), 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.

Laham, D. (1997). Latent Semantic Analysis approaches to categorization. In M. G. Shafto & P. Langley (Eds.), Proceedings of the 19th annual meeting of the Cognitive Science Society (p. 979). Mawhwah, NJ: Erlbaum.

Landauer, T. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes, 25*, 259-284.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.

Landauer, T.K., McNamara, D.S., Dennis, S., & Kintsch, W. (2007). Handbook of Latent Semantic Analysis, Lawrence Erlbaum Associates.

Lau, E.F., Holcomb, P.J., & Kuperberg, G.R. Dissociating N400 effects of prediction from association in single word contexts. Under review.

Li, X., Shu, H., Liu, Y., & Li, P. (2006). Mental representation of verb meaning: behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience, 18*, 1774-1787.

Malchukov, A. L. (2008). Animacy and asymmetries in differential case marking. *Lingua, 118*, 203 – 221.

Marslen-Wilson, W., Brown, C., & Tyler, L. K. (1988). Lexical representations in spoken language comprehension. *Language and Cognitive Processes, 3*, 1-16.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1-2), 71-102.

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*, 913-934.

McRae, K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language, 38*, 283-312.

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes, 12*, 137-176.

Minkoff, S. (2000). Animacy hierarchies and sentence processing. In A. Carnie, Guilfoyle, E. (Ed.), The Syntax of Verb Initial Languages (pp. 201–212). Oxford: Oxford University Press.

Morris, R. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 92-103.

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes, 26*, 131–157.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms., from http://www.usf.edu/FreeAssociation/.

Nieuwland, M.S., & Kuperberg, G.R. (2008). When the truth isn't too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science, 19*, 1213-1218.

Nieuwland, M. S., & Van Berkum, J. J.A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research, 24*, 691-701.

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

Otten, M., Nieuwland, M. S., & Van Berkum, J. J. A. (2007). Great expectations: specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience, 8*, 89.

Paczynski, M., Kuperberg, G.R. (2011). Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb argument processing. *Language and Cognitive Processes*. Special Issue: the Cognitive Neuroscience of Semantic Processing. *26(9),* 1402-1456

Rayner, K., Warren, T., Juhasz, B., & Liversedge, S. (2004). The effects of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*, 1290-1301.

Rosenbach, A. (2008). Animacy and grammatical variation—Findings from English genitive variation. *Lingua, 118*, 151-171.

Rösler, F., Pütz, P., Friederici, A. D., & Hahne, A. (1993). Event-related brain potentials while encountering semantic and syntactic constraint violations. *Journal of Cognitive Neuroscience, 5*, 345-362.

Rugg, M., D. (1984). Event-related potentials and the phonological processing of words and non-words. *Neuropsychologia, 22,* 435-443.

Sanford, A. J., & Garrod, S. C. (1998). The role of scenario mapping in text comprehension. *Discourse Processes, 26*, 159–190.

Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2011). Anomalies at the borderline of awareness: an ERP study. *Journal of Cognitive Neuroscience, 23*, 514-523.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures*. Hillsdale, NJ: L. Erlbaum.

Silva-Pereyra, J., Harmony, T, Villanueva, G, Fernandez, T, Rodriguez, M, Galan, L, Diaz-Comas, L, Bernal, J, Fernandez-Bouzas, A, Marosi, E, Reyes, A. (1999). N400 and lexical decisions: automatic or controlled processing? *Clinical Neurophysiology, 110*, 813–824.

Snider, N., Zaenen, A. (2006). Animacy and Syntactic Structure: Fronted NPs in English. In M. Butt, M. Dalrymple, and T.H. King, eds. Intelligent Linguistic Architectures: Variations on Themes by Ronald M. Kaplan. CSLI Publications: Stanford.

Stroud, C., & Phillips, C. (In Press). Examining the evidence for an independent semantic analyzer: An ERP study in Spanish. *Brain and Language*.

Tyler, L. K. (1992). Verb-argument structures. In L. K. Tyler (Ed.), *Spoken Language Comprehension: An Experimental Approach to Disordered and Normal Processing* (pp. 107-124). London: MIT Press.

Van Berkum, J. J.A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: evidence from the N400. *Journal of Cognitive Neuroscience, 11*, 657-671.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology. Learning, Memory and Cognition, 31*(3), 443-467.

van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010). Monitoring in language perception: mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience, 22*, 67-82.

van de Meerendonk, N., Kolk, H., Chwilla, D.J. & Vissers, C. T.W.M. (2009). Monitoring in language perception. *Language and Linguistics Compass, 3*, 1211-1224.

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

Van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. *Language and Cognitive Processes, 8*, 485-531.

Van Petten, C., & Luka, B. (In press). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*.

Van Petten, C., Weckerly, J., McIsaac, H. K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science, 8*, 238–242.

Van Valin, R. L., R. (1997). *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press.

Wang, S., Ditman, T., Choi, A., Kuperberg, G.R. (2010). The effects of task on processing real-world, animacy and syntactically violated sentences. Annual Meeting of the Cognitive Neuroscience Society.

Warren, T., & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review*, *14*, 770-775.

Wiese, H. (2003). Semantics as a gateway to language. In H. Härtl, Tappe, H. (Ed.), *Mediating Between Concepts and Language* (pp. 197–222). Berlin/New York: Mouton de Gruyter.

Figure Captions

Figure 1. Electrode montage. Analyses of variance were conducted at midline, medial, lateral and peripheral electrode columns shown (see Methods).

Figure 2. ERPs evoked by critical nouns in Control, semantically Related and Unrelated Real-World Knowledge violation sentences along midline electrode sites, as well as voltage maps in the 150-250ms, N400 (300-500ms) and P600 (700-900ms) time windows. Solid black lines and white bars indicate Control condition; dotted green line and bar indicate Related Real-World Knowledge Violation condition; dashed blue line and bar indicate Unrelated Real-World Knowledge Violation condition. The plots are shown using a -100-0ms pre-stimulus baseline. All voltage maps show differences between ERPs to the violations and control critical words, averaged across each time window. Bar graphs show the amplitude of ERPs to each condition averaged across each time window, across the four electrode sites where the effects were maximal (indicated below each bar graph). Error bars show standard errors.

Figure 3. ERPs evoked by critical nouns in Control, semantically Related and Unrelated Animacy Selection Restriction Violation sentences along midline electrode sites, as well as voltage maps in the 150-250ms, N400 (300-500ms) and P600 (700-900ms) time windows. Solid black lines and white bars indicate Control condition; dotted green line and bar indicate Related Animacy Selection Restriction Violation condition; dashed blue line and bar indicate Unrelated Animacy Selection Restriction Violation condition. The plots are shown using a -100-0ms pre-stimulus baseline. All voltage maps show differences between ERPs to the violations and control critical words, averaged across each time window. Bar graphs show the amplitude of ERPs to each condition averaged across each time window, across the four electrode sites where the effects were maximal (indicated below each bar graph). Error bars show standard errors.

Figure 4. ERPs evoked by sentence-final words in all conditions along midline electrode sites. Solid black lines and white bars indicate control Condition, dotted green lines and bar indicate Related Violation conditions, dashed blue lines indicate Unrelated Violation conditions. The plots are shown using a -100-0ms pre-stimulus baseline. Bar graphs show the amplitude of ERPs to each condition averaged across the 300-500ms time window and across Cz, Pz, CP1 and CP2 where the effects were maximal. Error bars show standard errors.

Table 1. Types of linguistic violations and example sentences.

Sentence Type	Example
1. Control The critical animate noun (e.g. *guitarist*) is semantically related to the group of content words in the preceding context (*pianist, played, music, bass, strummed*), *and* it conforms to expectations based on real-world knowledge about how likely it is for the Agent to be carrying out this action in this particular context.	The pianist played his music while the bass was strummed by the guitarist during the song.
2. Related Real-World Knowledge Violations The critical animate noun (e.g. *drummer*) is semantically related to the group of content words in its preceding context (*pianist, played, music, bass, strummed*), *but* it violates expectations based on real-world knowledge about how likely it is for the Agent to be carrying out this action in this particular context (a bass is unlikely to be strummed by a drummer in this situation). Note that this event is implausible but not impossible. Note also that the Agent is animate and therefore matches the animacy restrictions of the verb.	The pianist played his music while the bass was strummed by the drummer during the song.
3. Unrelated Real-World Knowledge Violations The critical animate NP (e.g. *gravedigger*) is not related to the group of content words in its preceding context (*pianist, played, music, bass, strummed*) *and* it violates expectations based on real-world knowledge about how likely it is for the Agent to be carrying out this action in this particular context (a bass is unlikely to be strummed by a gravedigger in this situation). Note that this event is implausible but not impossible. Note also that the Agent is animate and therefore matches the animacy restrictions of the verb.	The pianist played his music while the bass was strummed by the gravedigger during the song.
4. Related Animacy Selection Restriction Violations The critical inanimate noun (e.g. *drum*) is semantically related to the group of content words in its preceding context (*pianist, played, music, bass, strummed*), *but* it violates the animacy-based selection restrictions of the verb for an animate Agent (drums are inanimate and therefore cannot carry out the action of strumming). Note that this event is impossible, rather than simply implausible.	The pianist played his music while the bass was strummed by the drum during the song.
5. Unrelated Animacy Selection Restriction Violations The critical inanimate noun (e.g. *coffin*) is not semantically related to the group of content words in its preceding context (*pianist, played, music, bass, strummed*) *and* it also violates the animacy-based selection restrictions of the verb for an animate Agent (coffins are inanimate and therefore cannot carry out the action of strumming). Note that this event is impossible, rather than simply implausible.	The pianist played his music while the bass was strummed by the coffin during the song.

Table 2.

Sentence Type	CN Length		CN Frequency		SSV		Plausibility ratings*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
1. Control	7.54	2.08	2.45	0.89	0.22	0.02	6.3	0.3
2. Related Real-World Knowledge Violation	7.58	1.92	2.37	0.80	0.18	0.02	2.4	0.4
3. Unrelated Real-World Knowledge Violation	7.56	1.98	2.31	0.81	0.00	0.01	2.2	0.5
4. Related Selection Restriction Violation	6.23	1.98	2.74	0.74	0.18	0.02	1.8	0.5
5. Unrelated Selection Restriction Violation	6.23	1.98	2.74	0.74	0.01	0.01	1.3	0.3

CN: Critical noun
SD: Standard Deviation.
Length: number of letters.
Frequency based on the SUBTLEXus Corpus' log of word form frequency per million, $LgSUBTL_{WF}$ (Brysbaert and New 2009); available on the Internet through the English Lexicon Project http://elexicon.wustl.edu/
SSV: Semantic Similarity Values, as determined using a Latent Semantic Analysis (LSA; Landauer and Dumais 1997; Landauer et al. 1998; available on the internet at http://lsa.colorado.edu) between CN and preceding sentence context.
*Plausibility ratings on a 7-point Likert scale. Plausibililty ratings of the fillers: Mean: 6.2; SD: 0.4.

Table 3. Degree of match between participants' acceptability judgments during the ERP experiment and our prior categorizations of each sentence type.

Sentence Type	Degree of Match
Control	89% (5.7)
Related Real-World Knowledge Violations	79% (11.4)
Unrelated Real-World Knowledge Violations	89% (7.8)
Related Animacy Selection Restriction Violations	93% (4.9)
Unrelated Animacy Selection Restriction Violations	97% (4.3)
Fillers	93% (5.2)

Mean percentages are shown with SDs in brackets. A match was classifying a non-violated sentence as acceptable and a violated sentence as unacceptable.

Table 4. Omnibus ANOVAs in the 0-100ms, 150-250ms, N400 (300-500ms), and P600 (700-900) time windows, comparing ERPs to critical nouns across all five sentence types.

	Main effect of Sentence Type		*Interactions between Sentence Type and AP Distribution*	
	DoF	**F value**	**DoF**	**F value**
0-100ms				
Midline	4, 76	0.83	16, 304	0.53
Medial	4, 76	1.02	8, 152	0.98
Lateral	4, 76	0.81	12, 228	0.36
Peripheral	4, 76	0.62	16, 304	0.17
150-250ms				
Midline	4, 76	5.15**	16, 304	1.61
Medial	4, 76	5.44**	8, 152	1.15
Lateral	4, 76	5.15**	12, 228	0.75
Peripheral	4, 76	4.99**	16, 304	1.05
N400 (300-500ms)				
Midline	4, 76	8.53***	16, 304	2.93*
Medial	4, 76	10.10****	8, 152	1.18
Lateral	4, 76	9.10***	12, 228	1.78
Peripheral	4, 76	7.09***	16, 304	2.54**
P600 (700-900ms)				
Midline	4, 76	7.91***	16, 304	2.89*
Medial	4, 76	9.26****	8, 152	6.07***
Lateral	4, 76	7.69***	12, 228	3.20*
Peripheral	4, 76	6.00**	16, 304	1.66

DoF: Degrees of Freedom.
*p < .05 **p < .01. ***p < .001. ****p < .0001.

Table 5. Simple effects ANOVA comparing ERPs to each type of violated noun with the control critical nouns in the 150-250ms, N400 (300-500ms) and P600 (700-900ms) time windows

	150-250ms				***N400***				***P600***			
	Sentence Type		***Sentence Type x AP Distribution***		***Sentence Type***		***Sentence Type x AP Distribution***		***Sentence Type***		***Sentence Type x AP Distribution***	
	DoF	**F value**	**DoF**	**F value**	**DoF**	**F value**	**DoF**	**F value**	**DoF**	**F value**	**DoF**	**F value**
A. Related Real-World Knowledge Violations												
Midline	1,19	0.859	4,76	0.511	1,19	2.47	4,76	4.40*	1,19	0.06	4,76	1.46
Medial	1,19	0.078	2,38	0.7	1,19	3.24	2,38	0.46	1,19	0.51	2,38	0.31
Lateral	1,19	0.274	3,57	0.562	1,19	1.81	3,57	0.82	1,19	1.38	3,57	0.08
Peripheral	1,19	1.039	4,76	0.815	1,19	0.82	4,76	2.48	1,19	0.35	4,76	0.57
B. Unrelated Real-World Knowledge Violations												
Midline	1,19	14.461**	4,76	3.686*	1,19	10.69**	4,76	7.95****	1,19	0.79	4,76	1.77
Medial	1,19	16.301**	2,38	3.098*	1,19	14.37**	2,38	1.79	1,19	0.9	2,38	2.19
Lateral	1,19	13.554**	3,57	1.106	1,19	13.23**	3,57	3.31	1,19	1.54	3,57	2.46
Peripheral	1,19	13.516**	4,76	1.746	1,19	10.74**	4,76	6.69**	1,19	1.64	4,76	0.72
C. Related Animacy Selection Restriction Violations												
Midline	1,19	3.357	4,76	0.866	1,19	19.36***	4,76	4.50**	1,19	5.93*	4,76	1.43
Medial	1,19	3.029	2,38	1.37	1,19	24.93****	2,38	2.26	1,19	11.14**	2,38	7.45**
Lateral	1,19	2.088	3,57	0.368	1,19	17.91***	3,57	2.44	1,19	8.80**	3,57	3.28
Peripheral	1,19	1.121	4,76	0.333	1,19	12.76**	4,76	2.55	1,19	4.86*	4,76	0.71
D. Unrelated Animacy Selection Restriction Violations												
Midline	1,19	5.052*	4,76	0.567	1,19	11.90**	4,76	1.61	1,19	23.27***	1.61	6.31**
Medial	1,19	3.392	2,38	0.409	1,19	12.33**	2,38	1.07	1,19	23.68***	1.07	15.39***
Lateral	1,19	2.634	3,57	1.041	1,19	12.08**	3,57	0.41	1,19	34.88****	0.41	6.70**
Peripheral	1,19	1.823	4,76	2.058	1,19	11.36**	4,76	0.53	1,19	27.92****	0.53	3.13

*p < .05 **p < .01. ***p < .001. ****p < .0001.

Table 6. 2 x 2 ANOVAs: effects of Violation Type and Relatedness on violated nouns within the 150-250ms, N400 (300-500msec) and P600 (700-900msec) time windows.

	Main effect of Violation Type		*Main effect of Relatedness*		*Interaction between Violation Type and Relatedness*	
	DoF	**F value**	**DoF**	**F value**	**DoF**	**F value**
150-250ms						
Midline	1, 19	0.01	1, 19	9.58**	1, 19	9.390**
Medial	1, 19	0.02	1, 19	7.47*	1, 19	16.67**
Lateral	1, 19	0.54	1, 19	8.30*	1, 19	11.79**
Peripheral	1, 19	0.90	1, 19	7.81*	1, 19	9.74**
N400 (300-500ms)						
Midline	1, 19	5.17*	1, 19	4.20	1, 19	2.34*
Medial	1, 19	4.51*	1, 19	5.31*	1, 19	14.88**
Lateral	1, 19	4.39*	1, 19	6.22*	1, 19	16.39***
Peripheral	1, 19	3.78	1, 19	4.82*	1, 19	9.26**
P600 (700-900ms)						
Midline	1, 19	12.21**	1, 19	0.02	1, 19	1.46
Medial	1, 19	14.03**	1, 19	0.02	1, 19	2.21
Lateral	1, 19	10.06**	1, 19	0.00	1, 19	2.44
Peripheral	1, 19	11.01**	1, 19	0.20	1, 19	0.22

*p < .05 **p < .01. ***p < .001. ****p < .0001.

Fig. 1

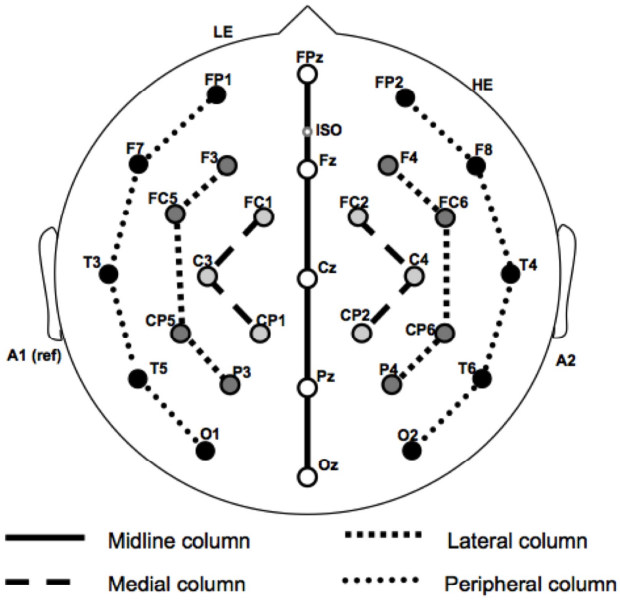

LE
FPz
FP1
FP2
HE
ISO
F7
F3
Fz
F4
F8
FC5
FC1
FC2
FC6
T3
C3
Cz
C4
T4
CP5
CP1
CP2
CP6
A1 (ref)
T5
P3
Pz
P4
T6
A2
O1
O2
Oz
Midline column
Lateral column
Medial column
Peripheral column

Fig. 2

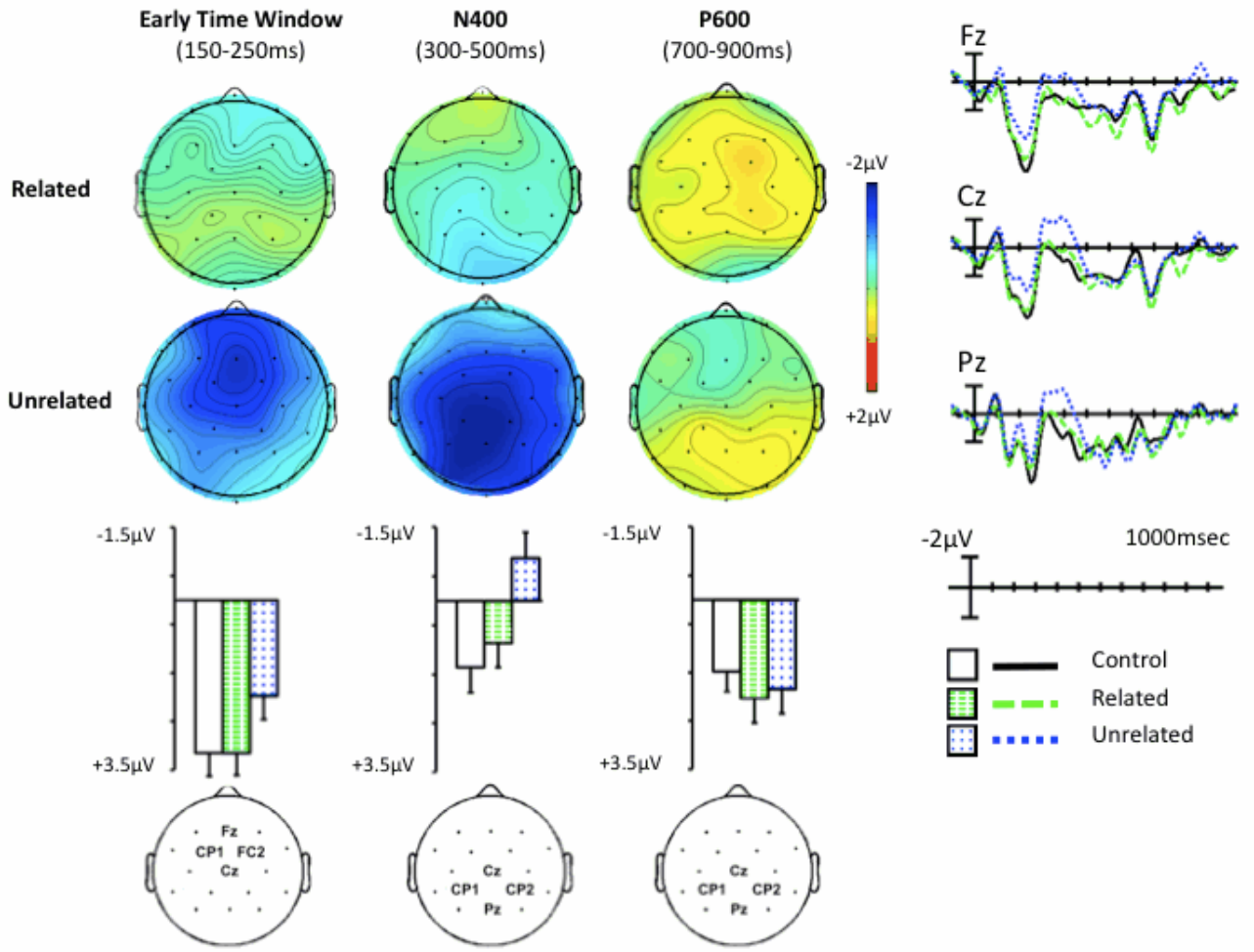

Early Time Window
(150-250ms)
N400
(300-500ms)
P600
(700-900ms)
Related
Unrelated
-2μV
+2μV
Fz
Cz
Pz
-1.5μV
+3.5μV
-2μV
1000msec
Control
Related
Unrelated
Fz
CP1 FC2
Cz
Cz
CP1 CP2
Pz
Cz
CP1 CP2
Pz

Fig. 3

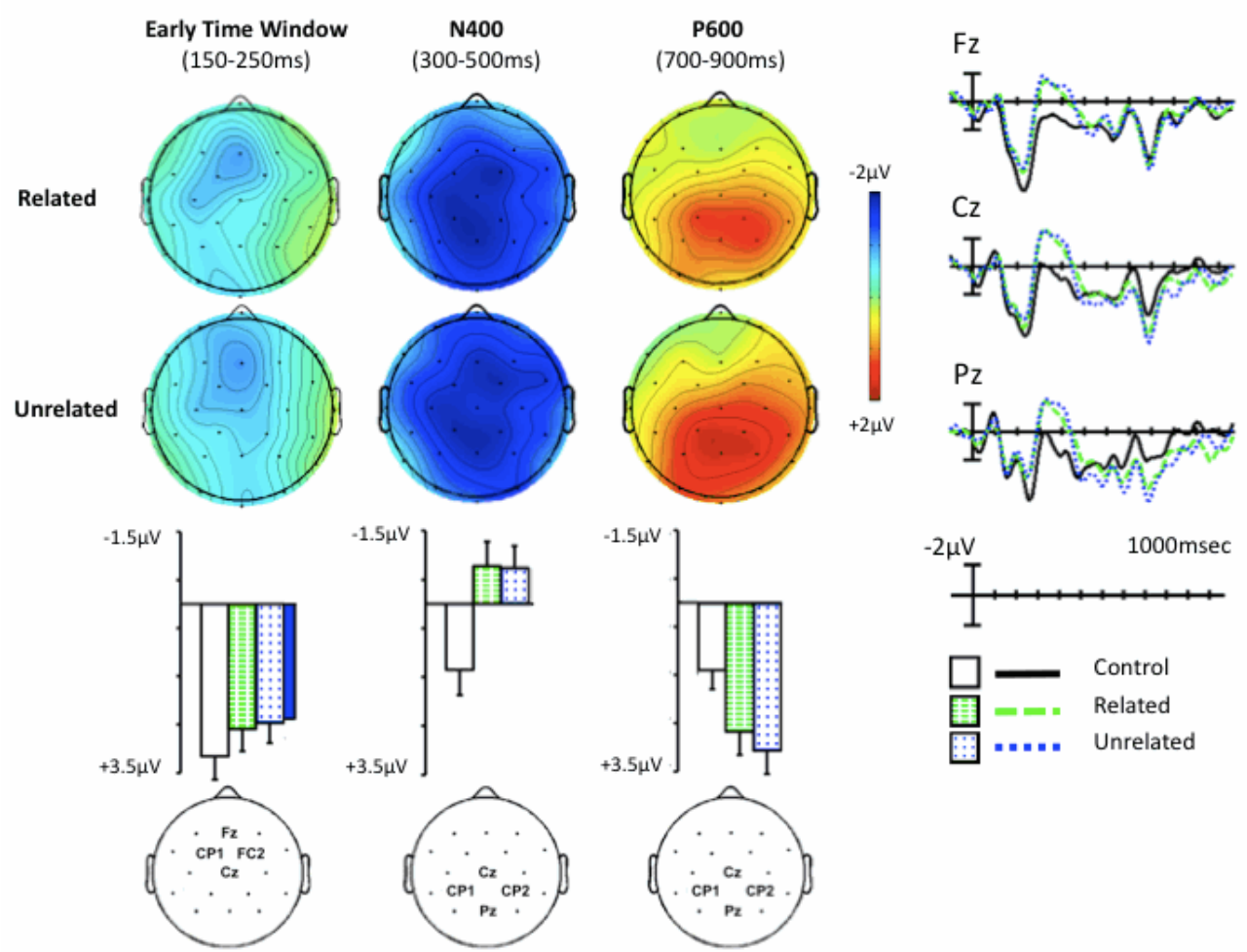
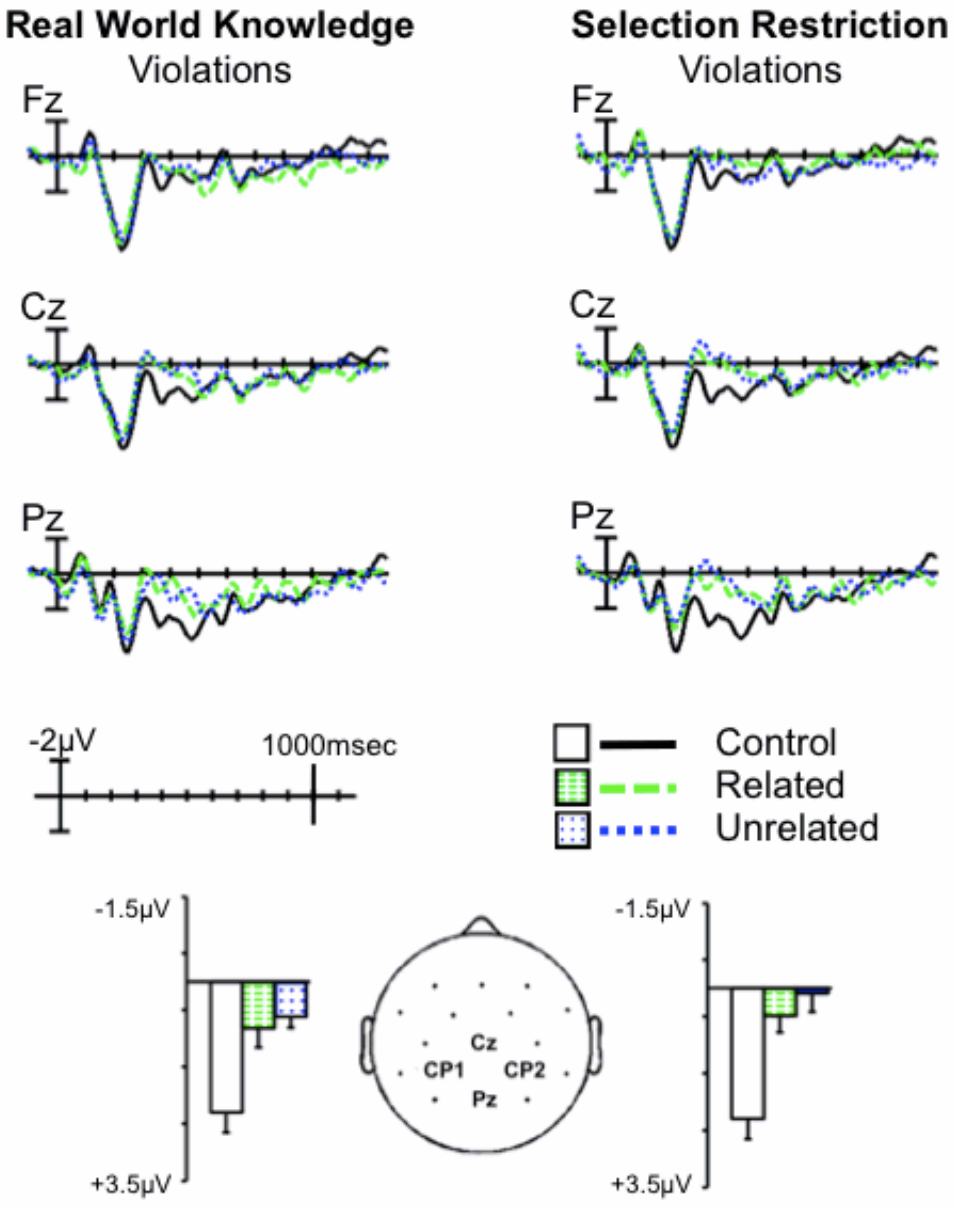

Early Time Window
(150-250ms)
N400
(300-500ms)
P600
(700-900ms)
Related
Unrelated
-2μV
+2μV
Fz
Cz
Pz
-1.5μV
+3.5μV
-2μV
1000msec
Control
Related
Unrelated
Fz
CP1 FC2
Cz
Cz
CP1 CP2
Pz
Cz
CP1 CP2
Pz

Fig. 4


Real World Knowledge
Violations
Selection Restriction
Violations
Fz
Cz
Pz
-2μV
1000msec
Control
Related
Unrelated
-1.5μV
+3.5μV
Cz
CP1
CP2
Pz
-1.5μV
+3.5μV

Neurophysiological Exploration of the Role of Animacy in Verb Argument Processing

Introduction.

That animacy information—whether a noun phrase describes an animate or inanimate object— plays a role in online verb argument processing is generally accepted. Yet despite over three decades of research, how this information is utilized by the language parser remains a topic of much debate. In the current study we contrasted two major categories of theories that have been put forth regarding animacy's role in verb argument processing: animacy as thematic role proxy versus animacy hierarchy based argument ordering. Below we briefly describe these two proposals in relation to experimental evidence in their support.

The first formal processing model that considered a role of animacy in thematic role assignment was introduced by Frazier and Rayner (1982). In their study, the authors found that the reading time penalty incurred when reading reduced relative clauses (e.g. The witness examined by the lawyer…) compared to their unreduced counterparts (e.g. The witness that was examined by the lawyer…), was substantially smaller when the subject argument was inanimate (The evidence examined by the lawyer…), a result later replicated by Rayner, Carlson, and Frazier (1983) as well as Ferreira and Clifton (1986). Frazier and Rayner (1982) suggested a thematic processing, relying on a small set of semantic features, such as animacy, facilitated the reassignment of the correct patient role to the subject noun phrase. Specifically, the authors proposed that animate subject arguments, being better prototypical agents, were more difficult to reassign to the patient role than were inanimate subject arguments, which were inherently more patient-like.

Subsequent studies have suggested that animacy may affect even earlier stages of processing (Trueswell, Tanenhaus & Garnsey, 1994; MacDonald, Pearlmutter, & Seidenberg, 1994). The

authors of these studies proposed that rather than aiding in thematic role *reanalysis*, as proposed by Frazier and Rayner (1982), animacy information was used immediately by readers, leading them to more readily interpret inanimate subject arguments as patients during the *initial* parse of the sentence.

More direct support for the immediate use of noun phrase animacy during online language processing comes from neurophysiological studies looking at event-related potentials (ERPs), a direct measure of neural activation. In an early study exploring effects of argument animacy in fully plausible sentences, Weckerly and Kutas (1999) found that animate subject arguments evoked an attenuated N400 compared to that evoked by inanimate subject arguments. Because the N400, a negative going deflection peaking approximately 400ms after stimulus onset, is thought to index semantic processing load, the authors proposed that the observed modulation of the N400 may have been driven by an association between animacy and thematic role assignment. Specifically, the authors suggested that since sentential subjects tend to be agents, animate subject noun processing was facilitated, as animate nouns tend to be more 'agent-like.' More recently, Nakano et al. (2010) replicated this modulation of the N400 by subject noun animacy. Similarly to Weckerly and Kutas (1999), Nakano et al. (2010) proposed that the language parser utilized animacy information for at least provisional thematic role assignment. The authors further suggested that inanimate subject nouns may have been used by the language parser to predict an upcoming non-canonical (i.e. passive) grammatical structure.

The role of animacy in online language parsing has recently been formalized by Bornkessel and Schlesewsky in their extended Argument Dependency Model (eADM; 2006). Within the model, the authors propose that several 'prominence' hierarchies are utilized to establish the likelihood of a particular thematic role being assigned to a given argument. Amongst these

prominence hierarchies is the animacy hierarchy; consistent with previous linguistic proposals (see Silverstein, 1976; Aissen, 2003), Bornkessel and Schlesewsky propose humans are high on this hierarchy, while inanimate objects are low. Within the eADM, N400 amplitude is seen as, in part, indexing the level of difficulty during initial thematic role assignment as well as difficulties in verb-argument linking.

Additionally, the eADM proposes that difficulties in the correct mapping of an argument to its correct role about a verb, as in the case of reduced relative clauses for example, would lead to additional processing indexed by the P600, a positive going ERP starting approximately 500-600ms post stimulus onset. The P600 was originally described as being evoked by violations of expected syntactic structure, whether or not such structures were anomalous on a sentential level (Ostnerhout & Holcomb, 1992), as well as outright morphosyntactic violations (Hagoort, Brown, & Groothusen, 1993). Though it has since been shown to be evoked by a variety of linguistic errors (for review, see Kuperberg, 2007), it is generally thought to reflect structural repair or reanalysis, whether the structure in question is syntactic or semantic in nature[7].

In contrast to studies of language comprehension, studies of language production have generally found support for verb argument animacy influencing language through the animacy hierarchy based linear ordering of verb arguments, sometimes referred to as the 'animate first' principle. This 'animate first' principle is formally codified within the rules of syntax of a number of languages. For example, in Navajo (Hale, 1972), transitive constructions in which an inanimate object acts upon an animate one must be expressed in the passive voice, e.g. "The man was warmed by the sun." Note that this is not a restriction on inanimate nouns in the subject position, as transitive constructions involving two inanimate entities can be expressed in the active voice, e.g. "The sun warmed the fields." Importantly, similar constraints on verb argument ordering can be

[7] See also Kolk's Error Monitoring Proposal (van de Meerendonk et al., 2010).

found in English, in ditransitive constructions. For example ditransitive constructions in which the recipient is inanimate while the theme is animate, are preferentially expressed in the propositional object alternation, e.g. “The President sent the ambassador to England,” rather than the double object construction, e.g. “??The President sent England the ambassador” (cf. Levin, 1989). Note, however, that again the double object construction becomes fully grammatical if both post-verbal arguments are inanimate, e.g. “The man bought his car a new set of tires” (see Paczynski & Kuperberg, 2011, for more in depth discussion of animacy hierarchy).

In laboratory settings in which verb argument animacy and thematic role assignment were systematically varied, evidence once again points toward animacy influencing production through an animate-first principle, rather than associations with thematic roles. In a series of studies, Bock demonstrated that, when asked to describe pictures containing an inanimate agent and an animate patient, participants employed the passive voice up to 74% of the time, compared to never when the agent was animate (Bock 1986; Bock & Loebell, 1990, Bock, Loebell & Morey, 1992). Ferreira (1994) likewise found that when the verb required an animate direct object argument (in canonically-ordered sentences), participants tended to use the passive voice when the arguments differed in animacy, placing the animate noun in subject position (e.g. “The man was alarmed by the news.” rather than “The news alarmed the man.”). When both arguments were animate (e.g. “The vagabond alarmed the man.”), participants again relied exclusively on the active voice. Similar biases have been observed in delayed recall tasks, in which participants frequently misremember the original grammatical form of sentences so as to place the animate argument before the inanimate one (Feleki & Branigan, 1997; McDonald, Bock & Kelly, 1993; Tanaka et al., 2005;). Indeed, Demuth et al. (2005) found that native Sesotho speakers attained adult proficiency in using animacy hierarchy constraints several years prior to attaining adult proficiency in using

thematic roles, offering further support for animacy influencing language independently of thematic roles.

Importantly, recent work from our group (Paczynski & Kuperberg, 2011) provided evidence suggesting that even during language comprehension, adult language users utilize animacy hierarchy based constraints on argument ordering rather than using animacy information to guide thematic role processing. In Experiment 1, we showed that inanimate direct object arguments assigned the patient role evoked an attenuated N400 compared to animate arguments assigned the patient role. In Experiment 2 we contrasted the processing of direct object arguments assigned the experiencer role, which requires the argument be animate, with those arguments assigned the patient role, more closely associated with inanimate nouns. We found that neither the N400 nor P600 were modulated by the thematic role assigned to the direct object argument. Taken together, our results argued against animacy playing a central role in facilitating thematic role processing at post-verbal arguments, at least in English. Instead, we interpreted our data as supporting animacy influencing verb argument processing through the animacy hierarchy, with animate arguments being preferred earlier in an utterance, while inanimate arguments being preferred later in the utterance.

The aim of the current study was to explore this question further. Although our previous results are intriguing, verb argument assignment in English is strongly driven by linear ordering. In our earlier study we examined thematic role processing at a post verbal argument, at which point the thematic role to be assigned to that argument has been fully determined by the preceding verb. One might argue that by this point in the construction of a coherent verb-argument structure, animacy information is of little utility to the language parser. Thus, in the current study we used ERPs to explore how the animacy of subject noun phrases influences the processing of verbs in

either active or passive structures in which the initial verb argument is assigned the agent or patient role, respectively.

Previous studies examining the influence of animacy on thematic role processing have relied on either complex, infrequent syntactic constructions (e.g. object relative clauses; Weckerly & Kutas, 1999; Traxler et al., 2002) or linguistic violations (e.g. Frisch & Schleswesky, 2001; Nakano et al., 2010). Such manipulations necessarily provide information about how the language parser functions under extreme conditions, which may not be representative of how animacy information is utilized in more common linguistic input. Thus, in the current study we examined how animacy influences verb argument processing in syntactically simple active and passive constructions (see Table 1 for example sentences). Note that we used the past perfect active voice. This was done for two reasons. First, in contrast to the simple past, the past perfect requires an auxiliary verb, like the passive voice does, thus making the two structures nearly syntactically identical. Second, the simple past is heavily overrepresented within daily discourse; as usage frequency of syntactic structure has previously been shown to influence processing, we chose a simple, yet less commonly used active construction[8].

If animacy impacts thematic role processing, this would lead to the following set of predictions. First, inanimate subject nouns should evoke a greater N400 amplitude than animate nouns, as seen by Weckerly and Kutas (1999) as well as Nakano et al. (2010). If this increase in N400 amplitude is due to a greater difficulty in assigning inanimate nouns the agent role, this would lead to the following predictions on the verb. There should be no difference in processing of verbs in the active voice, as the subject argument has already been assigned the correct thematic role prior to encountering the verb. However, a processing difference should be observed for verbs in the

[8] Note that in theory, participants could predict grammatical voice based solely on the auxiliary verb. In order to prevent participants from relying on auxiliary verbs to ascertain grammatical voice, two additional conditions were included as fillers: imperfective active voice (*was VERB-ing*) and past pluperfect (*had been VERB-ed*).

passive voice. Specifically, processes related to thematic role reassignment, as indexed by the P600 evoked by the verb, should be reduced in the case of inanimate subject noun phrases relative to animate subjects, as the reassignment of the subject to the patient role should be easier in the case of inanimate arguments which are more patient-like.

On the other hand, if the initial increase in N400 previously observed to inanimate subject nouns is related to the prediction of an upcoming non-canonical syntactic structure, as proposed by Nakano et al (2010), this would predict a somewhat different pattern of results. Specifically, for sentences in the active voice, inanimate subject nouns should lead to a larger P600 on the verb relative to active sentences in which subject noun phrases are animate, as the animacy based prediction of syntactic structure would be incorrect. A similar pattern of results should be observed for sentences in the passive voice with animate subjects, relative to inanimate subjects, as here, once again, the animacy based prediction of thematic role assignment is not met.

Finally, if animacy does not impact thematic role assignment until the verb, this would predict that in active sentences, verbs preceded by inanimate nouns should evoke a larger N400 relative to those preceded by animate nouns, as it should be easier to assign the animate nouns to the agent role. Again, the opposite pattern would be predicted for sentences in the passive voice; critical verbs preceded by animate nouns should evoke a larger N400 than those preceded by inanimate nouns, due to inanimate nouns being more easily assigned the patient role. Note that in all cases, an animacy by grammatical voice interaction would be predicted on the critical verb.

A somewhat different pattern of results would be predicted if animacy hierarchy biases on argument ordering impact processing. Similarly to the above proposal, this would predict a greater N400 to be evoked on inanimate subject nouns compared to animate ones, as these would violate the ‘animate first’ expectation. We would predict that, independent of grammatical voice, critical

verbs following inanimate nouns should evoke a reduced N400, compared to critical verbs following animate subject nouns. The reasoning behind this is as follows. By violating broad semantic expectations, inanimate subject nouns should become more salient than animate subject nouns. Previous research has demonstrated that increasing subject noun salience leads to these arguments remaining more active or more readily accessed within working memory (Johnson & Gordon, 2009; see Paczynski & Kuperberg, in press, for more extensive discussion). This would, presumably, facilitate lexico-semantic integration of the verb and the subject argument, leading to an attenuation of the N400. Additionally, it is possible that violating animacy hierarchy based constraints on verb argument ordering may lead to a greater processing cost within the P600 time window, as this waveform has been shown to be modulated by violations expectations. Note, that the animacy hierarchy proposal predicts a main effect of subject noun animacy on critical verb processing but *no* animacy by grammatical voice interaction.

Methods

Development of Materials

Two hundred forty (240) transitive verbs (e.g. *block, scratch*) were selected such that each one allowed for both the subject and direct object argument to be plausibly filled by either an animate or inanimate noun. Subsequently, 240 animate nouns and 240 inanimate nouns were chosen such that each noun could serve as a plausible agent for one or more of the transitive verbs selected as well as serve as a plausible patient for one or more of the transitive verbs selected. Note that due to this constraint, noun length and frequency was not matched between animate and inanimate nouns. Average length of animate nouns was significantly longer than that of inanimate ones (7.5 vs. 6.1 characters, respectively, $t(239)=1$, $p < 0.001$). Additionally, animate nouns were

slightly less frequent, based on SUBTITLE Frequency (available at http://elexicon.wustl.edu/), with $LgSUBTL_{CD}$ of 2.46 vs. 2.59 (log frequency per million) for animate vs. inanimate nouns, a difference that approached significance ($t(228)=0.94$, $p=0.07$). Note, however, that such differences would be expected to result in the N400 to inanimate nouns being *attenuated* compared to animate nouns, an effect opposite to the one predicted within the current experimental paradigm.

Each noun was paired with a verb for which it could serve as a plausible agent as well as a verb for which it could serve as a plausible patient. This resulted in each verb being paired with two animate nouns, one agent, one patient, and two inanimate nouns, one agent, one patient. Across all pairings, each noun served as both an agent and a patient, thus controlling for any inherent agent-like or patient-like associations that a given noun may have. Additionally, to control for effects of lexico-semantic association, semantic similarity values (SSVs) between each verb-noun pair were calculated using Latent Semantic Analysis (LSA), (Landauer and Dumais, 1997; Landauer et al., 1998), available on the internet at http://lsa.colorado.edu, using tasaALL space (1st year college student reading level). Subject arguments were swapped between scenarios until SSVs were identical between all noun-verb pairings (M=0.27).

For each verb, two experimental sentence frames were created: 1) *had VERB-ed*, pluperfect active, and 2) *was VERB-ed*, simple past passive. In order to prevent the auxiliary verbs *had* and *was* from predicting grammatical voice, two control sentence frames were also created for each verb: 1) *was VERB-ing* (past progressive active) and 2) *had been VERB-ed* (pluperfect passive). Thus both of the auxiliary verbs, *had* and *was*, could resolve to either an active or passive structure, preventing participants from potentially reanalyzing initial thematic role assignment prior to the presentation of the critical verb.

Each syntactic frame was then combined with appropriate subject nouns, animate and inanimate, to yield a total of eight conditions, four experimental and four control. See Table 1 for summary of all conditions along with example stimuli.

For each initial verb phrase a plausible sentence continuation was created. As previous research has suggested (see review by Kuperberg, 2007) that plausibility judgments lead to deeper processing of linguistic input, we also created anomalous sentence continuations. These were identical to the plausible sentence continuation with the exception that one word was substituted in order to render the overall meaning of the sentence anomalous. Substitutions could occur on any content word in the sentence to prevent participants from being able to anticipate location or nature of semantic anomaly. Anomalous words were chosen to be semantically related to the preceding sentential context (e.g. “The sheriff had blocked the highway with his truck/*badge…”).

Eight experimental lists were created, such that across all lists, each verb appeared in all eight of the conditions and that each list contained the same number of sentences from each of the eight different conditions. Within each list, items were pseudo-randomized such that no more than two scenarios in a row were from the same condition, with the additional constraint that no more than four scenarios in a row contained subject nouns of the same animacy nor that more than four scenarios in a row utilized the same grammatical voice. After these conditions were met, half of all sentence continuations were changed to their anomaly-containing counterpart. Again, pseudo-randomization was used to ensure that no more than three sentences in a row were of the same plausibility and that within each list, condition did not predict plausibility.

ERP Experiment

Participants

Twenty-nine undergraduate native English speakers from Tufts University were recruited for the experiment. Of these, four were rejected due to excessive ocular artifacts and one due to recording error. Of the twenty-four usable participants, age 18-25 (M=20.1, SD=2.2), 15 female, 9 male. All participants had normal or corrected-to-normal vision, were not taking psychoactive medications, had no learning disability, no history of neurological or psychiatric disorders and had not learned languages other than English before the age of 5. All were right handed as assessed through a modified version of the Edinburgh handedness inventory (Oldfield, 1971). Written consent was obtained from all subjects before participation according to the established guidelines of Tufts University. Participants were paid for their participation.

Electrophysiological Recording

Twenty-nine tin electrodes were held in place on the scalp by an elastic cap (Electro-Cap International, Inc., Eaton, OH), see Figure 1. Electrodes were also placed below the left eye and at the outer canthus of the right eye to monitor vertical and horizontal eye movements, and on the left and right mastoids. Impedance was kept below 2.5 kΩ for all scalp and mastoid electrode sites and below 10 kΩ for the two eye channels. The EEG signal was amplified by an Isolated Bioelectric Amplifier System Model HandW-32/BA (SA Instrumentation Co., San Diego, CA) with a bandpass of 0.01 to 40 Hz and was continuously sampled at 200 Hz by an analogue-to-digital converter. The stimuli and behavioral responses were simultaneously monitored with a digitizing computer.

Data Analysis

ERPs were formed by off-line averaging of artifact-free trials, and time-locked to the onset of critical words in each sentence. Each trial was baselined using average voltage between a -50

pre-stimulus to 50ms post stimulus onset. Three time-windows of interest were chosen a priori: the 300-500msec, 500-700msec and 700-900msec. The first time window encompasses the N400. The second and third time windows encompassed the early and late P600. We also carried out analyses of the 100-200msec, and 200-300msec time windows to ensure later effects were not driven by early effects, possibly due to artifact. The modulation of average ERPs within each of these time windows was examined using analyses of variance (ANOVAs) for repeated measures at each of four electrode columns (see Figure 1). The Midline column had five levels of electrode sites along the anterior-posterior distribution (AP Distribution) of the scalp (FPz, Fz, Cz, Pz, Oz), the Medial column had three levels for AP Distribution and two levels for Hemisphere (FC1/FC2, C3/C4, CP1/CP2), the Lateral column had four levels for AP Distribution and two levels for Hemisphere (F3/ F4, FC5/FC6, CP5/CP6, P3/P4), and the Peripheral column had five levels of AP Distribution and two levels of Hemisphere (FP1/FP2, F7,F8, T3/T4, T5/T6, O1/O2). For critical verbs, we carried out a 2 (subject noun animacy) x 2 (grammatical voice) ANOVAs, with AP Distribution and Hemisphere (for the medial, lateral, and peripheral columns) as additional within-subject factors. Similar analyses were carried out for the subject noun and semantic control noun, with animacy and semantic-anomaly as within-subject factors, respectively. Significance was set at alpha equal to 0.05. A Geisser-Greenhouse correction was applied to all repeated measures with more than one degree of freedom, for which original degrees of freedom and corrected probability levels are reported.

Linearly interpolated voltage maps showing the scalp distribution of differences in ERPs elicited by critical nouns were produced using EEGLAB v4.512 for MatLab software.

Results

Subject Noun

Visual inspection of the waveform seen in Figure 2 indicated that inanimate subject nouns evoked an attenuated P2 amplitude compared to animate subject nouns. This was followed by a robust N400 effect for inanimate, relative to animate, subject nouns that appeared to have a slightly anterior distribution.

Early Effects (100-200ms, 200-300ms)

Within the 100-200ms time window, neither was there a main effect of animacy, nor were there any effects involving animacy significant at any of the four electrode columns, all $Fs < 1$, $p > 0.72$. However, animacy modulated the amplitude of the P2 (200-300ms) as indicated by a significant main effect of animacy at all four electrode columns, all $Fs > 9.57$, $ps < 0.01$. No interactions involving animacy reached significance at any electrode columns, all $Fs < 1.82$, $ps > 0.19$.

N400 (300-500ms)

We found a highly significant main effect of animacy at all four electrode columns, all $Fs > 29.49$, $ps < 0.0001$. Additionally, there was a significant animacy by ap-distribution interaction at midline, $F(4,92)=6.13$, $p < 0.01$, medial, $F(2,46)=4.13$, $p\ 0 < 0.05$, and peripheral, $F(4,92)=5.28$, $p < 0.05$, due to the effect of animacy having a fronto-central distribution. Finally, there was an animacy by hemisphere interaction at the peripheral electrode column, $F(1,23)$, $p < 0.05$, due to the difference in N400 effect being greater over right hemisphere electrodes.

P600 (500-700ms, 700-900ms)

Neither in the early nor in the late P600 time windows did the main effect of animacy reach significance at an electrode column, all Fs < 1.92, ps > 0.18, nor were any interactions involving animacy significant at any electrode column, all Fs < 1.58, ps > 0.21.

Critical Verb

Early Effects (100-200ms, 200-300ms)

Within the 100-200ms time window, neither the main effect of animacy nor that of grammatical voice reach significance at any of the four electrode columns, Fs < 1, ps > 0.44, nor were any interactions involving either animacy or grammatical voice significant, Fs < 2.43, ps > 0.13.

There was, however, a significant main effect of grammatical voice at the medial electrode column, $F(1,23)=6.10$, $p < 0.05$, as well as a significant voice by ap-distribution by hemisphere interaction at the lateral electrode column, $F(3,69)= 3.56$, $p < 0.05$. The main effect of voice approached significance at the midline, $F(1,23)=3.37$, $p=0.08$, and lateral, $F(1,23)=3.44$, $p=0.08$, electrode columns. Although occurring within the P2 time window, the effect appeared to be driven by a small, negative going deflection, peaking approximately 225ms post stimulus onset over right parietal sites. See Figure 3.

Because of the novelty of this ERP component, we compared the ERPs evoked during this time window by the experimental passive condition with those evoked by the control active condition, the past progressive. As can be seen in Figure 3 inset, the waveforms overlap almost completely. Post-hoc comparison of the two conditions showed no main effect of grammatical voice, Fs < 1, ps > 0.87, nor any significant interactions involving voice, Fs < 1, ps > 0.75. This

would suggest the component was not related to grammatical voice. See Figure 3 inset, as well as footnote 9 for discussion.

N400 (300-500ms)

There was a main effect of animacy at the midline, $F(1,23)=5.67$, $p < 0.05$, medial, $F(1,23)=5.15$, $p < 0.05$, and lateral, $F(1,23)=5.75$, $p < 0.05$, electrode columns, approaching significance at the peripheral electrode column, $F(1,23)=4.19$, $p = 0.05$. The effect was due to critical verbs following inanimate subject nouns evoking an attenuated N400 compared to critical verbs following animate subject nouns. There was no main effect of voice at any electrode column, $Fs < 1$, $ps > 0.45$, nor did any interactions involving animacy or grammatical voice reach significance at any electrode column, $Fs < 1.9$, $ps > 0.14$. See Figures 3, 4, and 5.

P600 (500-700ms, 700-900ms)

Within the early P600 time window, there was a significant animacy by ap-distribution interaction at the medial, $F(2,46)=3.86$, $p < 0.05$, lateral, $F(3,69)=6.58$, $p < 0.01$, and peripheral, $F(4,92)=3.47$, $p < 0.05$, electrode columns, due to critical verbs following inanimate subject nouns evoking a greater positivity over posterior electrode sites, compared to critical verbs following animate subjects. Main effects of animacy only approached significance at the midline, $F(1,23)=3.02$, $p=0.10$, and medial, $F(1,23)=3.06$, $p=0.09$, electrode columns. No other effects involving grammatical voice or animacy reached significance, $Fs < 1.9$, $ps > 0.18$. See Figure 5.

Within the late P600 time window, the main effect of animacy reached significance at the midline, $F(1,23)=9.06$, $p < 0.01$, medial, $F(1,23)=8.58$, $p < 0.01$, lateral, $F(1,23)=6.64$, $p < 0.05$, and peripheral, $F(1,23)=7.71$, $p < 0.05$, electrode columns. The effect was due to critical verbs

following inanimate subjects evoking a broadly distributed positivity compared to that evoked by critical verbs following animate subjects. Main effect of voice did not reach significance at any electrode column, $Fs < 2.51$, $ps > 0.13$. No interactions involving animacy or grammatical voice reached significance, $Fs < 1.49$, $ps > 0.22$. See Figure 5.

Semantic Anomaly Control Word

Early Effects (100-200ms, 200-300ms)

Within the 100-200ms time window, there was no main effect of condition (control vs. semantic anomaly) at any electrode column, all $Fs < 1$, $ps > 0.86$, nor were any interactions involving condition significant, all $Fs < 1.82$, $ps > 0.19$.

Within the 200-300ms, there was a main effect of condition at the midline, $F(1,23)=6.09$, $p < 0.05$, and peripheral, $F(1,23)=4.55$, $p < 0.05$, approaching significance at the medial, $F(1,23)=3.98$, $p = 0.06$, and lateral, $F(1,23)=3.64$, $p = 0.07$, electrode columns. The effect was due to a slight attenuation of the P2 for semantically anomalous words compared to non-anomalous words. No interactions involving condition reached significance, all $Fs < 1.93$, $ps > 0.16$.

N400 (300-500ms)

There was a significant effect of condition at the midline, $F(1,23)=68.34$, $p < 0.0001$, medial, $F(1,23)=58.50$, $p < 0.0001$, lateral, $F(1,23)=66.78$, $p < 0.0001$, and peripheral, $F(1,23)=57.93$, $p < 0.0001$, electrode columns. There was also a significant condition by ap-distribution interaction at the midline, $F(1,23)=6.26$, $p < 0.01$, and lateral, $F(3,69)=4.98$, $p < 0.05$, electrode columns. The effect was due to anomalous words evoking a greater negativity compared

to non-anomalous words, especially over central electrode sites. No other interactions reached significance, Fs < 1.62, ps > 0.21. See Figure 6.

P600 (500-700ms, 700-900ms)

Within the early P600 time window, the main effect of animacy reached significance at the midline, $F(1,23)=6.61$, $p < 0.05$, medial, $F(1,23)=5.50$, $p < 0.015$, and peripheral, $F(1,23)=4.60$, $p < 0.05$, electrode columns, nearly reaching significance at the lateral, $F(1,23)=4.22$, $p = 0.05$, electrode column. The effect was due to the negativity observed in the N400 time window continuing into the early P600 time window. No interactions involving condition reached significance.

Within the late P600 time window, the main effect of condition reached significance at the lateral, $F(1,23)=4.74$, $p < 0.05$, electrode column. The effect was due to anomalous words evoking a slightly greater positivity compared to non-anomalous words. No interactions involving condition reached significance, Fs < 1.97, ps > 0.15.

Discussion

In the current study we contrasted two proposals regarding how verb argument animacy influences processing: through an association with thematic roles or through an animacy-hierarchy based ‘animate-first’ principle. We found that inanimate sentence subjects evoked a robust, anteriorly distributed N400 effect, relative to animate subjects. Additionally, we found that the N400 evoked by critical verbs was attenuated for verbs following inanimate versus animate subjects. Additionally, critical verbs following inanimate subject noun phrases evoked a broadly distributed P600 effect. Neither of these effects was modulated by grammatical voice. Finally,

grammatical voice failed to modulate ERPs evoked by critical verbs within any of the examined time windows. Below we discuss these findings in more detail.

Consistent with previous finding by Weckerly and Kutas (1999) as well as Nakano et al. (2010), inanimate subject nouns evoked a greater N400 effect compared to animate subject nouns. Because inanimate nouns used in our study were more frequent than the animate ones, the observed effect could not be due to low-level lexical properties of the items, as this would have resulted in the opposite pattern of N400 modulation (Lau, Philips, & Poeppel, 2008).

Recall from the introduction that the observed pattern of N400 modulation for the subject argument is consistent with predictions made by both models being tested in the current study, although they offer different interpretations of what this modulation indexes. Within the 'animacy as thematic role proxy' model, the larger N400 amplitude in response to inanimate subject nouns is due to the mismatch between the thematic role most often assigned to the subject noun, i.e. an agent, and the thematic role more closely associated with an inanimate noun, i.e. a patient (cf. Bornkessel & Schlesewsky, 2006). Nakano et al. (2010) go further and suggest that this additional processing may be associated with the language parser switching expectations from a canonical argument structure (SVO) to a non-canonical one (OVS). The 'animate-first' hypothesis, in contrast, suggests that the observed increase in N400 amplitude for inanimate, relative to animate, subject noun phrases is driven by a mismatch between broad expectations about the animacy of the sentence initial noun phrases (animate) and that of the subject noun actual presented (inanimate).

The predictions made by the two proposals diverge at the point of the critical verb. The 'animacy as thematic role proxy' set of proposals all predict some type of animacy by grammatical voice interaction, either within the N400 or P600 time window. On the other hand, the 'animate first' proposal predicts only a main effect of subject noun animacy. Our results are more consistent

with the latter; subject animacy had effect in both the N400 and late P600 time window but did not interact with grammatical voice in any time window examined.

We proposed that the violation of the 'animate first' principle makes the initial subject argument more salient. Recent work by Johnson and Gordon (2009) demonstrated that the reading time penalty normally associated with reading object relative clauses, compared to their subject relative clause counterparts, was reduced when the initial clause argument was made more salient by means of an adjectival modifier. Presumably, such increased salience would make the argument more readily accessible in working memory, either in facilitating the initial (correct) assignment of the noun phrase to the appropriate verb argument or during re-analysis of the verb argument structure.

Our results provide two pieces of evidence in support of our explanation. First, consistent with Nakano et al.'s (2010) findings, the N400 effect evoked by inanimate subject nouns in the current study was more anteriorly distributed than the classic N400 effect evoked by semantic anomalies. As noted by Nakano et al. (2010), such anteriorly distributed negativities have previously been associated with increased working memory load (cf. King & Kutas, 1995; Münte et al., 1998). Second, the amplitude of the N400 is thought to reflect the degree of semantic feature matching between an incoming word and its preceding context (cf. Kutas & Federmeier, 2011). As semantic similarity values between the critical verb and preceding subject NP were matched across conditions, we propose that the observed attenuation of the N400 evoked by critical verbs following inanimate subjects was due to the semantic features of the subject noun being more active within working memory, thus facilitating the matching of these features with the semantic features of the incoming critical verb.

Moving onto the P600 modulation observed in the current experiment. Recall that the P600 observed in the current study had a wide distribution. This contrasted considerably from the P600 effects previously reported in response to syntactic manipulation (e.g. Osterhout & Holcomb, 1992) as well as to violations of thematic role structure (see Kuperberg, 2007), in which the effect was more posteriorly distributed. As our paradigm did not include any manipulation that would evoke a more traditional P600 effect, the distinctness of distribution is necessarily speculative. Nonetheless, the more anterior distribution of the P600 in our experiment suggests that it may share functional overlap with an anteriorly distributed P600 reported by Federmeier, Wlotko, De Ochoa-Dewald, and Marta Kutas (2007). As in the current experiment, words which evoked the anterior P600 reported by Federmeier et al. (2007) were fully plausible within their preceding context. The effect was only found in response to words which were unexpected in highly constrained contexts. The authors interpreted their results as suggesting that the anterior positivity could be due to "surprise and/or increased resource demands entailed by the need to override or suppress a strong prediction" (p. 81). It may be the case, therefore, that even within an experimental paradigm in which half of all sentences presented begin with an inanimate subject noun, such inanimate first constructions are nonetheless linguistically surprising. This seems likely given the N400 effect observed for inanimate, versus animate, subject nouns.

Finally, we turn to the effects of grammatical voice. As noted above, we found no evidence for grammatical voice impacting critical verb processing within any of the examined time windows. While null results must necessarily be interpreted with caution, there are several reasons to believe that our data is not due to Type II error. First, visual inspection of Figure 3 fails to reveal any clear main effect of grammatical voice nor any interaction with subject noun animacy in any time

window[9]. Second, we found an N400 effect not only to sentence initial inanimate subject nouns but to semantically anomalous words embedded later within sentences. As the anomalous words were semantic related to the preceding context, detection of the violation relied on a full syntactic parse of the preceding sentence. Finally, during debriefing, nearly all participants commented on the large number of passives in the experiment without an explicit prompt from the experimenter. This suggests that participants found grammatical voice to be a salient feature of the experimental stimuli.

At first the lack of processing cost associated with online passive structure processing may seem surprising. Passive structures have often been proposed to be more syntactically complex (e.g. Baker, 1988) than active structures. Additionally, correct thematic role assignment has been found to be impaired for passive, versus active, structures (Ferreira, 2003), especially in patients with Broca's aphasia (e.g. Thompson & Lee, 2009). However, the view of passives as syntactically complex is not universal (e.g. Culicover & Jackendoff, 2005). Additionally, work comparing reading times for active versus passive structures found that *per character* reading times were in fact *shorter* for simple passives compared to simple past tense sentences.

We therefore interpret our findings as suggesting that simple passive structures are as easily computed during online language parsing as are canonical active structures. We propose that thematic role assignment may be relatively computationally simple even in passive structures, at least at the point of the verb in English sentences. It is possible that previously reported difficulties in appropriate thematic role assignment (e.g. Ferreira, 2003) in passive structures may have been

[9] The exception to this was a small negativity peaking at around 225ms post stimulus onset over right parietal sites for critical verbs in the passive voice, compared to those in the active voice. However, as noted in the results, this effect was also observed for control active sentences in the past progressive. In both cases the auxiliary verb required prior to the critical verb is *was*, with difference in voice being determined by the morphological suffice *–ed* vs. *–ing*. Thus, we suggest that rather than being related to grammatical voice processing itself, the observed negativity may be related to orthographic processing.

due to more downstream processes, such as the interpretation of the optional post-verbal agent argument.

Conclusions

Our study adds to a growing body of work suggesting that animacy may influence online language processing through its association with preferences for the linear ordering of verb arguments based on the animacy hierarchy. We suggest that at least some of the previously reported associations between animacy and thematic roles may be explained through the increase in the saliency of verb arguments which do not conform to general biases in surface sentence form.

References

Baker, M. (1988) Incorporation: A Theory of Grammatical Function Changing, University of Chicago Press, Chicago, Illinois.

Bock, J. K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. *12*, 575 – 586.

Bock, J. K., & Loebell, H. (1990). Framing Sentences. *Cognition, 35*, 1-39

Bock, J. K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*. 99, 150 –71.

Bornkessel, I, Schlesewsky, M (2006) The Extended Argument Dependency Model: A Neurocognitive Approach to Sentence Comprehension Across Languages. *Psychological Review*. 113, 787-821

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

Demuth, K., Machobane, 'M., Moloi, F., Odato, C. (2005) Learning Animacy Hierarchy Effects in Sesotho Double Object Applicatives. *Language*. 81, 421-447

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, 1146, 75-84.

Feleki, E. & Branigan, H. (1997). *Animacy effects in Greek sentence production: Evidence for single stage syntactic processing?* Poster presented at the Third Conference of Architectures and Mechanisms of Language Processing (AMLaP-97), Edinburgh, England.

Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language*, 33, 715-736.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47, 164-203.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*. 25, 348 – 368.

Frazier, L. & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*. 14, 178 – 210.

Frisch, S., & Schlesewsky, M. (2001). The N400 indicates problems of thematic hierarchizing. *Neuroreport*, 12, 3391 – 3394.

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

Hale, K. (1972) A note on subject-object inversion in Navajo. In: Kachru, B.,Lees, R.,Malkiel,Y.,Pietrangeli,A.,Saporta, S. (Eds.), *Issues in Linguistics: Papers in Honor of Henry and Renée Kahane*. University of Illinois Press, Urbana, 300–309.

Johnson, M., & Gordon, P. (2009). *Relative Salience of Noun Phrases in Object-Extracted Relative Clauses.* Poster Presented at Twenty Second Annual CUNY Conference, Davis, California, USA.

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

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato□s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211–240.

Landauer, T. K., Foltz, P. W., & Dumais, S. T. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes, 25*, 259–284.

Levin, B. (1993). English Verb Classes and Alternations: A Preliminary Investigation. University Of Chicago Press.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.

McDonald, J.L., Bock, K., & Kelly, M. (1993). Word and word order: semantic, phonological and metrical determinants of serial position. *Cognitive Psychology*. 25, 188 – 230.

Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, 395(6697), 71-3.

Nakano, H., Saron, C., Swaab, T.Y. (2010) Speech and Span: Working Memory Capacity Impacts the Use of Animacy but Not of World Knowledge during Spoken Sentence Comprehension. *Journal of Cognitive Neuroscience.* 22, 2886–2898

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

Paczynski, M., Kuperberg, G.R. (2011). Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb argument processing. *Language and Cognitive Processes. Special Issue: the Cognitive Neuroscience of Semantic Processing*. 26(9), 1402-1456

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

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*. 22, 358-374.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language,* 47,69—90.

Tanaka, M., Branigan, H. P., & Pickering, M. J. (2005). *The role of animacy in Japanese sentence production*. Paper presented at eighteenth annual CUNY conference, Tucson, AZ.

Trueswell, J., Tanenhaus, M., & Garnsey, S. (1994). Semantic influences on parsing: use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*. 33, 285 – 318.

van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010). Monitoring in language perception: mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, 22, 67-82.

Weckerly, J., & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology*, 36, 559 – 570

Figure Captions

Figure 1. Electrode montage. Analyses of variance were conducted at midline, medial, lateral and peripheral electrode columns shown (see Methods).

Figure 2. ERPs evoked by subject noun phrases at electrode Fz. Plot is shown using a -50ms pre-stimulus to 50ms post-stimulus baseline. Solid black line indicates ERPs evoked by animate noun phrases; dotted blue line indicates ERPs evoked by inanimate noun phrases. Voltage map shows average difference between the two conditions in the N400 (300-500ms) time-window.

Figure 3. ERPs evoked by critical verbs at select midline electrode sites. Plot is shown using a -50ms pre-stimulus to 50ms post-stimulus baseline. Solid black line indicates ERPs evoked by critical verbs in active voice, preceded by animate subject noun phrases. Dotted blue line indicates ERPs evoked by critical verbs in passive voice, preceded by animate subject noun phrases. Dashed red line indicates ERPs evoked by critical verbs in active voice, preceded by inanimate subject noun phrases. Dash-dotted green line indicates ERPs evoked by critical verbs in passive voice, preceded by inanimate subject noun phrases

Figure 4. ERPs evoked by critical verbs at select midline electrode sites. Plot is shown using a -50ms pre-stimulus to 50ms post-stimulus baseline. Solid black line indicates ERPs evoked by critical verbs in active voice. Dotted blue line indicates ERPs evoked by critical verbs in passive voice. Inset shows early ERPs evoked by critical noun at the P4 electrode site in the two conditions described above as well as those evoked by critical verbs in the control active (i.e. progressive) condition (dashed red line).

Figure 5. ERPs evoked by critical verbs at select midline electrode sites as well as voltage maps contrasting ERPs evoked by critical verbs following animate vs. inanimate subject noun phrases across the three time-windows of interest, N400 (300-500ms), early P600 (500-700ms) and late

P600 (700-900ms). Plot is shown using a -50ms pre-stimulus to 50ms post-stimulus baseline. Solid black line indicates ERPs evoked by critical verbs following animate subject noun phrases. Dotted blue line indicates ERPs evoked by critical verbs following inanimate subject noun phrases.

Figure 6. ERPs evoked by semantic control word at electrode Cz. Plot is shown using a -50ms pre-stimulus to 50ms post-stimulus baseline. Solid black line indicates ERPs evoked by non-anomalous control words; dotted red line indicates ERPs evoked by anomalous control words. Voltage map shows average difference between the two conditions in the N400 (300-500ms) time-window. Note: Voltage map scale is double that of preceding voltage maps.

Table 1. Experimental conditions with example verb phrases.

Condition	**Example**
Animate Subject/ Active Voice	The chef had cut…
Inanimate Subject/ Active Voice	The knife had cut…
Animate Subject/ Passive Voice	The victim was cut...
Inanimate Subject/ Passive Voice	The carrot was cut…
Control Conditions	**Example**
Animate Subject/ Active Voice	The chef was cutting…
Inanimate Subject/ Active Voice	The knife was cutting …
Animate Subject/ Passive Voice	The victim had been cut..
Inanimate Subject/ Passive Voice	The carrot had been cut…

Fig. 1

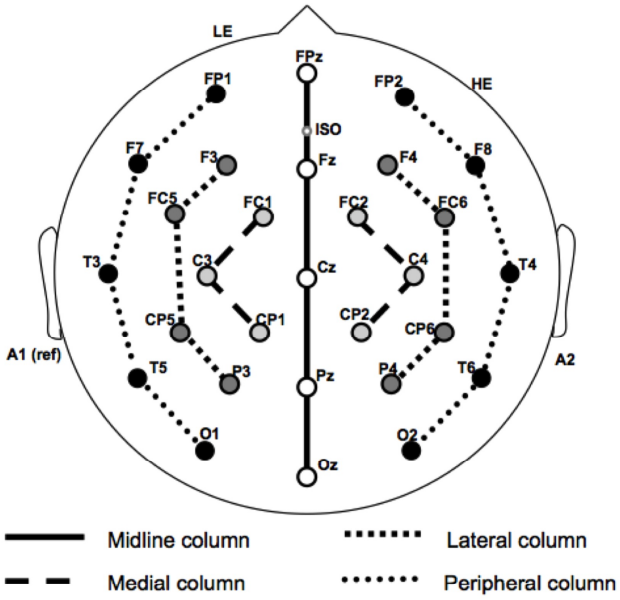
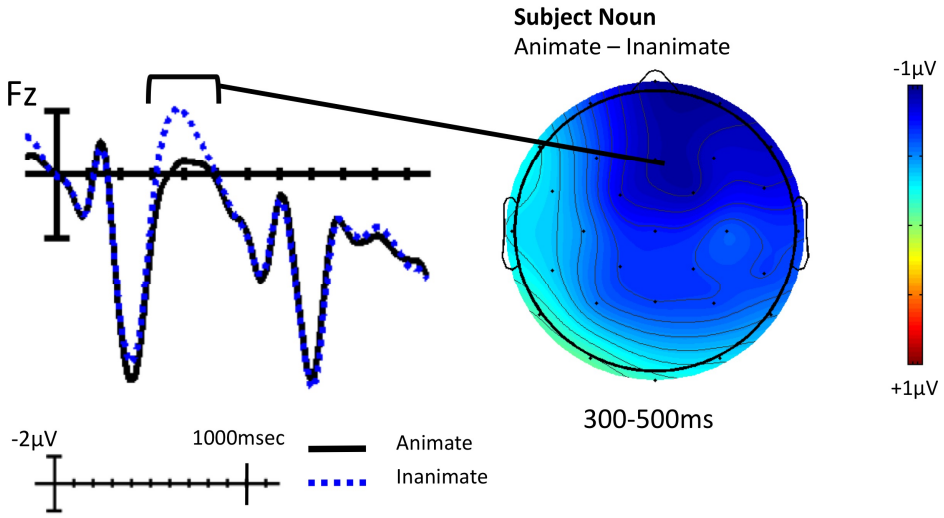

LE
FPz
FP1
FP2
HE
ISO
F7
F3
Fz
F4
F8
FC5
FC1
FC2
FC6
T3
C3
Cz
C4
T4
CP5
CP1
CP2
CP6
A1 (ref)
T5
P3
Pz
P4
T6
A2
O1
O2
Oz
Midline column
Lateral column
Medial column
Peripheral column

Fig. 2


Subject Noun
Animate – Inanimate
Fz
-1μV
+1μV
300-500ms
-2μV
1000msec
Animate
Inanimate

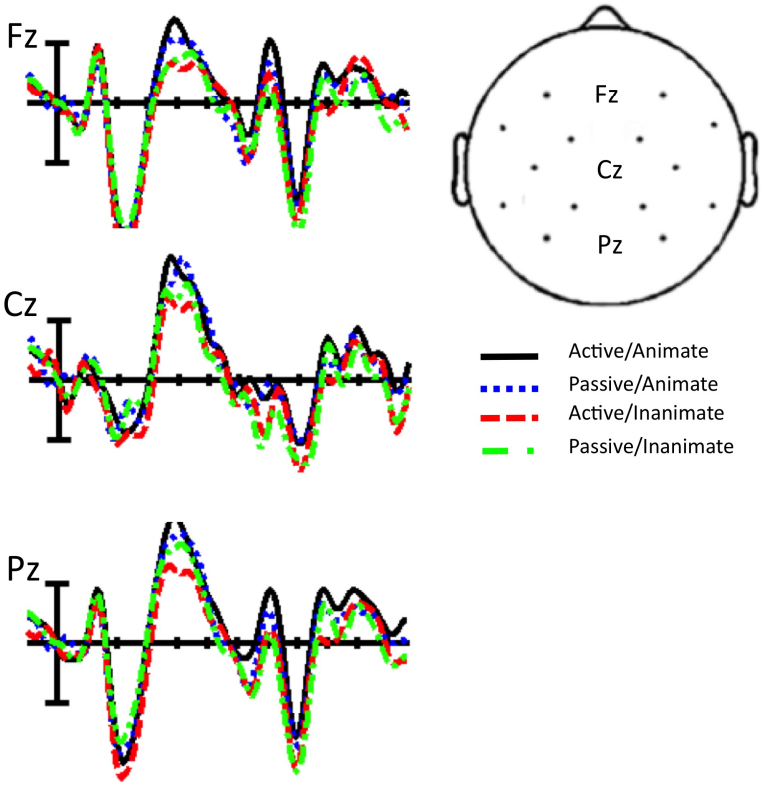

Fz
Cz
Pz
Fz
Cz
Pz
Active/Animate
Passive/Animate
Active/Inanimate
Passive/Inanimate

Fig. 3

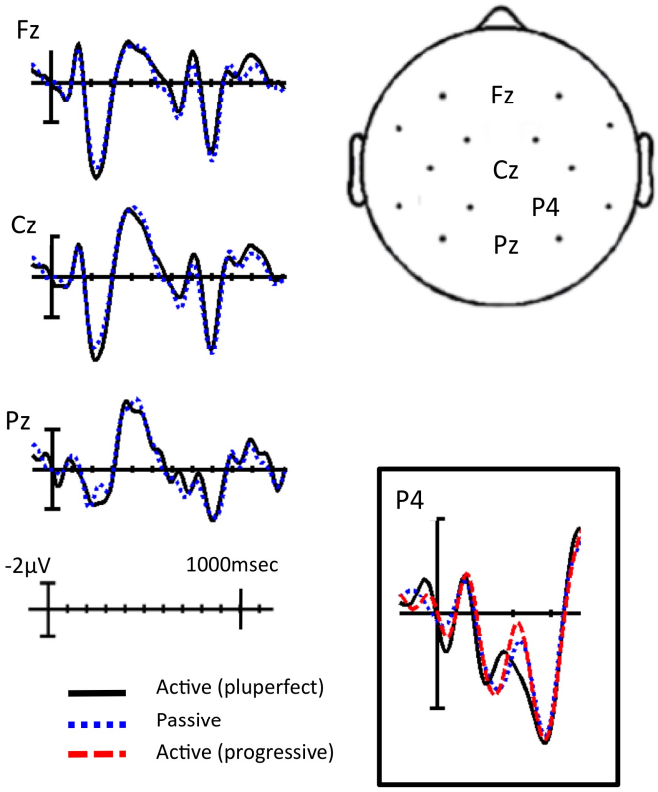

Fz
Cz
Pz
-2μV
1000msec
Active (pluperfect)
Passive
Active (progressive)
Fz
Cz
P4
Pz
P4

Fig. 4

Fig. 5

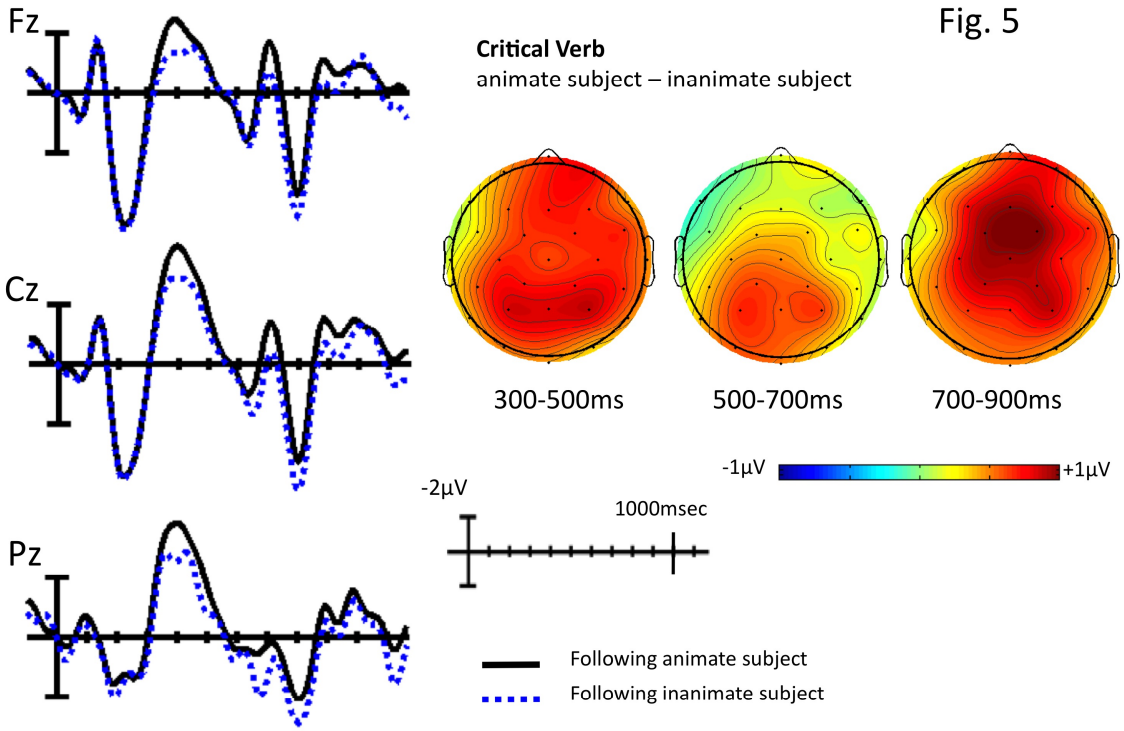
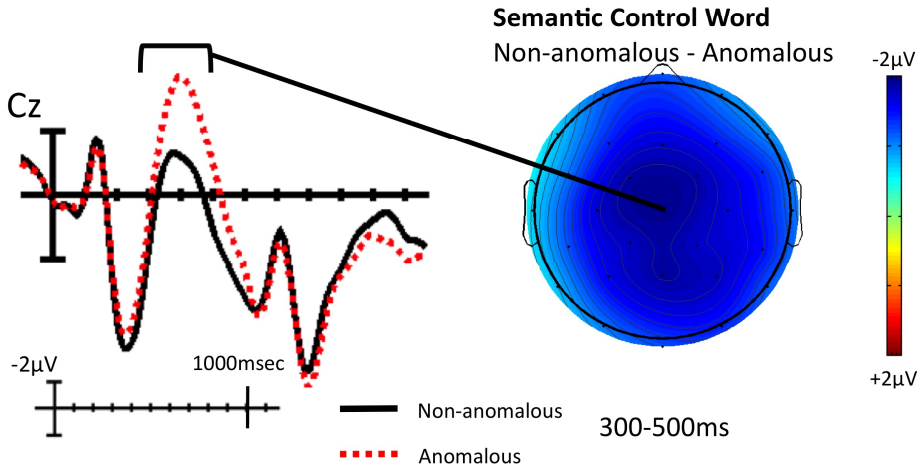

Fz
Cz
Pz
Critical Verb
animate subject – inanimate subject
300-500ms
500-700ms
700-900ms
-1μV
+1μV
-2μV
1000msec
Following animate subject
Following inanimate subject

Fig. 6


Semantic Control Word
Non-anomalous - Anomalous
-2μV
+2μV
Cz
-2μV
1000msec
Non-anomalous
Anomalous
300-500ms

Part II: Online Aspectual Interpretation within Short Narratives

When Events Change Their Nature

Introduction

Upon reading a sentence such as "The cat prowled for several minutes," we intuitively understand that the event constitutes a single action—a cat prowling—which took several minutes to complete. Yet upon encountering the sentence, "The cat pounced for several minutes," we readily understand that the sentence describes multiple pounces, rather than a single pounce lasting several minutes, despite this information not being overtly specified by any of the words in the sentence. As the both *prowl* and *pounce* refer to a singular action, and because the two sentences are syntactically identical, the difference in event representation must lie in compositionality—the combination of semantic, syntactic, and potentially pragmatic, information to construct the overall meaning of an utterance. In this study, we examine the mechanisms and time course of constructing these two different event representations during word-by-word sentence processing, using Event-Related Potentials (ERPs)—a direct measure of online brain activity.

In linguistics, the temporal structure of events described by a verb phrase is referred to as *aspect*. In the current study, we look at verbs within two broad aspectual categories: durative and punctive. Durative verbs, such as *prowl,* are temporally unbounded in that they describe actions without an intrinsic endpoint. In contrast, punctive verbs, such as *pounce*, are temporally bounded, with endpoints inherent in their definition. In the example sentences B and E (see Table 1), the durative adverbial phrase, "For several minutes" defines the length of time over which the event occurs. Because *prowl* is also durative, it readily combines with this durative adverbial phrase. The same, however, is not true for the punctive verb, *pounce*. To accommodate the aspectual constraints of the durative adverbial phrase, *pounce* must be interpreted as referring to a series of pounces, becoming an event that lacks an intrinsic endpoint. Because this interpretation is not motivated by

overt syntactic or morphological markings, such as the progressive *–ing* in English, the event is said to be *aspectually coerced* (Moens & Steedman, 1988).[10] Aspectual coercion, as found in sentence B. (Table 1), has generally been viewed as an operation within the *semantic compositional system* (Jackendoff, 1997; Pustejovsky, 1991; Smith, 1991; de Swart, 1998). Below we provide a brief overview of some proposals that have been put forth to explain the difference in interpretation between A. and B., as well as between B. and E. (Table 1).

According to a first set of proposals, the mismatch between the aspectual information carried by the verb and the temporal modifier is resolved through the engagement of an additional *semantic operator* that resolves this aspectual mismatch. This semantic operator has been termed either a plurality (Jackendoff, 1997) or iterativity (de Swart, 1998) operator, which repeats the punctive action to fill the time specified by the modifier. Importantly, this operator is syntactically unrealized, unlike the progressive construction *was V–ing* in English, which overtly allows for an iterative interpretation of a punctive verb (e.g. *was* pouncing). We shall refer to this proposal as *operator-based iterative coercion.*

In contrast to the above, Dölling (1995; 1997; 2003) proposed that initial semantic composition proceeds normally in sentences such as D (Table 1), with no aspectual mismatch between the durative adverbial phrase and the punctive verb. Nonetheless, the resulting meaning is incongruous with real-world knowledge (e.g a single pounce cannot last several minutes). As a result of this incongruity, the initial interpretation must be revised in order to render the sentence

[10] It should be noted that the term, 'aspectual coercion' is used to refer to any aspectual representation of an event which does not follow from transparent composition. Other examples include *additive coercion*, *subtractive coercion* and *inchoactive coercion*. For example, as discussed by Brennan and Pylkkänen (2010), in the sentence, "Within ten minutes the child was asleep," the action described is not simply that of a child sleeping but also the transition from wakefulness to sleep. For more in depth discussion see Hamm and van Lambalgen (2005) or Croft (2012). For simplicity, in the current manuscript we will primarily use the term *aspectual coercion* to describe the differences in interpretation between A. and B, and between B. and E in Table 1.

coherent (e.g. multiple pounces must have occurred). Following Brennan and Pylkkänen (2008) we will refer to this proposal as *pragmatic iterative coercion*.

Both the *operator-based* and *pragmatic iterative coercion* hypotheses assume that the punctive meaning of verbs like *pouce* is primitive and stored within the lexicon. However, Rothstein (2004) has suggested that it is actually the *iterative* interpretation which is primative (as in E), and that the *punctive* interpretation is derived (as in A). This point can best be illustrated in 1a. and 2b. below, in which both sentences can readily be understood as referring to repetitions of a punctive action:

1a. After several minutes the queen waved to the people.

2b. After several minutes the dog wagged its tail.

According to Rothstein's proposal, additional semantic operations must be engaged in order to arrive at the punctive meaning of *pounced* in A. We will refer to this as *punctive coercion.*

Finally, some authors have proposed that neither the punctive nor the iterative interpretation of punctive verbs like *pounced* is derived (Croft, 2012). Instead, all meanings of the verbs are stored within the lexicon, in this case both punctive and iterative—a type of lexical polysemy (Klepousniotou, 2002). According to these proposals, aspecual coercion is unnecessary, as the meaning of both A. and B. both arise from simple lexical access, requiring no additional operations for compositionality to proceed. We will refer to this as *aspectual polysomy*.

Psycholinguistic exploration of aspectual coercion

To summarize, theoretical proposals that attempt to describe how we arrive at different interpretations of A. vs. B. and B. vs. D. can be divided into three broad groups, each of which makes different predictions about processing. First *iterative coercion,* either *operator-based* or *pragmatic* imply that additional processing is engaged during the interpretation of B. vs. A. The

second proposal, *punctive coercion*, makes the opposite prediction: that the interpretation of A. requires more processing than D. Finally, the *aspectual polysomy* proposal predicts no difference in processing costs for A. vs. B., as in both cases the meaning is primitive, requiring no engagement of additional operators during composition.

Several investigators have begun to test these hypotheses in psycholinguistic experiments. In the first psycholinguistic study of this phenomenon, Piñango and colleagues (Piñango, Zurif, & Jackendoff, 1999) contrasted processing on the durative adverbial modifier, *until*, while participants listened to sentences such as "The insect hopped/glided effortlessly until it reached the garden." The preceding verb was either punctive (*hopped*) or durative (*glided*). At certain points during auditory presentation, letter-string probes appeared on the screen and participants were required to make a lexical decision. Piñango et al. found that participants were slower to respond to probes appearing 250ms after durative adverbs following punctive verbs, than after durative adverbs following durative verbs. They interpreted this slower response time as reflecting the cognitive costs associated the engagement of a plurality/iteration operator within the semantic composition system (Jackendoff, 1997; Pustejovsky, 1991; Smith, 1991; de Swart, 1998). Piñango et al. replicated these findings in a later study (Piñango, Winnick, Ullah, & Zurif, 2006) that varied the timing of lexical decision probe presentation. The authors showed an effect of aspectual coercion only when probes were presented 250ms after participants heard the critical adverb (e.g. *until*), but not when the probe was presented concurrently with the critical word. They therefore suggested that semantic composition does not occur immediately upon syntactic licensing, but rather that it develops more slowly. Additional support for an increase in processing cost associated with *aspectual coercion* came from a functional imaging study by the same group: more activity was seen in several areas of the temporal lobe when participants read sentences like C. than when they read sentences like D.

(Piñango et al, 2006). Finally, Piñango and Zurif (2001) found that patients with Wernicke's aphasia, but not those with Broca's aphasia, had trouble comprehending sentences such as B., which the authors argued gave further support for aspectual coercion involving additional processing within the semantic composition system.

Todorova, Straub, Badecker, & Frank (2000a) found converging evidence for immediate costs associated with aspectual coercion. Participants read sentences, presented segment by segment, such as "Even though / Howard sent / a large check / to his daughter / for many years, / she refused to accept his money", and made acceptability judgments about the evolving sentence fragments. The authors found that participants took longer to make judgments on durative adverbial phrases (*for many years*) than punctive ones (*last year*), but only when the initial verb-phrase described punctive events (*sent a large check*). No such reaction time differences between durative and punctive adverbial phrases were seen when the first event described was durative (*sent large checks*). The authors suggested that this effect was driven either by immediate costs of aspectual coercion, similar to those described by Piñango et al. (1999), or by a reanalysis of the initial aspectual interpretation of the critical verb-phrase. These results were replicated by Husband, Bretta, and Stockall (2006) using a standard reading time paradigm, in which participants read for comprehension rather than making plausibility judgments.

More recently, Brennan and Pylkkänen (2008) examined the online cost of aspectual coercion using both behavioral and neuroimaging (magneto-encephalography, MEG) methods. Unlike the studies discussed above, but following Pickering and colleagues (Pickering, McElree, Frisson, Chen, & Traxler, 2006), considered below, the authors placed the adverbial modifier prior to the critical verb, as in sentences like "For/After 45 seconds the computer beeped in the busy lab," to examine the potential costs of aspectual coercion in the absence of any possible aspectual

reanalysis. In both experiments participants made plausibility judgments following each sentence. In the self-paced reading experiment, the authors reported longer reading times on critical verbs (e.g. “beeped”) following durative adverbial phrases (e.g. “For 45 seconds…”) compared to those following punctive adverbial phrases (e.g. “After 45 seconds…”). Again, these data were interpreted as supporting the *iterative coercion* hypothesis, rather than the *punctive coercion* or the *aspectual polysomy* hypothesis.

In the MEG experiment, Brennan and Pylkkänen (2008) reported an early effect (340-380ms post critical verb onset) which localized to right lateral frontal, anterior temporal and posterior temporal brain regions, as well as a later effect (440-460ms post critical verb onset) which localized to an anterior midline region, on punctive verbs following durative versus punctive adverbial modifiers. The early effects partially overlapped a region previously associated with the N400 (Halgren et al., 2002), an ERP associated with violations of expectations based on real-world knowledge (Hagoort, Hald, Bastiaansen, & Petersson, 2004), while the later effects overlapped with a region previously found to be active in an MEG study of another type of semantic coercion (complement coercion). The authors thus interpreted their MEG results as being most consistent with the *pragmatic iterative coercion* proposal, which, as discussed above, suggests that initial composition yields a meaning which violates real-world knowledge and subsequently must be revised (Dölling, 1995).

Not all studies, however, have found evidence supporting *iterative coercion,* whether operator-based or pragmatic. Pickering and colleagues (2006) used both reading time and eye-tracking methods to explore aspectual coercion using stimuli from Piñango et al. (1999) (Experiments 1 & 2), although with the addition of a second sentence to improve readability. They also examined stimuli from Todorova, Straub, Badecker, & Frank (2000b) (Experiments 3 & 4).

Unlike most of the studies described above (but see Husband, Bretta, & Stockall, 2006), no superimposed task was used: participants simply read the sentences for comprehension. Additionally, in order to test whether Piñango et al.'s initial results were due to an aspectual mismatch between the adverb and the verb, or resulted from a reanalysis of an initial aspectual interpretation, the authors also included modified versions of the stimuli, in which the adverb preceded the critical verb ("**Until** it reached the far end of the garden, the insect glided/hopped effortlessly under the moonlight."). Contrary to the results of Piñango et al. (1999, 2006) and Todorova et al. (2000b), Pickering et al. (2006) failed to find any effect of the experimental manipulation. The authors concluded that, under more natural reading conditions, readers may leave aspectual information underspecified. Importantly, such underspecification is not equivalent to *aspectual polysomy*. While the *aspectual polysomy* hypothesis proposes that all aspectual interpretations of a verb/verb phrase are stored within the lexicon, the *underspecification* hypothesis instead proposes that aspectual information does need to be computed but that this computation can be delayed. Because keeping the raw linguistic information in working memory is costly, Pickering and colleagues argue that when working memory demands are high, as might be the case during a dual task paradigm, comprehenders will immediately compute the full aspectual interpretation of the verb/verb phrase, leading to the costs observed by Tororova et al. (2000a, 2000b) and Pinango et al. (1999, 2006).

A similar null result was reported by Bott (2007) in German using a standard reading time paradigm. Similar to Pickering et al. (2006), Bott (2007) also used sentences in which an adverbial modifier preceded a punctive verb. Participants, however, made plausibility judgments after each sentence. Bott found no significant effects of adverbial phrase (punctive vs. durative), either at the verb or at any subsequent segment, up to and including the sentence-final word. Unlike Pickering et

al. (2006), Bott (2007) argued that the results could not be attributed to underspecification, as he found reading time increases in association with a more complex accomplishment-to-achievement aspectual coercion of aspectual coercion. Bott argued that simple types of aspectual coercions, such as in B. are computationally easy, at least within *Event Calculus* theory (Hamm & van Lambalgen, 2005).

Additional possible mechanisms of aspectual coercion

In sum, all previous studies have failed to find evidence supporting the *punctive coercion* hypothesis, while results have been mixed with regards to support for the *iterative coercion* and *aspectual polysomy* proposals, For example, Pinango et al (1999, 2006a, 2006b), Todorova et al. (2000a, 2000b), Husband et al. (2006) as well as Brennan and Pylkkänen (2008) all found evidence supporting the iterative coercion hypotheses, neither Pickering et al. (2006) nor Bott (2007) found any difference in processing between sentences such as A. vs. B. and D. vs. B. Although neither group used their null results to argue against aspectual coercion, and iterative coercion in particular, their results are compatible with the aspectual polysomy proposal of Croft (2012).

In addition to the mechanisms and interpretations discussed above, there are also some additional thus far unexplored alternative hypotheses regarding the potential costs associated with aspectual coercion. The first possibility is that, rather than arising through coercion per se, previously reported costs associated with aspectual coercion were in fact drive by the iterativity of the action itself. Consider conditions A., B., D., and E. in Table 1. Previous studies have primarily considered pair-wise contrasts, either looking at punctive verbs combined with either punctive or durative adverbial phrases (e.g. A. vs. B., Table 1), or examining differences in how punctive and durative verbs combined with durative adverbial phrases (e.g. D. vs. E., Table 1). The one exception to this thus far has been Todorova et al. (2000a), who examined all four conditions types

of conditions (e.g. A., B., D., & E., Table 1). In these experiments, additional processing costs observed for sentence B. (vs. A. or vs. E.) have been interpreted as evidence for aspectual coercion. However, sentence B (*For several minutes the cat pounced*…) differs from the other three sentences types in one other crucial way: it is the only sentence that describes an event consisting of multiple iterations of an action. This makes it impossible to distinguish effects of aspectual coercion independently of event iterativity per se. Consistent with the hypothesis that iterativity itself can drive processing costs that have previously been attributed to coercion, Husband et al. (2006) found longer reading times on verb-phrases describing multiple actions (e.g. *sent large checks*) than those describing single actions (e.g. *sent a large check*). Similarly, Todorova et al. (2000b) reported longer reading times for frequentative adverbial phrases (e.g. *every year*) compared to punctive adverbial phrases (e.g. *last year*), at least following punctive verb phrases. We term the hypothesis that iterativity itself leads to processing costs, the *uncoerced iterativity* hypothesis.

A second unexamined hypothesis is that previously observed costs of aspectual coercion were driven by a simple mismatch of semantic features: between the temporal adverbial phrase and the critical verb, independent of enriched composition or event iterativity. Again consider Table 1. In sentence A. both the temporal features of adverbial phrase and the verb are punctive and thus match. In contrast, in B. the adverbial phrase is durative and therefore mismatches the punctive feature of the verb. An analogous situation exists in D (*After several minutes the cat prowled*…), where the temporal features of the adverbial phrase and verb once again mismatch: the former is punctive while the latter is durative. Importantly, while the interpretation of B. requires enriched composition, semantic composition in D is completely transparent. This sort of simple semantic feature mismatch could, once again, be the driving factor for previously observed costs associated

with aspectual coercion, in which only sentences like A and B or B and E were contrasted. We term this the *semantic feature mismatch* hypothesis.

Finally, a third hypothesis that has not been explicitly explored (but see Todorova et al., 2000b) is that there is a cost associated with shifts in aspectual event representation. In aspectual coercion, the shift in aspectual representation is due to enriched composition; a punctive event is coerced into an ongoing, temporally non-trivial activity due to combining a punctive verb with a durative adverbial phrase. However, a similar shift in aspectual interpretation occurs when combining a punctive verb with a frequentative adverbial phrase (e.g. *Several times*). Here again, a punctive event shifts representation to describe a temporally non-trivial activity. However, rather than arising from enriched composition, in this case the interpretation is semantically transparent. We shall call this hypothesis the *aspectual shift hypothesis*.

The present study

In the present study, we used event-related potentials (ERPs) to investigate the mechanisms driving the costs associated with aspectual coercion. ERPs are a direct measure of neural activity that can be indexed on the surface of the scalp and that can be divided into several components (of different latencies, morphologies and polarities) that reflect distinct neurocognitive processes.

The ERP components that have been most commonly described in association with language processing are the N400 and P600. The N400 is a negative-going deflection peaking over centro-parietal sites at around 400ms post-stimulus onset, and is generally associated with lexico-semantic processing. Of particular importance to the present study, the amplitude of the N400 is attenuated (less negative) when an incoming word matches (versus mismatches) expected semantic features (Kutas & Federmeier, 1999; Paczynski & Kuperberg, in revision) or when is congruous (versus incongruous) with real-world knowledge (Bicknell, Elman, Hare, McRae, & Kutas, 2010;

Hagoort, et al, 2004; Kuperberg et al., 2006; Paczynski & Kuperberg, in revision). The P600 is a positive-going component peaking over parietal electrode sites between 500 and 700ms post-stimulus onset. It is thought to reflect continued analysis and/or reanalysis of incoming linguistic information and that can, under some circumstances, be triggered by syntactic and/or semantic violations (for review, see: Kuperberg, 2007; van de Meerendonk et al., 2010).

Finally, there are several negativities associated with sentence processing which are distinct from the N400. Although these negativities vary in time of onset, starting as early as 200ms post stimulus onset (King & Kutas, 1995) to as late as 600ms (Baggio, Lambalgen, & Hagoort, 2008), they tend to be sustained, on the order of several hundred milliseconds, and have a somewhat anterior distribution. The exact functional significance of these negativities has yet to be determined, however they have been proposed to index working-memory costs (King & Kutas, 1995). Recent work from our own group (Wittenberg, Paczynski, Weise, Jackendoff, & Kuperberg, 2011) has reported a late, sustained negativity to semantically complex verb argument structures (light verb constructions). Most relevant to the current study, these negativities have been associated with aspectual interpretation (Bott, 2007; Baggio, Lambalgen, & Hagoort, 2008). Bott (2007) reported a sustained anteriorly distributed anterior negativity to a different type of aspectual coercion (additive coercion) while Baggio, Lambalgen, and Hagoort (2008) reported a similar, although more widely distributed, sustained negativity in response to semantically transparent, rather than coerced, aspectual interpretation.

In the present study, we measured ERPs as participants viewed six types of sentences (see Table 1 for examples). As noted above, previous studies of aspectual coercion have usually contrasted only two conditions, manipulating either the aspectual class of the verb or the adverbial phrase (punctive or durative). Only Todorova et al. (2000a) used a 2 x 2 design in which both verb

and adverbial phrase were manipulated, but even this was not sufficient to explore all potential sources of costs associated with aspectual coercion. Here, we used a 3 x 2 design in which we fully crossed adverbial phrase type (punctive, durative, frequentative) with verb type (punctive, durative). Following Pickering et al. (2006), Brennan and Pylkkänen (2008) and Bott (2007), temporal adverbial phrases occurred prior to the critical verb, ensuring that any effects observed would be due to *initial* aspectual interpretation, rather than revision (see Pickering et al, 2006, for more in depth discussion). This design enabled us to distinguish between the previously examined mechanisms of coercion – *operator-based iterative coercion*, *pragmatic iterative coercion*, *punctive coercion* and *aspectual polysomy* -- as well as to test the three unexplored, alternative hypotheses described above: *uncoerced iterativity, semantic feature mismatch,* and *aspectual shift*.

Both types of *iterative coercion* hypotheses predict greater cost associated with punctive verbs following durative adverbial phrases (*For several minutes the cat <u>pounced</u>…*) than following either punctive (*After several minutes.…*) or frequentative (*Several times…*) adverbial phrases. The *pragmatic iterative coercion* hypothesis predicts that the processing cost would manifest as an N400 effect, as the initial semantic interpretation of the sentence would mismatch expectations based on real-world knowledge (e.g. a pounce cannot last several minutes; cf. Hagoort et al., 2004). This N400 effect might then be followed by a P600 effect due to a need for semantic reanalysis arrive at a coherent interpretation of the sentence. The *operator-based iterative coercion* hypothesis predicts that the processing costs would manifest either as a sustained negativity, due to increase working memory demands of engaging an additional semantic operator (King & Kutas, 1995) and/or a shift in aspectual representation Bott (2007) and Baggio, Lambalgen, and Hagoort (2008).

The *punctive coercion* hypothesis predicts costs in the opposite direction: increased costs associated with punctive verbs preceded by punctive adverbial phrases (*After several minutes the*

cat pounced.…), relative to both durative (*For several minutes.*…) and frequentative (*Several times.*…) adverbial phrases. This cost would most likely manifest as a sustained late negativity, similar to the one reported by Baggio et al. (2008), as Rothstein's (2005) punctive coercion proposal suggests that the coercion shifts the aspectual representation of the event.

In contrast to the above, the *uncoerced iterativity* hypothesis predicts that the type of adverbial phrase will modulate ERPs evoked not only in sentences with punctive verbs (e.g. *pounced*), but also in sentences with durative verbs (e.g. *prowled*). Looking at Table 1, one can see that, in three out of the six experimental conditions, the event described involves multiple iterations of an activity: punctive verbs with durative adverbial modifiers (*For several minutes the cat pounced.*…) and both punctive and durative verbs with frequentative modifiers (*Several times the cat pounced/prowled.*…). If event interativty is itself costly, this would suggest that this is due to embodied cognition[11], with the simulation of multiple action requiring an increase in working memory demands over the simulation of a single action. This predict a sustained negativity in sentences in conditions B, C, and F, compared to sentences in condictions A, D, and E.

The *semantic feature mismatch* hypothesis predicts that adverbial phrase type would modulate ERPs evoke in sentences with punctive verbs and those with durative verbs, although this modulation would differ from that predicted by the *uncoerced iterativity* hypothesis. The *semantic feature mismatch* hypothesis predicts that more cost should be incurred by punctive verbs preceded by durative (versus punctive) adverbial phrases, while durative verbs should show the opposite pattern of effects i.e. durative verbs following punctive adverbial phrases incurring more cost than

[11] In making our predictions, we do not differentiate between the strong embodied cognition hypothesis (event simulation is necessary for language comprehension) and the weak embodied cognition hypothesis (language comprehension is augmented by event simulation), see Mahon and Caramazza (2008) for discussion. In both cases, simulation would be predicted to occur, with both hypotheses predicting that the simulation of multiple events would be more computationally costly.

those following durative adverbial phrases. The cost would be expected to manifest as an N400 effect, due to semantic feature mismatch between the adverbial phrase and the verb (cf. Kutas & Federmeier, 2011). It is less clear how the N400 would be modulated by frequentative adverbial phrases; while they necessarily define a non-punctive amount of time, this is not overtly specified, thus it is unclear whether the 'durative' semantic feature would be activated by these phrases or not.

Finally, the *aspectual shift* hypothesis predicts that adverbial phrase will only modulate ERPs evoked in sentences with punctive verbs. In sentences B and C, in contrast to A, the event described is takes time to complete. Thus, the aspectual interpretation of the sentences as a whole must be shifted from punctive to durative. Note that for sentences with durative verbs (D, E & F), the adverbial phrase does not modify the aspectual interpretation of the sentence; in all cases the event remains durative in nature. This hypothesis therefore predicts a shift in aspectual representation of punctive verbs preceded by either durative (*For several minutes*) or frequentative (*Several times*) adverbial phrases, but not for punctive verbs preceded by punctive adverbial phrases (*After several minutes)*. This cost might manifest as a broadly distributed late negativity, similar to that reported by Bott (2007) and Baggio, Lambalgen, and Hagoort (2008). However, as noted above, the durative nature of frequentative adverbial phrases is not overtly specified. Thus, the processing cost may either be incurred later for frequentative adverbial phrases compared to durative ones, or the cost may be smaller, due to temporal smearing if the durative interpretation takes variable time to complete.

Note that the null hypothesis in this experiment is the *lexical polysemy* hypothesis. The reason for this is that this hypothesis suggests that all possible aspectual representations of a verb are stored explicitly in the lexicon, thus precluding the need for shifts in aspect, (re)computations of representations, or semantic feature mismatches.

Finally, in order to examine Pickering et al.'s (2006) hypothesis that, under natural reading conditions, aspectual information remains underspecified, we embedded our critical aspectual manipulation within a three-sentence discourse context. Participants were asked to read for comprehension only, with intermittent comprehension questions throughout the experiment. Our critical manipulation always occurred in the second sentence, thus making it necessary for participants to create links between this information with both the preceding and subsequent sentences. If readers make an immediate commitment to the full aspectual interpretation as soon as it becomes licensed, this would predict effects of our manipulation should be apparent on the critical verb. However, if readers can leave such information underspecified until it become necessary, this would predict effects of our manipulation on the sentence final word, the point at which the final meaning of a sentence is computed. A summary of the predictions made by each of the hypotheses can be found in Table 2.

Methods

Development of Materials

Two-hundred-and-seventy scenarios were initially created, each with six experimental conditions. Each scenario consisted of three sentences. The first sentence provided an introductory context and was the same across experimental conditions (e.g. "Lilly's kitty was always having small adventures."). The second sentence always began with an adverbial phrase, which, depending on the experimental condition, was either 1) punctive, indicating a point in time when an event occurred (e.g. "After several minutes"), 2) durative, indicating a duration of time that encompassed the event (e.g. "For several minutes"), or 3) frequentative (i.e. explicitly iterative), indicating that more than one iteration of the event occurred (e.g. "Several times"). The adverbial phrase was then

followed by a subject noun phrase (e.g. “the cat”), followed by a critical verb which was either 1) punctive, describing a short, near-instantaneous action (e.g. “pounced”), or 2) durative, describing an action that could continue for an indeterminate amount of time (e.g. “prowled”). The critical verb was followed by a short (1-5 word) conclusion that made the sentence plausible (e.g. “on the rubber mouse,” “about the backyard”). Note that this conclusion necessarily differed based on verb type. However, its length was matched between the two verb types. A third sentence was added to provide a plausible continuation to the situation. Again, to maintain plausibility, the third sentence differed for a given scenario depending on the critical verb type, although its length was matched across verb types. The three types of adverbial phrases in the first sentence were crossed with the two types of verbs in the second sentence, resulting in six conditions of each scenario. Examples are given in Table 1.

In order to narrow down the stimulus set to those scenarios in which the combination of adverbial phrases and verbs resulted in the intended final sentence interpretation, a rating study was conducted. Forty undergraduate participants recruited from Tufts University completed the study, which was hosted online through SurveyMonkey.com. Six lists where created using a Latin Square design, such that each list contained one version of each scenario and, across all six lists, each scenario appeared in all six conditions. Sentences in each list were pseudo-randomized to ensure that no more than two consecutive sentences belonged to the same condition. Participants read only the first two sentences of each scenario. Each scenario was accompanied by a question asking how many times an action was performed (e.g. How many times did the cat pounce?). Participants indicated their response by selecting one of four options: 1) once, 2) one long time, 3) 2-10 times, or 4) more than 10 times.

The following exclusion criteria were then used to eliminate scenarios: 1) a punctive verb following a punctive adverbial phrase being rated as occurring more than once, 2) a punctive verb following a durative adverbial phrase being evaluated as occurring either once or for one long time, 3) a durative verb following either a punctive or a durative adverbial phrase being rated as occurring more than once (2-10 times or more than 10 times), and 4) a punctive or a durative verb following a frequentative adverbial phrase being evaluated as occurring either once or for one long time. In total, 57 scenarios were eliminated as one or more participant gave ratings meeting exclusion criteria for one or more versions of a scenario. In order to allow for an even distribution of the remaining scenarios across 6 lists, an additional 3 scenarios were eliminated, yielding a total of 210 scenarios in the final stimulus set.

Six counterbalanced lists were created, using a Latin Square Design, each containing 210 scenarios. Each list contained an equal number of scenarios from each of the six conditions (35 scenarios per condition per list), with each scenario appearing once in each list and, across all 6 lists, each scenario appearing in each of the 6 conditions. Each list was ordered pseudo-randomly to ensure that no more than two consecutive scenarios belonged to the same condition.

Characterization of the final stimulus set

Plausibility ratings

A plausibility rating study of the final stimulus set was carried out by participants who did not take part in the ERP experiment. Participants were recruited through online postings to Amazon Mechanical Turk (www.mturk.com), and were compensated $1.00 if they properly completed the survey, regardless of eligibility. However, only participants who met all of the eligibility criteria used in the ERP study (described below) were included in the analysis. The rating study was again hosted through SurveyMonkey.com. The scenarios were presented up until and including the

critical verb, followed by three periods ("...") to indicate that the sentence continued after the critical verb. Participants were asked to rate these 'story beginnings' for plausibility on a scale of 1 (Least Sense) to 5 (Most Sense). Again, 6 lists were used such that, across all lists, each scenario appeared in all 6 experimental conditions. Additionally, we included 10 catch questions to ensure that participants were attending to the task. In total, 80 people participated; 13 were rejected because they did not meet eligibility criteria or performed poorly on catch questions. Because the randomization procedure for list assignment (based on month of birth) did not provide a uniform distribution of responses across lists, only the first ten participants from each list were included in the analysis.

Plausibility ratings are shown in Table 1. There were main effects of Verb Type ($F_s(1,59)=47.88$, $F_i(1,209)=17.3$; Adverbial Phrase Type: $F_s(2,118)=23.59$, $F_i(2,418)=15.95$; all ps < 0.01). These were due to scenarios with punctive verbs being rated as slightly more plausible than those with durative verbs (3.7 vs. 3.6, respectively) and scenarios with durative adverbial phases being rated as slightly more plausible than those with punctual adverbial phrases, which, in turn, were rated as more plausible than those with frequentative adverbial phrases (3.8 vs. 3.7 vs. 3.6). Additionally there was a significant Verb Type by Adverbial Phrase Type interaction, again, on both items and participants analyses, $F_s(2,118)=8.5$, $F_i(2,418)=11.78$, ps < 0.01. As follow-up one-way ANOVAs examining effects of Adverbial Phrase Type for each Verb Type separately indicated a significant main effect for both types of verbs, pairwise comparisons were carried out. Scenarios with punctive verbs following durative adverbial phrases were rated as more plausible than those in which punctive verbs followed frequentative adverbial phrases, at least on items analysis, $F_i(1,418)=4.39$, $p < 0.05$, see Table 1. Additionally, scenarios with durative verbs following durative adverbial phrases were rated as more plausible than those in which durative verbs followed

punctive adverbial phrases, again only on the items analysis, $F_i(1,418)=4.70$, $p < 0.05$. No other effects reached significance, $Fs < 2.44$, $ps > 0.11$. In all cases, these differences between conditions in plausibility were small (maximum difference less than 0.4 on a 5 point scale) and they did not pattern with any of the predictions. Therefore, no attempts were made to further cut the stimulus set to match fully for plausibility across conditions.

Latent Semantic Analysis

Semantic similarity values (SSVs) were calculated using Latent Semantic Analysis (LSA) (Landauer, 1998; Landauer & Dumais, 1997), available on the internet at http://lsa.colorado.edu, using tasaALL space (1st year college student reading level), comparing the critical verb with its preceding context (i.e. sentence 1, adverbial phrase and subject noun phrase), as well as the sentence-final word of sentence 2 with its preceding context (i.e. sentences 1 and 2). For both types of word, a document-to-document comparison was conducted.

At the critical verb, an overall ANOVA revealed no significant main effect of Adverbial Phrase Type, $F(2,418) < 1$, $p = 0.986$. However, the main effect of Verb Type approached significance, $F(1,209) = 3.56$, $p = 0.06$, as LSA values were slightly higher on punctive than on durative verbs, see Table 3. The interaction between Verb Type and Adverbial Phrase Type was significant, $F(2,418)=4.122$, $p < 0.05$. To determine the source of this interaction, 3-way ANOVAs were run on each verb type separately. The effect of Adverbial Phrase Type approached significance on punctive verbs, $F(2,418) = 2.51$, $p = 0.08$, and was non-significant on durative verbs, $F(2,418)=1.71$, $p = 0.18$, see Table 3. Again, because LSA value differences between conditions were small and did not pattern with predictions made by any hypothesis being tested, the experimental material was not further altered.

For the final word of the second sentence in each scenario, an overall ANOVA revealed no significant main effect of Adverbial Phrase Type, $F(2,418)<1$, $p = 0.61$, nor a significant interaction between Verb Type and Adverbial Phrase Type, $F(2,418) < 1$, $p = 0.952$. There was a significant main effect of Verb Type, $F(1,209) = 7.90$, $p < 0.01$, which was due to sentence-final words following punctive verbs having higher LSA values than those following durative verbs [0.133(0.12) and 0.098(0.09) respectively].

Length and Frequency

Although the punctive verbs were significantly shorter (number of letters: 6.7, SD: 1.4) than the durative verbs (7.0, SD: 1.5), $t(419)=2.12$, $p < 0.05$, this difference was less than half a letter, vs. respectively. Sentence-final words following punctive verbs were significantly longer (number of letters: 6.3, SD: 2.4) than those following durative verbs (5.7, SD: 2.3), $t(419)=2.49$, $p < 0.05$, although again, this difference was only about half a letter.

Finally, we compared the two types of critical verbs, as well as the sentence-final words of the second sentence in each scenario, on frequency, based on the SUBTLEXus (http://expsy.ugent.be/subtlexus/) database of film and television subtitles. The specific frequency measure we used was $LgSUBTL_{WF}$, the log frequency per million words within the corpus. There was no significant difference between the frequency of the punctive (2.43 SD: 0.9) and the durative (2.25, SD: 0.9) verbs, $t(419)<1$, 2.43, nor between the sentence-final words following the punctive (3.04, SD: 2.4) and the durative verbs (3.06, SD: 4.0), $t(419)<1$.

ERP Experiment

Participants

Thirty-seven undergraduate native English speakers from Tufts University were initially recruited. Data from seven participants were subsequently excluded from the analysis: six due to

excessive ocular or muscular artifacts, and one due to recording error, leaving a total of thirty participants (17F/13M), age 18-25 (M=19.1, SD=1.7). All participants had normal or corrected-to-normal vision, were not taking psychoactive medications, had no learning disability, no history of neurological or psychiatric disorders, and had not learned languages other than English before the age of 5. All were right-handed as assessed through a modified version of the Edinburgh handedness inventory (Oldfield, 1971). Written consent was obtained from all subjects before participation according to the established guidelines of Tufts University. They were paid for their participation.

Stimulus presentation

Each participant was randomly assigned to one of the six counterbalanced lists (five usable participants per list). Participants sat in a comfortable chair in a dimly-lit room, separate from the experimenter. Stimuli were presented on a video monitor. Each trial was preceded by the word "READY" and participants pressed a button to initiate the trial. Each trial began with the first sentence of a scenario, which was presented as a whole in the middle of the screen. Once participants finished reading the sentence they pressed a button at which point the sentence was replaced by a fixation-cross displayed in the center of the screen for 500ms followed by 100ms blank screen. The second and third sentences of the scenario were then presented one word at a time (450ms with 100ms blank screen interstimulus interval, except for the final word of sentence 2, which was followed by a 500ms blank screen, and the final word of sentence 3, which was followed by a 750ms blank screen). In 10% of the trials, the blank screen at the end of sentence 3 was followed by a comprehension question about the preceding scenario, which participants answered with a button press. Questions were unrelated to aspectual information. They were randomly distributed throughout the experiment, with half requiring a "yes" response and half requiring a

"no" response. Participants were asked to refrain from blinking or moving during the word-by-word presentation portion of each trial. Stimuli were presented in seven blocks of 30 trials, with short breaks between blocks. Participants viewed five practice trials before the start of the experiment.

Electrophysiological Recording

Twenty-nine tin electrodes were held in place on the scalp by an elastic cap (Electro-Cap International, Inc., Eaton, OH), see Figure 1. Electrodes were also placed below the left eye and at the outer canthus of the right eye to monitor vertical and horizontal eye movements, and on the left and right mastoids. Impedance was kept below 2.5 kΩ for all scalp and mastoid electrode sites and below 10 kΩ for the two eye channels. The EEG signal was amplified by an Isolated Bioelectric Amplifier System Model HandW-32/BA (SA Instrumentation Co., San Diego, CA) with a bandpass of 0.01 to 40 Hz and was continuously sampled at 200 Hz by an analogue-to-digital converter. The stimuli and behavioral responses were simultaneously monitored with a digitizing computer.

Data Analysis

ERPs were formed by averaging artifact-free trials off-line. They were time-locked to the onset of the critical verbs in each sentence as well as the sentence-final words, using a -50 pre-stimulus to +50 post-stimulus onset baseline. ERPs were analyzed in 100ms intervals from 100ms to 1200ms post stimulus onset in order to determine the time course of effects. In order to examine how the modulation of the waveforms varied across the scalp, the scalp was subdivided into regions along its anterior–posterior distribution, at both mid and peripheral sites (each region contained three electrode sites, see Figure 1). Two omnibus repeated-measures ANOVAs, one covering mid regions, designated by A, B, C, D, and E, in Figure 1 and another covering peripheral regions, designated by F, G, H, and I in Figure 1, were conducted in each time window.

In the mid-regions omnibus ANOVA, the within-subject variables were Verb Type (2 levels: punctive, durative), Adverbial Phrase Type (3 levels: punctive, durative, frequentative), as well as Region (5 levels: prefrontal, frontal, central, parietal, occipital), see Figure X. Interactions between Verb Type and Adverbial Phrase Type were followed up by examining the effects of Adverbial Phrase Type at each level of Verb Type.

In the peripheral regions omnibus ANOVA, the within-subjects variables were Verb Type (2 levels: punctive, durative), Adverbial Phrase Type (3 levels: punctive, durative, frequentative), Region (2 levels: Frontal, Parietal) and Hemisphere (2 levels: Left, Right). Interactions were followed up as described above.

The Geisser-Greenhouse correction was used in cases with more than one degree of freedom in the numerator (Greenhouse & Geisser, 1959) to protect against Type 1 error resulting from violations of sphericity assumption. In these cases, we report the original degrees of freedom with the corrected p value. In all analyses, significance level was set at 0.05. Only effects that were significant in omnibus ANOVAs over two consecutive 100ms intervals (e.g. 300-400ms and 400-500ms) were considered for further analysis. Linearly interpolated voltage maps showing differences in ERP scalp distributions within the time windows of interest were produced using EEGLab (MatLab).

Results

Critical Verb

We found no significant Verb Type by Adverbial Phrase Type interactions on either the mid-regions or periperal regions omnibus ANOVAs within the 100-200ms, or 200-300ms time windows, see Table 3.

In the early half of the N400 time window (300-400ms), there was a Verb Type by Adverbial Phrase Type interaction that approached significance in both the mid-regions and peripheral regions ANOVAs, but this failed to persist into the subsequent time-window (400-500ms).

Starting at 500ms and continuing through to 1200ms post-critical verb onset, the mid regions ANOVAs revealed significant Verb Type by Adverbial Phrase interactions that reached or approached significance at each 100ms time window (except between 900-1000ms), see Table 3. No effects involving either factor were seen in the peripheral regions ANOVA, all $Fs < 2.98$, $ps > 0.08$. To follow up the interaction in the mid-regions ANOVA, we carried out ANOVAs examining the effect of Adverbial Phrase Type on the punctive and the durative verbs separately.

There were no effects of Adverbial Phrase Type for durative verbs (all $Fs < 2.18$, $ps > 0.16$). In contrast, the main effect of Adverbial Phrase Type was significant, or approached significance, for punctive verbs in all 100ms time windows from 500ms, except between 900-1000ms, see Table 4. Further follow-up pair-wise ANOVAs showed that punctive verbs following durative adverbial phrases evoked a larger negativity than punctive verbs following punctive adverbial phrases in all time windows examined (except between 900-1000ms), see Table 4. Punctive verbs following durative adverbial phrases also evoked a larger negativity than punctive verbs following frequentative adverbial phrases in the 800-900ms, 1000-1100ms and 1100-1200mec time windows, see Table 4. Although, punctive verbs following frequentative adverbial phrases appeared to evoke a slightly larger negativity than punctive verbs following punctive adverbial phrases, this was not significant in any time window, except between 500-600ms, see Table 4.

There were no main effects of either Verb Type or Adverbial Phrase Type, and no interactions between either or both of these factors and Region, or, in the peripheral regions ANOVA, with Hemisphere, in any time window, all $Fs < 1.7$, $ps > 0.23$.

Final Word of Sentence 2

Within the recording epoch following the final word in sentence two, there were no significant Verb Type by Adverbial Phrase Type interactions, and no other effects involving either of these factors, at either midline or peripheral regions (all $Fs < 1.20$ and $ps > 0.1$).

Discussion

In this study, we used ERPs to distinguish between several hypotheses regarding the online processing cost associated with iterative coercion. We found that punctive verbs in durative contexts evoked a sustained negativity between 500-1200msec, relative to punctive verbs in punctive contexts, and a sustained negativity between 800-1200ms, relative to punctive verbs in frequentative contexts. In contrast, adverbial phrase type did not modulate ERPs evoked by durative verbs. These data demonstrate that iterative coercion is associated with sustained neural processing cost and that this activity may be primarily driven by aspectual coercion itself, rather than more general shifts in aspectual representation. Moreover, they suggest that, while reading for comprehension, participants can make an immediate commitment to full aspectual interpretation, as soon as it becomes syntactically licensed.

In the following discussion we first deal with potential confounds that might have driven these effects. We then discuss the timing of the costs of aspectual coercion and how this relates to previous proposals of aspectual underspecification. We then consider these results inform our understanding of the costs associated with iterative coercion. We wrap up with open questions and conclusions.

Potential Confounds

As noted in the Methods section, there was a significant context by verb type interaction in the plausibility ratings. However, there are several reasons why this cannot account for our ERP results. First, the differences in plausibility were very small. Second, effects of plausibility are usually seen on the N400 component, which has a more posterior-central scalp distribution and is maximal between 300-500ms. In contrast, the ERP effects observed here were widespread, began later, and were more sustained. Third, our plausibility ratings did not pattern with our ERP effects. Specifically, while scenarios with punctive verbs following durative adverbial phrases were rated as the most plausible, they evoked the largest sustained negativity; in addition, despite differences in plausibility between the scenarios with durative verbs, there were no associated ERP effects.

Similarly, while there was a significant verb type by adverbial phrase interaction in semantic similarity values (SSVs) for critical verbs, this too cannot account for the pattern of ERP results. Again, differences in semantic relatedness between words generally modulate N400 amplitude (for review, see Kutas & Federmeier, 2011), rather than the widely distributed later component we saw here. And again, they did not pattern with the ERP results: the effect of adverbial phrase type on SSVs was larger on durative verbs than on punctive verbs, but there was only an ERP effect of adverbial phase type on punctive verbs.

Sustained neural processing costs associated with Iterative Coercion

Our findings show that iterative coercion is associated with sustained neural processing cost from 500-1200ms after an aspectual interpretation becomes available. This is consistent with the hypothesis that iterative coercion is computationally costly. As noted in the Introduction, two proposals have been put forth regarding the source of increased processing in association with iterative coercion: pragmatic (Brennan & Pylkkänen, 2008) and semantic operator-based

(Jackendoff, 1997; de Swart, 1998). A pragmatic-based iterative interpretation seems unlikely. This account proposes that sentences such as "For several minutes the cat pounced…" are initially interpreted as describing a single punctive event lasting a non-trivial amount of time, which is inconsistent with real-world knowledge; the sentence is subsequently reinterpreted as describing several iterations of the event occurring over the specified time period. We know, however, from many previous studies that initial violations of real-world knowledge evoke an N400 effect (Hagoort et al. 2004; Kuperberg et al. 2003; Sanford et al, 2010; Paczynski & Kuperberg, under revision), and reanalysis is often reflected by a P600 (Kuperberg, 2007). As noted above, we found little impact of our manipulation on modulation of either of these components.

Our results are more consistent with the *operator-based iterative coercion* hypothesis. Unlike the pragmatic iterative coercion hypothesis, this account proposes that transparent semantic composition cannot proceed in sentences such as "For several minutes the cat pounced…" due to the mismatch between the aspectual information carried by the verb and that of the adverbial phrase. As a result, the language parser must engage a morphosyntactically unrealized semantic operator (Jackendoff, 1997; de Swart, 1998). We suggest that the sustained, broadly distributed negativity, starting at 500ms post stimulus onset through to the end of the recording epoch, reflected the engagement of this semantic operator, potentially increasing working memory demands.

Importantly, we were able to rule out several alternative hypotheses regarding the factors driving increased processing in association with aspectual coercion. First, consistent with previous behavioral (Piñango et al., 1999) and MEG (Brennan & Pylkkänen, 2008) research, our study was able to rule out the possibility that the iterative meaning of punctual verbs is a primitive and the punctual one is derived – the *punctive coercion* hypothesis (Rothstein, 2004). There were no

differences in neural activity to punctive verbs preceded by punctive adverbial phrases relative to durative and frequentative adverbial phrases.

Second, our results suggest that effects associated with iterative coercion are not simply reducible to the detection of a mismatch between the lexical aspectual (semantic) features of the adverbial phrase and the critical verb – the *semantic feature mismatch* hypothesis. This proposal would have predicted modulation within the N400 time window for both punctive and durative verbs. Third, our results are incompatible with the proposal that iterative coercion is driven by event iterativity more generally—the *uncoerced iterativity* hypothesis. Were this the case, we should have observed an ERP effect to punctive (and durative) verbs in frequentative (versus punctive) contexts, similar to that seen to punctive verbs in durative contexts. Instead, there were no robust differences in ERPs evoked by punctive verbs in frequentative versus durative contexts, and there was no effects of adverbial phrases on durative verbs at all.

We found very limited support for the proposal that the costs of iterative coercion may in part be driven by a more general shift of representation events described by punctual verbs as ongoing activities. This would have predicted that punctive verbs following either durative phrases or frequentative ones would have incurred similar increases in neural processing costs relative to punctive verbs following punctive adverbial phrases. Although ERPs evoked by punctive verbs following frequentative adverbial modifiers did not differ significantly from those following durative modifiers until 800ms post stimulus onset, neither did they differ from those evoked by punctive verbs following punctive adverbial phrases, except within the 500-600msec time window. Thus, while the current results cannot rule out the *aspectual shift hypothesis*, neither do they provide strong support for it.

Timing of Aspectual Interpretation

Our findings of an immediate cost of aspectual coercion on the critical verb, the first place where the full aspectual interpretation is syntactically licensed, contradicts Pickering et al.'s (2006) proposal that aspectual interpretation remains underspecified when task demands are low. Similarly to Pickering et al., our participants only read for comprehension, with only intermittent questions about the content of the short scenarios, none of which were related to full aspectual interpretation. What could account for the difference between our findings and those of Pickering et al. (2006)? We believe that the difference in results is related to differences in study design, specifically in the type of stimuli used. Unlike previous studies, our critical manipulation was embedded in the middle of a longer linguistic context. Thus in order to form a coherent discourse representation, participants needed to integrate the information presented in the second sentence with the critical manipulation not only with the preceding discourse but subsequent discourse as well. We suggest that because discourse information needed to be integrated over multiple sentences, aspectual interpretation was computed immediately in order to reduce cognitive load and/or potential ambiguity of storing multiple event representations simultaneously. While it may be the case that when linguistic input is relatively simple, comprehenders may be able to leave aspectual interpretation underspecified, we would argue that most linguistic input requires more complex manipulation of information and thus believe that our stimluli more closely approximate natural reading contexts.

Open questions

Our study leave open several questions. The first question concerns the exact circumstances under which the costs of iterative coercion are manifested immediately. We have argued that the reason why we observed an immediate cost of aspectual interpretation while Pickering et al. (2006) did not was due to the greater discourse complexity of the stimuli used. We believe that the greater

discourse complexity increased readers cognitive load as they attempt to integrate information over multiple sentences. Presumably increases in cognitive load would increase working memory demands as well, which, Pickering et al. (2006) argued, would increase the likelihood that readers would commit immediately to full aspectual interpretation. However, the experimental paradigms utilized by Pickering et al. (2006) and our own are considerably different. Although participants readily adjust to visual word-by-word presentation, because this method of presenting linguistic input is relatively uncommon we cannot rule out the possibility that task demands may have increased working memory demands sufficiently to push comprehenders towards an immediate commitment to aspectual interpretation. Thus future studies will need to explore this question using stimuli of similar complexity as those utilized in our experiment using alterative methodology.

Another open question is how these ERP findings relate to MEG findings reported by Brennan and Pylkkänen (2008). As discussed in the Introduction, using MEG, Brennan and Pylkkänen (2008) found that some processing cost for iterative coercion at 340-380ms and 440-460ms post stimulus onset. This is earlier than the onset of the negativity effect reported here. One possibility for this discrepancy is that MEG and EEG provide somewhat different information regarding neural activation. Specifically, unlike EEG, the MEG signal is not distorted by the skull and is maximally sensitive to processing within fissures and sulci, thus making it more sensitive to smaller, short-lived effects. These factors may also explain why the MEG effect was more localized than the more widespread ERP effect. Additionally, Brennan and Pylkkänen (2008)'s recording epoch ended at 600ms post stimulus onset, the point at which we saw maximal divergence between conditions. Interestingly, visual inspection of the MEG waveforms suggests a potential divergence between the coerced and uncoerced condition at approximately 550ms post stimulus onset. It will

therefore be important for future studies to directly compare ERP and MEG effects to the same experimental stimuli in the same participants.

The current experiment also raises questions about the functional significance of the sustained negativity that we saw in association with iterative operator-based coercion. As noted in the introduction, some previous ERP studies have reported similar sustained negativities in association with semantic compositional processes. Bott (2007) reported a more anteriorly-distributed sustained negativity between approximately 650-1100ms in response to a different type of aspectual coercion—additive coercion, in which a punctual event is re-represented as encompassing not only the event itself but the actions taken to realize the event, e.g. "Within five minutes the man found his keys," where *found* also indicates a *searching* action. Baggio, Lambalgen, and Hagoort (2008) reported a widely distributed, sustained negativity between 500 and 800ms in response to shifts in aspectual representation which did not require coercion. Although such findings suggest a common processing network for aspectual interpretation, future studies will need to contrast different types of aspectual processing more directly.

Finally, as discussed above, it will be important to further examine the precise timing of when aspectual coercion occurs. We have argued that, when linguistic input is relatively simple, as in Pickering et al's (2006) study, comprehenders can leave aspectual interpretation temporarily underspecified (but see Brennan & Pylkkänen, 2008, Experiment 1). In a richer discourse context, however, we have suggested that aspectual coercion can occur as soon as the relevant aspectual information becomes available. It will be important for future studies to test this hypothesis directly using both ERPs and eye-tracking techniques.

Conclusions

In sum, the results of the current experiment are consistent with previous findings that iterative coercion is associated with neurocognitive costs during online word-by-word processing. Importantly, our data suggest that these costs are driven by the engagement of a morphosyntactically unrealized semantic operator, used by the language parser to arrive at a coherent representation of the linguistically described event. Additionally, our study was able to rule out a number of alternative proposals regarding the processing of interative coercion. Taken together with previous findings, our results add to the growing body of evidence that sustained negativities may be associated with enriched semantic compositional processes that are necessary to reach a fully coherent representations of meaning.

References

Baggio, G., Lambalgen, M. V., & Hagoort, P. (2008). Memory and Language Computing and recomputing discourse models : An ERP study. *Journal of Memory and Language*, *59*, 36-53.

Bott, O. (2007). Temporal Processing and Shifts in Aktionsart - Evidence from ERPs. *Poster presented at the 13th Annual Conference on Architectures and Mechanisms for Language Processing*. Turku, Finland.

Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language, 63*, 489-505.

Brennan, J., & Pylkkänen, L. (2008). Processing events: behavioral and neuromagnetic correlates of Aspectual Coercion. *Brain and language*, *106*, 132-43.

Brennan, J. and Pylkkänen, L. (2010). Processing Psych Verbs: Behavioral and MEG Measures of Two Different Types of Semantic Complexity. *Language and Cognitive Processes*, 25, 77-807.

Chierchia, G. (1984). *Topics in the syntax and semantics of infinitives and gerunds*.

Cooper, R., Winter, A. L., Crow, H. J., & Walter, W. G. (1965). Comparison of subcortical, cortical and scalp activity using chronically indwelling electrodes in man. *Electroencephalography and Clinical Neurophysiology*, *18*, 217-228.

Croft, W. (2012). *Verbs: aspect and causal structure*. Oxford: Oxford University Press.

de Swart, H. (1998). Aspect shift and coercion. *Natural Language & Linguistic Theory*, 16, 347–385.

DeLucchi, M. R., Garoutte, B., & Aird, R. B. (1962). The scalp as an electroencephalographic averager. *Electroencephalography and Clinical Neurophysiology, 14*, 191-196.

Dölling, J. (1995). Ontological domains, semantic sorts and systematic ambiguity. *International Journal of Human–Computer Studies*, 43, 785–807.

Dölling, J. (1997). Semantic Form and Abductive Fixation of Parameters. In R. Der Van Sandt, R. Blutner, & M. Bierwisch (Eds.), *From underspecification to interpretation, Working papers of the institute for logic and linguistics* (p. 113–138). Heidelberg: IBM Deutschland.

Dölling, Johannes (2003). Flexibility in adverbal modification: Reinterpretation and contextual enrichment. In E. Lang, C. Maienborn, & C. Fabricius-Hansen (Eds.), *Modifying adjuncts* (pp. 511–552). Berlin & New York: Mouton de Gruyter.

Elk, M. V., Schie, H. T. V., Zwaan, R. A., & Bekkering, H. (2010). NeuroImage The functional role of motor activation in language processing : Motor cortical oscillations support lexical-semantic retrieval. *NeuroImage*, *50*, 665-677.

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: a review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology*, *61*, 825-50.

Friederici, A. D., Mecklinger, a, Spencer, K. M., Steinhauer, K., & Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: a spatio-temporal analysis of event-related brain potentials. *Brain Research. Cognitive Brain Research*, *11*, 305-23.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Brain Research. Cognitive Brain Research*, *1*, 183-92.

Geisler, C., & Gerstein, G. L. (1961). The surface EEG in relation to its sources. *Electroencephalography and Clinical Neurophysiology*, *13*, 927-934.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences*, 4th ed. (pp. 819-836). Cambridge, MA: MIT Press

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*, 438-41.

Halgren, E., Dhond, R., Christensen, N., Vanpetten, C., Marinkovic, K., Lewine, J., et al. (2002). N400-like Magnetoencephalography Responses Modulated by Semantic Context, Word Frequency, and Lexical Class in Sentences. *NeuroImage, 17*, 1101-1116

Hamm, F. & van Lambalgen, M (2005). The proper treatment of events. Blackwell, Malden, MA.

Husband, E. M., Beretta, A., & Stockall, L. (2006). Aspectual Computation : Evidence for Immediate Commitment. *Talk given at Architectures and Mechanisms of Language Processing 12th Annual Conference.* Nijmegen, The Netherlands.

Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.

King, J. W., & Kutas, M. (1995). Who Did What and When? Using Word- and Clause-Level ERPs to Monitor Working Memory Usage in Reading. *Journal of Cognitive Neuroscience, 7*, 376-395.

Klepousniotou, E. (2002). The Processing of Lexical Ambiguity: Homonymy and Polysemy in the Mental Lexicon. *Brain and Language,* 223, 205-223.

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

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.

Landauer, T. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*, 259-284.

Landauer, T. K., & Dumais, S. T. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition , Induction , and Representation of Knowledge. *Psychological Review*, *104*, 211-240.

Mahon, B.Z., & Caramazza, A. (2008). A critical look at the Embodied Cognition Hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology - Paris*, 102, 59-70.

Moens, M., & Steedman, M. (1988). Temporal Ontology and Temporal Reference. *Computational Linguistics*, *14*, 15-28.

Nieuwland, M. S., & Van Berkum, J. J. (2008). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. *Brain and language*, *106*, 119-31.

Oldfield, R. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, 9, 97–113.

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

Paczynski, M. & Kuperberg, G. (under revision). Impact of Aspectual and Animacy Information on Semantic and Syntactic Processing: Evidence from Event-Related Potentials.

Pickering, M. J., McElree, B., Frisson, S., Chen, L., & Traxler, M. J. (2006). Underspecification and Aspectual Coercion. *Discourse Processes*, *42*, 131-155.

Piñango, M. M., Winnick, A., Ullah, R., & Zurif, E. B. (2006). Time-course of semantic composition: the case of aspectual coercion. *Journal of psycholinguistic research*, *35*, 233-44.

Piñango, M. M., Zurif, E. B., & Jackendoff, R. (1999). Real-time processing implications of enriched composition at the syntax-semantics interface. *Journal of psycholinguistic research*, *28*, 395-414.

Piñango, M. & Zurif, E. (2001). Semantic operations in aphasic comprehen sion: implications for the cortical organization of language. *Brain and Language,* 79, 297–308.

Pustejovsky, J. (1991). The syntax of event structure. *Cognition*, *41*, 47-81.

Rothstein, S. (2004). *Structuring events: A study in the semantics of lexical aspect*. Oxford: Blackwell

Smith, C. S. (1991). *The parameter of aspect.* Dordrecht: Kluwer.

Todorova, M., Straub, K., Badecker, W., & Frank, R. (2000a). Processing correlates of aspectual computation. *Presented to theWorkshop on Events and Paths, ESSLLI XII*. Birmingham, England.

Todorova, M., Straub, Kathy, Badecker, William, & Frank, Robert. (2000b). Aspectual Coercion and the Online Computation of Sentential Aspect. *Proceedings of the Twenty-second annual conference of the cognitive science society* (pp. 3-8). Mahwah, NJ, USA: Lawrence Erlbaum Associates.

van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010). Monitoring in language perception: mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, *22*, 67-82.

Wittenberg, E., Paczynski, M., Wiese, H., Jackendoff, R., & Kuperberg, G. (2011). Light Verbs Don’t Make Light Work. *Annual CUNY Sentence Processing Conference. Stanford University, Stanford, CA.*

Figure Captions

Figure 1. Electrode montage. Analyses of variance were conducted at mid regions (A, B, C, D, & E) and peripheral regions (F, G, H, & I, see Methods).

Figure 2. ERPs evoked by punctive critical verbs at select frontal and central electrode sites. Critical verbs preceded by punctive adverbial phrases are indicated by solid black line. Critical verbs preceded by durative adverbial phrases are indicated by dashed green line. Critical verbs preceded by frequentative adverbial phrases are indicated by dotted blue line.

Figure 3. ERPs evoked by durative critical verbs at select frontal and central electrode sites. Critical verbs preceded by punctive adverbial phrases are indicated by solid black line. Critical verbs preceded by durative adverbial phrases are indicated by dashed green line. Critical verbs preceded by frequentative adverbial phrases are indicated by dotted blue line.

Figure 4. Voltage maps showing average differences in ERPs evoked by punctive critical verbs preceded by durative adverbial phrases relative to those preceded by punctive adverbial phrases in 100ms time-windows from 100ms to 1100ms post critical word onset.

Table 1. Experimental conditions

Condition	**Example**	**LSA at CV** Mean (SD)	**LSA at SFW** Mean (SD)	**Plausibility** Mean (SD)
A. Punctive Adverb, Punctive Verb	After several minutes the cat pounced…	0.058 (0.11)	0.132 (0.11)	3.7(0.06)
B. Durative Adverb, Punctive Verb	For several minutes the cat pounced…	0.064 (0.11)	0.134 (0.13)	3.9 (0.06)
C. Frequentative Adverb (i.e. explicitly iterative), Punctive Verb	Several times the cat pounced…	0.061 (0.11)	0.133 (0.11)	3.6 (0.06)
D. Punctive Adverb, Durative Verb	After several minutes the cat prowled…	0.047 (0.08)	0.099 (0.10)	3.5 (0.06)
E. Durative Adverb, Durative Verb	For several minutes the cat prowled…	0.041 (0.08)	0.098 (0.09)	3.7 (0.06)
F. Frequentative Adverb (i.e. explicitly iterative), Durative Verb	Several times the cat prowled…	0.044 (0.08)	0.098 (0.09)	3.6 (0.05)

LSA : Semantic similarity value based on document-to-document Latent Semantic Analysis (http://lsa.colorado.edu)
CV : Critical Verb (e.g. *pounced*, *prowled*)
SFW: Sentence Final Word following critical verb.
SD : Standard Deviation
Plausibility rating based on a five point scale: 1 = Highly Implausible/Impossible, 5 = Highly Plausible

Table 2. Summary of predictions made by the seven hypotheses regarding costs of aspectual processing.

	N400	**Sustained Negativity**	**P600**
Operator-based iterative coercion	— —	B. > A., C.	— —
Pragmatic iterative coercion	B. > A., C.	— —	?? B. > A., C.
Punctive coercion	— —	A. > B., C.	— —
Uncoerced iterativity	— —	B., C. > A. F. > D. E.	— —
Semantic feature mismatch	B. > A. D. > E.	— —	— —
Aspectual shift	— —	B. => C. > A.	— —
Aspectual polysemy	— —	— —	— —

X. > Y. : Amplitude of specified ERP (N400, Sustain Negativity, or P600) evoked by critical word in condition X. predicted to be greater than in condition Y.
X. => Y. : Amplitude of specified ERP (N400, Sustain Negativity, or P600) evoked by critical word in condition X. predicted to be equal to or greater than in condition Y.
?? : Prediction is uncertain
— — : No modulation of specified ERP is expected between any of the conditions.

Table 3. F values for ANOVAs on Critical Verb.

Time window (ms)	*Mid Regions ANOVA*				*Peripheral Regions ANOVA*		
	V	*Adv*	*V X Adv*		*V*	*Adv*	*V X Adv*
100-200	< 1	1.04	< 1		< 1	< 1	< 1
200-300	< 1	< 1	< 1		< 1	< 1	< 1
300-400	1.03	< 1	**2.85^**		< 1	< 1	**2.75^**
400-500	1.66	< 1	2.35		2.37	< 1	1.81
500-600	< 1	1.16	**4.22***		< 1	< 1	**2.98^**
600-700	< 1	< 1	**4.06***		< 1	< 1	**3.13^**
700-800	< 1	1.55	**2.51^**		2.10	< 1	1.45
800-900	< 1	< 1	**4.43***		< 1	< 1	1.94
900-1000	< 1	< 1	1.73		1.68	< 1	< 1
1000-1100	2.34	< 1	**3.51***		2.65	< 1	**2.60^**
1100-1200	1.26	< 1	**4.00***		1.28	< 1	**2.71^**

V: Verb Type
Adv: Adverbial Phrase Type
Degrees of Freedom:
Verb Type: 1, 29
Adverbial Phrase Type: 2, 58
Verb Type by Adverbial Phrase Type: 2, 58
^ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 4. F values for ANOVAs on Critical Punctive Verbs Only

	Mid Regions			
	Adv	*D v P*	*D v F*	*F v P*
500-600	**5.82****	**12.77****	2.30	**4.77***
600-700	**2.61^**	**4.64***	2.83	1.94
700-800	**3.96***	**4.93***	2.79	3.89
800-900	**4.46***	**9.47***	**12.31****	< 1
900-1000	1.99	— —	— —	— —
1000-1100	**4.62***	**8.14***	**12.39****	< 1
1100-1200	**4.49***	**8.47***	**5.52***	1.89

Adv: Adverbial Phrase Type
P v D: Puntive vs. Durative Adverbial Phrase Type
P v F: Punctive vs. Frequentative Adverbial Phrase Type
D v F: Durative vs. Frequentative Adverbial Phrase Type
Degrees of Freedom:
Adv: 2, 58
P v D, P v F, D v F: 1, 29
— — : No pair-wise comparison done due to lack of main effects of Adverbial Context Type

^ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Fig. 1

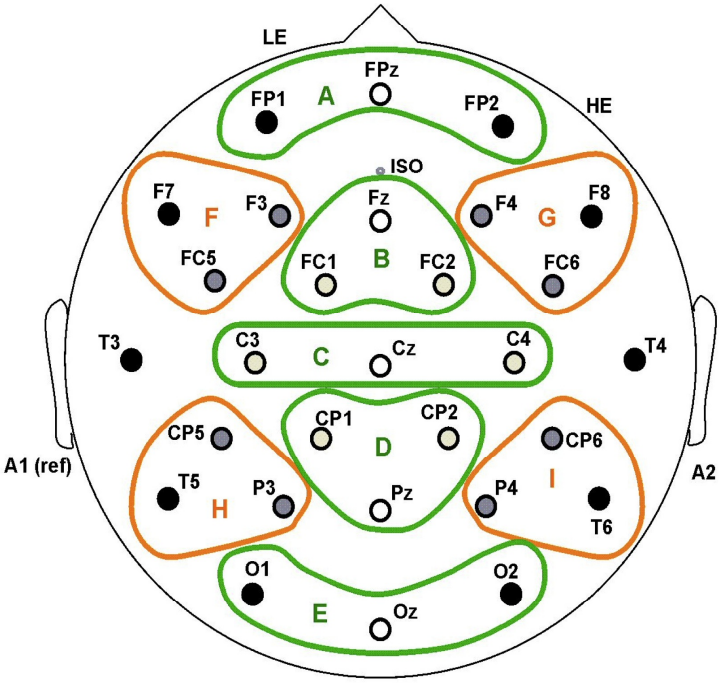

LE
HE
FPz
FP1
FP2
A
ISO
F7
F3
F
Fz
F4
F8
G
FC5
FC1
B
FC2
FC6
T3
C3
C
Cz
C4
T4
CP1
CP2
CP5
CP6
D
A1 (ref)
T5
P3
H
Pz
P4
I
A2
T6
O1
O2
E
Oz

Fig. 2

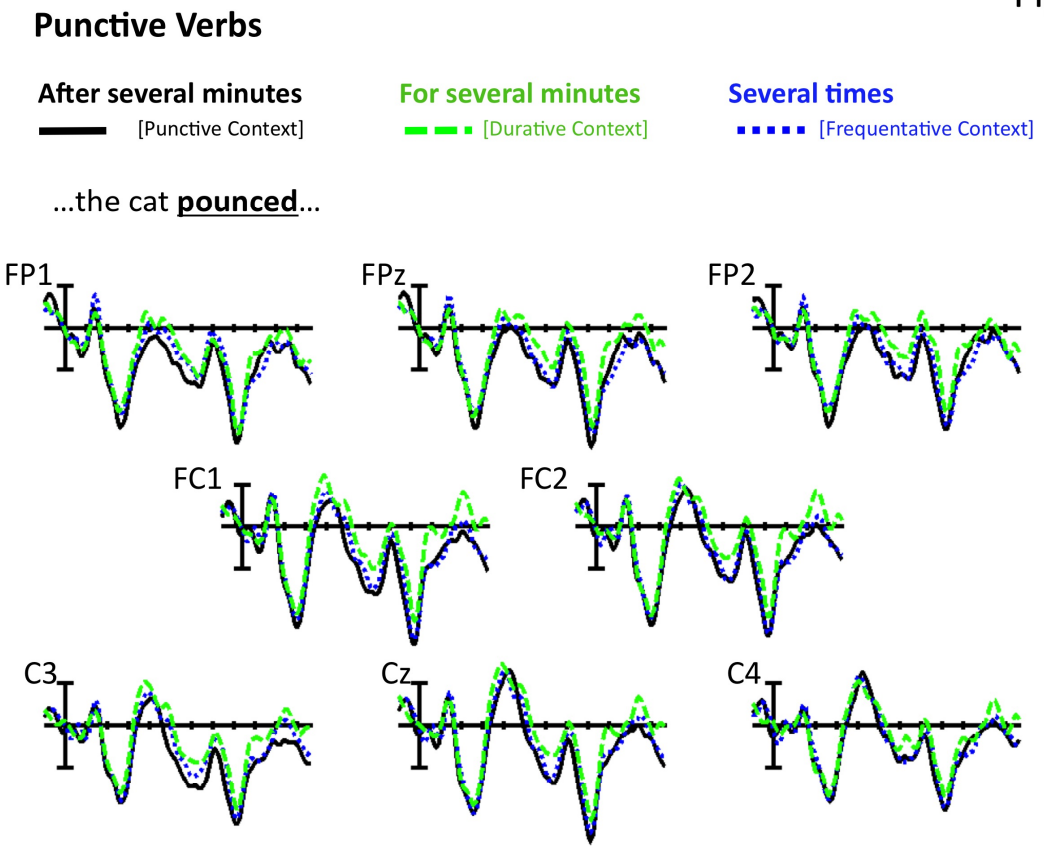

Punctive Verbs
After several minutes
[Punctive Context]
For several minutes
[Durative Context]
Several times
[Frequentative Context]
...the cat pounced...
FP1
FPz
FP2
FC1
FC2
C3
Cz
C4

Fig. 3

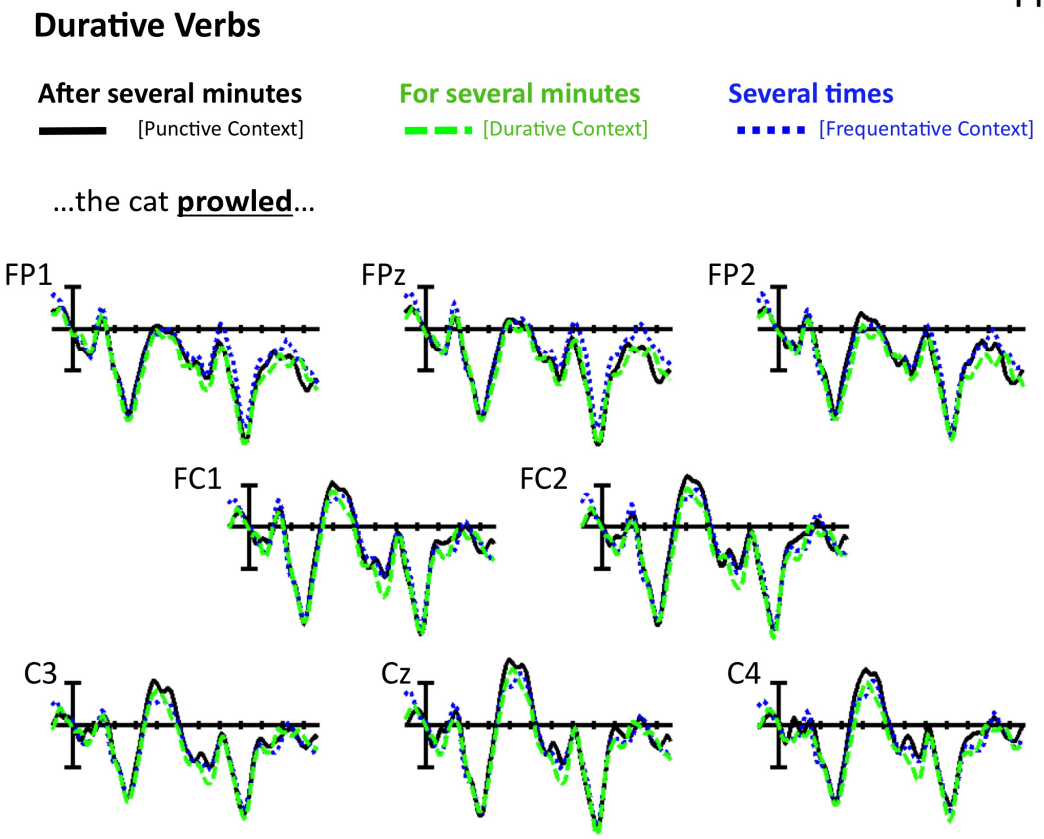
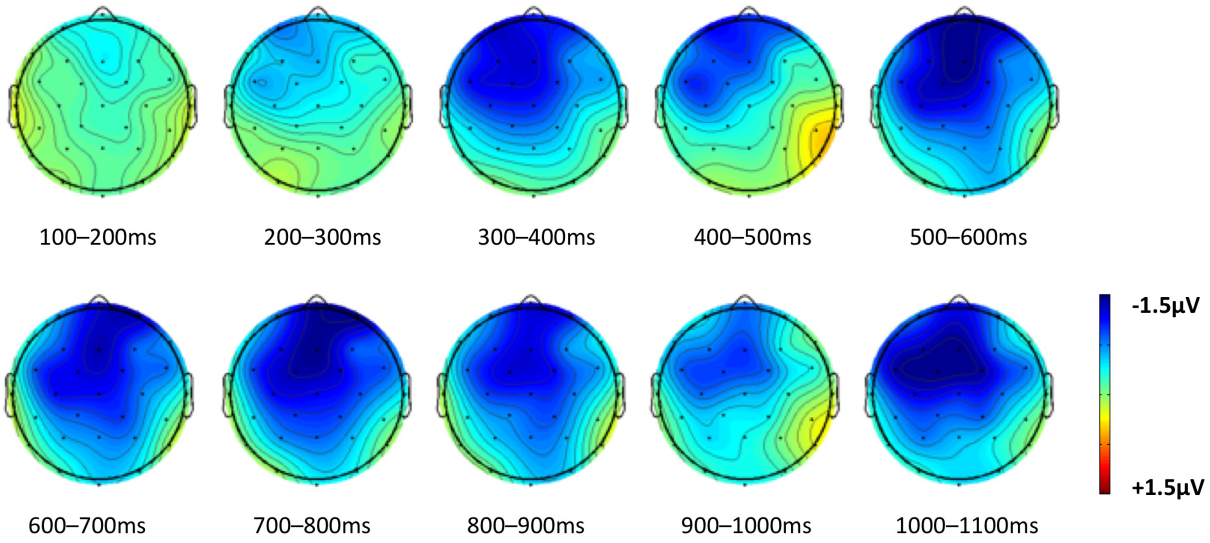

Durative Verbs
After several minutes
[Punctive Context]
For several minutes
[Durative Context]
Several times
[Frequentative Context]
...the cat prowled...
FP1
FPz
FP2
FC1
FC2
C3
Cz
C4

Fig. 4

Punctive Verbs

Punctive Context - Durative Context



Enriched and Transparent Aspectual Interpretation: A Neuropsychological Study

Introduction

Although aspect has been the topic of heated theoretical debates in linguistics for decades, it remains an understudied topic within the realm of neurolinguistics. The aim of the current study is to explore how aspectual information is constructed online under two distinct circumstances: 1) aspectual coercion, in which the full aspectual interpretation requires enriched composition, and 2) construction of a complex aspectual interpretation which is compositionally transparent.

Below we briefly review the theoretical background on lexical and grammatical aspect as well as previous research which examined how these sources of aspectual information are utilized in language comprehension. We then discuss the questions addressed by the current study.

Aspect

Within linguistics two forms of aspect are defined: lexical and grammatical. Vendler (1957) initially described four lexical aspect categories: States, such as *live*, which are ongoing events that do not change unless acted upon by an external force, Activities, such as *run*, that are ongoing with no set endpoint but require continuous application of force, Accomplishments, such as *run a mile*, which are similar to Activities except have an inherent endpoint, and Achievements, such as *die*, which are events that are instantaneous or nearly so, and involve a change of state. A fifth category, Semelfactive, which are like Achievements but do not entail any change of state, such as *blink*, was later added by Comrie (1976). Although these categories have, to some degree, wide spread acceptance within the field, they remain controversial both in terms of definition as well as number (see Croft, 2012, for an extensive recent discussion).

Although, as the name suggests, lexical aspect was initially viewed as being an inherent part of a word within the lexicon, it has become clear that lexical aspect is compositional, being defined

over a verb phrase or, potentially, over an entire utterance. At its most basic, this can be seen for the difference between an activity and an accomplishment. To “write” is an activity. To “write a letter” is an accomplishment. Yet combined with a bare plural, “to write letters” becomes, once again, an activity.

Importantly, lexical aspect appears to be dependent on adverbial phrases that serve as modifiers of the verb phrase. For example, the lexical aspect of the event described by the verb phrase “the light flashed” is semelfactive. Yet, combined with an adverbial modifier delimiting a length of time, it becomes an activity: “the light flashed for several hours.” Note that the statement does not describe an event in which a single ‘flash’ lasted for several hours. Instead, proficient language users will readily understand the statement describes an event consisting of multiple flashes. As this interpretation is not motivated by overt syntactic or morphological markings, such as the progressive *–ing* in English, the event is said to be *aspectually coerced* (Moens & Steedman, 1988). This type of aspectual coercion is termed *iterative coercion* to differentiate it from other types, such as *additive coercion*, in which a preparatory action is implied, though not overtly stated (e.g. “In ten minutes the train arrived at the station” where the ‘approaching’ action is necessary for a coherent interpretation, yet is not specified within the text). The mismatch between the aspectual information carried by the verb phrase and the adverbial modifier has been proposed to be resolved through the engagement of a morphosyntactically unrealized *semantic operator* that resolves this aspectual mismatch, termed either a plurality (Jackendoff, 1997) or iterativity (de Swart, 1998) operator, whose function is to ‘repeat’ the action to fill the time specified.

Like lexical aspect, grammatical aspect also defines the temporal dimension of an event. However, grammatical aspect is sometimes referred to as *viewpoint* aspect as it is used to focus attention on different temporal dimensions of an event. For example, the perfective aspect (e.g.

wrote) describes the event as done, from the point of the narrative, and thus is temporally bound. On the other hand, the imperfective aspect, also known as the progressive in English (e.g. *was writing*), instead places the narrative focus 'inside' the event. From this narrative perspective, the event is not yet completed, independent of whether the event has or has not terminated 'in real life.' For example, a statement "Yesterday the dog was walking through the park" need not, and indeed rarely does, imply that the dog in question is continuing its rounds about the park when the statement is made. Rather, it focuses the reader/listener on the action itself, excluding any terminal stage of an event. Finally, the perfect aspect (e.g. *had written*), like the perfective, describes the event as done. However, it more explicitly emphasizes the outcome/resultant state of an event and its continued relevance to the overall situation being described (Comrie,1976).

Previous Research on Aspect

Aspectual Coercion

Although, as already indicated, several types of aspectual coercion exist, the one which has received most wide-spread attention is *iterative coercion*, in which a punctive event, such as *blink* or *flash* is combined with a durative adverbial modifier, such as *for several hours* or *until dawn*, to yield an interpretation of an event in which multiple iterations of the action are understood to have occurred. Because such interpretation is semantically enriched, requiring the engagement of an additional semantic operator (e.g. Jackendoff, 1997), interpreting such phrases has been proposed to incur additional processing costs.

The first evidence of such cost was demonstrated by Piñango and colleagues (Piñango, Zurif, & Jackendoff, 1999) who contrasted processing on durative adverbial phrases while participants listened to sentences such as "The insect hopped/glided effortlessly until it reached the garden," in which the main verb was either punctive (*hopped*) or durative *activities* (*glided*).

Participants made lexical decisions to visually presented probes when presented 250ms after the auditory presentation of the durative adverbial phrase. Participants were slower to respond to probes following punctive verb, compared to durative verbs, a finding later replicated by Piñango, Winnick, Ullah, and Zurif (2006). The authors interpreted their data as supporting the theory that *iterative coercion* requires additional semantic compositional processing. This conclusion was further supported by the findings that patients with Wernicke's aphasia, but not those with Broca's aphasia, had difficulties with the correct interpretation of sentences requiring iterative coercion (Piñango & Zurif, 2001).

Studies from several other groups have likewise found evidence supporting an online processing cost to *iterative coercion*. These have included behavioral measures, (Todorova, Straub, Badecker, & Frank, 2000a,b; Husband, Bretta, & Stockall, 2006; Brennan & Pylkkänen, 2008, Experiment 1), as well as neurophysiological measures (fMRI, Piñango, Zurif, Palumbo, Gruber, & Yurgeluntodd, 2006; MEG, Brennan & Pylkkänen, 2008, Experiment 2). Most recently an ERP study by our group (Paczynski & Kuperberg, in prep.) reported that *iterative coercion* evokes a sustained, widely distributed negatity, on critical verbs in sentences such as "For several minutes the cat pounced…," in comparison to the processing of their semantically transparent counterparts, "After several minutes the cat pounced…" Importantly, our experiment was able to rule out several alternative proposals put forth regarding *iterative coercion* (see Paczynski & Kuperberg, in prep, for more in depth discussion).

However, not all studies investigting iterative coercion have yeilded positive results. In a extensive study, Pickering, McElree, Frisson, Chen, and Traxler (2006) attempted to find evidence for online processing costs for *iterative coercion* across four experiments utilizing several different simulus sets and experimental paradigms (reading time and eye-tracking). The authors failed to find

evidence for *iterative coercion* impacting processing in any of the experiments. Because they did find evidence for online costs of *complement coercion*, another type of enriched semantic composition, the authors concluded that, at least under conditions where no additional tasks are imposed on participants, readers may leave aspectual interpretation underspecified. They suggest that previously observed costs associated with *iterative coercion* were due to increase working memory demands imposed by additional tasks, which biased participants to make an immediate commitment to the full aspectual interpretation.

Although Pickering et al. (2006)'s proposal is intriguing, it does not readily lend itself to explaining why Husband et al. (2006) observed immediate costs of *iterative coercion* using stimuli similar to those used by Pickering et al. (2006) in which participants likewise only read for comprehension. Nor does the underspecification proposal readily explain why we observed a sustained negativity (Paczynski & Kuperberg, in prep), where, once again, participants' only task was to read for comprehension. We return to this question in the *Current Experient* section below.

Transparent Aspectual Composition

As noted above, lexical aspect is not determined exclusively by the lexical properties of the verb itself, but is achieved compositionally. Thus, the activity 'write' becomes an accomplishment when combined with a singular noun, e.g. 'write a letter.' Importantly, such changes in lexical aspect, more often than not, occur through transparent, rather than enriched composition. A punctive action such as 'flash' can become an activity without recourse to coercion, simply by combining with a frequentative adverbial phrase such as 'several times,' as in "several times the light flashed." Because of its transparent nature, aspectual interpretation without coercion has generally not been studied.

One exception to this is a recent ERP study by Baggio, Lambalgen, & Hagoort (2008). The authors examined sentence final verbs in sentences such as:

1a. Het meisje was een brief aan het schrijven toen haar vriendin koffie op het tafelkleed **morste** (*The girl was writing a letter when her friend **spilled** coffee on the tablecloth.*)

1b. Het meisje was een brief aan het schrijven toen haar vriendin koffie op het papier **morste**. (*The girl was writing a letter when her friend **spilled** coffee on the paper.*)

1c. Het meisje was brieven aan het schrijven toen haar vriendin koffie op het tafelkleed **morste**. (*The girl was writing letters when her friend **spilled** coffee on the tablecloth.*)

1d. Het meisje was brieven aan het schrijven toen haar vriendin koffie op het papier **morste.** (*The girl was writing letters when her friend **spilled** coffee on the paper.*)

In sentences such as 1a and 1c, the act of spilling coffee does not necessarily interrupt the letter writing of the girl. On the other hand, the coffee spilling on the paper, as in 1b and 1d, clearly puts an end to the writing event. Nonetheless, there is fundamental difference in the aspectual representation of the basic event described in 1b vs. 1d. As already described above, 'writing a letter' has a set end-point, and is thus an accomplishment, 'writing letters' on the other hand is an activity, with no defined endpoint. Thus, it is only in 1b that the coffee spilling incident not only terminates the activity, but prevents an implied goal from being accomplished. Thus, the aspectual interpretation of the two events is different. The authors found that while there was no difference in ERPs evoked by the critical verb in the terminating vs. non-terminating context of 1c vs. 1d, the terminating context of 1b, in which a implied goal was not met, evoked a sustained anterior negativity. The authors interpret their finding as suggesting that it is not the termination of the event specified earlier in the sentence that incurs additional neural activation. Rather, the processing load is due to a need to update the event representation. We return to this study in the *Current Study* section below.

Grammatical Aspect

Numerous behavioral studies have examined the effect of grammatical aspect on event representation. Morrow (1985) examined how readers resolved references to ambiguous locations in short discourse. Prior to testing, participants studied the layout of a house. After memorizing the layout, participants read short narratives in which a character moved through this house. Morrow found that grammatical aspect impacted how participants resolved ambiguous referents (e.g. "She walked/was walking past the dining room to the kitchen. She noticed that the light had been left on." Question: "In which room was the light left on?"). Participants were more likely to interpret an ambiguous referent as a Goal (e.g. kitchen) when reading descriptions in perfective aspect (*walked*) and as Sources or Paths (e.g. dining room) following descriptions in the imperfective aspect (*was walking*). Similar Goal over Source preference was also reported by Rohde, Kehler, and Elman (2006) using sentences containing verbs of transfer, such as *hand*. When participants read sentences in perfective aspect (e.g. *John handed a book to Bob*.), they were more likely to interpret an ambiguous pronoun (e.g. *He*) as referring to the Recipient/Goal (e.g. *Bob*) rather than Source (e.g. John) compared to when sentences were in the imperfective aspect (e.g. *John was handing a book to Bob*). These behavioral findings were replicated by Ferretti, Rodhe, Kehler, and Crutchley (2009, experiment 1). Such results have generally been seen as supporting the position that events expressed in the imperfective aspect are interpreted as unfinished, at least from a narrative perspective.

While behavioral studies such as those described above indicate that offline interpretation of events is influence by grammatical aspect, a recent ERP study by Ferretti, Kutas and McRae (2007) demonstrated that grammatical aspect can have an immediate impact during online language processing. The authors compared the processing of typical versus atypical locations for actions that

were presented either in perfect ("The diver had snorkeled in the…") versus imperfective aspect ("The diver was snorkeling in the…"). While typical locations (e.g. *ocean*) evoked an attenuated N400 amplitude compared to atypical locations (e.g. *pond*) following verbs in the imperfective aspect, no such difference in N400 amplitude was noted following verbs in the perfect aspect. Recall that the perfect aspect focuses narrative attention on the outcome of an action, rather than on the action itself. Because the N400 is thought to index the degree of match between the semantic features of an incoming word and its preceding context (cf. Kutas & Federmeier, 2011), the authors interpreted their results as indicating that semantic features of typical location for an action were more activated when the action was considered ongoing rather than completed.

Converging neurophysiological evidence for the immediate use of grammatical aspect information in event representation comes from Ferretti, Rodhe, Kehler, and Crutchley (2009, experiment 2). Following Rodhe et al. (2006), participants read sentences such as "Sue(Source) handed/was handing a timecard to Fred(Goal)" in which the grammatical aspect of a verb of transfer was either in perfective or imperfective aspect. Note that unlike Rodhe et al. (2006), Ferretti et al. (2009) used sentences in which the two characters involved in the event were of different genders. This allowed the second sentence read by participants to start with a prounoun (*he, she*) which either matched or mismatched the expected referent. The authors found an anterior negativity for pronouns which mismatched the expected referent (e.g. "Sue handed a timecard to Fred. ?<u>She</u>…"), compared to those which were expected. Such results thus indicate that readers utilized grammatical aspect information as soon as it becomes available to guide their processing.

<u>Current Study</u>

As already noted, our study aimed at examining online aspectual information processing in two distinct contexts: 1) aspectual coercion, requiring enriched composition, and 2) aspectually

complex yet compositionally transparent constructions. We describe these two aims in more detail below.

Our recent findings (Paczynski & Kuperberg, in prep.) demonstrated that even when participants read only for comprehension, an immediate neural processing cost can be found for *iterative coercion,* as demonstrated by a sustained, widely distributed negativity to critical punctive verbs (*pounced*) preceded by durative adverbial phrases (*For several minutes*) compared to those preceded by punctive adverbial phrases (*After several minutes*). Our first aim was to replicate these findings. Additionally, we sought to expand on them by contrasting constructions in the perfective aspect (*For/After several minutes the cat pounced…*) with those in the imperfective aspect (*For/After several minutes the cat was pouncing…*). As discussed above, only the former construction should require *iterative coercion*, as in the latter construction, event iterativity should be specified by the imperfective, i.e. progressive, form (Moens & Steedman, 1988). Nonetheless, this is a supposition that has not been tested empirically. Grammatical constructions such as the imperfective are generative in nature; given a novel verb such as *wug*, English speakers will readily generate the imperfective form *wugging*. Thus, it is possible that the activation of the base form of a punctive verb (e.g. *pounce*) when preceded by a durative adverbial phrase may itself activate the iterativity/plurality operator, despite such activation being redundant, given the imperfective aspect of the overall verb construction. Alternatively, it is possible that the imperfective construction may itself trigger the activation of an iterativity/plurality operator to allow for the multiple iteration interpretation of punctive verbs (e.g. *was pouncing*).

We move onto the second aim of examining complex, yet compositionally transparent, aspectual constructions. Recall that Baggio et al. (2008) found that verbs that interrupted an event evoked a sustained negativity only when this interruption terminated an action before the

accomplishment of an implied goal (e.g. writing a letter). Although suggestive, the study leaves open several questions. First, the participants' task was to respond to probe questions after each sentence, which focused attention on event completion or lack there of (e.g. "The girl has written a letter"). Thus it is unclear whether participants would make a similar computation without such task demands. Second, the critical manipulation fell on the sentence final word, thus the effects of experimental manipulation occurred during sentence wrap-up processing. Therefore it remains an open question whether a similar pattern of neural activation would be observed when the manipulation is embedded within a sentence.

In order to address these concerns, we utilized a more subtle manipulation of premature termination of an event using the combinatorial nature of aspectual interpretation. Recall that the perfect aspect strongly biases towards an interpretation of the event as not only done but completed. This can best be illustrated by contrasting the sentence "Within an hour the man baked cookies for the sale" with "Within an hour the man had baked cookies for the sale." Compared to the former sentence in perfective aspect, the latter in perfect aspect more unambiguously conveys that by the end of the hour the man had finished baking all of the cookies he intended to bake for the sale. Contrast this with "For an hour the man had baked cookies for the sale." Here the adverbial phrase and grammatical aspect form a more complex predicate. While it is clear that the baking event has terminated, the event lacks the sense of accomplishment of the intended goal. Notice that this sense of premature event termination is lacking in "For an hour the man was baking cookies for the sale."

Thus, our study consisted of two complementary experiments. In the first, we explored the interaction between temporal adverbial modifiers and grammatical aspect of punctive verbs to examine the neural processing related to enriched aspectual interpretation. In the second, we explored the interaction between temporal adverbial modifiers and grammatical aspect of activity

verbs to examine the neural processing related to complex but semantically transparent aspectual interpretation. A summary of all eight experimental conditions can be found in Table 1. In keeping with our earlier study (Paczynski & Kuperberg, in prep.), sentences containing critical manipulations were embedded within a three-sentence discourse with participants reading for comprehension.

Additionally, we decided to determine participants' language proficiency using the TOAL-4 (Hammill, Brown, Larsen, & Wiederholt, 1994). Several studies have demonstrated that language proficiency can influence neural processing in relation to the processing of complex syntactic structures (e.g. King & Kutas, 1995; Weckerly & Kutas, 1999). As no previous work has examined the effects of language proficiency on aspectual interpretation, our analysis was exploratory and we did not include language proficiency in determining our predictions.

Predictions

Based on our recent work (Paczynski & Kuperberg, in prep), we predicted that punctive verbs in perfective aspect preceded by durative adverbial phrases would evoke a sustained negativity relative to punctive verbs in perfective aspect preceded by punctive adverbial phrases. With regards to punctive verbs in the imperfective aspect, we made the following set of predictions. If, as is usually assumed, punctive verbs in the imperfective aspect already have iterative interpretations, we would predict that adverbial phrase type should not modulate ERPs evoked by these verbs. If the iterative interpretation is based on the automatic activation of an iterativity/plurality operator, this would predict that ERPs evoked by punctive verbs in the imperfective aspect would be similar to those evoked by punctive verbs in perfective aspect following durative adverbial phrases; if, on the other hand, the iterative interpretation is lexicalized, we would predict that they more closely match ERPs evoked by punctive verbs in perfective aspect

following punctive adverbial phrases. If punctive verbs in imperfective aspect nonetheless activate the base verb form (e.g. pounce), this might predict that durative adverbial phrases may impact the processing of punctive verbs even in imperfective aspect.

Moving on to predictions for transparent aspectual interpretation, as already noted, the durative adverbial phrases, such as “For an hour,” imply an ongoing activity, while the perfective aspect, “had baked,” implies a completion of an activity. Thus sentences such as “For an hour the man had baked cookies for a sale” contain a conflict between the aspectual implications of the adverbial phrase and the verb phrase. If, as we have proposed, these sentences imply that an event was terminated prior to completion, this would result in an aspectual interpretation similar to the one reported by Baggio et al. (2008) and would thus predict that critical verbs in this condition should evoke a sustained negativity in comparison to the other three conditions (see Table 1), as none of these imply a premature termination of an implied goal. Contrariwise, critical verbs in this construction may evoke an enhanced N400 effect, due to the aspectual feature mismatch between the critical verb and preceding adverbial phrase.

Methods

Development of Materials

Iterative Coercion

Stimuli were based off of those developed for Paczynski and Kuperberg (in prep). Briefly, each scenario consisted of three sentences. The first sentence provided an introductory context (e.g. “Lilly’s kitty was always having small adventures.”). The second sentence, in which the critical manipulation occurred, began with a temporal adverbial modifier, followed by a subject noun phrase, followed by the critical verb. In our previous study, six conditions were developed, of

which two were used in the current study: 1) punctive verbs preceded by punctive adverbial phrases ("After several minutes the cat pounced…") and 2) punctive verbs preceded by durative adverbial phrases ("For several minutes the cat pounced…"). The critical verb was followed by a short (1-5 word) conclusion that made the sentence plausible ("on the rubber mouse"). Finally, a third sentence was added to provide a plausible conclusion. From these two conditions, two new conditions were created in which the critical verb was changed from the perfective aspect ("pounced") to the imperfective aspect ("was pouncing"). See Table 1a for examples of each condition.

The original stimuli, with 210 scenarios, had been normed such that scenarios in the punctive adverbial condition were interpreted as indicating that the action specified by the verb occurred once, while scenarios in the durative adverbial condition were interpreted as indicating that the action occurred multiple times. In order to be able to address our question regarding the neural activation, if any, associated with the iterative interpretation of punctive verbs in the imperfective aspect, it was necessary to use only scenarios in which this was the default interpretation.

We therefore conducted an online rating study, recruiting participants through postings to Amazon Mechanical Turk (www.mturk.com). Participants were compensated $1.00 if they properly completed the survey, regardless of eligibility. However, only native speakers of English between ages 18 and 34 were included in the analysis. The rating study was hosted on SurveyMonkey.com. Four lists where created using a Latin Square design, such that each list contained one version of each scenario and, across all four lists, each scenario appeared in all four conditions. We also included 40 additional scenarios, which were identical to the experimental sentences in structure except utilized durative verbs (e.g. "walk"). These were included as catch trials as, regardless of

adverbial phrase or grammatical aspect, each should be interpreted as describing an action occurring only once or one long time. Sentences in each list were pseudo-randomized to ensure that no more than two consecutive sentences belonged to the same condition. Participants read only the first two sentences of each scenario. Each scenario was accompanied by a question asking how many times an action was performed (e.g. “How many times did the cat pounce?”). Participants indicated their response by selecting one of four options: 1) once, 2) one long time, 3) 2-10 times, or 4) more than 10 times. In total, 60 people participated; 7 were rejected because they did not meet eligibility criteria or performed poorly on catch questions. Because the randomization procedure for list assignment (based on month of birth) did not provide a uniform distribution of responses across lists, only the first twelve participants from each list were included in the analysis. The following exclusion criteria were then used to eliminate scenarios: 1) a punctive verb in perfective aspect following a punctive adverbial phrase being rated as occurring more than once, 2) a punctive verb in perfective aspect following a durative adverbial phrase being evaluated as occurring either once or for one long time, and 3) a punctive verb in imperfective aspect being evaluated as occurring either once or for one long time, regardless of adverbial phrase. A total of 88 scenarios were rejected as one or more participant gave ratings meeting exclusion criteria for one or more versions of a scenario. Of the remaining scenarios, 100 were chosen so as to be matched on plausibility based on a previously conducted rating study (see Paczynski & Kuperberg, in prep).

Transparent Aspectual Interpretation

One hundred forty four scenarios were created using a similar format to those used for the interative coercion stimuli. Again, each scenario consisted of three sentences, in which the first provided an introductory context, the second the experimental manipulation, and the third a plausible continuation. One hundred forty four transitive activity verbs were chosen (e.g. circle) and

combined with direct object noun phrase (e.g. mistakes). Scenarios were designed around each verb-noun combination (e.g. “Wendy took care of proof-reading for one of her friends. Within forty minutes she had circled the mistakes in the manuscript.”). Four versions of each scenario were created crossing temporal adverbial phrase, *within* vs. *for*, and grammatical aspect, perfect (“had circles”) vs. imperfective (“was circling”). Importantly, the time referenced by the adverbial modifier was kept constant across all four conditions and was chosen such that the duration was sufficient to complete the described task. This was done to ensure that differences between conditions could not be attributed to differences in narrative time passed, a factor previously shown to impact semantic processing by our group (Ditman, Holcomb, & Kuperberg, 2008). Examples of all four experimental conditions can be seen in Table 1b.

As the combination of the “for” adverbial phrase and perfect aspect was, by design, aspectually complex, we conducted a naturalness rating study. Once again, we recruited participants through postings to Amazon Mechanical Turk (www.mturk.com). Participants were compensated $1.00 if they completed the survey, regardless of eligibility. However, only native speakers of English between ages 18 and 34 were included in the analysis. The rating study was hosted on SurveyMonkey.com. Four lists where created using a Latin Square design, such that each list contained one version of each scenario and, across all four lists, each scenario appeared in all four conditions. Additionally, sixteen scenarios, using a similar structure as the experimental scenarios, were constructed so as to describe highly semantically anomalous situations (e.g. “The tennis player was shipwrecked about his first grand slam win. Within fifteen minutes he had admired...”). These were used as catch trials to ensure participants were attending to the task. Each list was pseudo-randomized such that no more than two consecutive scenarios belonged to the same condition. The scenarios were presented up until and including the critical verb, followed by three periods (“...”) to

indicate that the sentence continued after the critical verb. Participants were asked to rate these ‘story beginnings’ for naturalness on a scale of 1 (not at all natural) to 5 (very highly natural).

In total, 60 people participated; 6 were rejected because they did not meet eligibility criteria or performed poorly on catch questions. Because the randomization procedure for list assignment (based on month of birth) did not provide a uniform distribution of responses across lists, only the first twelve participants from each list were included in the analysis. We selected one hundred scenarios such that the average naturalness rating was as high as possible while at the same time allowing naturalness to be matched across conditions, all ps > 0.34. See Table 1b for naturalness ratings.

Finally, in order to determine that our critical manipulation was effective, we conducted one final rating study in which participants rated the ‘completeness’ of the event described. As with other rating studies, participants were recruited through postings to Amazon Mechanical Turk (www.mturk.com). Participants were compensated $1.00 if they properly completed the survey, regardless of eligibility. However, only native speakers of English between ages 18 and 34 were included in the analysis. The rating study was hosted on SurveyMonkey.com. Four lists were created, using a Latin Squares design, such that each list contained the same number of items from each condition and across lists, each scenario appeared in each of the four conditions. Each scenario was presented up through the end of sentence two. Participants rated each sentence on a five point scale how much of the intend goal was accomplished, where 1 indicated ‘none/not at all’ and 5 indicated ‘completed/fully accomplished.’ Additionally, four catch questions were included (e.g. ‘Please rate this sentence with a 1’) to ensure participants were attending to the task.

A total of 60 people participated of whom 7 were rejected because they did not meet eligibility criteria or performed poorly on catch questions. Because the randomization procedure for

list assignment (based on month of birth) did not provide a uniform distribution of responses across lists, only the first twelve participants from each list were included in the analysis. Consistent with our intentions, scenarios which combined “within” adverbial phrases with verbs in the perfect aspect were rated as indicating the intended goal as being most completed of all four experimental conditions (Table 1b for means). Most importantly, the difference in completeness rating between the ‘within’ and ‘for’ versions of sentences in the perfective aspect was significantly different, $t(198)=12.75$, $p<0.00001$. This demonstrated that despite the described period of time being sufficient to complete the task described, the use of the “for” rather that “within” adverbial phrase was sufficient to bias readers towards interpreting the event as terminating before completion.

Final Stimulus Set

We created the final stimulus set by combining the two sets of stimuli described above. Yes/No comprehension questions were created for a quarter of scenarios in each of the two lists. Questions referred to information presented in the first, second, or third sentence of each scenario and never focused on the aspectual interpretation of critical second sentence. We created four lists such that across all four lists, each scenario appeared within each of the scenario appropriate conditions, and that within each list, there were the same number of items for all eight conditions. Thus each list contained a total of 200 scenarios, with 25 items per condition per list. Lists were divided into eight blocks of 25 items each, pseudo-randomized to ensure that no more than two consecutive scenarios belonged to the same condition and that each block had approximately the same number of scenerios for from each condition.

ERP Experiment

Participants

Thirty-six undergraduate native English speakers from Tufts University were initially recruited. Data from four participants were subsequently excluded from the analysis due to excessive ocular or muscular artifacts, leaving a total of thirty two participants (17F/15M), age 18-25 (M=20.5, SD=1.3). All participants had normal or corrected-to-normal vision, were not taking psychoactive medications, had no learning disability, no history of neurological or psychiatric disorders, and had not learned languages other than English before the age of 5. All were right-handed as assessed through a modified version of the Edinburgh handedness inventory (Oldfield, 1971). Written consent was obtained from all subjects before participation according to the established guidelines of Tufts University. They were paid for their participation.

All participants were administered the “Word Similarities” and “Sentence Combining” subtests of the Test of Adolescent and Adult Language, Fourth Ed. (TOAL-4, Hammill, Brown, Larsen, & Wiederholt, 2007) as a measure of language proficiency. These subtests were selected as the first served as a measure of lexical knowledge while the second of combinatorial semantic interpretation. The tests were administered using LimeSurvey (Schmitz, 2010) hosted on one of our laboratory servers and took between 30 and 45 minutes to complete. We separated participants based on a median split of the combined score for the two subtests (i.e. sixteen participants per group). The mean score for the ‘high proficiency’ group was 56.7 (SD: 3.1), while the mean score for the ‘low proficiency’ group was 47.4 (SD: 5.1). Differences in group scores were significant, $t(30)=5.86$, $p < 0.0001$.

Stimulus presentation

Each participant was randomly assigned to one of the four lists, such that each list was viewed by eight participants. Participants sat in a comfortable chair in a dimly-lit room, separate from the experimenter. Stimuli were presented on a video monitor. Each trial was preceded by the word “READY” and participants pressed a button to initiate the trial. Each trial began with the first sentence of a scenario, which was presented as a whole in the middle of the screen. Once participants finished reading the sentence they pressed a button at which point the sentence was replaced by a fixation-cross displayed in the center of the screen for 500ms followed by 100ms blank screen. The second and third sentences of the scenario were then presented one word at a time. To more closely parallel natural reading times, words were presented using a variable presentation procedure (see also Nieuwland & Kuperberg, 2008; Nieuwland, Ditman, & Kuperberg, 2010). Word duration in milliseconds was computed as ((number of letters x 27) + 187), up to a 10 letter maximum. Each word was followed by a100ms blank screen interstimulus interval, except for the final word of sentence two, which was followed by a 500ms blank screen, and the final word of sentence 3, which was followed by a 750ms blank screen. Twenty five percent of trials were followed by a comprehension question, which participants answered with a button press. Questions were randomly distributed throughout the experiment, with half requiring a “yes” response and half requiring a “no” response. Participants were asked to refrain from blinking or moving during the word-by-word presentation portion of each trial. Stimuli were presented in eight blocks of 25 trials, with short breaks between blocks. Participants viewed five practice trials before the start of the experiment.

Electrophysiological Recording

Twenty-nine tin electrodes were held in place on the scalp by an elastic cap (Electro-Cap International, Inc., Eaton, OH), see Figure 1. Electrodes were also placed below the left eye and at the outer canthus of the right eye to monitor vertical and horizontal eye movements, and on the left and right mastoids. Impedance was kept below 2.5 kΩ for all scalp and mastoid electrode sites and below 10 kΩ for the two eye channels. The EEG signal was amplified by an Isolated Bioelectric Amplifier System Model HandW-32/BA (SA Instrumentation Co., San Diego, CA) with a bandpass of 0.01 to 40 Hz and was continuously sampled at 200 Hz by an analogue-to-digital converter. The stimuli and behavioral responses were simultaneously monitored with a digitizing computer.

Data Analysis

ERPs were formed by averaging artifact-free trials off-line. The trials were time-locked to the onset of the critical verbs in each sentence, using a -50 pre-stimulus to +50 post-stimulus onset baseline. ERPs were analyzed in the following time intervals: 50-100ms, 100-200ms (N1), 200-300ms (P2), 300-500ms (N400), 500-700ms (early P600), 700-900ms (late P600), and 900-1100ms. In order to examine how the modulation of the waveforms varied across the scalp, the scalp was subdivided into regions along its anterior–posterior distribution, at both mid and peripheral sites (each region contained three electrode sites, see Figure 1). Two omnibus repeated-measures ANOVAs, one covering mid regions (indicated by A, B, C, D, and E in Figure 1) and another covering peripheral regions (indicated by F, G, H and I in Figure 1), were conducted in each time window.

In the mid-regions omnibus ANOVA, the within-subject variables were adverbial phrase (2 levels), grammatical aspect (2 levels), as well as Region (5 levels: prefrontal, frontal, central, parietal, occipital). Interactions between adverbial phrase and grammatical aspect were followed up

by pairwise comparisons between the four conditions. Note that for the *iterative coercion* portion of the experiment, the two levels of adverbial phrase was *punctive* and *durative,* and the two levels of grammatical aspect was *perfective* and *imperfective*; for the *transparent aspectual interpretation* portion of the experiment, the two levels of adverbial phrase were 'for' and 'within', and grammatical aspect was *perfect* and *imperfective.*

In the peripheral regions omnibus ANOVA, the within-subjects variables were were adverbial phrase (2 levels), grammatical aspect (2 levels), Region (2 levels: Frontal, Parietal) and Hemisphere (2 levels: Left, Right). Interactions were followed up as described above.

The Geisser-Greenhouse correction was used in cases with more than one degree of freedom in the numerator (Greenhouse & Geisser, 1959) to protect against Type 1 error resulting from violations of sphericity assumption. In these cases, we report the original degrees of freedom with the corrected p value. In all analyses, significance level was set at 0.05. We first conducted a whole group analysis of our data. We then analyzed each of the two proficiency groups separately. Linearly interpolated voltage maps showing differences in ERP scalp distributions within the time windows of interest were produced using EEGLab (MatLab).

Results

Iterative Coercion

We report any main effects of adverbial phrase and interactions involving adverbial phrase. See Figure 2 and Figure 3 for ERPs and voltage maps, respectively.

Whole Group Analysis

Within the entire recording epoch, no effects of interest reached significance, save a main effect of *adverbial phrase* at peripheral regions analysis in the N400 (300-500ms) which

approached significance, $F(1,31)=3.07$, $p=0.09$. This was due to critical verbs following durative adverbial phrases evoking a slightly more negative N400 than did critical verbs following punctive adverbial phrases.

Low Proficiency Group

Within the N400 time window, there was a significant *adverbial phrase* by *grammatical aspect* interaction at both the mid and peripheral regions analyses, $F > 4.70$, $ps < 0.05$. This was due to a significant modulation of the N400 evoked by critical verbs in the perfective aspect, $Fs > 20.40$, $ps < 0.001$, but not in the imperfective aspect, $Fs < 1$, $ps > 0.93$. This effect was due to critical verbs in the perfective aspect following durative adverbial phrases evoking a greater *positivity* compared to those following punctive adverbial phrases.

Within the 500-700ms time window, an *adverbial phrase* by *grammatical aspect* by *region* interaction approached significance. This was due to a continuation of the effect from the N400 time window at anterior frontal, $F(1,15)=4.87$, $p < 0.05$, and frontal, $F(1,15)=5.07$, $p < 0.05$, mid regions.

High Proficiency Group

In the N400 time window, there was a main effect of *adverbial phrase* at the mid and peripheral region analyses, $F(1,15)=11.5$, $p < 0.01$, $F(1,15)=6.3$, $p<0.05$, respectively, due to critical verbs evoking a greater negativity following durative adverbial phrases versus punctive adverbial phrases. Additionally, the *adverbial phrase* by *grammatical aspect* interaction approached significance at mid region analysis, $F(1,15)=2.88$, $p=0.07$, due to the afore mentioned negativity being significant for critical verbs in the perfective, $F(1,15)=10.3$, $p<0.01$, but not imperfective aspect, $F(1,15)=1.23$, $p=0.29$. There was also an *adverbial phrase* by *grammatical aspect* by *region* by *hemisphere* interaction in the peripheral region analysis. This was due to a significant *adverbial*

phrase by *grammatical aspect* interaction over the right frontal peripheral region. Again, this was due to critical verbs in the perfective aspect evoking a greater negativity following durative adverbial phrases compared to punctive ones, $F(1,15)=7.9$, $p<0.01$, while no such modulation was observed for critical verbs in the imperfective aspect, $F(1,15)<1$, $p=0.99$.

Within the 500-700ms time window, a similar pattern emerged, with the main effect of adverbial phrase being significant, $F(1,15)=8.85$, $p<0.01$, and the *adverbial phrase* by *grammatical aspect* interaction approaching significance at the mid region analysis, due to adverbial phrase significantly modulating critical verbs in the perfective, $F(1,15)=25.40$, $p<0.001$, but not imperfective, $F(1,15)<1$, $p=0.45$, aspect. Additionally, there was a significant *adverbial phrase* by *grammatical aspect* by *region* by *hemisphere* interaction, $F(1,15)=5.13$, $p<0.05$ in the peripheral regions analysis. Again, this was due to a significant *adverbial phrase* by *grammatical aspect* interaction at the front right peripheral region, $F(1,15)=5.43$, $p<0.05$, due to critical verbs in the perfective aspect evoking a greater negativity following durative adverbial phrases versus punctive ones, $F(1,15)=6.83$, $p<0.05$, while adverbial phrase did not modulate ERPs evoked by critical verbs in imperfective aspect, $F(1,15)<1$, $p=0.73$.

Within the 700-900ms time window, there was a marginally significant *adverbial phrase* by *grammatical aspect* by *region* interaction at the mid region analysis, $F(1,15)=2.88$, $p=0.08$. Although this appeared to be due to a continuation of the previously described negativity over anterior electrode regions, no follow-up analyses reached significance, $Fs < 2.72$, $ps > 0.12$. Additionally, there was a significant *adverbial phrase* by *grammatical aspect* by *region* by *hemisphere* interaction at peripheral regions, $F(1,15)=5.05$, $p<0.05$. However, once again, no follow-up analyses yielded significant results, $Fs < 2.27$, $ps > 0.14$.

Within the 900-1100ms time window, there was a significant *adverbial phrase* by *grammatical aspect* by *region* interaction at the peripheral regions analysis, $F(1,15)=4.59$, $p<0.05$. However, no follow-up comparisons were significant, $Fs < 1.69$, $ps > 0.21$.

Transparent Aspectual Interpretation

Only effects involving interactions between the two main factors, *adverbial phrase* and *grammatical aspect,* in the overall ANOVAs at mid and peripheral regions analysis, are reported. See Figure 4 and Figure 5 for ERPs and voltage maps, respectively.

Whole Group Analysis

Visual inspection of the waveform indicated that critical verbs in perfect aspect elicited a greater negativity compared to critical verbs in the other three conditions within the N400 time window that was somewhat right anteriorly distributed. Statistical analysis within the 300-500ms time window confirmed a significant *adverbial phrase* by *grammatical aspect* by *hemisphere* interaction at peripheral regions analysis, $F(1,31)=4.29$, $p<0.05$. Follow-up pairwise comparisons revealed a significant *adverbial phrase* by *hemisphere* interaction for critical verbs in the perfect aspect, $F(1,31)=5.37$, $p<0.05$, and *grammatical aspect* by *hemisphere* interaction for critical verbs following 'For' adverbial phrases. Follow-up comparisons at each hemisphere revealed that while the effect of *adverbial phrase* was significant for critical verbs in perfect aspect over both left, $F(1,15)=7.50$, $p<0.05$, and right, $F(1,15)=8.70$, $p<0.05$, hemispheres, the effect of *grammatical aspect* for critical verbs following 'For' adverbial phrases approached significance at the right hemisphere, $F(1,15)=3.70$, $p=0.07$, but not left hemisphere, $F(1,15)=1.53$, $p=0.24$, suggesting that the effect was slightly right lateralized.

As the *adverbial phrase* by *grammatical aspect* interaction only approached significance at mid region analysis, $F(1,15)=2.48$, $p=0.13$, no follow-up comparisons were carried out.

Low Proficiency Group

Within the 200-300ms time window, the interaction between adverbial phrase and grammatical aspect approached significance at the peripheral regions analysis, $F(1,15)=4.52$, $p=0.05$. However, follow-up comparisons failed to reveal a significant *adverbial phrase* by *grammatical aspect* interaction in either the left or right hemisphere, $F(1,15) = 1.03$, $p=0.32$, $F(1,15) < 1$, $p=0.79$, respectively.

As in whole group analysis, visual inspection once again showed that critical verbs in the perfect aspect eliciting a greater negativity compared to the other three conditions, especially over anterior electrode locations. This impression was confirmed statistically by a significant *adverbial phrase* by *grammatical aspect* by *region* interaction at both mid, $F(4,60)=5.18$, $p<0.05$, and peripheral, $F(1,15)=5.47$, $p<0.05$, region analysis. Follow-up pairwise comparisons indicated that this interaction was due to a significant *adverbial phrase* by *region* interaction for critical verbs in perfect aspect at both mid, $F(4,60)=6.11$, $p<0.01$, and peripheral, $F(1,15)=7.19$, $p<0.05$, analysis. Analyses within individual regions revealed that *adverbial phrase* significantly modulated the ERPs evoked by critical verbs in the perfect aspect within the N400 time-window at: Anterior Frontal Mid, Frontal Mid, Central Mid, and Frontal Peripheral regions, $Fs > 4.83$, $ps < 0.05$.

Although visual inspection suggested an adverbial phrase by grammatical aspect interaction on waveforms elicited in the late time window, statistical analysis failed to reveal any significant *adverbial phrase* by *grammatical aspect* interactions at either mid or peripheral regions analysis, $Fs < 1.96$, $ps > 0.18$.

High Proficiency Group

Within both the 500-700ms and 700-900ms time windows, there was a significant *adverbial phrase* by *grammatical aspect* by *hemisphere* interaction in the peripheral regions analysis, Fs> 6.56, ps < 0.05. Follow-up comparisons indicated a significant *adverbial phrase* by *hemisphere* interaction for critical verbs in perfective aspect in both time-windows, Fs >4.72, ps <0.05. Additionally, there was a *grammatical aspect* by *hemisphere* interaction for critical verbs following "For" adverbial phrases, again in both time windows, Fs>11.62, ps<0.01.

Follow-up comparisons for each hemisphere revealed that the *adverbial phrase* by *grammatical aspect* interaction was significant over the right hemisphere at both time windows, Fs > 4.26, p<0.05, but not over the left, Fs < 1, p>0.33. Although the effect appeared to be due to critical verbs following 'For' adverbial phrases evoking a greater negativity when in perfect aspect imperfective aspect, this effect failed to reach significance at either time-window, Fs < 3.82, ps > 0.07.

Summary of Findings

Iterative Coercion

Whole group analysis failed to reveal significant effects of experimental manipulation. Intriguingly, this was due to a profound difference between the low and high proficiency groups in the processing of critical verbs in the perfective aspect. In the low proficiency group punctive verbs following durative adverbial phrases evoked a greater positivity, compared to those following punctive adverbial phrases, between 300 and 700ms. The effect had a slightly anterior distribution. In contrast to this, the high proficiency group showed the opposite pattern of ERP modulation, with punctive verbs following durative adverbial phrases evoking a sustain, right lateralized negativity

between 300-700ms, compared to punctive verbs following punctive adverbial phrases. Importantly, neither group showed modulation of critical punctive verbs in the imperfective aspect.

Transparent Aspectual Interpretation

Whole group analysis revealed a greater N400 effect, with a somewhat right anterior distribution, for critical verbs following 'For' adverbial phrases compared to the other three conditions. Follow-up comparison revealed that the effect was primarily driven by the low proficiency group, in which the effect was more pronounced and more clearly anteriorly distributed. The high proficiency group, on the other hand, showed only modest effects which did not reach significance in any pair-wise comparisons.

Discussion

In this study, we examined the online processing of aspectual information as participants reach short stories for comprehension. Our two main aims were to further investigate the neural processes associated with *iterative coercion*, a compositionally enriched form of aspectual interpretation and to examine the neural processes associated with the interpretation of complex, yet compositionally transparent, aspectual representations. As a secondary aim, we investigated what role, if any, participants' language proficiency plays in aspectual interpretation. Below, we will first discuss our findings for *iterative coercion*, followed by a discussion of our findings for transparent semantic interpretation. We then discuss the broader implications of our findings.

Iterative Coercion

Our study sheds important light on the question of online processing of iterative coercion. As noted in the introduction, there has been considerable discrepancy in previous studies regarding the processing costs, or lack there of, when participants attempt to parse sentences such as "For

several minutes the cat pounced…" While several groups have found an immediate processing cost in behavioral studies either with a concurrent task (Piñango et al. 1999, 2006; Todorova et al. 2000a,b; Brennan & Pylkkänen, 2008, Experiment 1) or without (Husband et al., 2006), the findings are not universal. In a particularly in-depth look at iterative coercion, Pickering et al. (2006) found no evidence for online processing costs for iterative coercion across four studies, using both reading time and eye-tracking paradigms. The authors suggested that the lack of effects was due to participants leaving aspectual interpretation underspecified under reading conditions in which no additional task was required.

While Pickering et al. (2006)'s interpretation is intriguing, it does not account for why Husband et al. (2006) did find an immediate cost of aspectual coercion in a reading time experiment in which, like Pickering et al. (2006), participants simply read for comprehension. Nor does it explain our recent findings of a sustained negativity in response to iterative coercion (Paczynski & Kuperberg, in prep) in an ERP paradigm where, once again, participants read for comprehension only.

The current results offer a way of consolidating these disparate findings. Similarly to Pickering et al. (2006), on whole group analysis we found likewise found no effect of iterative coercion on processing. However, a separate examination of our high and low proficiency groups revealed a significant difference in ERPs for punctive verbs following durative adverbial phrases (i.e. iterative coercion) and those following punctive adverbial phrases.

Consistent with our previous findings (Paczynski & Kuperberg, in prep), the high proficiency group showed a sustained, broadly distributed negativity for iterative coercion. We have previously argued that the negativity may reflect the additional working memory load incurred when engaging an iterativity/plurality operator. The current findings lend further support to this

hypothesis. Although we examined language proficiency, rather than working memory more specifically, previous research has suggested that the two may be intimately related (cf. Caplan & Waters, 1999), as there is a strong correlation with increase working memory span and ability to parse complex syntactic structures. Previous research has indicated that anterior negativities are more pronounced in high, versus low proficiency comprehenders (King & Kutas, 1999) and this has been interpreted as suggesting that high proficiency comprehenders are able to mobilize working memory resources more effectively than low proficiency comprehenders for the purposes of facilitating language processing.

On the other hand, the low proficiency group showed the opposite effect: a sustained broadly distributed positivity. Importantly, like the high proficiency group, the low proficiency group did not show any difference in ERPs evoked by punctive verbs in the imperfective aspect. Thus, both groups demonstrated a sensitivity to the semantically enriched versus semantically transparent parse of sentences with punctive verbs. What might account for this anterior positivity? Clearly, it is not a matter of event iterativity per se. Punctive verbs in the imperfective aspect, which, in a prerating study, were rated as referring to multiple iterations of an action regardless of adverbial modifier, evoked ERPs similar to those evoked by punctive verbs in punctive contexts, which were rated as referring to a single iteration of an action. Thus, the positivity clearly relates to processing related in some way to the interpretation of enriched composition. Unfortunately, to our knowledge this is the first report of a *positivity* to enriched composition of any sort, thus making interpretation difficult. However, late anterior positivities have previously been related to violations of semantic expectancy (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007). In the current experiment, the majority of verbs following non-punctive adverbial phrases referred to actions which were temporally unbounded, whether due to the use of the imperfective aspect (e.g.

pouncing) or basic lexical properties of the verb (e.g. *bake*). Thus the presentation of a temporally bounded verb, (e.g. *pounced*) in a non-punctive adverbial context may have violated predictions made by low proficiency participants during online comprehension. Such an explanation is at least consistent with our previous findings (Paczynski & Kuperberg, in prep.), in which we found that iterative coercion evoked a sustained negativity even in whole group analysis. Given that the sample pool of participants was the same for both experiments, it seems reasonable to assume a similar distribution of language proficiencies across the two experiments. However, in our previous experiment, adverbial phrases offered much lower predictive value in terms of boundedness or unboundedness of the described event. Thus, one would anticipate that violations of expectancy would be much lower, leading to an attenuation of the sustained positivity in low comprehenders, preventing this positivity from attenuating the sustained negativity elicited by iterative coercion in high proficiency participants in whole group analysis.

No matter what the explanation for the observed group difference is, one might anticipate that such differences in neural activation may manifest in behavioral measures. This would offer a different interpretation of the null results reported by Pickering et al. (2006). Rather than indicating that participants leave aspectual interpretation underspecified, it is possible that the overall group analysis failed to pick up on individual differences in reading speed based on their proficiency. Although speculative, this hypothesis is easily testable in future studies in which participant language proficiency is taken into consideration.

Transparent Aspectual Interpretation

Recall that the perfect aspect is strongly associated with the successful completion of an accomplishment. This is easiest to observe in the contrast between “Within an hour Jane read the article.” vs. “Within an hour Jane had read the article.” While in the former sentence is ambiguous

as to whether or not Jane finished reading the article in the specified amount of time, the latter unambiguously signals that the action is completed. On the other hand, the “For” adverbial phrase is a marker that an action is not done, or, to be more precise, that if the action has some end-point, that end-point has not been reached (contrast: “For an hour the girl ran around the park.” vs. “For an hour the boy wrote a letter.”). Thus, the “For” adverbial phrase conveys aspectual information which is, to some extent, in opposition with the aspectual information conveyed by the perfect aspect.

Note, however, that despite aspectual opposition between the adverbial phrase and grammatical aspect in the following sentence, “For an hour the man had baked cookies for a sale,” the statement nonetheless conveys a coherent message. Specifically, there is an implied termination of the event prior to its ‘successful’ completion, in this case the man not having made all of the cookies he had intended to make.

One of the questions we asked was whether participants would be sensitive to this opposition between adverbial phrase and grammatical aspect. Our whole group analysis indicated that readers may indeed detect this aspectual opposition as indicated by a somewhat right anteriorly distributed N400 effect. When groups were compared separately, it appeared that the effect observed in whole group analysis was driven by low proficiency comprehenders. In this group, critical verbs in perfect aspect following ‘For’ adverbial phrases evoked a robust negativity between 300-500ms compared to all three of the other conditions. On the other hand, no significant effects of experimental manipulation were observed in the high proficiency group. Because the distribution of the negativity observed for the low proficiency group was more anteriorly distributed and right lateralized, we proposed that, rather than an N400 effect, it may instead be a right anterior negativity (RAN).

What might be the functional significance of a RAN in the current context? Because the RAN has most frequently been reported to violations of prosody (e.g. Eckstein & Friederici, 2006), one possibility is that the effect we observed was a byproduct of prosodic errors in participants internal 'speech' as they read. It is possible that low proficiency language users rely more heavily on internal 'speech' during reading than do high proficiency language users. As constructions such as "For forty minutes she had circled…" are less frequent then their imperfective counterparts ("For forty minutes she was circling…"), it is possible that the low proficiency comprehenders were less able than their more proficient counterparts from adjusting their internal articulation to match the (relatively) novel construction.

However, we believe that an alternative interpretation of the results may be more appropriate. In a recent study, Ye and Zhou (2008) reported the presence of a right anterior negativity with a similar time of onset and scalp distribution in response to certain semantic violation in native Chinese speakers. The author's compared the processing of critical verbs in sentences such as 2a,b below:

2a. minjing ba xiaotou juliu zai paichusuo

the policeman BA the thief kept in the police station

2b. xiaotou ba minjing juliu zai paichusuo

the thief BA the policeman kept in the police station

Note: the experiment also included passive version of sentences 2a,b, which differ in construction only in using 'bei' instead of 'ba'. The authors found that the critical verb 'kept' in 2b elicited a right anterior negativity compared to the same critical verb in 2a. Based on similar negativities have previously been reported in the Stroop task (e.g Liotti et al., 2000), during task-switching (Brass et al., 2005), as well as mismatch between the content of a picture and a subsequently presented

sentence (Wassennar & Hagoort, 2007), the authors propose that the negativity may index the suppression or inhibition of interfering sources of information.

In the current study, we specifically designed construction such as “For forty minutes she had circled…” to provide conflicting aspectual information. The perfect aspect suggests a completion of a task or activity, while ‘for’ adverbial phrases denote an ongoing process. Thus, in order to arrive at the appropriate ‘event was terminated prior to completion’ interpretation (see rating study in Methods), participants needed to inhibit the default reading of the perfect aspect as indicating that the event was, in fact, completed.

Note that this interpretation is compatible with the results of Baggio et al. (2008) in which the authors compared sentences such as (critical word underlined in bold):

3a. Het meisje was een brief aan het schrijven toen haar vriendin koffie op het tafelkleed **morste** (*The girl was writing a letter when her friend **spilled** coffee on the tablecloth.*)

3b. Het meisje was een brief aan het schrijven toen haar vriendin koffie op het papier **morste**. (*The girl was writing a letter when her friend **spilled** coffee on the paper.*)

The authors found an anterior negativity for the critical word in 3b vs. 3a, with a similar onset time as that negativity found in the current study. Baggio et al. (2008) suggested that the effect was due to the event model requiring updating, as the critical verb in 3b prevented the accomplishment of *writing a letter* from being completed successfully (the letter could no longer be written). We suggest that this updating may have been driven by the expected event representation needing to be inhibited in order to arrive at the correct, final interpretation of the sentence.

Implications

Our study has two broad implications. First, it underscores the importance of examining individual differences in the study of language processing. Second, it adds to a growing body of

literature suggesting that anterior negativities may index processes related to more complex aspectual interpretation. We discuss each in turn below.

Our findings of both qualitative and quantitative difference in neural activation between high and low proficiency language users are consistent with several previous studies. For example, differences in comprehension proficiency were found to affect the processing of verb arguments as well as the parsing of complex syntactic structures (Weckerly & Kutas, 1999). The authors found that in good comprehenders, matrix verbs elicited a sustained anterior negativity, while in poor comprehenders, the matrix verb elicited a late positivity. Nakano, Saron, and Swaab (2010) found that while violations of animacy based verb selection restrictions (*The mailbox *bit…*) evoke a P600 effect in good comprehenders, the same violation evoked an N400 effect in poor comprehenders. Finally, Malaia, Wilbur, and Weber-Fox (2009) showed that high proficiency comprehenders were significantly more sensitive to the lexical aspect of verbs in the processing of reduced relative clauses compared to lower proficiency comprehenders.

The results of our current study add to these previous findings. First we found that iterative coercion evoked a sustained anterior negativity in our high proficiency participants, but a sustained anterior positivity in our low proficiency participants. In both cases, ERPs evoked by critical verbs in perfective aspect in the uncoerced condition patterned with the ERPs evoked by critical verbs in imperfective aspect, suggesting that in both high and low proficiency participants, the observed effect index processing that was qualitatively different from that involved in transparent aspectual interpretation. Importantly, whole group analysis did not reveal any effects of our experimental manipulation. Thus, without looking at individual differences, our results would have led to the incorrect conclusion that enriched compositional processing is not required for the proper interpretation of iterative coercion.

Furthermore, while we did find significant modulation of ERPs in response to the parsing of semantically transparent, yet complex, aspectual constructions on whole group analysis, our results clearly indicate that the two groups did not processes these constructions in an identical fashion. Specifically, our whole group effect was primarily driven by the low proficiency group, with the high proficiency group shown only a trend in a similar direction. This suggests that even when high and low proficiency comprehenders engage similar underlying neural processes, the degree to which these processes are engaged nonetheless differs under some circumstances.

Our study also sheds new light on the functional significance of negativities previously observed in response to aspectual. In the current study, we found that the computation of iterative coercion, at least in high proficiency participants, was associated with a sustained negativity, lasting several hundred milliseconds. This is consistent with our recent work (Paczynski & Kuperberg, in prep) were we also found a wide spread negativity in response to iterative coercion that lasted for several hundred milliseconds. It is also consistent with a study by Bott (2007) who reported a sustained anterior negativity in response to a different kind of aspectual coercion known as 'additive' coercion. Additive coercion is so named as it 'adds' a preparatory event that is not overtly specified in the discourse. For example, in the sentence "In five minutes the man found his keys" a 'searching' event is implied, though not mentioned.

On the other hand, we found a much shorter, though also anteriorly distributed negativity to complex transparent aspectual interpretation. As noted above, this result is similar to that of Baggio et al. (2008) who reported a negativity in response to another instance of complex, yet transparent, aspectual interpretation. Although the negativity reported by Baggio et al. (2008) was of slightly longer duration than that observed in the current study, it was nonetheless much shorter lasting than

the negativities we report here to iterative coercion as well as previous findings by both our group (Paczynski & Kuperberg, in prep) and others (Bott, 2007).

We suggest that what all complex aspectual interpretations have in common is a need to inhibit or suppress a 'default' aspectual interpretation. We suggest that such inhibitory processing lasts approximately 200ms, although potentially taking longer depending on the degree to which the information must be suppressed and/or how much information requires suppression. Thus, in cases of complex, yet transparent, aspectual interpretation, such as in the current study or the study conducted by Baggio et al. (2008), no additional processing is required, hence the associated negativity only lasts on the order of 200-300ms. On the other hand, in the case of enriched composition, as is the case for iterative or additive coercion, additional information must be generated in order to construct the appropriate overall aspectual interpretation. We suggest that this type of processing requires additional time, resulting in the sustained negativities observed in the current study, as well as in Paczynski and Kuperberg (in prep) and Bott (2007).

Conclusions

Our study indicates that the computation of aspectual interpretation is not a unitary processes but instead relies on several distinct processes. Furthermore, our findings suggest that comprehenders with different levels of language proficiency may rely on different processing strategies to achieve the same aim of correctly understanding linguistic input. Our study underscores the importance of taking such individual differences into account when attempting understand human language processing.

References

Baggio, G., Choma, T., van Lambalgen, M., & Hagoort, P. (2010). Coercion and compositionality. *Journal of Cognitive Neuroscience*, 22(9), 2131-40.

Baggio, G., Lambalgen, M. V., & Hagoort, P. (2008). Computing and recomputing discourse models□: An ERP study. *Journal of Memory and Language*, *59*, 36-53.

Bott, O. (2007). Temporal Processing and Shifts in Aktionsart - Evidence from ERPs. *Poster presented at the 13th Annual Conference on Architectures and Mechanisms for Language Processing*. Turku, Finland.

Brass, M., Ullsperger, M., Knoesche, T.R., von Cramon, D.Y., & Phillips, N.A. (2005). Who comes first? The role of the prefrontal and parietal cortex in cognitive control. *Journal of Cognitive Neuroscience, 17*,1367–1375.

Brennan, J. R., & Pylkkänen, L. (2008). Processing events: behavioral and neuromagnetic correlates of Aspectual Coercion. *Brain and language*, *106*, 132-43.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*, 77-94.

Comrie, B. (1976) *Aspect*. Cambridge: Cambridge University Press.

Croft, W. (2012) *Verbs: aspect and causal structure.* Oxford: Oxford University Press.

de Swart, H., 1998. Aspect shift and coercion. *Natural Language and Linguistic Theory, 16*, 347–385.

Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2008). Time travel through language: Temporal shifts rapidly decrease information accessibility during reading. *Psychonomic Bulletin & Review*, *15*, 750-756.

Eckstein, K., & Friederici, A.D. (2006). It's Early: Event-related potential evidence for initial interaction of syntax and prosody in speech comprehension. *Journal of Cognitive Neuroscience*. 18, 1696-1711.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain research, 1146*, 75-84.

Ferretti, T. R., Kutas, M., & Mcrae, K. (2007). Verb Aspect and the Activation of Event Knowledge. *Cognition, 33*, 182-196.

Ferretti, T. R., Rohde, H., Kehler, A., & Crutchley, M. (2009). Verb aspect, event structure, and coreferential processing. *Journal of Memory and Language*, 61, 191-205.

Hammill, D. D., Brown, S. C., Larsen & Wiederholt, J. L. (2007). *Test of adolescent and adult language*. Austin, Texas: Pro-Ed.

Husband, E. M., Beretta, A., & Stockall, L. (2006). Aspectual Computation : Evidence for Immediate Commitment. *Talk given at Architectures and Mechanisms of Language Processing 12th Annual Conference* (pp. 2000-2000). Nijmegen, The Netherlands.

Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.

King, J. W., & Kutas, M. (1995). Who Did What and When? Using Word- and Clause-Level ERPs to Monitor Working Memory Usage in Reading. *Journal of Cognitive Neuroscience, 7*, 376-395.

Kuperberg, G. R., Choi, A., Cohn, N., Paczynski, M., & Jackendoff, R. (2010). Electrophysiological correlates of complement coercion. *Journal of Cognitive Neuroscience*, 22, 2685-701.

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.

Liotti, M., Woldorff, M.G., Perez III, R., & Mayberg, H.S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia, 38,* 701–711.

Malaia, E., Wilbur, R. B., & Weber-Fox, C. (2009). ERP evidence for telicity effects on syntactic processing in garden-path sentences. *Brain and Language*, *108*, 145-58.

Moens, M., & Steedman, M. (1988). Temporal Ontology and Temporal Reference. *Computational Linguistics*, *14*, 15-28.

Morrow, D.G. (1985). Prepositions and verb aspect in narrative understanding. Journal of Memory & Language, *24*, 390-404.

Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of cognitive neuroscience*, *22*, 2886-98.

Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science, 19,* 1213–1218

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, *63*, 324-346.

Oldfield, R. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia,* 9, 97–113.

Paczynski, M., & Kuperberg, G. R. (in prep). When events change their nature.

Pickering, M. J., McElree, B., Frisson, S., Chen, L., & Traxler, M. J. (2006). Underspecification and Aspectual Coercion. *Discourse Processes, 42,* 131-155

Piñango, M. M., Winnick, A., Ullah, R., & Zurif, E. B. (2006). Time-course of semantic composition: the case of aspectual coercion. *Journal of psycholinguistic research, 35,* 233-44.

Piñango, M. M., Zurif, E. B., & Jackendoff, R. (1999). Real-time processing implications of enriched composition at the syntax-semantics interface. *Journal of Psycholinguistic Research, 28,* 395-414.

Piñango, M. M., Zurif, E., Palumbo, C., Gruber, S., & Yurgeluntodd, D. (2006). Cortical focus of semantic composition: An fMRI study. *Brain and Language, 99,* 183-184.

Piñango, M. M., Zurif, 2001. Semantic Operations in Aphasic Comprehension: Implications for the Cortical Organization of Language. *Brain and Language,* 79, 297–308.

Rohde, H., Kehler, A., Elman, J. (2006). Aspectual Effects on Pronoun Interpretation. *Poster presented at the 19nd Annual CUNY Sentence Processing Conference*. New York, NY

Schmitz, C. (2010). Limesurvey [computer software]. http://www.limesurvey.org

Todorova, M., Straub, K., Badecker, W., & Frank, R. (2000a). Processing correlates of aspectual computation. *Presented to theWorkshop on Events and Paths, ESSLLI XII*. Birmingham, England.

Todorova, M., Straub, Kathy, Badecker, William, & Frank, Robert. (2000b). Aspectual Coercion and the Online Computation of Sentential Aspect. *Proceedings of the Twenty-second annual conference of the cognitive science society* (pp. 3-8). Mahwah, NJ, USA: Lawrence Erlbaum Associates.

Vendler, Z. (1957). Verbs and times. *The Philosophical Review, 66,* 143–160

Weckerly, J., & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology*, 36, 559 – 570

Ye, Z., & Zhou, X. (2008). Involvement of cognitive control in sentence comprehension: Evidence from ERPs. *Brain Research*, *1203*, 103–115.

Figure Captions

Figure 1. Electrode montage. Analyses of variance were conducted at mid regions (A, B, C, D, & E) and peripheral regions (F, G, H, & I, see Methods).

Figure 2. ERPs evoked by critical verbs in the Iterative Coercion portion of the experiment at select frontal and central electrode sites for whole group, high proficiency group, and low proficiency group. Critical verbs in perfective aspect preceded by punctive adverbial phrases are indicated by solid black line. Critical verbs in imperfective aspect preceded by punctive adverbial phrases are indicated by dashed blue line. Critical verbs in perfective aspect preceded by durative adverbial phrases are indicated by dotted red line. Critical verbs in imperfective aspect preceded by durative adverbial phrases are indicated by dashed-dotted green line.

Figure 3. Voltage maps showing average differences in ERPs evoked by critical verbs in perfective aspect preceded by durative adverbial phrases relative to those preceded by punctive adverbial phrases in the 300-500ms, 500-700ms, and 700-900ms time-windows.

Figure 4. ERPs evoked by critical verbs in the Transparent Aspectual Composition portion of the experiment at select frontal and central electrode sites for whole group, high proficiency group, and low proficiency group. Critical verbs in perfect aspect preceded by ‘within’ adverbial phrases are indicated by solid black line. Critical verbs in imperfective aspect preceded by ‘within’ adverbial phrases are indicated by dashed blue line. Critical verbs in perfect aspect preceded by ‘for’ adverbial phrases are indicated by dotted red line. Critical verbs in imperfective aspect preceded by ‘for’ adverbial phrases are indicated by dashed-dotted green line.

Figure 5. Voltage maps showing average differences in ERPs evoked by critical verbs in perfect aspect preceded by ‘for adverbial phrases relative to those preceded by ‘within’ adverbial phrases in the 300-500ms, 500-700ms, and 700-900ms time-windows.

Table 1. Experimental conditions

a. Conditions for *iterative coercion.* First sentence used as context for the examples below:
Lilly's kitty was always having small adventures.

Condition	**Example**
A. Punctive Adverb, Perfective Aspect	After several minutes the cat pounced on the rubber mouse.
B. Durative Adverb, Perfective Aspect	For several minutes the cat pounced on the rubber mouse.
C. Punctive Adverb, Imperfective Aspect	After several minutes the cat was pouncing on the rubber mouse.
D. Durative Adverb, Imperfective Aspect	For several minutes the cat was pouncing on the rubber mouse.

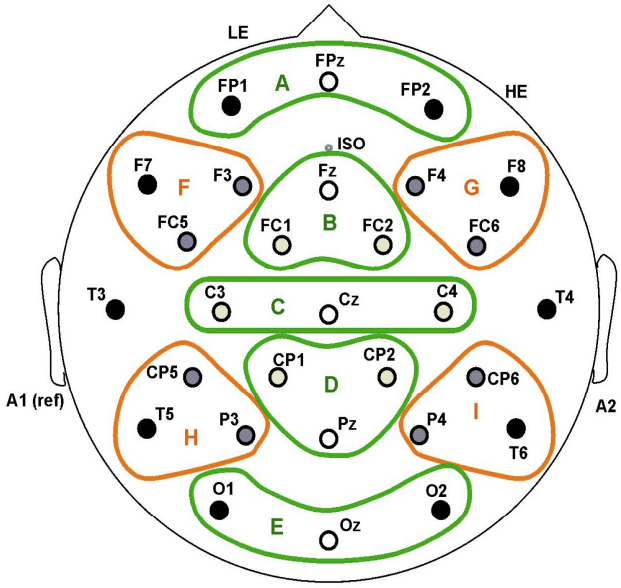
See Paczynski & Kuperberg (in prep) for descriptive statistics.

b. Conditions for *transparent aspectual interpretation.* First sentence used as context for the examples below:
Wendy took care of proof-reading for one of her friends.

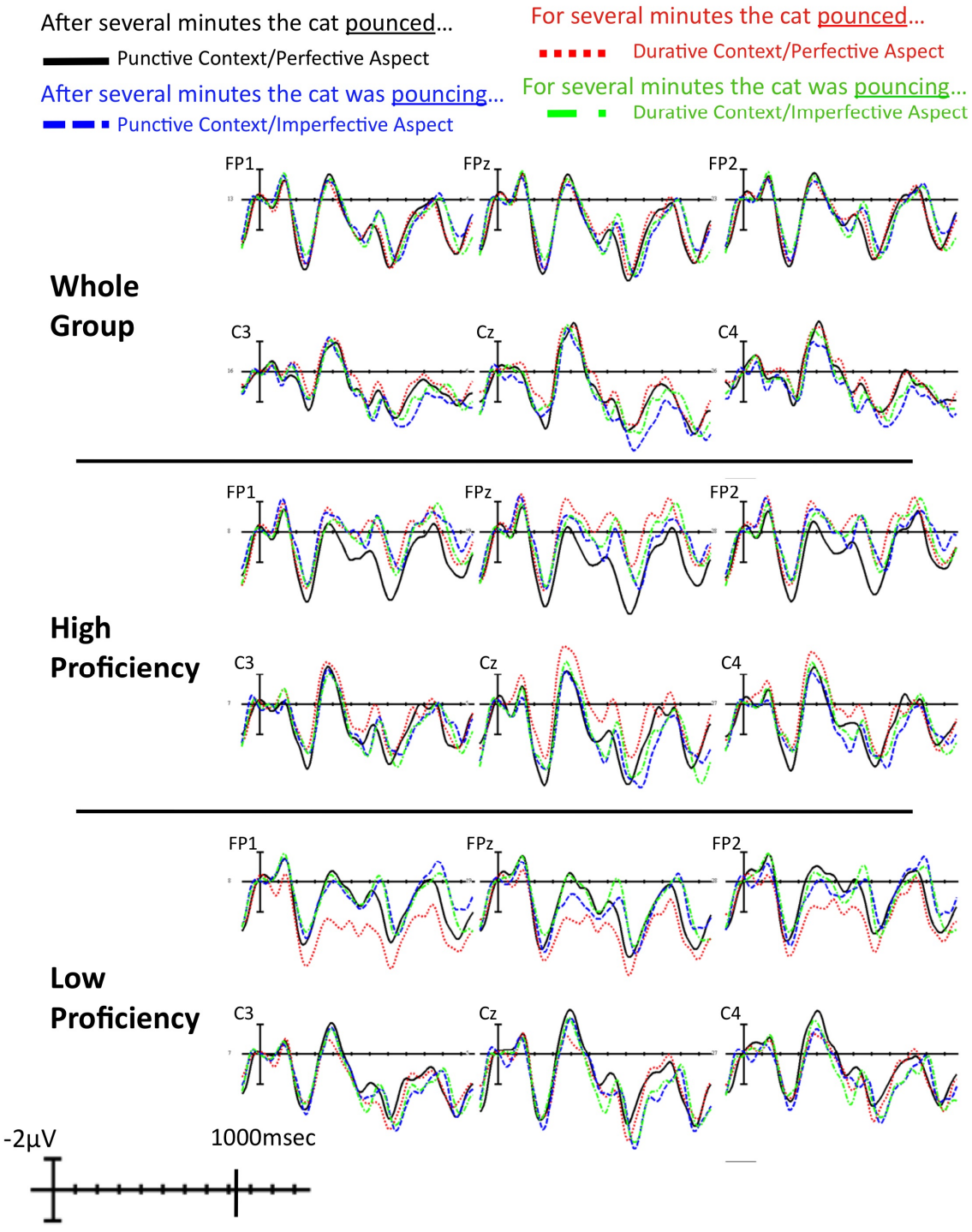
Condition	**Example**	**Naturalness** Mean (SD)	**Completeness** Mean (SD)
A. Within Adverb, Perfect Aspect	Within forty minutes she had circled the mistakes in the manuscript.	3.86 (0.33)	4.62 (0.18)
B. For Adverb, Perfect Aspect	For forty minutes she had circled the mistakes in the manuscript.	3.82 (0.26)	4.16 (0.26)
C. Within Adverb, Imperfective Aspect	Within forty minutes she was circling the mistakes in the manuscript.	3.76 (0.29)	3.04 (0.26)
D. For Adverb, Imperfective Aspect	For forty minutes she was circling the mistakes in the manuscript.	3.88 (0.30)	3.61 (0.24)

Naturalness rating based on a five point scale: 1 = not at all natural, 5 = very highly natural
Completeness rating based on a five point scale: 1 = none/not at all, 5= completed/fully accomplished

Fig. 1


LE
HE
FPz
FP1
FP2
A
ISO
F7
F3
Fz
F4
F8
F
G
FC5
FC1
B
FC2
FC6
T3
C3
C
Cz
C4
T4
CP5
CP1
CP2
CP6
D
I
A1 (ref)
T5
P3
Pz
P4
A2
H
T6
O1
O2
E
Oz

Iterative Coercion Fig. 2

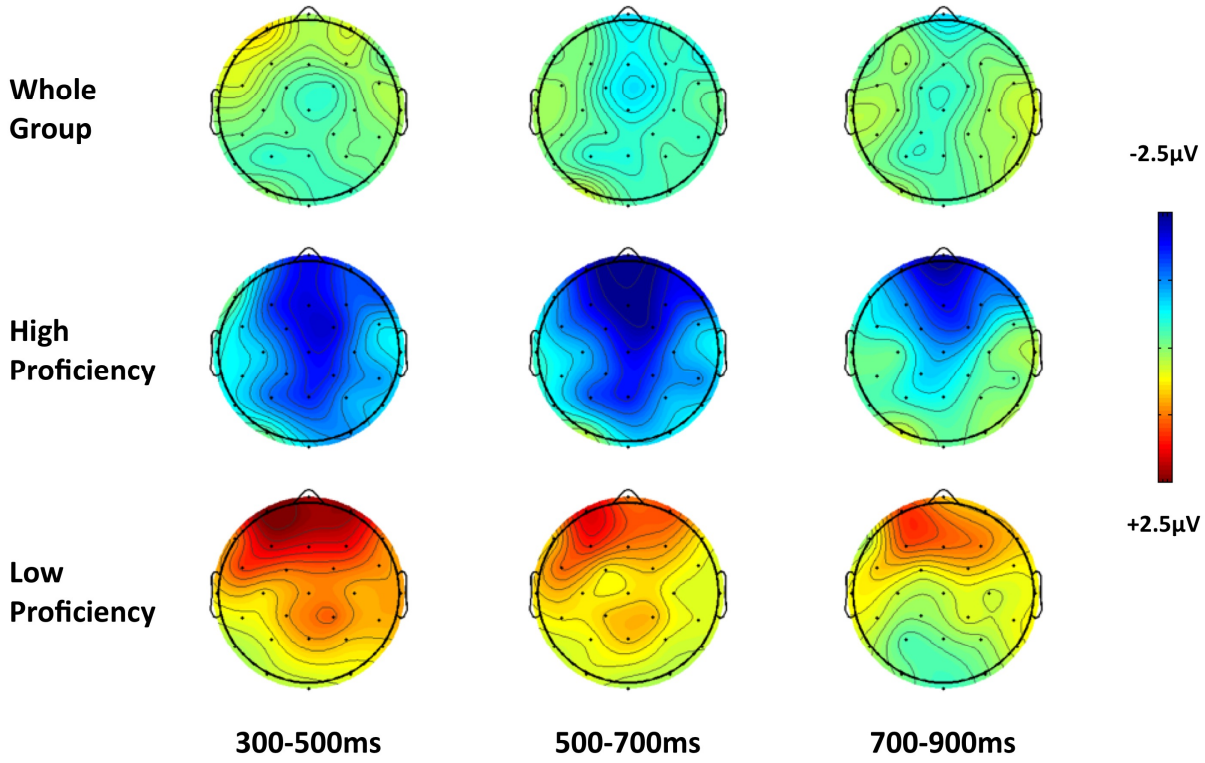


Iterative Coercion

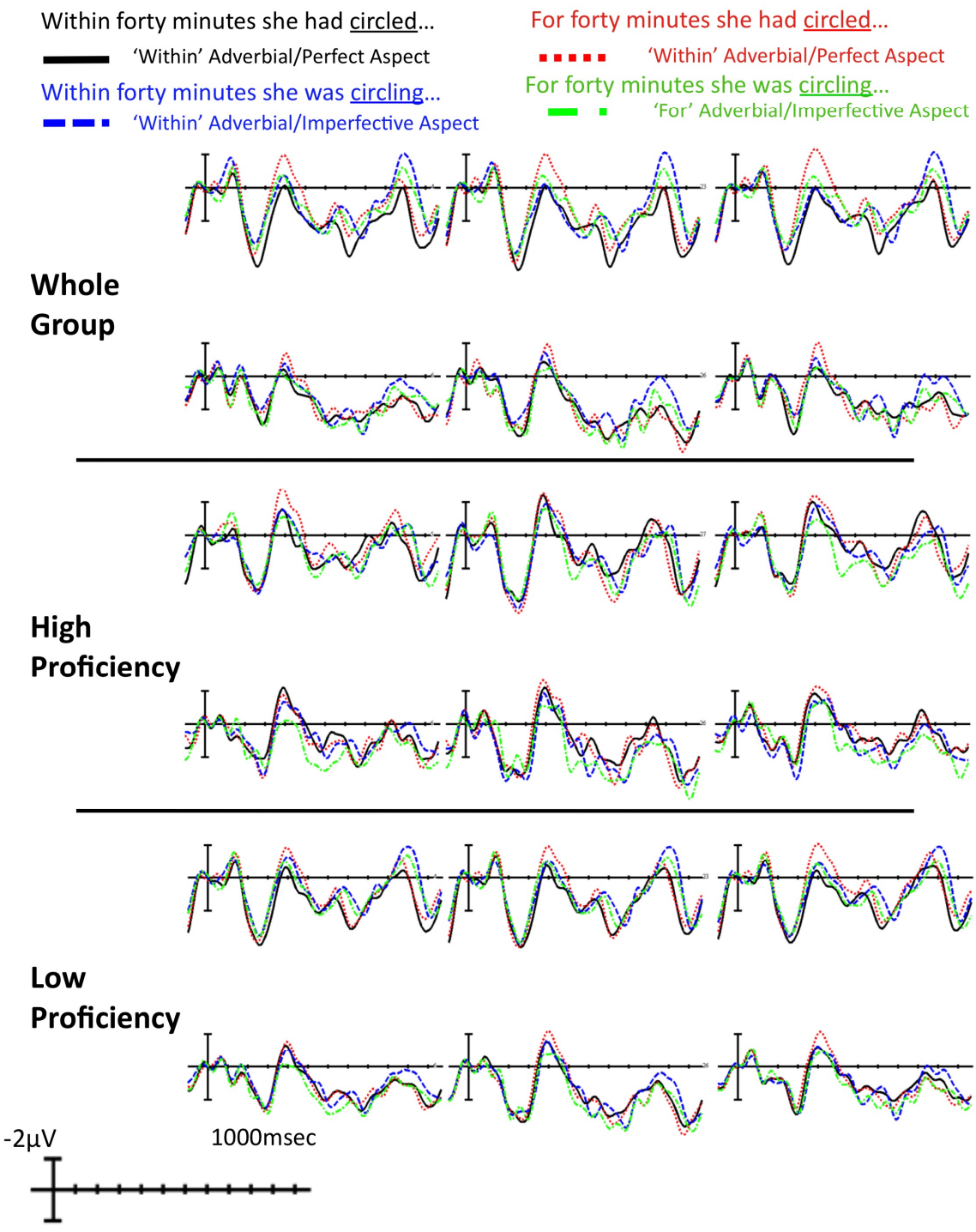
Fig. 3

Punctive Verbs in Perfective Aspect

Durative Context – Punctive Context



Transparent Aspectual Composition Fig. 4



Transparent Aspectual Composition

Fig. 5

Critical Verbs in Perfect Aspect

'For'' Adverbial Context – 'Within' Adverbial Context