Title

Information Structure in Language Comprehension: Cognitive Mechanisms Involved in Syntactic, Prosodic and Discourse Focus

Author
Beier, Eleonora Judith

Publication Date
2022

Peer reviewed|Thesis/dissertation

Information Structure in Language Comprehension:
Cognitive Mechanisms Involved in Syntactic, Prosodic and Discourse Focus

By

ELEONORA JUDITH BEIER
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Fernanda Ferreira, Chair

Tamara Swaab

Emily Morgan

Committee in Charge

2022

ABSTRACT

Information structure (IS) refers to the organization of new (essential) and given (redundant) information in language. Psycholinguistic theories hold that during comprehension, people infer the timing of the new, important information in order to facilitate processing. Several linguistic cues can be used to track IS and the timing of new information, such as the discourse context, syntactic focus constructions and prosodic focus. Previous studies have explained the effects of IS on language processing through a variety of potential cognitive mechanisms, such as attention, depth of processing, and the interaction of IS with good-enough processing and semantic prediction. However, many of these studies have produced contradictory findings that are hard to reconcile. To date, there is no clear picture of how discourse, syntactic and prosodic focus each contribute to language comprehension.

The goal of this dissertation is to begin to disentangle the cognitive mechanisms engaged by IS. Chapter 1 provides an extensive review of literature on IS, from linguistic theories to the psycholinguistic findings supporting different cognitive accounts. Chapter 2 presents an eye-tracking reading study assessing the interaction of syntactic focus with semantic prediction. Chapter 3 describes a self-paced reading replication of the same experiment, with additional analyses revealing the developing effect of syntactic focus over the sentence. Chapter 4 consists of a successful online replication of a foundational study by Cutler and Fodor (1979), which set the stage for this dissertation and much of the literature on IS. Lastly, in Chapter 5 I interpret my findings in the context of the linguistic and cognitive theories reviewed in Chapter 1, and describe two planned follow-up experiments. I conclude with a cognitive view of IS informed by this multidisciplinary literature and by the findings described in this dissertation, and highlight many open questions that will guide my future research and move the study of IS forward.

ACKNOWLEDGMENTS

I would like to thank Dr. Fernanda Ferreira for being the best mentor I could have asked for and for her unwavering support from beginning to end of my PhD journey. Our brainstorming sessions and deep dives into the history of psycholinguistics have played an essential role in my development as a cognitive psychologist. I would also like to extend my gratitude to my dissertation committee members, Drs. Tamara Swaab and Emily Morgan, for their expert guidance and kind comments throughout the writing of this dissertation. I thank the many other mentors who have shaped and strengthened my research in the past years, including Drs. Petr Janata, Gwen Rehrig, Lee Miller, Michelle Cohn, and Georgia Zellou, as well as Drs. Philippe Rast and Emilio Ferrer for their generous advice on statistical analysis. I am forever grateful to the cooperative environment of the UC Davis Psychology Program that has made these fruitful collaborations possible.

A big thanks goes to all present and past members of the Ferreira Lab: I have really appreciated our friendly and supportive lab culture, and I will miss our stimulating lab meetings. I am also indebted to the many undergraduate research assistants whose hard work has been essential to all aspects of my research. And I could not imagine completing this doctoral program without the companionship of my dearest friends and colleagues, Nene, Zoe and Tanya.

Last but not least, I owe my deepest gratitude to the most important people in my life. Mamma, thank you for always being kind and generous, and teaching me positivity. Pa, thank you for always being humorous and up for adventure, and teaching me resilience. Adolfo, thank you for always being there for me through the ups and downs, whether at 0 or 2833 miles away. Your constant love and support are the source of my strength.

TABLE OF CONTENTS

CHAPTER 1: GENERAL INTRODUCTION

One of the main purposes of language is to communicate information. As such, utterances typically contain some information that is known to the speaker, but not to the listener. The speaker's goal is therefore to communicate this new information in such a way that the listener will be able to quickly acquire it and integrate it within their own store of knowledge. Given how rapidly comprehenders are able to process the input and formulate a response in a conversation, despite the noisy environment through which communication operates (Gibson et al., 2013; Goodman & Frank, 2016; Traxler, 2014), the mechanisms behind language production and comprehension have likely been optimized towards this goal. In this dissertation, I review the literature on this topic and describe three experiments aimed at disentangling the cognitive mechanisms involved in the search for new information during language comprehension.

One way that speakers can facilitate the transfer of new information is to help the listener connect it to facts that they already know. For example, if a speaker says "It's red", the listener might not know what this utterance is referring to, and will need further clarification. However, if the speaker says "My new car is red", the listener will be able to connect this new property to an already stored representation, namely the fact that the speaker has a new car. The part of the utterance that refers to facts already known by the listener can be called *given*, as opposed to *new* information not known by the listener (Haviland & Clark, 1974; see the next section for notes on terminology). According to Haviland and Clark's *given-before-new* strategy, speakers tend to begin their utterance with given information and insert new information at the end. This strategy facilitates both comprehension and production. In comprehension, the given information allows the listener to retrieve the long-term representations to which new concepts will be appended. In turn, the speaker can begin their utterance with easily retrievable ideas, which gives them time to formulate the more critical new information (Bock & Irwin, 1980; in accordance with the Easy First strategy in language production, MacDonald, 2013). *Information structure* (hereafter, IS) refers to the positioning of new and given information within an utterance, as well as whether the utterance includes all new or all given information.

The given-before-new strategy matches the canonical word worder of English and it has been observed across languages (Bock & Irwin, 1980; Christianson & F. Ferreira, 2005; V. S. Ferreira & Yoshita, 2003). However, speakers can also use other tools to facilitate the listener's processing of new information. Listeners typically expect the new information to appear later in the sentence, but this expectation can be altered through syntax and prosody (Carlson et al., 2009). New information can be highlighted through syntactic cues such as focusing constructions (e.g., "It was Amy who ate the apple") or through prosodic prominence (e.g., "AMY ate the apple"). Different languages vary in the extent to which they employ prosody or syntax to convey the IS of sentences (Lambrech & Polinsky, 1997; Samek-Lodovici, 2005), which influences how comprehenders rely on these cues to infer the location of new information. For languages where IS is marked primarily through syntax, listeners weigh syntactic cues more than prosodic ones, while the opposite is true for languages where IS is marked primarily through prosody (Calhoun et al., 2021; Kember et al., 2019; Yan & Calhoun, 2020). In addition to syntactic and prosodic cues, comprehenders can also track IS through the discourse context. For example, the IS of the sentence "Amy ate the apple" depends on which question precedes it: "Who ate the apple?" ("Amy" is new), or "What did Amy eat?" ("the apple" is new).

The goal of this dissertation is to draw an integrated view of how syntactic, prosodic and discourse cues to IS affect the processes involved in language comprehension. In the remainder of this chapter, I review the multidisciplinary literature on IS, starting with the theoretical linguistic background behind the concepts of IS and focus. I then move to psycholinguistic findings and group them into three main theoretical accounts of IS: under the More Resources account, IS guides attention and processing resources to new information; under the Deeper Processing account, IS leads to deeper processing of focused new information; and under the Altered Predictions account, IS changes the strength or content of semantic predictions. While these theoretical views are not incompatible with each other, findings explained through each account are sometimes contradictory. Thus, in this dissertation I begin to tease apart these potential mechanisms.

In two experiments using eye-tracking (Chapter 2) and self-paced reading (Chapter 3) I measure the interaction of syntactic focus with semantic prediction, testing the opposing hypotheses of the More Resources and Deeper Processing accounts. I then provide evidence for the robustness of the facilitatory effect of discourse focus on reaction times (Chapter 4), replicating an influential finding by Cutler and Fodor (1979). Lastly, in Chapter 5 I relate these findings to the theoretical background reviewed in this chapter and describe several future directions for research on IS, aimed at integrating the many divergent lines of research on this topic into a comprehensive cognitive account.

Focus and information structure

Psycholinguistic studies of IS have their roots in early theoretical frameworks from the fields of linguistics, formal semantics, and pragmatics (e.g., Halliday, 1968; Roberts, 1996; Rochemont, 1986; Rooth, 1992). A full review of these frameworks is beyond the scope of this dissertation, which will primarily focus on the cognitive processes associated with IS. However, it is important to note that across these traditions, a large variety of terms have been used to refer to various aspects of IS, which are sometimes used interchangeably, and sometimes argued to reflect fundamentally different concepts. What most theories have in common is the idea that the IS of a sentence depends on the relationship between the knowledge of two interlocutors and what each interlocutor knows the other knows, or the *common ground,* established throughout the discourse (Krifka, 2008; Vallduví, 2016). Terms such as *given*, *theme*, *background* and *presupposition* have been used to refer to constituents that have already been established in the discourse and/or are already known by the listener. Terms such as *new*, *rheme* and *focus* have been used to refer to constituents that have not been previously mentioned, and that move the discourse forward by answering a *question-under-discussion* (QUD; Roberts, 1996). In my first example, “My new car is red”, “red” answers the QUD “What color is your new car?”, which could be either explicit or implicit in the discourse.

For example, in Robert’s (1996) formal pragmatics view of IS, discourse is like a cooperative game where two interlocutors share the common goal of communicating “the way things are” in the world. The game is played through a series of questions and answers. The question-under-discussion (QUD) is the immediate sub-topic being discussed, which implies a possible set of alternative answers. Because the QUD is not always expressed explicitly in discourse, speakers may use prosodic focus in English (and equivalent strategies in other languages) to let the listener infer what the QUD is. In other words, under Roberts’ view, prosodic focus is a cooperative strategy used by speakers to facilitate communication by reminding the listener what it is that they are talking about (what is the current QUD). Similarly, Vallduví (2016) reviews theories claiming that the main reason to include given information in utterances is to raise one of the many possible QUDs to the state of max-QUD, or the most prominent question the discussion is about.

Different theories vary in their definitions of these terms (see Vallduví, 2016), and by far the most inconsistently used term is *focus*. In most psycholinguistics literature and under some linguistic perspectives, any new information in a sentence can be considered in focus (Breen et al., 2010; F. Ferreira & Lowder, 2016; Halliday, 1968; Hornby, 1974; Krifka, 2008). However, under other linguistic accounts (e,g., Goodhue, 2022; Kratzer & Selkirk, 2020; Rooth, 1992), focus denotes a specific situation, where a constituent is placed in contrast to an antecedent in the discourse. For example, in "My new car is red", "red" would only be considered in focus insofar as the QUD implies a contrast with possible alternatives (e.g., "Is your new car green?"). In particular, Kratzer and Selkirk (2020) argue that in language, "newness" is the default, and it therefore does not need to be marked through focus. On the other hand, there are morphosyntactic markers for contrast and givenness (I return to this account in Chapter 5).

Regardless of how these concepts are best defined with respect to syntactic and semantic structures, the psychological reality is that in any conversation, some information will be new to the listener, some will be given, and some will be in contrast to previously established ideas. Therefore, throughout this dissertation I will adopt a less formal use of the word *focus*, and follow the definitions used in recent psycholinguistic work (e.g., Breen et al., 2010; Calhoun et al., 2021; F. Ferreira & Lowder, 2016). This literature has distinguished between *presentational focus*, or simply new information, and *contrastive focus*, or new information that additionally stands in contrast with some alternatives. Both are in opposition to *given* information already known to the listener. Presentational focus can be further divided into *broad* (or *sentence*-) focus, where all constituents in a sentence are new; and *narrow* focus, where only one constituent (e.g., only the subject or the object of the sentence) answers the QUD. As the majority of the literature

I will review concerns narrow presentational focus, this is what I will refer to as simply focus, and specify contrastive or broad focus when relevant.

However, it is important to keep in mind that many of the syntactic and prosodic cues associated with narrow focus are similar (and sometimes indistinguishable) to those used to mark contrastive focus (Breen et al., 2010; Rochemont, 1986; Roettger et al., 2019). This will be relevant when inferring which cognitive mechanisms best explain the psycholinguistic findings on IS, which I turn to in the next section.

Psycholinguistic findings and cognitive theories

Psycholinguistic work has studied the effects of IS on language processing by manipulating the discourse directly (e.g., through question-answer pairs), as well as through syntactic and prosodic manipulations. An early influential finding that set the stage for much of the following literature was a phoneme monitoring experiment by Cutler and Fodor (1979). Cutler and Fodor measured reaction times to an early and a late target word (beginning with a target phoneme) within acoustically identical sentences (e.g., "The man on the <u>corner</u> was wearing a <u>blue</u> hat"; early and late target words are underlined). The sentence was preceded by one of two questions, placing focus on either the early or the late target (e.g., "Which man was wearing the hat?" / "What hat was the man wearing?"). They found faster reaction times when the target words were focused by the preceding question, suggesting that the discourse questions guided listeners' attention towards the new, focused information in the sentence.

This finding was significant because it helped explain previously reported effects of prosody on phoneme monitoring reaction times (Cutler, 1976; Shields et al., 1974). Cutler (1976) had found faster reaction times for acoustically identical target words embedded in sentences where the preceding prosody led listeners to expect the target word to be stressed. This demonstrated that listeners actively predicted upcoming prosody, but it was unclear how this could be beneficial to comprehension. Cutler and Fodor interpreted this finding in light of IS, claiming that by predicting upcoming prosody, listeners were really trying to predict the timing of the focused information.

There is a close correspondence between prosody and IS (Aylett & Turk, 2004; Breen et al., 2010; Calhoun, 2010). New information is typically marked acoustically through stress and pitch accents, both of which relate to the broader concept of prosodic prominence. Specifically, new information receives greater intensity, higher pitch and longer duration than given information (Breen et al., 2010; Calhoun, 2010; Roettger et al., 2019). The match between these acoustic characteristics and different types of IS (e.g., presentational vs. contrastive focus) is not strictly one-to-one (Grice et al., 2017), but different prosodic patterns can be distinguished computationally and are thought to probabilistically inform listeners as to the location of new information (Calhoun, 2010; Roettger et al., 2019). Listeners prefer sentences where the new information is also prosodically prominent to sentences where information structure and prosody do not match (Nooteboom & Kruyt, 1987) and are sensitive to missing or superfluous pitch accents (Dimitrova et al., 2012).

Various aspects of prosody help listeners predict the timing of upcoming stressed words (Beier & F. Ferreira, 2018). First, the default prosody of sentences in English is driven by the Nuclear Stress Rule (Chomsky & Halle, 1968), whereby the rightmost lexical item in a sentence receives the strongest (nuclear) stress. This matches the given-before-new strategy, as listeners expect new information to occur at the end of the sentence and for it to receive the highest prosodic prominence. In addition, the distribution of stressed and unstressed syllables is rhythmic, and can be described through hierarchical metrical grids (Liberman & Prince, 1977; Martin, 1972). Martin (1972) proposed that listeners internalize these metrical structures to predict the timing of upcoming stressed syllables and direct attention to them, an idea later termed the Attentional Bounce Hypothesis (ABH; Pitt & Samuel, 1990). Overall, in accordance with Cutler and Fodor's (1979) early claims, listeners are likely to use prosody as a cue to the location of focused new information.

Cutler and Fodor also briefly discussed syntactic focus as an additional tool used by listeners to track IS. The most commonly studied syntactic focus constructions are it- and wh-clefts (e.g., "It was the girl who saw the cat"; "What the girl saw was the cat"), although other syntactic devices such as preposing and dislocation can also be used to mark focused constituents (Birner & Ward, 2009). Words focused through syntax are read for longer (Lowder & Gordon, 2015) and are remembered better (Birch et al., 2000; Birch & Garnsey, 1995). Additionally, comprehenders are more likely to notice incongruent information when it is focused syntactically than when it is presupposed (Bredart & Modolo, 1988; Erickson & Mattson, 1981; Hornby, 1974). Similarly to prosody, comprehenders are sensitive to the match between syntactic focus and the IS established through the discourse context (Banon & Martin, 2019), as well as to the match between syntactic and prosodic focus (Calhoun et al., 2021). In summary, these findings also support Cutler and Fodor's claim that syntactic focus is an additional cue to the IS of a sentence, which listeners use in conjunction with prosody and discourse to predict the timing of upcoming new information.

Why is predicting the timing of new information important for language comprehension? In the next sections, I explore various cognitive mechanisms that have been theorized to account for these findings and the evidence supporting each.

Attention and processing resources (the More Resources account)

Cutler and Fodor's (1979) original explanation for their phoneme monitoring findings was that discourse, prosody and syntax are all cues used by listeners to allocate attention and processing resources towards the focus of the sentence. This contributes to comprehension by facilitating the processing of the most essential parts of the sentence, resulting in faster reaction times. This interpretation extended previous claims that listeners could track the rhythmic properties of prosody to predict the timing of stressed words and direct attention to those points in time (Martin, 1972; Shields et al., 1974; Pitt & Samuel, 1990).

There is some evidence to support both the involvement of attention by IS, and its facilitatory effects on processing. For example, an fMRI study found that words focused through pitch accents recruited a network of brain areas associated with domain general attention (Kristensen et al., 2013). Similarly, ERP studies that manipulated IS through question-answer pairs (Yang et al., 2019) and syntax (Chen et al., 2014) found that focused words elicited a larger P2 component, which has been associated with attention (Luck & Hillyard, 1994; Singhal et al., 2002). However, a limitation of these studies is that they only indexed attentional processes occurring at or after the presentation of the focused word. A critical aspect of Cutler and Fodor's view is that cues in the preceding context pre-allocate attention in a predictive manner, such that any facilitatory effect of focus could not be explained by the local bottom-up saliency of focused words (e.g., the acoustic saliency of the pitch accents used by Kristensen et al., 2013).

To date there is little evidence for the predictive pre-allocation of attention through cues of IS. Recent years have seen a growing interest in cortical tracking of speech, whereby neural oscillations appear to track periodicities in spoken language (for a review, see Meyer, 2017).

Outside of speech, neural oscillations have been shown to entrain to rhythmically predictable stimuli, functioning as a temporal prediction mechanism that allocates attention to the expected timing of important information (Lakatos et al., 2008). Neural entrainment has been suggested to help listeners track prosodic rhythm as well, which could allow listeners to predict the timing of future stressed words and direct attention to them. However, there is still ample debate on whether this mechanism is at play in language comprehension and whether it is engaged by prosodic metrical structure (Beier et al., 2021; Ding & Simon, 2014; Kösem & van Wassenhove, 2017; Meyer et al., 2020). Overall, more research is needed to clearly identify the attentional predictive mechanisms engaged by speech rhythm and other cues of IS (see *Future Directions* in Chapter 5).

Several lines of research also support the idea that the processing of focused information is facilitated thanks to the availability of more processing resources. Cutler and Fodor's (1979) and Cutler's (1976) phoneme monitoring findings have been replicated and extended through several follow-ups (Akker & Cutler, 2003; Cutler & Darwin, 1981; Cutler & McQueen, 2014; Ip & Cutler, 2020; Rysling et al., 2020; also see Beier & F. Ferreira, 2022 in Chapter 4). On the other hand, a study specifically designed to test whether prosodic rhythm facilitated processing of expected stressed syllables using phoneme monitoring only found this effect for a string of unrelated words and not for more naturalistic speech (Pitt & Samuel, 1990).

Later studies switched from the phoneme monitoring task to more direct measures of processing such as ERPs (F. Ferreira & Anes, 1994). Words focused through discourse and syntax have been shown to elicit a smaller N400 component (or N400-like negative deflection in Yang et al., 2019), suggesting that focus facilitates semantic processing (Chen et al., 2014; Yang et al., 2019).

Similarly, the N400 effect of semantically unpredictable words was reduced for sentences with greater rhythmic regularity (Rothermich et al., 2012). This was interpreted as support for the ABH (Pitt & Samuel, 1990), or the idea that a regular prosodic structure lets listeners predict the temporal location of upcoming stressed syllables. As focused words are also prosodically prominent, greater rhythmic regularity in the prosody is a cue to the location of focused words, leading to easier integration of semantic violations falling on those locations (thus, a smaller N400 amplitude to unpredicted words).

In summary, there is evidence to support Cutler and Fodor's (1979) original claim that cues of IS pre-allocate attention and processing resources to information expected to be focused, facilitating processing. I will refer to this view as the More Resources account.

Depth of processing (the Deeper Processing account)

A related, yet contrasting view is that focused constituents are processed more deeply than non-focused constituents, which I will refer to as the Deeper Processing account. This is supported by the behavioral finding that comprehenders are more likely to notice semantic incongruencies when they are focused than when they are presupposed (Bredart & Modolo, 1998; Erickson & Mattson, 1981; Hornby, 1974). Additionally, eye-tracking results show that words are read for longer when they are focused through question-answer pairs (Benatar & Clifton, 2014) and through syntactic focusing constructions (Lowder & Gordon, 2015).

This view is not inconsistent with the More Resources view, in that words that are more deeply processed likely require more processing resources and would therefore benefit from the pre-allocation of attention. However, studies interpreted through this view have led to contrasting results. Using the RSVP technique, Wang et al. (2009) observed the N400 response to plausible or implausible words contained in the answer to question-answer pairs, such that the target word would either be focused or de-focused based on the preceding question. They found that the N400 effect was largest when the target was in focus, due to an increased reduction of the N400 to the plausible focused word. This was interpreted as evidence that focused words were processed more deeply than de-focused words. This finding was later expanded in an auditory version of the task where the target word could also either be accented or not (Wang et al., 2011). It was found that the largest N400 effect (due to the largest reduction of the N400 to plausible words) occurred for words that were both focused and accented. This suggests that prosodic marking of IS plays an additive role, leading to further deeper processing of the focused accented word.

Although the Deeper Processing account is not inconsistent with the More Resources view, these findings are hard to reconcile with those of Rothermich et al. (2012). Rothermich et al. found a *smaller* N400 effect for stressed words in sentences with rhythmically regular prosody, which they interpreted as reflecting the fact that more resources were allocated to words in prosodically prominent positions. On the other hand, Wang et al. (2009; 2011) found a *larger* N400 effect for discourse-focused and accented words, which they interpreted as reflecting the fact that those words were processed more deeply. Thus, it is unclear whether focus leads to *facilitated* processing (faster RTs, smaller N400 effect) or *deeper* processing (better memory encoding, longer reading times, larger N400 effect) or a combination of both.

Reconciling findings across accounts

It is difficult to reconcile these discrepant findings because different studies employed very different tasks and manipulated focus through different cues of IS. For example, participants in Rothermich et al. (2012) were instructed to detect semantic anomalies, while those in Wang et al. (2011) simply listened to all stimuli without an explicit task. Additionally, the two experiments were conducted in different languages (German and Dutch, respectively), and it has been shown that listeners weigh cues of IS differently across languages (Calhoun et al., 2021; Kember et al., 2019). Lastly, some recent findings suggest that contrary to what was theorized by Cutler and Fodor (1979), different cues to IS may not be interchangeable, but rather have individual effects on different aspects of language processing (Yan & Calhoun, 2020). Thus, the fact that Rothermich et al. manipulated IS through prosodic rhythmic regularity while Wang et al. did so through question-answer pairs and pitch accents may account for the discrepant results.

Another explanation comes from studies showing a dissociation between newness and focus (Chen et al., 2012, 2014). Chen et al. (2014) independently manipulated newness (whether a word had already been mentioned in the preceding discourse or not) and syntactic focus (through focus particles in Mandarin Chinese). They found that while new words elicited a larger N400 than given words, focused words led to a smaller N400 than de-focused words. Critically, Rothermich et al. did not manipulate newness per se (as there was no discourse context prior to the target sentences) but only the degree of rhythmic regularity (a proxy for prosodic focus if we follow the claims of the ABH; Pitt & Samuel, 1990). On the other hand, the target words in Wang et al. were both new and focused, by virtue of answering the preceding question. Thus, the different patterns observed by these two studies could be due to focus and newness influencing processing in different ways: focus may facilitate processing (thanks to attention allocation) while new words are more difficult to process than given words.

Lastly, a critical difference in the results of these two studies is what drives the size of the N400 effect (the difference in the response to predictable vs. unpredictable words). In Rothermich et al., a smaller N400 effect was found thanks to a reduction of the N400 response to focused unpredictable words, while the N400 to predictable words stayed the same across conditions. In Wang et al. the opposite was true: a larger N400 effect was found thanks to a reduction of the N400 to focused predictable words, while the N400 to unpredictable words stayed the same. This supports the idea that the cues of IS used in these two studies influenced language processing in different ways.

This points to the importance of determining how exactly IS interacts with semantic prediction. The original finding that the N400 amplitude varies as a function of a word's predictability (Kutas & Hillyard, 1984) has been interpreted in at least two ways: (1) The N400 reflects difficulty of lexical access; predictable words are partially pre-activated thanks to the constraining context, so that lexical access is facilitated at the word; (2) The N400 reflects the difficulty of integrating the word with its context after it has been accessed; predictability effects arise not thanks to pre-activation, but because constraining contexts allow easier integration (for reviews, see Lau et al., 2008; Swaab et al., 2012; Van Petten & Luka, 2012).

Rothermich et al. seem to assume the second interpretation, whereby unpredictable words in rhythmically irregular contexts are harder to integrate with the context because of a lack of processing resources. Wang et al.'s results fit best with the first view of the N400: it may be that the focus condition led to stronger pre-activation, which in turn led to easier lexical access of the predictable word. This introduces the possibility that IS has a complex interaction with semantic prediction, potentially influencing both pre-activation and integration.

Semantic prediction (the Altered Predictions account)

A third account posits that comprehenders engage in more semantic prediction for parts of the sentence that are expected to carry new information, while they engage in good-enough processing for parts of the sentence expected to carry given information (F. Ferreira & Lowder, 2016). In this view, semantic prediction facilitates comprehension by pre-activating a set of likely lexical candidates; thus, comprehenders engage in prediction for parts of the sentence that are expected to be most important in order to optimize comprehension. As a result of increased prediction for focused constituents, processing of predicted words should be facilitated thanks to greater pre-activation (as in Wang et al.). On the other hand, stronger predictions may lead to a greater processing cost for unpredicted words. This would result in a larger N400 effect in focused positions, which is consistent with the Deeper Processing view but is in contrast with results from the More Resources view.

However, IS may influence not just the strength of semantic prediction, but also the content of what is predicted. While I have primarily reviewed studies on presentational focus, cues associated with newness can also be interpreted as marking a contrast with possible alternatives (Breen et al., 2010; Rochemont, 1986; Roettger et al., 2019). Literature on contrastive focus shows that pitch accents and syntactic focusing devices (e.g., focus particles such as “not only… but also…”) lead to the activation of semantic alternatives (Husband & F. Ferreira, 2016; Kurumada et al., 2014; Lowder et al., 2021; Lowder & F. Ferreira, 2016; Weber et al., 2006). Thus, if listeners interpret all focus cues as potentially contrastive (Kratzer & Selkirk, 2020), then they may be more likely to expect a focused word to be less predictable that if it were not in focus. In other words, focus may lead listeners to expect the unexpected. Following this logic, listeners may have a reduced N400 to focused unpredictable words (because they are now more predictable) and an increased N400 to focused predictable words (because they are now less predictable). This would lead to a smaller N400 effect for focused words, which is consistent with the More Resources account but at odds with the Deeper Processing findings.

It is clear that, given the current data, it is not possible to disentangle these two alternative ways that IS could interact with semantic prediction. I will refer to both possibilities under the umbrella of the Altered Predictions account, or the idea that cues to IS alter the strength and/or the content of semantic predictions during online processing.

Summary of cognitive mechanisms

In this introduction, I have reviewed several lines of research addressing how different aspects of IS influence language comprehension. Many of these studies have followed Cutler and Fodor's (1979) claim that syntactic, prosodic and discourse cues of IS guide the comprehender's attention towards the most important elements of a sentence. The processing of focused information is therefore facilitated, thanks to the pre-allocation of processing resources. Other studies have explained the effects of IS by claiming that focused constituents receive deeper processing. Lastly, IS has been shown to influence other important components of language comprehension, including semantic prediction. However, there is no consensus on how exactly it does so, such as whether it facilitates the integration of unpredictable words, whether it leads to stronger pre-activation of the most predictable words, or whether it alters the content of semantic predictions.

It is likely that all of these mechanisms are at play. However, to date there is no clear picture, as different studies have provided inconsistent findings that are hard to reconcile due to the use of different methodologies and tasks. Studies differ in the ways they manipulate IS, such as through discourse, pitch accents, speech rhythm, or syntactic focus constructions, and they differ in the languages used to do so. As comprehenders' use of cues varies across languages, it is hard to relate findings across these studies. Thus, there is a need to clarify the exact role of IS in influencing attention, depth of processing and semantic prediction at different stages of comprehension; whether prosodic, syntactic and discourse cues of IS affect these mechanisms similarly or have independent effects; and how this varies for speakers of different languages.

This dissertation

The goal of this dissertation is to begin to disentangle these mechanisms. Chapter 2 (Experiment 1) presents an eye-tracking reading study assessing the interaction of syntactic focus with semantic predictability. This study was designed to tease apart two contrasting patterns of results that would be expected based on either the More Resources or the Deeper Processing accounts. In my original dissertation plan, this study was followed by two other experiments: an auditory eye-tracking study using the visual world paradigm to test the influence of syntactic focus on semantic prediction prior to the focused word; and an ERP study comparing the effects of syntactic focus and prosodic pitch accents on the N400 effect. However, due to the Covid-19 pandemic this plan was revised to account for the inability to test participants in person, which would have been necessary for both of these experiments. The visual world study is currently in development, and it is described under *Future Directions* in Chapter 5. The same section also outlines a planned experiment that will identify the temporal attention mechanisms involved by speech rhythm and discourse focus using ERPs and time-frequency analyses of EEG.

As an alternative to the originally planned experiments, in this dissertation I present the results of two different follow-up studies conducted online in accordance with safety protocols in place during the pandemic. Chapter 3 (Experiment 2) reports the findings of a replication of Experiment 1 using self-paced reading, which reveals new insights into the developing effect of syntactic focus throughout the sentence. Chapter 4 (Experiment 3) consists of a replication of Cutler and Fodor's (1979) influential study, which was recently published as a registered report in a special replication issue of the *Journal of Memory and Language* (Beier & F. Ferreira, 2022). I chose to conduct a replication of Cutler and Fodor (1979) given its prominence in the psycholinguistic literature on IS, and its role in shaping the ideas explored in this dissertation. Additionally, as recent evidence puts into question their original claims for the interchangeability of discourse, syntactic and prosodic cues of IS, it is critical to scrutinize whether their findings (which related the effects of prosody to those of discourse) are robust.

The next three chapters provide more details about the motivation, methods and results of each experiment. I then relate these findings to the cognitive and linguistic theories described in this introduction in the *General Discussion* (Chapter 5).

References

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition*, *6*(2), 81–96. https://doi.org/10.1017/s1366728903001056

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56. https://doi.org/10.1177/00238309040470010201

Banon, J. A., & Martin, C. (2019). Anticipating information structure: An event-related potentials study of focus assignment via the it -cleft. *Neuropsychologia*, *134*, 107203. https://doi.org/10.1016/j.neuropsychologia.2019.107203

Beier, E. J., Chantavarin, S., Rehrig, G., Ferreira, F., & Miller, L. M. (2021). Cortical tracking of speech: Toward collaboration between the fields of signal and sentence processing. *Journal of Cognitive Neuroscience*, *33*(4), 574–593. https://doi.org/10.1162/jocn_a_01676

Beier, E. J., & Ferreira, F. (2018). The temporal prediction of stress in speech and its relation to musical beat perception. *Frontiers in Psychology*, *9*, 431. https://doi.org/10.3389/fpsyg.2018.00431

Beier, E. J., & Ferreira, F. (2022). Replication of Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. Cognition, 7(1), 49–59. *Journal of Memory and Language*, *126*, 104339. https://doi.org/10.1016/j.jml.2022.104339

Benatar, A., & Clifton, C. (2014). Newness, givenness and discourse updating: Evidence from eye movements. *Journal of Memory and Language*, *71*(1), 1–16. https://doi.org/10.1016/j.jml.2013.10.003

Birch, S. L., Albrecht, J. E., & Myers, J. L. (2000). Syntactic Focusing Structures Influence Discourse Processing. *Discourse Processes*, *30*(3), 285–304. https://doi.org/10.1207/S15326950dp3003_4

Birch, S. L., & Garnsey, S. M. (1995). The effect of focus on memory for words in sentences. *Journal of Memory and Language*, *34*(2), 232–267.

Birner, B. J., & Ward, G. (2009). Information Structure and Syntactic Structure. *Language and Linguistics Compass*, *4*, 1167–1187.

Bock, K., & Irwin, D. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 467–484.

Bredart, S., & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychologica*, *67*(2), 135–144. https://doi.org/10.1016/0001-6918(88)90009-1

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7), 1044–1098. https://doi.org/10.1080/01690965.2010.504378

Calhoun, S. (2010). How does informativeness affect prosodic prominence? *Language and Cognitive Processes*, *25*(7–9), 1099–1140. https://doi.org/10.1080/01690965.2010.491682

Calhoun, S., Wollum, E., & Kruse Va'ai, E. (2021). Prosodic Prominence and Focus: Expectation Affects Interpretation in Samoan and English. *Language and Speech*, *64*(2), 346–380. https://doi.org/10.1177/0023830919890362

Carlson, K., Dickey, M. W., Frazier, L., & Clifton, C. (2009). Information structure expectations in sentence comprehension. *Quarterly Journal of Experimental Psychology*, *62*(1), 114–139. https://doi.org/10.1080/17470210701880171

Chen, L., Li, X., & Yang, Y. (2012). Focus, newness and their combination: Processing of information structure in discourse. *PLoS ONE*, *7*(8). https://doi.org/10.1371/journal.pone.0042533

Chen, L., Wang, L., & Yang, Y. (2014). Distinguish between focus and newness: An ERP study. *Journal of Neurolinguistics*, *31*, 28–41. https://doi.org/10.1016/j.jneuroling.2014.06.002

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. Harper and Row.

Christianson, K., & Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition*, *98*(2), 105–135. https://doi.org/10.1016/j.cognition.2004.10.006

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics*, *20*(1), 55–60.

Cutler, A., & Darwin, C. J. (1981). Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception & Psychophysics*, *29*(3), 217–224. https://doi.org/10.3758/BF03207288

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, *7*(1), 49–59. https://doi.org/10.1016/0010-0277(79)90010-6

Cutler, A., & McQueen, J. M. (2014). How prosody is both mandatory and optional. In J. Caspers, Y. Chen, W. Heeren, N. Pacilly, N. Schiller, & E. van Zanten (Eds.), *Above and Beyond the Segments* (pp. 71–82). John Benjamins.

Dimitrova, D. V., Stowe, L. A., Redeker, G., & Hoeks, J. C. J. (2012). Less is not more: Neural responses to missing and superfluous accents in context. *Journal of Cognitive Neuroscience*, *24*(12), 2400–2418. https://doi.org/10.1162/jocn_a_00302

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, *8*, 1–7.

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 540–551. https://doi.org/10.1016/S0022-5371(81)90165-1

Ferreira, F., & Anes, M. (1994). Why study spoken language? In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 33–56). Academic Press.

Ferreira, F., & Lowder, M. W. (2016). Prediction, information structure, and good-enough language processing. *Psychology of Learning and Motivation*, *65*, 217–247. https://doi.org/10.1016/bs.plm.2016.04.002

Ferreira, V. S., & Yoshita, H. (2003). Given-New Ordering Effects on the Production of Scrambled Sentences in Japanese. *Journal of Psycholinguistic Research*, *32*(6), 669–692. https://doi.org/10.1023/A:1026146332132

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056. https://doi.org/10.1073/pnas.1216438110

Goodhue, D. (2022). All Focus is Contrastive: On Polarity (Verum) Focus, Answer Focus, Contrastive Focus and Givenness. *Journal of Semantics*, *39*(1), 117–158. https://doi.org/10.1093/jos/ffab018

Goodman, N. D., & Frank, M. C. (2016). Pragmatic Language Interpretation as Probabilistic Inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. https://doi.org/10.1016/j.tics.2016.08.005

Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*, *64*, 90–107.

Halliday, M. A. K. (1968). Notes on transitivity and theme in English Part 3. *Journal of Linguistics*, *4*(2), 179–215. https://doi.org/10.1017/S0022226700001882

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, *13*, 512–521.

Hornby, P. A. (1974). Surface structure and presupposition. *Journal of Verbal Learning and Verbal Behavior*, *13*(5), 530–538. https://doi.org/10.1016/S0022-5371(74)80005-8

Husband, E. M., & Ferreira, F. (2016). The role of selection in the comprehension of focus alternatives. *Language, Cognition and Neuroscience*, *31*(2), 217–235. https://doi.org/10.1080/23273798.2015.1083113

Ip, M. H. K., & Cutler, A. (2020). Universals of listening: Equivalent prosodic entrainment in tone and non- tone languages. *Cognition*, *202*, 104311. https://doi.org/10.1016/j.cognition.2020.104311

Kember, H., Choi, J., Yu, J., & Cutler, A. (2019). The processing of linguistic prominence. *Language and Speech*, *64*(2), 1–24. https://doi.org/10.1177/0023830919880217

Kösem, A., & van Wassenhove, V. (2017). Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. *Language, Cognition and Neuroscience*, *32*(5), 536–544. https://doi.org/10.1080/23273798.2016.1238495

Kratzer, A., & Selkirk, E. (2020). Deconstructing information structure. *Glossa: A Journal of General Linguistics*, *5*(1), 1–53. https://doi.org/10.5334/gjgl.968

Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, *55*(3–4), 243–276.

Kristensen, L. B., Wang, L., Petersson, K. M., & Hagoort, P. (2013). The interface between language and attention: Prosodic focus marking recruits a general attention network in spoken language comprehension. *Cerebral Cortex*, *23*(8), 1836–1848. https://doi.org/10.1093/cercor/bhs164

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, *133*(2), 335–342. https://doi.org/10.1016/j.cognition.2014.05.017

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. https://doi.org/10.1038/307161a0

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, *320*(5872), 110–113. https://doi.org/10.1126/science.1154735

Lambrech, K., & Polinsky, M. (1997). Typological Variation in Sentence-Focus Constructions. *CLS*, *33*(1), 141–165.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933. https://doi.org/10.1038/nrn2532

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*(2), 249–336.

Lowder, M. W., & Ferreira, F. (2016). Prediction in the processing of repair disfluencies: Evidence from the visual-world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(9), 1400–1416.

Lowder, M. W., & Gordon, P. C. (2015). Focus takes time: Structural effects on reading. *Psychonomic Bulletin & Review*, *22*(6), 1733–1738. https://doi.org/10.3758/s13423-015-0843-2

Lowder, M. W., Ryan, G., Opie, J., & Kaminsky, E. (2021). Effects of contrastive focus on lexical predictability during sentence constructions. *Quarterly Journal of Experimental Psychology*, *74*(1), 179–186. https://doi.org/10.1177/1747021820949155

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, *31*(3), 291–308. https://doi.org/10.1111/j.1469-8986.1994.tb02218.x

MacDonald, M. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4,* 226.

Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, *79*(6), 487–509. https://doi.org/10.1037/h0033467

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, *48*(7), 2609–2621. https://doi.org/10.1111/ijlh.12426

Meyer, L., Sun, Y., & Martin, A. E. (2020). Synchronous , but not entrained: Exogenous and endogenous cortical rhythms of speech and language processing. *Language, Cognition and Neuroscience*, *35*(9), 1089–1099. https://doi.org/10.1080/23273798.2019.1693050

Nooteboom, S. G., & Kruyt, J. G. (1987). Accents, focus distribution, and the perceived distribution of given and new information: An experiment. *The Journal of the Acoustical Society of America*, *82*(5), 1512–1524. https://doi.org/10.1121/1.395195

Pitt, M., & Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology. Human Perception and Performance*, *16*(3), 564–573. https://doi.org/10.1037/0096-1523.16.3.564

Roberts, C. (1996). Information Structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, *5*, 1–69. https://doi.org/10.3765/sp.5.6

Rochemont, M. S. (1986). *Focus in Generative Grammar*. Benjamins.

Roettger, T. B., Mahrt, T., & Cole, J. (2019). Mapping prosody onto meaning – the case of information structure in American English. *Language, Cognition and Neuroscience*, *34(7)*, 841-860. https://doi.org/10.1080/23273798.2019.1587482

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics: An International Journal of Semantics and Its Interfaces in Grammar*, *1*(1), 75–116. https://doi.org/10.1007/BF02342617

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia*, *50*(2), 232–244. https://doi.org/10.1016/j.neuropsychologia.2011.10.025

Rysling, A., Bishop, J., Clifton, C., & Yacovone, A. (2020). Preceding syllables are necessary for the accent advantage effect. *The Journal of the Acoustical Society of America*, *148*, EL285. https://doi.org/10.1121/10.0001780

Samek-Lodovici, V. (2005). Prosody-Syntax Interaction in the Expression of Focus. *Natural Language & Linguistic Theory*, *23*(3), 687–755. https://doi.org/10.1007/sl

Shields, J. L., McHugh, A., & Martin, J. G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology*, *102*(2), 250–255. https://doi.org/10.1037/h0035855

Singhal, A., Doerfling, P., & Fowler, B. (2002). Effects of a dual task on the N100–P200 complex and the early and late Nd attention waveforms. *Psychophysiology*, *39*(2), 236–245. https://doi.org/10.1111/1469-8986.3920236

Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. A. (2012). Language-related ERP components. In S. J. Luck & E. S. Kappenman (Eds.), *Oxford Handbook of Event-Related Potential Components* (pp. 397–440). Oxford University Press.

Traxler, M. J. (2014). Trends in syntactic parsing: Anticipation, Bayesian estimation, and good-enough parsing. *Trends in Cognitive Sciences*, *18*(11), 605–611. https://doi.org/10.1016/j.tics.2014.08.001

Vallduví, E. (2016). Information structure. In M. Aloni & P. Dekker (Eds.), *The Cambridge Handbook of Formal Semantics* (pp. 728–755). Cambridge University Press.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Wang, L., Bastiaansen, M., Yang, Y., & Hagoort, P. (2011). The influence of information structure on the depth of semantic processing: How focus and pitch accent determine the size of the N400 effect. *Neuropsychologia*, *49*(5), 813–820. https://doi.org/10.1016/j.neuropsychologia.2010.12.035

Wang, L., Hagoort, P., & Yang, Y. (2009). Semantic illusion depends on information structure: ERP evidence. *Brain Research*, *1282*, 50–56. https://doi.org/10.1016/j.brainres.2009.05.069

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, *49*(Pt 3), 367–392. https://doi.org/10.1177/00238309060490030301

Yan, M., & Calhoun, S. (2020). Rejecting false alternatives in Chinese and English: The interaction of prosody, clefting, and default focus position. *Laboratory Phonology*, *11*(1), 1–39. https://doi.org/10.5334/labphon.255

Yang, C. L., Zhang, H., Duan, H., & Pan, H. (2019). Linguistic focus promotes the ease of discourse integration processes in reading comprehension: Evidence from event-related potentials. *Frontiers in Psychology*, *9*, 1–14. https://doi.org/10.3389/fpsyg.2018.02718

CHAPTER 2: EYE-TRACKING READING STUDY

Introduction

This chapter describes the results of an eye-tracking experiment investigating the interaction of semantic predictability and syntactic focus constructions on reading times. Previous literature shows that both semantic predictability and syntactic focus constructions influence reading times: readers spend more time fixating semantically unpredictable than predictable words (Clifton et al., 2007), and they read syntactically focused words for longer than de-focused words (Lowder & Gordon, 2015). However, to the best of our knowledge, the interaction between these two factors has not been explored. Characterizing how the effect of predictability varies as a function of focus is a first step towards understanding the mechanisms through which information structure (IS) influences language processing.

In this study, target words were embedded in short "stories" which led the target word to be either predictable or unpredictable (yet plausible). The target word was presented in a "neutral" construction ("He ate a <u>sandwich"</u>), a focusing wh-cleft construction ("What he ate was a <u>sandwich"</u>), or a de-focusing subordinate clause ("After he ate a <u>sandwich</u>,..."). Reading times (First Fixation Duration, First-Pass Duration, and Go-Past Duration, further described in the *Method*) were measured on the target word as participants read each story. Our analyses aimed to determine whether the size of the predictability effect (longer reading times for unpredictable compared to predictable words) was influenced by the syntactic focus construction.

Critically, different theoretical accounts for how IS influences language processing (see Chapter 1) lead to different predictions for this interaction. The More Resources account predicts that compared to the neutral condition, there will be a larger effect of predictability for de-focused words and a smaller effect for focused words. According to this account, focus cues (such as syntactic focus) lead comprehenders to allocate attention and processing resources to the expected timing of the focused words, while paying less attention and devoting fewer resources to de-focused words. Thus, it may be easier to integrate unpredictable words when they are focused than when they are de-focused, resulting in a smaller difference in reading times between predictable and unpredictable words (similarly to the smaller N400 effect found by Rothermich et al., 2012 for metrical prosody cues). On the other hand, the Deeper Processing account leads to the opposite prediction: compared to the neutral condition, there will be a larger effect of predictability for focused words and a smaller effect for de-focused words. In this view, focused words receive deeper processing than de-focused words, which are more likely to be shallowly processed (Wang et al., 2009, 2011). Thus, readers may be less likely to notice the unpredictable words when they are de-focused, leading to a smaller effect of predictability, while they may spend more time processing them when they are focused, leading to a larger effect (similarly to the larger N400 effect found by Wang et al., 2011 for words in discourse and prosodic focus). These two alternative patterns can also be explained under the Altered Prediction account (F. Ferreira & Lowder, 2016), according to which the strength and/or content of semantic predictions are altered as a function of focus. This possibility will be explored in a subsequent experiment (see *Future Directions* in Chapter 5).

In summary, the present experiment will begin to disentangle these mechanisms by testing whether the effect of predictability is greater for syntactically focused or for de-focused words. These two opposite patterns would support the predictions of either the More Resources account or the Deeper Processing account. Of course, it is likely that more than one mechanism is engaged by syntactic focus, and that language processing involves a combination of all three accounts. Thus, while this experiment provides a first investigation into this question, more research is needed to discern how these mechanisms interact at different stages of processing, and as reflected through different measures.

Method

Participants

Participants consisted of 141 undergraduate students at the University of California, Davis recruited through the SONA participant pool and compensated with class credit. All participants were native English speakers and reported normal or corrected-to-normal vision. A total of 9 participants were removed from the analyses due to track or calibration failures, 5 were removed due to failure to complete the experiment or experimenter error, 5 were removed due to failure to respond correctly to at least 75% of the comprehension questions, and 2 were removed because of exceedingly high word skipping rate (above 80%). Thus, data was analyzed for a total of 120 participants.

Stimuli

The experimental stimuli were 48 short “stories” consisting of an introductory sentence (1), a context sentence (2), and a target sentence (3-5). The introductory sentence established the broader discourse context. The context sentence included a context word (e.g., “lunch”) that set a narrower context leading to particular predictions (e.g., setting up the context of making lunch, leading to the prediction of plausible lunch items). The target sentence began with either a wh-cleft focus construction (3; “What he made was…”), a “neutral” construction (4; “He made…”) or a subordinate clause de-focusing construction (5; “After he made…”). The object of this construction was the target word, which was either predictable given the preceding context (e.g., “sandwich” in the context of “lunch”) or unpredictable but plausible (“sandwich” in the context of “breakfast”). In all conditions, the target word was preceded by the same adjective (e.g., “tasty”), and followed by the same parenthetical clause; this ensured that the target was surrounded by the same words in all conditions.

Example stimulus:

(1) John was trying to concentrate on his homework, but kept getting distracted.

(2) He realized it was probably because he was hungry, so he made himself some **lunch/breakfast**.

(3) <u>Focus</u>: *What he made was* a tasty **sandwich/omelet**, which didn’t take long. He then decided to take a break.

(4) <u>Neutral</u>: *He made* a tasty **sandwich/omelet**, which didn’t take long. He then decided to take a break.

(5) <u>De-focus</u>: *After he made* a tasty **sandwich/omelet**, which didn’t take long, he decided to take a break.

Each story was constructed so that it could lead to two different possible targets (“sandwich” / “omelet”) associated with two different context words (“lunch” / “breakfast”), and in such a way that switching the context word would lead the same target to be either predictable or unpredictable, yet plausible. Thus, as each story could be constructed around either of two target words, there were a total of 96 unique targets. For each story, the target was manipulated for Predictability (Predictable, Unpredictable) and Focus construction (Focused, Neutral, De-focused), leading to 6 experimental conditions. Each participant was presented with only one of the two versions of each story (based on either of the two target words), thus reading a total of 48 experimental items. Participants saw each item in one of the six experimental conditions, which were counterbalanced across twelve lists.

Additionally, 48 filler items matched for length were adapted from the experimental stimuli of previous studies (Boudewyn et al., 2012; Cook & Myers, 2004; Fraundorf et al., 2013) and consisted of plausible and predictable short discourses.

Stimulus norming

All experimental items were normed for cloze probability and plausibility. Both norming studies were conducted in two rounds, as items that were deemed unusable through the results of the first round of norming (e.g., were too unconstraining) were edited or replaced in the second round until we reached a total of 48 usable items. The norming studies were conducted through Qualtrics surveys administered in person.

In the cloze norming task, items were cropped at the target word and participants were asked to fill in the blank with the word that best fit the preceding context. Items were counterbalanced for context word and focus condition across participants. The results of the cloze task were used to select the target word for each item in each context; the selected targets always consisted of the word with the highest cloze probability in the predictable context, while presenting a low (below 20%) cloze probability in the unpredictable context. The selected target words presented an average cloze probability of 49% (*SD* = 23%) in the predictable context, and of 3% (*SD* = 4%) in the unpredictable context.

In the plausibility norming task, a different set of participants read the completed items and were asked to rate "how much sense" they made on a Likert scale from 1 (Makes no sense at all) to 7 (Makes perfect sense). Items were counterbalanced for target word, focus condition and predictability across participants. The target words presented an average plausibility rating of 6.04 (*SD* = 1.32) in the predictable context, and of 4.37 (*SD* = 2.06) in the unpredictable context. Thus, while the targets were on average rated as less plausible in the unpredictable context, they were not rated as entirely implausible. None of the participants who took part in the norming studies participated in the eye-tracking experiment.

Procedure

In the eye-tracking experiment, participants were instructed to read each story silently at a normal pace. Eye-movements were recorded using an Eye-link 1000 eye-tracking system (SR Research, Osgoode, Ontario, Canada) at a sampling rate of 1000Hz. While viewing was binocular, only the right eye was tracked. A headrest and chinrest prevented participants from moving during the recording sessions.

Each item was presented in its entirety on a 1,024 x 768 display monitor. The target sentence always began on a new line to ensure that the target word was presented at approximately the same location on the screen for all items. Each trial began with a fixation cross positioned at the top left corner of the screen where the stimulus appeared once the participant's gaze was stable. On one third of the trials (16 experimental items, 16 filler items), the story was followed by a comprehension question to which participants were prompted to respond either "true" or "false" using a handheld button box. Comprehension questions were constructed to be relatively challenging, requiring some inferencing, and never specifically tested recollection of the target word but rather aspects of the discourse context (e.g., "John went back to his homework right after making his meal. True/False").

The experiment began with a calibration procedure followed by five practice trials. Next, participants read all stimuli presented in random order. Recalibration between trials was performed when necessary. The entire experiment lasted approximately 45 minutes.

Analysis

Eye-tracking measures

The dependent variable in this experiment was reading time for the target words (e.g., “sandwich”). However, there are multiple ways to quantify reading time through eye-tracking, as each word can be either fixated once, multiple times, or skipped (not fixated) altogether; readers also frequently regress to previously fixated words (Clifton et al., 2007). The data analysis focused on three well-established eye-tracking measures, associated with early processing, late-early processing, and late processing, respectively.

First Fixation Duration (FFD), or the duration in ms of the first fixation on the target word. FFD is thought of as a measure of early lexical processing and it has been shown to be longer for unpredictable compared to predictable words (Clifton et al., 2007).

First Pass Duration (also called *Gaze Duration*), or the sum of the durations of all fixations on the target word the first time the target word was read. For example, if the reader’s eyes fixate the target word on two locations and then continue reading the next words, First Pass will be the sum of the durations of those two fixations (in the case when the reader only fixates the word once before passing onto the next word, First Pass is equivalent to FFD). First Pass also reflects early processing (but not as early as FFD) and it has been shown to be longer for unpredictable than predictable words (Clifton et al., 2007), as well as for syntactically focused compared to de-focused words (Lowder & Gordon, 2015).

Go-Past Duration (also called *Regression Path Duration*), or the sum of the duration of all fixations the first time the target word is read (First Pass) plus the duration of any additional fixations made during the regression to previous words, before passing onto the next words. Go-Past reflects late processing and it has been shown to be longer for syntactically focused than de-focused words (Lowder & Gordon, 2015).

While these were chosen as the three primary measures for the analyses, we also conducted exploratory analyses on other eye-tracking measures, which typically follow similar patterns. These include *Total Time Duration*, or the sum of the durations of all fixations on the target word, including when the reader goes on to read the next word and then regresses back to fixate the target word again. *Skipping Rate*, or the percentage of trials where the reader skipped the target word, thus not fixating it. Skipping Rate has been shown to be higher for predictable than unpredictable words (Clifton et al., 2007). Lastly, *Regression-Out Rate*, or the percentage of trials where the reader made a regression to previous words after having fixated the target word the first time.

Data pre-processing

Following standard pre-processing procedures for eye-tracking data, all data was manually "cleaned" by trained research assistants. Using data filters in EyeLink software, nearby fixations shorter than 80 ms and within 1 character from each other were combined. Fixations longer than 1500 ms were discarded from the analyses. The research assistants corrected the vertical location of fixations that were displaced due to calibration errors; the position of fixations was only adjusted vertically and never horizontally, which could result in fixations being associated with different target regions. All of the trained research assistants were naïve to the hypotheses of the study. A total of 15 trials were discarded from the analyses due to excessive calibration error.

Additionally, the analyses for all eye-tracking duration measures except for Skipping Rate and Regression-Out Rate were conducted only on trials where the target word was not skipped during first pass reading. The target word was skipped for a total of 2157 trials (37% of all trials). Thus, only data from the remaining 63% of trials was used in the analyses for all measures of fixation duration.

Models

All measures were analyzed through mixed effects models including fixed effects of Focus construction (Focused, Neutral, De-focused) and Predictability (Predictable, Unpredictable) and their interaction. The fixed effects were sum-coded, such that each condition was compared to the grand mean. The factors were re-leveled in order to estimate all contrasts (dropped = Neutral or dropped = Focused).

The models initially presented maximal random effects structure with by-subject and by-item random intercept and random slopes for Focus and Predictability. If this maximal model failed to converge or led to singular fit, the random effects structure was simplified by sequentially removing the random effects associated with the least amount of variance (Barr et al., 2013).

In addition to the interaction term included in the models, the interaction between Focus construction and Predictability was further tested by comparing a model with an interaction to a model without an interaction using a log likelihood ratio test. If the model with the interaction presented a significantly better fit of the data, we considered the interaction to be significant.

Lastly, the two percentage measures (Skipping Rate and Regression-Out Rate) were analyzed through logistic mixed effects models using a binomial distribution, as each data point (for each trial) would either be 0 (e.g., target was not skipped) or 1 (e.g., target was skipped).

Results

Primary analyses

The results of the primary analyses on First Fixation Duration, First Pass Duration and Go-Past Duration are visualized in Figure 2.1.

First Fixation Duration

The mixed effects model on First Fixation Duration found a significant effect of Predictability, with faster reading times in the Predictable compared to the Unpredictable condition (β = -5.05, *SE* = 1.24, $p < 0.0001$) and no effect of Focus construction. There was a marginally significant interaction, with a smaller effect of Predictability in the Focused condition (β = 3.08, *SE* = 1.76, $p = 0.08$), while all other interaction contrasts were not significant. Model comparison revealed that the model with the interaction did not present a significantly better fit of the data compared to the model without the interaction, indicating that the interaction was overall not significant.

First Pass Duration

The mixed effects model on First Pass Duration showed a significant effect of Predictability, with faster reading times in the Predictable compared to the Unpredictable condition (β = -8.93, *SE* = 1.88, $p < 0.0001$). There were no significant effects of Focus construction, but reading times were marginally faster in the Neutral condition (β = -4.52, *SE* = 2.67, $p = 0.09$). There were significant interactions, with a smaller effect of Predictability in the

Focused condition (β = 5.98, *SE* = 2.66, $p < 0.05$), and a larger effect of Predictability in the De-focused condition (β = -5.38, *SE* = 2.65, $p < 0.05$). Model comparison revealed that the model with the interaction presented a significantly better fit than the model without the interaction ($\chi^2(2) = 6.17$, $p < 0.05$), indicating that the interaction was overall significant.

Post-hoc pairwise comparisons were conducted using the *emmeans* package in R with Tukey HSD correction (Lenth et al., 2018).The comparisons revealed that Predictable words received significantly shorter reading times than Unpredictable words in the De-focused condition (β = -28.6, *SE* = 6.49, $p < 0.0001$) and Neutral condition (β = -19.1, *SE* = 6.54, $p < 0.01$), but not in the Focused condition (β = -5.9, *SE* = 6.5, $p = 0.36$).

Go-Past Duration

The mixed effects model on Go-Past Duration showed a significant effect of Predictability (β = -20.55, *SE* = 6.29, $p < 0.01$), no effect of Focus construction, and no significant interactions. Model comparison revealed that the model with the interaction did not present a significantly better fit of the data compared to the model without the interaction, indicating that the interaction was overall not significant.

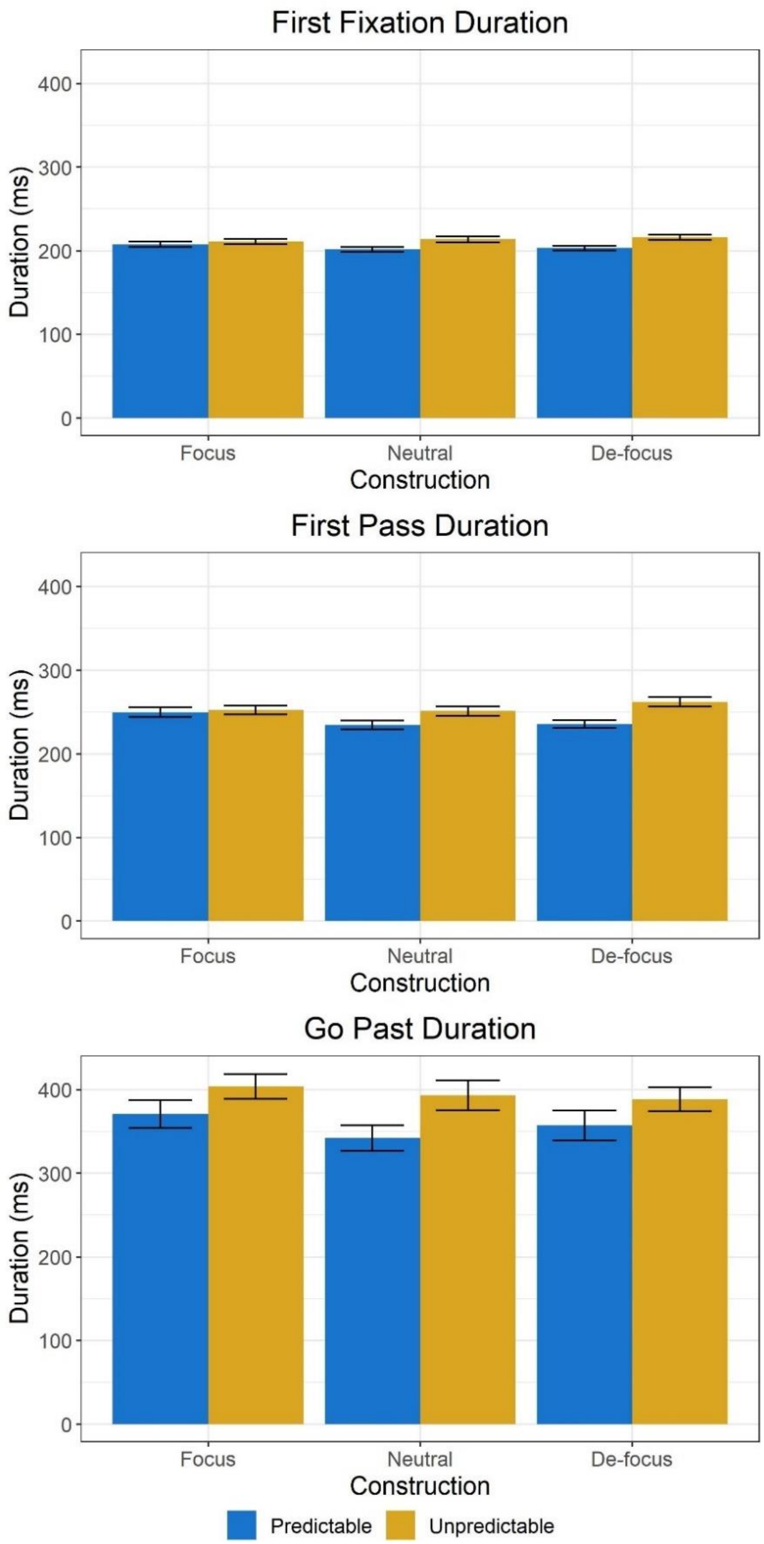


Figure 2.1. Effects of Predictability and Focus on First Fixation Duration, First Pass Duration, and Go-Past Duration. Error bars represent Standard Error of the Mean.

Exploratory analyses

The results of the exploratory analyses on Total Time Duration, Skipping Rate, and Regression-Out Rate are visualized in Figure 2.2.

Total Time Duration

The mixed effects model on Total Time Duration showed a significant effect of Predictability (β = -24.4, *SE* = 3.42, $p < 0.0001$) and no effect of Focus construction. There was a significant interaction, with a smaller effect of Predictability in the Focused condition (β = 11.03, *SE* = 4.84, $p < 0.05$), similarly to First Pass Duration. Model comparison revealed that the model with the interaction presented only a marginally better fit than the model without the interaction ($\chi^2(2) = 5.24$, $p = 0.07$).

Skipping Rate

The mixed effects model on Skipping Rate showed a significant effect of Predictability, with greater skipping for Predictable compared to Unpredictable words (β = 0.12, *SE* = 0.03, $p < 0.0001$), no effect of Focus construction and no interactions. Model comparison indicated that the interaction was overall not significant.

Regression-Out Rate

The mixed effects model on Regression-Out Rate found no effects of either Predictability or Focus construction, and no interactions. Model comparison indicated that the interaction was overall not significant.

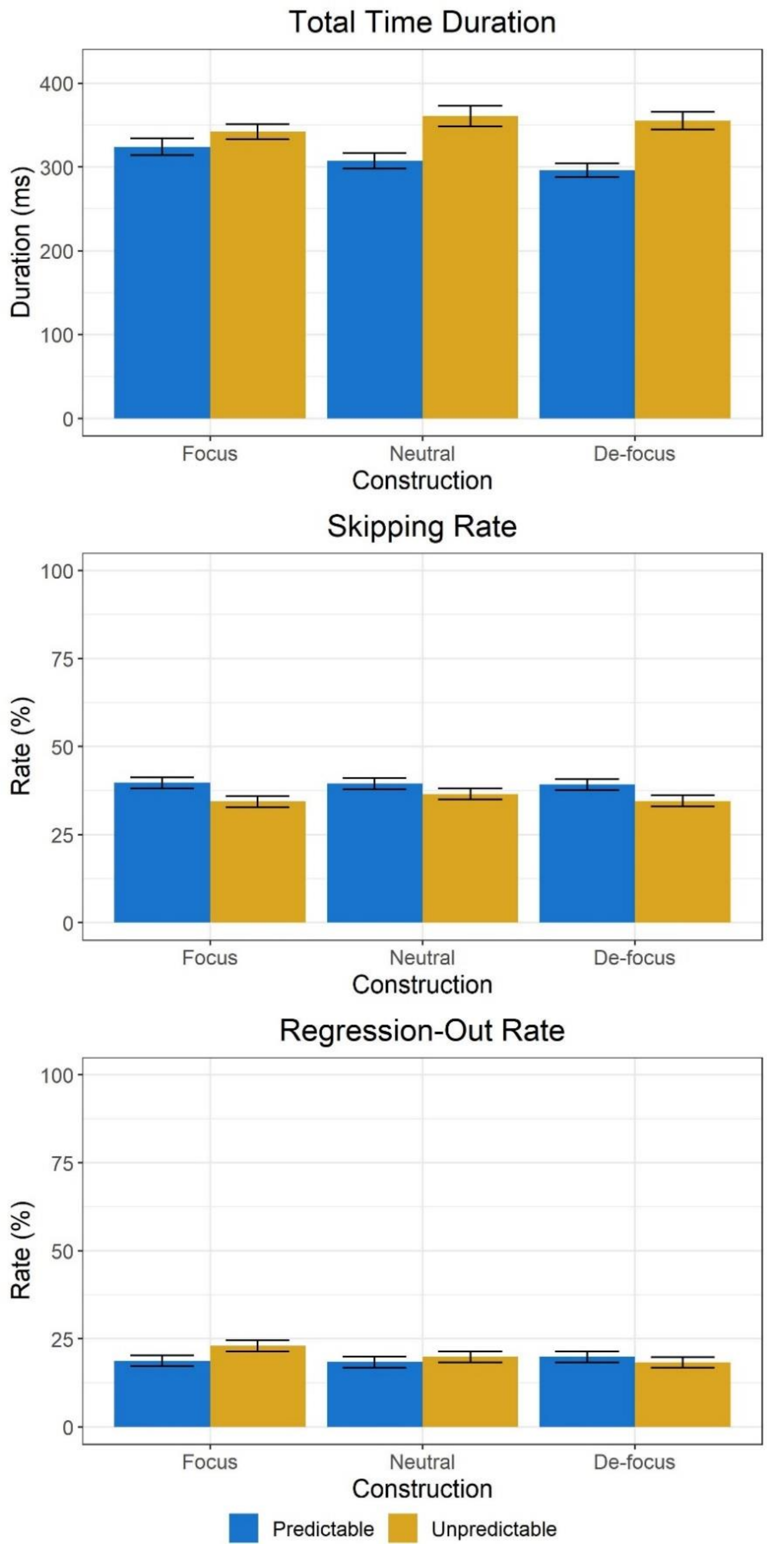


Figure 2.2. Effects of Predictability and Focus on Total Time Duration, Skipping Rate, and Regression-Out Rate. Error bars represent Standard Error of the Mean.

Discussion

The results of Experiment 1 replicated the well-established effect of predictability on reading times. They did not replicate previous findings for longer reading times on syntactically focused words (Lowder & Gordon, 2015), although there was a marginally significant trend for predictable words to receive longer reading times when focused; it is possible that this study did not have enough power to detect this small difference. Most importantly, there was an interaction between predictability and focus for First Pass Duration, the sum of the durations of all fixations on the target word the first time it was read. This measure reflects early processing (but not as early as First Fixation Duration) and it has been shown to be influenced by both predictability (Clifton et al., 2007) and syntactic focus (Lowder & Gordon, 2015); thus, it is not surprising that this measure reflected their interaction most strongly.

The interaction revealed that the effect of predictability was largest for de-focused words and smallest for focused words. At first glance, this pattern fits best with the More Resources account, implying that readers' processing of the target words was facilitated in the focusing condition thanks to the pre-allocation of attention and processing resources, which led to a smaller effect of predictability. On the other hand, processing was hindered for de-focused words, leading to a larger effect of predictability.

However, other contributing factors may explain this pattern. The smaller effect of predictability observed for focused words appears to be driven by longer reading times for predictable words. This suggests that rather than the processing of *unpredictable* words being facilitated thanks to greater processing resources, this smaller effect could be due to deeper processing of *predictable* words induced by the focusing of attention. Another possibility is that the focusing construction altered the types of predictions participants made about upcoming words, thus rendering the predictable word less predictable than in the neutral construction.

At the same time, the greater effect of predictability found for the de-focusing construction, compared to the neutral construction, appears to be driven by longer reading times on the unpredictable word. Thus, it may be that multiple mechanisms produced this pattern of results: compared to the neutral construction, the focusing construction led to longer reading times on predictable words (either as a result of deeper processing or altered semantic predictions); while the de-focusing construction led to longer reading times on unpredictable words, reflecting the cost of reduced processing resources.

In summary, Experiment 1 established that syntactic focus constructions interact with the effect of semantic predictability on reading times during early processing. In particular, there is a greater effect of predictability for de-focused words and a smaller effect for focused words. While this pattern of results fits best with the More Resources account, it does not fully rule out the contribution of other hypothesized mechanisms, including depth of processing and altered semantic predictions. Thus, more research is needed to isolate each of these potential mechanisms in order to estimate their relative contribution at different stages of processing, and with respect to the processing of predictable as opposed to unpredictable words. I further elaborate on these considerations and on some planned future directions in the *General Discussion* (Chapter 5).

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Boudewyn, M. A., Gordon, P. C., Long, D., Polse, L., & Swaab, T. Y. (2012). Does discourse congruence influence spoken language comprehension before lexical association? Evidence from event-related potentials. *Language and Cognitive Processes*, *27*(5), 698–733. https://doi.org/10.1080/01690965.2011.577980

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. *Eye Movements*, 341–371. https://doi.org/10.1016/B978-008044980-7/50017-3

Cook, A. E., & Myers, J. L. (2004). Processing discourse roles in scripted narratives: The influences of context and world knowledge. *Journal of Memory and Language*, *50*(3), 268–288. https://doi.org/10.1016/j.jml.2003.11.003

Ferreira, F., & Lowder, M. W. (2016). Prediction, information structure, and good-enough language processing. *Psychology of Learning and Motivation*, *65*, 217–247. https://doi.org/10.1016/bs.plm.2016.04.002

Fraundorf, S. H., Benjamin, A. S., & Watson, D. G. (2013). What happened (and what did not): Discourse constraints on encoding of plausible alternatives. *Journal of Memory and Language*, *69*(3), 196–227. https://doi.org/10.1016/j.jml.2013.06.003

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). *Emmeans: Estimated marginal means, aka least-squares means. R package version 1.1.2.*

Lowder, M. W., & Gordon, P. C. (2015). Focus takes time: Structural effects on reading. *Psychonomic Bulletin & Review*, *22*(6), 1733–1738. https://doi.org/10.3758/s13423-015-0843-2

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia*, *50*(2), 232–244. https://doi.org/10.1016/j.neuropsychologia.2011.10.025

Wang, L., Bastiaansen, M., Yang, Y., & Hagoort, P. (2011). The influence of information structure on the depth of semantic processing: How focus and pitch accent determine the size of the N400 effect. *Neuropsychologia*, *49*(5), 813–820. https://doi.org/10.1016/j.neuropsychologia.2010.12.035

Wang, L., Hagoort, P., & Yang, Y. (2009). Semantic illusion depends on information structure: ERP evidence. *Brain Research*, *1282*, 50–56. https://doi.org/10.1016/j.brainres.2009.05.069

CHAPTER 3: SELF-PACED READING STUDY

Introduction

This chapter describes the results of a replication attempt of Experiment 1 (Chapter 2) using the self-paced reading paradigm. We attempted to replicate the interaction between semantic predictability and focus construction using a larger sample with greater statistical power, in order to provide stronger support for this pattern of results. Self-paced reading was selected as an alternative to eye-tracking due to the following reasons. First, the onset of the Covid-19 pandemic in early 2020 demanded a switch from in-person to online methods; self-paced reading can easily be implemented online, while online eye-tracking methods are less available and produce less precise data than in-person collection. Second, self-paced reading data is more easily acquired and requires fewer time-consuming pre-processing steps (e.g., data cleaning; see Chapter 2), allowing us to collect a much larger sample of data than in Experiment 1.

Self-paced reading is an established measure of on-line processing difficulty (Jegerski, 2014; Marsden et al., 2018). In this task, participants read the same sentences as Experiment 1 one word at a time, pressing a button to reveal each word. Self-paced reading times have been shown to be influenced both by semantic predictability (De Vincenzi et al., 2003; Ditman et al., 2007; Smith & Levy, 2013) and syntactic manipulations (e.g., De Vincenzi et al., 2003; Ditman et al., 2007). While it is thought to reflect overall similar processes as reading times measured through eye-tracking (Smith & Levy, 2013), processing effects are differently distributed over the course of the sentence for self-paced and eye-tracking data. Morphosyntactic effects are observed at the target word for both self-paced and eye-tracking data, while semantic predictability effects are delayed for self-paced data compared to eye-tracking data; for self-paced data, lower predictability leads to longer reading times for the words following the target word (spillover regions) and for the sentence-final word (De Vincenzi et al., 2003; Ditman et al., 2007; Jegerski, 2014; Smith & Levy, 2013). Self-paced reading data is also related to ERP measures, such that larger N400 amplitudes at the target word predict longer sentence-final reading times (Ditman et al., 2007). Thus, self-paced reading is an appropriate task to attempt to replicate the interaction between semantic predictability and syntactic focus found in Experiment 1.

Method

Participants

Participants consisted of 339 undergraduate students at the University of California, Davis recruited online through the SONA participant pool and compensated with class credit. All participants were native English speakers. A total of 87 participants were removed from the analyses due to failure to respond correctly to at least 75% of the comprehension questions. Thus, data was analyzed for a total of 252 participants. Data collection took place from May to November 2020.

Stimuli

The stimuli consisted of the same set of 48 short "stories" and 48 filler items used in Experiment 1. Please refer to Chapter 2 for details.

Procedure

The experiment was conducted online through the Ibex Farm server (Drummond et al., 2016), with code from PCIbex (Zehr & Schwarz, 2018). Following the experiment instructions and five practice trials, participants read each item one word at a time, pressing the spacebar to reveal the next word. Prior to being revealed, each word was masked by underlining corresponding to the word's length; thus, participants were aware of the number and length of upcoming words before revealing them. By pressing the spacebar and revealing the next word, the previously revealed word was again masked with underlining, such that participants could not re-read previously revealed words. Like in Experiment 1, one third of the trials (16 experimental items, 16 filler items) was followed a comprehension question, to which participants could answer "true" or "false" using two keys on their keyboard or by clicking the answer on the screen. The entire experiment lasted approximately 45 minutes.

Analysis

Critical regions

Based on prior research, we expected the effect of semantic predictability to be observed not on the target word, but in the words immediately following it (spillover regions) and in the sentence or phrase-final word (De Vincenzi et al., 2003; Ditman et al., 2007; Jegerski, 2014; Smith & Levy, 2013). To the best of our knowledge, the effect of syntactic focus constructions on self-paced reading times has not been investigated, so it is unclear where during the sentence this effect will be observed; however, as it constitutes a syntactic manipulation, focusing constructions may influence reading times at the target word (De Vincenzi et al., 2003; Ditman et al., 2007).

As such, in addition to the reading times on the target word (e.g., Target region: "sandwich"; see (1) for full example), we also analyzed reading times for each of the three words following it (e.g., Spillover region 1: "which"; Spillover region 2: "didn't"; Spillover region 3:

“take”), which were the same across all conditions. Lastly, we analyzed reading times for the word that ended the parenthetical phrase following the target word (e.g., Phrase-final region: “long”) as well as for the word that ended the entire story/discourse (e.g., Discourse-final region: “break”).

(1) John was trying to concentrate on his homework, but kept getting distracted. He realized it was probably because he was hungry, so he made himself some lunch. What he made was a tasty sandwich, which didn’t take long. He then decided to take a break.

Data pre-processing

Reading times in a self-paced reading experiment constitute a form of reaction times (RTs), as they reflect the time participants take to respond to each word by pressing the spacebar, moving on to the next word. Therefore, they present the right-skewed distribution typical of RTs across different tasks and paradigms (Baayen & Milin, 2010; Payne & Federmeier, 2017). To deal with this non-normal distribution (which breaks the assumptions of the linear models we planned to use), we performed the analyses on log-transformed RTs, which present a more normal distribution. Prior to transformation, we also removed any outliers (3 SDs above or below the mean RTs for each participant) as well as any RTs below 100 ms or above 2000 ms. These pre-processing steps (trimming and log transformation) follow current practice and recommendations for self-paced reading paradigms (Dempsey et al., 2020; Jegerski, 2014; Marsden et al., 2018; Prasad & Linzen, 2021).

Models

Log-transformed RTs for each critical region were separately analyzed using linear mixed effects models. Similarly to Experiment 1, the models included fixed effects of Focus construction (Focused, Neutral, De-focused) and Predictability (Predictable, Unpredictable) and their interaction. The fixed effects were sum-coded, such that each condition was compared to the grand mean. The factors were re-leveled in order to estimate all contrasts (dropped = Neutral or dropped = Focused). The determination of the random effects structure and the testing for the significance of the interaction using model comparison were performed similarly to Experiment 1 (see Chapter 2).

Results

Reading times for all critical regions are visualized in Figure 3.1.

Target region

Reading times on the Target region were slower for Focused words ($\beta = 0.009$, $SE = 0.004$, $p < 0.05$) and faster for De-Focused words ($\beta = -0.01$, $SE = 0.004$, $p < 0.01$), while there was no effect of Predictability. However, there were significant interactions between Predictability and Focus constructions. Reading times were slower for Predictable words in the Neutral condition ($\beta = 0.007$, $SE = 0.004$, $p < 0.05$) while they were faster for Predictable words in the De-Focused condition ($\beta = -0.007$, $SE = 0.004$, $p < 0.05$). Model comparison revealed that the model with the interaction presented only a marginally better fit than the model without the interaction ($\chi^2(2) = 5.39$, $p = 0.07$).

Spillover region 1

Reading times in the Spillover region 1 (one word after the target) were faster for words in the Neutral condition ($\beta = -0.008$, $SE = 0.004$, $p < 0.05$), while there were no other effects of Focus and only a marginally significant effect of Predictability, with faster reading times for words in the Predictable target condition ($\beta = -0.005$, $SE = 0.003$, $p = 0.07$). There was a marginally significant interaction, with a greater effect of Predictability in the De-Focused condition ($\beta = -0.006$, $SE = 0.004$, $p = 0.09$). Model comparison revealed that the overall interaction was not significant.

Spillover region 2

Reading times in the Spillover region 2 (two words after the target) were faster in the Neutral condition ($\beta = -0.01$, $SE = 0.003$, $p < 0.001$) and slower in the De-Focused condition ($\beta = 0.01$, $SE = 0.003$, $p < 0.01$). There was a significant effect of Predictability, with faster reading times for words in the Predictable target condition ($\beta = -0.01$, $SE = 0.002$, $p < 0.0001$). There was also a marginally significant interaction, with a smaller effect of Predictability in the Neutral condition ($\beta = 0.006$, $SE = 0.003$, $p = 0.06$). Model comparison revealed that the overall interaction was not significant.

Spillover region 3

There were no effects of Focus construction on reading times in the Spillover region 3 (three words after the target), while there was an effect of Predictability, with faster reading times for words in the Predictable target condition ($\beta = -0.005$, $SE = 0.002$, $p < 0.05$) and no interactions. Model comparison revealed that the overall interaction was not significant.

Phrase-final region

Reading times in the Phrase-final region were faster for words in the Focused condition (β = -0.008, *SE* = 0.003, $p < 0.05$) and slower for words in the De-Focused condition (β = 0.01, *SE* = 0.003, $p < 0.01$). There was also an effect of Predictability, with faster reading times for words in the Predictable target condition (β = -0.006, *SE* = 0.002, $p < 0.01$). Lastly, there was a marginally significant interaction, with a smaller effect of Predictability in the Neutral condition (β = -0.006, *SE* = 0.003, $p = 0.0507$). Model comparison revealed that the overall interaction was not significant.

Discourse-final region

There was only an effect of Predictability on reading times in the Discourse-final region, with faster reading times for words in the Predictable target condition (β = -0.007, *SE* = 0.002, $p < 0.01$), while there were no effects of Focus and no interactions.

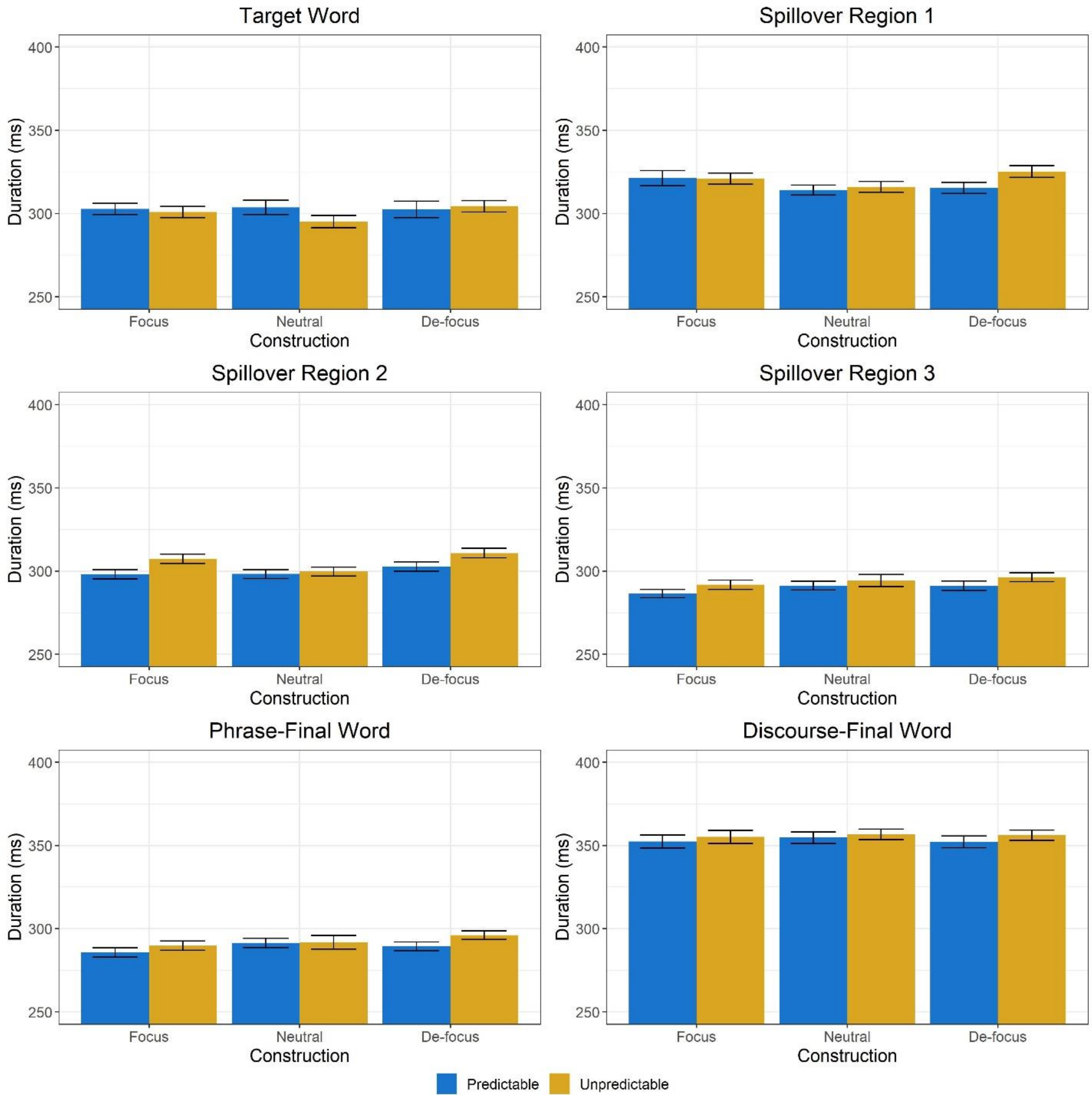


Figure 3.1. Effects of Predictability and Focus on reading times for all critical regions. Error bars represent Standard Error of the Mean. Note that the y-axis begins at 250 ms to facilitate visualization of small differences in means.

Discussion

We attempted to replicate the pattern of results found in Experiment 1 (Chapter 2) using the self-paced reading paradigm. Overall, some but not all the effects were replicated. First, as expected, we replicated the well-established effect of semantic predictability, whereby more predictable words receive shorter reading times than unpredictable words. Like in previous self-paced reading studies, this effect was significant not on the target word itself, but in the spillover regions and in the phrase- and discourse-final words. Notably, the discourse-final word showed an effect of predictability despite being on average 12.45 words from the predictable or unpredictable target word; moreover, in the focused and neutral conditions, the discourse-final word concluded a new sentence following the sentence containing the target word. To the best of our knowledge, previous self-paced reading studies have only reported the effect of predictability at a closer distance (e.g., the sentence-final word was on average only 2.3 words following the critical word in Ditman et al., 2007). Thus, our findings extend prior research by showing that the effect of semantic predictability persists throughout the discourse across multiple sentences. This sustained effect may be due to readers adopting a more attentive reading strategy upon encountering an unpredictable word, leading to greater processing difficulty and more time needed to integrate the target word within the context.

We also found several effects of focus on reading times. Focused target words received longer reading times while de-focused target words received shorter reading times; this pattern was only marginally significant in Experiment 1, and it replicates previous eye-tracking findings (Lowder & Gordon, 2015). It also further establishes the importance of including a neutral baseline, as the effect was driven by both the focusing and the de-focusing constructions (Birch & Garnsey, 1995; Lowder & Gordon, 2015). Focus constructions also affected reading times in the spillover and phrase-final regions, although in a different direction than for the target word. In the spillover regions 1 and 2, words in the neutral condition received faster reading times compared to words in both the focused and de-focused conditions. In the phrase-final region, the pattern was opposite that of the target word: reading times were fastest in the focused condition and slowest in the de-focused condition. One possible interpretation is that focusing constructions led readers to pay more attention to the target word (resulting in longer reading times); this effect extended to the two spillover regions immediately following it but eventually reversed as readers paid less attention to the parenthetical. When the target word was de-focused, on the other hand, readers paid less attention to it (resulting in shorter reading times) and instead focused their attention on the parenthetical following it. This would fit with the idea that the de-focused subordinate clause was shallowly processed and led readers to look for the new, important information in the phrase following it (F. Ferreira & Lowder, 2016). However, this interpretation is only speculative and would need to be further tested. Overall, these results suggest that focus constructions influence self-paced reading times in a complex manner, potentially reflecting multiple mechanisms and the dynamic allocation of attention throughout the sentence.

Lastly, the interaction between semantic predictability and focus construction was not replicated in this experiment. While the overall interaction was never significant when tested through model comparison, there were a few significant and marginal interaction contrasts in the model outputs. For the target word, reading times were slower for predictable words in the neutral condition; however, as the effect of predictability is not typically observed on the target word itself, it is unclear whether this interaction is meaningful. In the spillover region 1, the pattern of results is visually similar to the pattern found for First Pass in Experiment 1; however, both the effect of predictability and the interaction were only marginally significant. Lastly, for the spillover region 2 and the phrase-final region, the effect of predictability was marginally smaller in the neutral construction compared to both the focused and de-focused conditions. Overall, this data did not provide clear evidence for an interaction.

There are several possible explanations for the failure to replicate the interaction found in Experiment 1. First, although Experiment 2 had a larger sample size, the statistical power may still have been too low to detect this interaction. This might be because, compared to the eye-tracking data, the data acquired through the self-paced reading paradigm may have been more noisy due to a variety of reasons. Self-paced reading is generally an unnatural, constrained task which can lead to task-specific effects (Jegerski, 2014). Compared to the natural reading patterns measured with eye-tracking, which involve frequent skipping and regressions to previously read words, the self-paced paradigm forces readers to fixate each word sequentially. Moreover, the parafoveal preview of upcoming words can influence natural reading patterns (Schotter et al., 2012), but this preview is masked for self-paced reading. Thus, the unnaturalness of the task may have interfered with the effects of predictability and focus. The data may also have been made noisier due to the experiment being conducted online rather than in person. In Experiment 1, the exact location of the target word on the screen was carefully controlled; this was not the case for Experiment 2 because of differences in screen size and resolution across participants. It has been shown that previous inconsistent findings with respect to the effect of focus on eye-tracking reading times could be due to the confound between focus constructions and the location of target words on the screen, leading to oculomotor differences (Lowder & Gordon, 2015); thus, it is possible that the effect of focus and its interaction with predictability were obscured by the inability to control the location of the target word across participants. Lastly, participants in this online study may also have been less attentive and engaged than participants in Experiment 1. This is supported by the average comprehension question accuracy, which was 88.2% (*SD* = 7.3%) in Experiment 1 and 80.7% (*SD* = 14%) in Experiment 2, before data exclusions. In summary, it is possible that, despite the larger sample size, there was still not enough power to detect the interaction between focus and predictability due to the data being noisier.

A second potential reason for the failure to detect the interaction could be the fact that focus and predictability influenced reading times at different points throughout the sentence. The expected effect of focus (slower reading times for focused compared to de-focused words) was only observed at the target word; however, reading times for the target word were not influenced by predictability. On the other hand, the expected effect of predictability was found for the spillover, phrase-final and discourse-final regions; but the effect of focus construction for these regions reflected a more complex and varying pattern. Thus, it may be that the self-paced paradigm is simply not an appropriate task to measure this interaction, as semantic and syntactic manipulations have been found to influence reading times at different regions (De Vincenzi et al., 2003; Ditman et al., 2007). The only region with a visually similar pattern to Experiment 1 was the spillover region 1; this area would be a good candidate for a potential interaction, as it is the first region conforming to the expected effect of predictability, and it may exhibit some spillover effect of focus. A replication with greater power is needed to determine whether the pattern found in the spillover region 1 reflected a true interaction.

Lastly, it is possible that the interaction found in Experiment 1 was a false positive. To rule out this possibility, it would be necessary to attempt to replicate Experiment 1 using in-person eye-tracking data, in order to avoid the issues introduced by running the experiment online using self-paced reading. Nonetheless, the self-paced reading paradigm revealed a new set of interesting patterns that could be further explored, such as the developing effect of focus over the sentence.

Overall, while it did not fully replicate Experiment 1, this self-paced reading study provided new insights into the effects of focus and predictability. First, we extended previous literature by showing that the effect of semantic predictability persists throughout the discourse, reflecting greater processing difficulty and a more attentive reading strategy induced by the unpredictable target word. Second, we established that syntactic focusing and de-focusing constructions influence self-paced reading times, which, to the best of our knowledge, has not been shown before. Similarly to previous eye-tracking studies (Lowder & Gordon, 2015), focused target words were read more slowly while de-focused target words more quickly, suggesting that both focusing and de-focusing constructions influence depth of processing and attention. However, this effect reversed throughout the parenthetical following the target word, suggesting that readers pay more attention to words following the de-focused than focused target words. This dynamic allocation of attention and processing resources would need to be further studied in order to discern the specific mechanisms involved (e.g., see *Future Directions* in Chapter 5). Lastly, we did not find strong support for an interaction between focus and predictability using self-paced reading, but our results suggest that there may also be nuanced and evolving interactions that might be detected with more power and in a more controlled experimental setting.

References

Baayen, R. H., & Milin, P. (2010). Analyzing Reaction Times. *International Journal of Psychological Research*, *3*(2), 12–28.

Birch, S. L., & Garnsey, S. M. (1995). The effect of focus on memory for words in sentences. *Journal of Memory and Language*, *34*(2), 232–267.

De Vincenzi, M., Job, R., Di Matteo, R., Angrilli, A., Penolazzi, B., Ciccarelli, L., & Vespignani, F. (2003). Differences in the perception and time course of syntactic and semantic violations. *Brain and Language*, *85*(2), 280–296. https://doi.org/10.1016/S0093-934X(03)00055-5

Dempsey, J., Liu, Q., & Christianson, K. (2020). Convergent probabilistic cues do not trigger syntactic adaptation: Evidence from self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(10), 1906–1921. https://doi.org/10.1037/xlm0000881

Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2007). An investigation of concurrent ERP and self-paced reading methodologies. *Psychophysiology, 44*, 927–935. https://doi.org/10.1111/j.1469-8986.2007.00593.x

Drummond, A., von der Malsburg, T., Erlewine, M. Y., & Vafaie, M. (2016). *Ibex farm.* GitHub. https://github.com/addrummond/ibex.

Ferreira, F., & Lowder, M. W. (2016). Prediction, information structure, and good-enough language processing. *Psychology of Learning and Motivation*, *65*, 217–247. https://doi.org/10.1016/bs.plm.2016.04.002

Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 20–49). Routledge.

Lowder, M. W., & Gordon, P. C. (2015). Focus takes time: Structural effects on reading. *Psychonomic Bulletin & Review*, *22*(6), 1733–1738. https://doi.org/10.3758/s13423-015-0843-2

Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics*, *39*(5), 861–904. https://doi.org/10.1017/S0142716418000036

Payne, B., & Federmeier, K. D. (2017). Pace Yourself: Intraindividual Variability in Context Use Revealed by Self-paced Event-related Brain Potentials. *Journal of Cognitive Neuroscience*, *29*(5), 837–854.

Prasad, G., & Linzen, T. (2021). Rapid syntactic adaptation in self-paced reading: Detectable, but only with many participants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(7), 1156–1172. https://doi.org/10.1037/xlm0001046

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics*, *74*(1), 5–35. https://doi.org/10.3758/s13414-011-0219-2

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319. https://doi.org/10.1016/j.cognition.2013.02.013

Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX).* https://doi.org/10.17605/OSF.IO/MD832

CHAPTER 4: CUTLER & FODOR (1979) REPLICATION

Introduction

This chapter describes the results of a replication attempt of a foundational study by Cutler and Fodor (1979). This study sparked several lines of research into the role of information structure in language comprehension, which together provide the theoretical and methodological background for this dissertation, as reviewed in Chapter 1. This replication provides a stronger foundation to these diverging lines of research, and it is likely to spark new investigations into information structure. It was originally published as a registered report in a special issue on "Replicating Influential Findings in Memory and Language" in the *Journal of Memory and Language* (Beier & Ferreira, 2022). The *Journal of Memory and Language* has given permission to include these published materials as part of this dissertation.

Cutler and Fodor (1979; hereafter CF79) suggested that there may be several ways for speakers to communicate the location of new information, guiding the listener's attention to it and leading to more efficient communication. CF79 provided a theoretical explanation for previous findings from the literature on prosody, which showed that words receiving prosodic stress/accent were processed faster, as measured by reaction times (RTs) in a phoneme monitoring task (Cutler & Foss, 1977; Shields et al., 1974). For example, one study supported Martin's (1972) claim that listeners could pre-allocate attention to points in time where a stressed syllable was predicted to occur based on the somewhat predictable alternation of stressed and unstressed syllables corresponding to a metrical structure (Shields et al., 1974), an idea later termed the Attentional Bounce Hypothesis (Pitt & Samuel, 1990).

While it could be argued that stressed syllables are easier to process because they are acoustically salient and clearer (Lehiste, 1970), Cutler (1976) demonstrated that faster RTs were found even for acoustically identical words that had been spliced into sentences whose prosodic contour prior to the word was consistent with the presence of stress on the target word, rather than inconsistent with it. This result suggested that listeners used the prosodic structure to predict the temporal location of the word expected to be stressed, which guided their attention to it and resulted in faster RTs. However, it was still unclear what listeners would gain by tracking and predicting prosodic structure, and how this process contributed to comprehension.

CF79 posited that by tracking prosody, listeners were really searching for focused, new information. Prosody closely matches information structure, with new information tending to be marked with higher pitch, greater intensity, and longer duration than given information (Breen et al., 2010; Calhoun, 2010a; Roettger et al., 2019). Therefore, CF79 argued that the previous phoneme monitoring effects induced by prosody were a result of listeners using prosody as a cue to the temporal location of new information, which would facilitate comprehension. If this was the case, then other, non-prosodic cues to the location of new information should similarly lead listeners to pre-allocate their attention to those points in time and result in faster phoneme monitoring RTs.

CF79 tested this prediction by manipulating information structure through question-answer pairs such as those in the example above. Words starting with target phonemes (e.g., /k/ in “corner”, and /d/ in “dark”) were embedded in acoustically identical sentences (e.g., “The man on the corner was wearing the dark hat”). The sentences were preceded by a question that placed focus on the later target (e.g., “What hat was the man wearing?”) or on the earlier target (e.g., “Which man was wearing the hat?”). The main finding was an interaction between target and focus position, whereby RTs were faster when the preceding question placed focus on the target word. This pattern indicated that the preceding question led listeners to pre-allocate attention to the temporal location of information expected to be new, resulting in faster RTs, similarly to the previously reported effects of prosody (Cutler, 1976). The results of CF79 explained these previous findings, suggesting that they were in fact due to listeners using prosody to track information structure. CF79 also briefly discussed syntactic focus as an additional strategy that could be used by the speaker to guide the listener's attention to new information (e.g., "It was the man on the corner who was wearing the hat" vs "What the man was wearing was a dark hat"). Thus, information structure, or the timing of given and new information in language, might be marked by a variety of cues including discourse context (e.g., preceding question), prosodic stress/accent, and syntactic focus.

Following CF79's claim that discourse characteristics, prosody, and syntax might all play a similar role in guiding listeners' attention to new information, several lines of research developed investigating each of these potential influences on processing. However, as reviewed in Chapter 1, beyond a few exceptions, the literature has tended to diverge for each of these features.

The goal of the study described in this chapter is to attempt to replicate the original finding reported in CF79 of the effect of discourse focus on phoneme monitoring RTs using current methods for statistical analysis. To the best of our knowledge, there has not been a replication of CF79's study using the same stimulus materials and paradigm that did not introduce additional manipulations (such as a prosodic manipulation in Akker & Cutler, 2003), and that adopted modern standards of data analysis with a large enough sample size to ensure sufficient power. For example, both CF79 and Akker & Cutler (2003) analyzed non-transformed

RT data using ANOVAs; while this was the norm at the time, this is no longer considered optimal due to the positively right-skewed distribution of RT data, which violates the assumptions of an ANOVA (Baayen & Milin, 2010; Ratcliff, 1993; Whelan, 2008). Additionally, the emergence of linear mixed effects models as an alternative to running two separate ANOVAs on participants and items allows for a more representative and powerful analysis (Baayen et al., 2008; Stevens & Brysbaert, 2016). In this study, we compare the results obtained through an updated statistical analysis (mixed effects models on transformed RTs) to the results obtained through an analysis closer to the original (ANOVA on non-transformed RTs) to assess whether any differences in results are attributable to the kind of analysis performed. If the original findings reflected a true effect of focus, we expect to see compatible results with both kinds of analyses. On the other hand, if the original findings are attributable at least in part to the nature of the statistical analyses that were conducted, we expect the effects to be significant in the ANOVA results but not in the mixed effects models.

A replication of CF79 using modern statistical practices, whether the replication is successful or not, is a first step towards a reconciliation of the diverging literatures sparked by CF79. A successful replication of CF79 would provide a stronger foundation for these lines of research and may lead to new investigations into CF79's claim that discourse characteristics, prosody, and syntactic focus all play similar roles. An unsuccessful replication would also prove useful, as it would lead to a re-evaluation of the effect of discourse focus on language processing and further examination of focus in different types of paradigms. In particular, an unsuccessful replication would make CF79's interpretation of Cutler's (1976) prosodic effects as a search for new information less convincing, requiring further investigation into the distinct effects of prosodic and discourse focus. More broadly, we hope that this replication will revive discussion around CF79's central claims and provide guidance to future research, leading to more direct tests of how discourse, prosodic and syntactic focus interact.

Predictions

If our replication is successful, we predict we will replicate the following effects reported by CF79 for RTs in the phoneme monitoring task:

a. A main effect of Target position, such that RTs to targets in later positions will be faster than RTs in earlier positions.

b. A main effect of Focus position, such that RTs will be faster when the preceding question focused on the earlier part of the sentence than when it focused on the later part of the sentence.

c. Critically, an interaction between Target position and Focus position, such that RTs to the earlier target will be faster in the early focus condition, while RTs to the later target will be faster in the late focus condition.

d. Pairwise comparisons will show a significant difference in RTs between the focused and non-focused conditions for both the early target and the late target.

We also predict we will replicate the following effect on the number of errors in the memory test at the end of the experiment (see *Procedure*):

e. A main effect of whether the tested word had been focused, such that participants will be more likely to remember (make fewer errors for) words that had been focused during the experiment.

Method

Participants

Sample size determination

The sample size was determined through a power analysis based on the smallest effect size of interest. A full description of the effect size estimation and details of the power analysis can be found under *Power analysis* (we opted to report it after a full description of the analysis plan, as knowledge of the statistical analyses we plan to employ is required to understand the power analysis). Through the power analysis, we determined that a sample size of 180 participants would lead to more than 90% power using an alpha level of 0.02.

Sampling criteria

Participants were recruited online through the University of California, Davis SONA participant pool, in order to avoid in-person testing during the Covid-19 pandemic. Previous studies have been successful in replicating in-person, lab-based RT effects with auditory stimuli using online versions of the same experiments (Slote & Strand, 2016). Additionally, because the current study design is entirely within-subjects, any differences in hardware between participants that may lead to differences in RTs are not confounded with the effects of interest. Selection criteria were set so that only undergraduate students who resided in the US or Canada, were native English speakers, and had no hearing impairments were able to access the study. Data collection took place from November 2021 to January 2022.

Exclusion criteria

As the data was collected online and we did not have control over the experimental conditions (e.g., whether participants were paying attention and performing the task in a quiet environment), we employed several criteria to ensure data quality. We excluded participants based on the following criteria:

a. Participants who did not complete the entirety of the study.

b. Participants who did not meet the inclusion criteria (reside in the US or Canada, native English speakers, no hearing impairments), as self-reported by the participants.

c. Participants who performed the experiment in a noisy environment or who reported not having paid close attention during the entirety of the experiment, as self-reported by the participants in a questionnaire administered at the end of the experiment.

d. Participants who did not pass the headphones screening test at the beginning of the experiment (see *Procedure*)

e. Participants who reported not being able to hear the stimuli or having any other technical issues during the phoneme monitoring task, including not having fully understood the task instructions, as self-reported by the participants.

f. Participants who did not perform with at least 80% accuracy on comprehension questions administered throughout the experiment (see *Procedure* for details).

g. Participants whose reaction time data exceeded the limits for exclusion (or who failed to respond altogether) on more than 25% of experimental trials (following Rysling et al., 2020; see *Data exclusions and pre-processing* for data exclusion criteria).

Additionally, we excluded participants who restarted the experiment multiple times. We did not foresee this possibility in our Stage 1 Registered Report, but we believe that this criterion falls under the pre-registered exclusion criteria. While it is not clear why some participants restarted the experiment, we suspect this could be due to either: 1) technical issues or computer error forcing participants to start over (falling under criterion e); or 2) participants being informed that their answers to the initial questionnaire, attesting inclusion criteria, did not qualify them for participation; some participants may have started the experiment over, changing their answers in order to qualify (falling under criterion b). More importantly, participants who restarted the experiment half-way through would have been exposed to the same stimuli multiple times, leading to critical confounds with the experimental manipulations and a significant break from CF79, who only presented each stimulus once. Therefore, we believe this additional exclusion criterion follows from the pre-registered ones and is necessary to the integrity of the replication.

After we collected data for 180 participants, we used these criteria to determine how many participants would have to be removed from the analysis. We then continued to run participants until we reached our goal sample size and a minimum of 1600 observations per cell of our design (see *Power analysis*).

Final sample characteristics

A total of 295 participants took part in the experiment; of these, 200 were retained for the analyses while the rest were excluded due to one of the criteria described above (see Table 4.1). Our final sample size was therefore larger than the originally proposed sample of 180, but it was the minimum sample size required to achieve a minimum of 1600 observations per cell after RT data exclusions. Participants in the final sample had an average age of 20 years (median = 19) and included 139 females, 51 males, 5 participants who entered another gender category (e.g., nonbinary), and 5 participants who did not report gender.

Table 4.1 breaks down the number of participants excluded due to each criterion. While the number of exclusions was higher than we expected (possibly due to the online mode of data collection), we strove to follow our pre-registered exclusion criteria precisely. We believe these conservative criteria allowed us to ensure that only high-quality data from attentive participants were included in the analyses.

Table 4.1.
Number of participants excluded from the analyses for each experimental list.

Reason for exclusion	List 1	List 2	List 3	List 4	Total
Did not complete the experiment	7	3	2	3	15
Reported not having paid attention to the whole experiment and/or having performed the experiment in a noisy environment	7	9	5	14	35
Reported technical issues during the phoneme monitoring task	1	2	1	3	7
Comprehension question accuracy below 80%	1	1	0	2	4
More than 25% of RTs outside exclusion bounds	2	1	3	3	9
Restarted the experiment multiple times	8	5	6	6	25

Design

Following the original design of CF79, this study design consisted of a 2 (Target position: early, late) X 2 (Focus position: early, late) factorial design. While CF79 analyzed their data using 2 X 2 ANOVAs, our primary analysis used mixed effects models to achieve a better characterization of by-item and by-subject random variance, leading to greater power (Baayen & Milin, 2010; Brysbaert & Stevens, 2018; Stevens & Brysbaert, 2016). The model included fixed effects of Target position and Focus position and their interaction. We also performed two 2 X 2

ANOVAs (one with subjects and the other with items as random effect), consistent with what was done in the original analyses, in order to test whether any differences in the results are due to the use of different analyses (see *Analysis plan* for further details). Each stimulus was presented in one of the 4 possible experimental conditions resulting from this 2 X 2 design. Similarly to CF79, stimuli were counterbalanced across four lists such that each participant was only presented with each stimulus once, and with an equal number of stimuli in each condition.

Stimuli

Experimental items

We used the list of 32 experimental items reported by CF79 with the addition of four structurally similar stimuli reported by Akker & Cutler (2003), for a total of 36 stimuli. We opted to increase the original number of stimuli to 36 to increase power. We also chose not to go beyond those 36 items to ensure the integrity of this replication attempt. Each item consisted of a sentence preceded by one of two questions, each placing focus on a different target word in the sentence, such as in the following example (target words are bolded):

Early focus question: Which man was wearing the hat?

Late focus question: What hat was the man wearing?

Sentence: The man on the **corner** was wearing the **dark** hat.

In this example, the early target (“corner”) is focused through the early focus question, while the late target (“dark”) is focused through the late focus question. All target words began with one of three target phonemes to be detected: /k/ (e.g., “corner”), /d/ (e.g., “dark”) or /b/ (e.g., “bag”). Additionally, we modified a few of the original stimuli in which the target phoneme was presented as part of a consonant cluster (e.g., “Britain”), which has been found to influence phoneme detection (Cutler et al., 1987). We replaced those items with words starting with a single consonant which constituted the target phoneme (e.g., “Belgium”)[1].

Distractors

CF79 report presenting participants with an additional 32 distractor items (often referred to as “filler sentences”), but they did not provide a list of these items. We therefore created a new list of 36 distractor items. Distractor items consisted of a sentence preceded by a question, in order to be structurally similar to the experimental items. Following the same criteria as CF79, distractor items were designed to prevent participants from noticing that target phonemes always occur in the modifiers in the experimental items. Therefore, target phonemes occurred at the beginning of verbs, head nouns of noun phrases, adverbs, adjectives and prepositions. Additionally, 12 distractor items did not include the target phoneme. An example distractor item is provided below (word containing target phoneme is bolded):

Distractor question: What was the pet playing with?

Distractor item: The pet **cat** with gray fur was playing with the red ball.

Stimulus recording

CF79 presented all stimuli auditorily by recording both questions and experimental sentences. The same recording of each sentence was used in response to both questions, such that the stimuli were acoustically identical in all conditions. To avoid any effects of prosodic focus from confounding the design, CF79 report that “Care was taken in recording the experimental sentences to keep the intonation contours as neutral as possible; in particular, no potential target phrase was assigned emphatic accent.” (CF79, p. 52). CF79 do no report other details of the

[1] We thank Anne Cutler for suggesting this improvement to the original stimuli.

recording, such as the gender, age, or regional accent of the speaker, but while the original recordings are no longer available, they were likely recorded by a male native English speaker[2]. Therefore, our stimuli were also recorded by a male native English speaker specifically instructed not to place particular emphasis on the two target phrases. All recordings are available online through the Open Science Framework (https://osf.io/ru4y8/).

Procedure

We replicated the procedure of CF79 as closely as possible based on the details Cutler and Fodor provided, and we inferred any non-reported details based on similar work (e.g., Akker & Cutler, 2003). A few changes to the procedure addressed the fact that data were collected online rather than in person due to the Covid-19 pandemic.

The experiment was constructed using PsychoPy3 (Peirce et al., 2019) and then converted to PsychoJS, powered by the online server Pavlovia and distributed through the UC Davis SONA participant pool. Participants were tested individually online, rather than in person in groups of 2 or 3 at a time, as reported by CF79.

The experiment began by having the participant provide informed consent. Upon confirming their willingness to participate, participants were instructed to perform the experiment in a quiet environment with no distractions, wearing headphones and adjusting the volume so the sound is loud enough. Participants answered a few questions certifying that they met the selection criteria (resided in the US or Canada, that they were native English speakers with no hearing impairments) and that they were currently in a quiet environment with no

[2] Again, we thank Anne Cutler for providing this information.

distractions. Before they were permitted to continue to the experiment, participants were required to pass a headphone screening test developed for online experiments (Milne et al., 2021). This test involves an auditory detection task based on the perception of Huggins Pitch, which can only be completed successfully when sounds are heard dichotically through headphones rather than through speakers; the test was specifically designed to be integrated with online experiments using PsychoJS and Pavlovia (Milne et al., 2021).

Participants were then given instructions for the phoneme monitoring task. The instructions were presented both visually and auditorily, read by the same male speaker who recorded the stimuli. This made the procedure more closely resemble an in-person experiment, where the experimenter reads the instructions to the participants. The use of both oral and written instructions is also similar to Rysling et al.'s (2020) replication of Cutler (1976).

Following CF79, participants were instructed to pay close attention to the sentences in anticipation of a memory test to be administered at the end of the experiment, as well as to allow accurate responses to comprehension questions during the experiment (further described below). Secondly, they were instructed to listen for the occurrence of a target sound and press a button as soon as they heard a word beginning with that sound. The instructions included several examples of words beginning with each possible target sound (e.g., "These are some examples of words that start with the sound /k/: car, kettle, kite, cold, question.") to ensure that the task was clear to participants (as they were not able to ask for clarifications as they would in an in person experiment). Participants were told that each trial would start with a specification of which target sound to listen for, and that the target sound would only occur in the sentence and not the question preceding it. After the instructions, participants went through 5 practice trials; practice items were constructed using the same criteria as the distractor items and they were not repeated as part of the experiment. Following these practice trials, participants moved on to the main experiment.

Each trial began with a specification of the target sound. CF79 note that the target sound "was specified immediately prior to the question which preceded the sentence." (CF79, p. 53) but do not report how exactly this was implemented. Akker and Cutler's (2003) replication and expansion of CF79 reports that target sounds were specified visually on the screen, while Cutler and McQueen (2014) report that "standard phoneme detection instructions" involve "the auditory specification of the target segment (e.g., "/b/ as in boy")" (Cutler & McQueen, 2014, p. 77). As CF79 indicate that all instructions and materials were presented auditorily, we inferred that they likely used a procedure similar to that of Cutler and McQueen (2014)[3]. We therefore started each trial with an auditory specification of the target phoneme (e.g., "/b/ as in boy"). Similarly to Akker & Cutler (2003), this specification was followed by two seconds of silence, the auditory presentation of one of the two possible questions, another two seconds of silence, and finally the presentation of the sentence. A trigger was synchronized with the onset of the sentence in order to record the participant's reaction time to the target phoneme. The order of experimental and distractor items was pseudo-randomized for each participant, such that no more than three instances of the same target phoneme, target position, focus position, or stimulus type (experimental or distractor) were presented in succession (similarly to Akker & Cutler, 2003). While CF79 did not specify whether the order of items was randomized or not across participants, using a pseudo-random order avoided potential confounds of stimulus order.

[3] This was later confirmed by Anne Cutler.

Upon completion of the phoneme monitoring task, participants received a memory test similarly to CF79 and Akker and Cutler (2003). In the test, participants were visually presented with all 36 experimental sentences and were asked to select the correct choice between four alternative words replacing one word in each sentence. For half of the judgments, the choice was between alternatives to the word that had been in focus (the answer to the preceding question), while for the other half it was between alternatives to the word that had not been in focus. For half of the judgments, the choice was between alternatives of the target-bearing word (the word containing the target phoneme for that trial) while in the other half, the choice was between alternatives of the non-target-bearing word.

A few additional measures were used to ensure data quality, given that the study was performed online. To ensure that participants attended to and comprehended the stimuli during the phoneme monitoring task, a comprehension question was presented after half of the distractor items (for a total of 18 trials). While comprehension questions were not included in the original CF79 experiment, we added this measure as a data quality check given that the experiment was conducted online, which resulted in the experimenter having less control over whether participants were actively engaged in the task than if the experiment were conducted in person, as it was in CF79. As the comprehension questions followed only distractor items and not the experimental items, we believe that the addition of this task did not interfere with the participants' performance in such a way that it would constitute a significant deviation from the original CF79 experiment. Comprehension questions required a yes or no response. Half of the questions should have been answered "yes" and half "no". Half of the questions tested comprehension of the focused constituent and half of the non-focused constituent. Below is an example question:

Distractor item: The pet cat with gray fur was playing with the red ball.

Comprehension question: Was the pet cat playing with the yellow ball? Yes/no

Additionally, a final questionnaire presented after the completion of the memory task collected basic demographic information (e.g., age and gender) and asked participants to report whether they experienced any technical issues during the experiment (e.g., not being able to hear the stimuli, not understanding the instructions). Participants were also asked to self-report whether they paid attention to the entirety of the experiment, and to confirm that they performed the experiment in a quiet environment without distractions using headphones. Participants' answers to these questions were used as part of the exclusion criteria (see *Exclusion criteria*).

Analysis

Data exclusions and pre-processing

The analysis of reaction time (RT) data must deal with the fact that RT data are not normally distributed and so violate assumptions of linear models. There are several alternatives available to deal with the positively right-skewed distribution of RT data, including transforming the data (log or inverse transformations), excluding outliers (e.g., extremely quick or slow responses), and using statistical models that fit non-transformed data to a non-normal distribution such as a log-normal or ex-Gaussian distribution (Baayen & Milin, 2010; Fernández & Vadillo, 2020; Ratcliff, 1993; Whelan, 2008). Each alternative presents advantages and disadvantages, and multiple procedures may be used in conjunction; what is most important is to pre-determine which alternative(s) to select prior to data collection to minimize flexibility and the likelihood of false positives (Fernández & Vadillo, 2020)

In their original analyses, CF79 report that "16 very long reaction times were excluded as possible reprocessing ("double take") and that 73 very short reaction times were excluded as anticipations." (CF79, p. 54), but they do not report the specific cutoffs for determining what was considered a very long or very short RT. However, other studies by Cutler and colleagues report discarding any RT shorter than 100 ms, the minimum value for genuine RTs (Luce, 1986), and any RT longer than 1500 ms (e.g., Cutler, 1976; Akker & Cutler, 2003; Cutler & McQueen, 2014; more recently, Rysling et al., 2020). We opted to maintain this criterion for data exclusion in order to more closely replicate the analyses of CF79, and as an additional data quality check. Similarly to Rysling et al., 2020, we discarded all data from any participants whose RTs exceeded the limits for exclusion (or failed to respond altogether) on more than 25% of experimental trials. In our final sample (after participant exclusions), we discarded data from 3% of trials due to failure to respond, and 6% of trials with RTs outside the bounds; thus, 91% of data was used in the analyses.

After data exclusion, CF79 analyzed raw (non-transformed) RT data using ANOVAs. However, this method is no longer considered optimal as ANOVAs (and the linear mixed effects models we plan to use) assume the data to be distributed normally. Therefore, we chose to perform the analyses on inverse-transformed RTs, defined as invRT = − 1000 / RT (for an explanation of this transformation, see Baayen & Milin, 2010; Brysbaert & Stevens, 2018). We picked this particular alternative (e.g., as opposed to log-transformed RTs) because simulations have shown it to lead to more power (Brysbaert & Stevens, 2018; Ratcliff, 1993) and because this allowed for a more straightforward power analysis following Brysbaert and Stevens (2018).

Analysis plan

RT data

The inverse-transformed RTs were analyzed using linear mixed effects models in R (version 4.0.3, R Core Team, 2020) using the *lme4* and *lmerTest* packages (Bates et al., 2014; Kuznetsova et al., 2017). CF79 report the results of two repeated measures ANOVAs, one for subjects and one for items, combined into the *min F'* statistic (Clark, 1973). This statistic was developed to interpret the results of these two analyses combined but was later dropped in favor of reporting the *F* statistic of each ANOVA independently (*F1* and *F2*). More recently, the development of linear mixed effects models has allowed psycholinguists to account for by-subject and by-item random variance using a single analysis (Baayen et al., 2008). We therefore chose mixed effects models over the original *min F'* statistic given that they provide better characterization of the data and generally lead to greater power (Stevens & Brysbaert, 2016).

The model included fixed effects of Target position, Focus position and their interaction to replicate the contrasts of the 2 X 2 ANOVA used by CF79 (testing predictions a-c; see *Predictions*). To best approximate the interpretation of the effects to those of an ANOVA, we used sum-coded contrasts (similarly to Rysling et al., 2020). As such, the size of the interaction was assessed through the model. We assessed the significance of the interaction by comparing a model that includes the interaction to the same model without the interaction, by means of likelihood ratio test performed with the *anova*() function. If the interaction is significant, we expected the model that includes the interaction to present a significantly better fit of the data than the model without the interaction.

We started with a maximal random effects structure (including by-subject and by-item intercepts and slopes for all fixed effects) following Barr et al. (2013). In case of convergence issues (including singular fit), we first checked whether the convergence issues could be resolved through a change in the model optimizer. If the convergence issues persisted, we progressively dropped the random effects associated with the smaller variance until the model converged (Barr et al., 2013). When testing the significance of the interaction through model comparison, we adopted the same random effects structure for both models (with and without the interaction); in case of convergence issues, we simplified the random effect structure further until both models converged.

If the interaction turned out to be significant, we planned to perform pairwise comparisons to test whether the difference between focused and non-focused conditions was significant at both target positions (testing prediction d). We did so using the package *emmeans* (Lenth et al., 2018) along with Bonferroni corrections.

Lastly, we reported descriptive statistics to characterize the effect size and variance found in the data (see Table 4.2 in section *Results*), including the mean, median, standard deviation, standard error of the mean, 1st quantile and 3rd quantile for each condition (similarly to Rysling et al., 2020).

Replication of original ANOVA

While our primary analysis of the RT data conformed to current statistical standards, we also performed an exact replication of the original analyses conducted by CF79, using two by-subject and by-item ANOVAs on non-transformed RT data, combined into a single *min F'* statistic. This analysis allowed us to evaluate whether any differences in the significance of the effects were due to the use of different statistical methods. If the results reported by CF79 are attributable to the analysis of non-transformed RT data using linear models, then we would expect the effects to be significant as a result of the original ANOVAs but not significant in the mixed effects models.

Memory test data

Memory test responses were coded as zeros (incorrect response) and ones (correct response) and analyzed through a logistic mixed effects model (testing prediction e). To best approximate the analysis reported by CF79, conditions were re-coded based on whether the question was about a target-bearing word (WasTarget factor; not significant in CF79) and whether it was about a focused or non-focused word (WasFocused factor). Thus, the model included fixed effects of WasTarget and WasFocused, and no interaction. The model also included random effects for subject and question; in case of convergence issues, we simplified the maximal random effects structure following the same procedure as described for the RT analyses.

Power analysis

We performed a power analysis to determine the sample size. The power analysis was performed via simulations, following guidelines by Brysbaert and Stevens (2018) for power analysis of RT data using linear mixed effects models, and by adapting code and data provided by Brysbaert and Stevens (2018) on OSF (https://osf.io/fhrc6/).

Minimally interesting effect size and variance estimation

We first determined what we should consider as the smallest effect size of interest (Albers & Lakens, 2018; Dienes, 2021) as well as the expected variance in the data. While we predicted that we would replicate the pattern in means reported by CF79, we performed the power analysis on the assumption that their reported effect size might be inflated due to publication bias (Kühberger et al., 2014). As such, we selected a difference of 10 ms between conditions as the smallest effect size we wished to be able to detect, which is a very small effect size for RT data varying from 100 ms to 1500 ms (Brysbaert & Stevens, 2018). As for the variance, unfortunately neither CF79 nor Akker & Cutler (2003) reported a measure of variance in their data nor a standardized effect size such as Cohen's *d*. However, we were able to infer from the standard errors reported by Rysling et al.'s (2020) replication of Cutler (1976; a phoneme monitoring experiment using a similar design) that the standard deviation would likely be in the range of 150 to 200 ms (the actual *SD* in our sample was found to be close to this range, at 203 ms). This allowed us to perform the power analysis adapting the script provided by Brysbaert and Stevens (2018) to run a simulation power analysis based on RT data from a lexical decision task (Perea et al., 2015), which has a similar positively right-skewed distribution and similar variance (*SD* = 162 ms) as the data we expected to collect.

Power analysis method

The script was edited to reflect our proposed data analysis plan. Power was estimated using the *simr* package in R (Green & MacLeod, 2016) for a model that includes two sum-coded fixed effects (of two levels each) and their interaction, as well as random effects of participants and items. To deal with convergence issues, the random effects structure was simplified following the same procedure as described in section *Analysis plan*. Following guidelines by Brysbaert and Stevens (2018), we manipulated the pattern of means in Perea et al. (2015)'s lexical decision data by adding constants to each condition to reflect our expected pattern of means and the smallest effect size of interest. Because we predicted that we would find a significant difference in RTs for focused and non-focused words to both early and late targets, the minimum difference between those conditions was initially set to 10 ms (originally -69 ms and 16 ms in CF79). Additional constants were added to produce a pattern of means that would lead to a main effect of target position of 15 ms (58.5 ms in CF79), a main effect of focus position of -5 ms[4] (-26.5 ms in CF79) and an interaction effect (measured as the difference of the difference between focused and non-focused at each target position) of -30 ms (-85 ms in CF79). The pattern of means we used for the power analysis simulation and the original pattern reported by CF79 are illustrated in Figure 4.1. The selected pattern more than halved the effects found by CF79 while maintaining the minimally interesting effect size at 10 ms to focused and non-focused words at the two target positions; as such we believe it represents the minimally interesting pattern of means we predict to find in our data if a replication of CF79 is successful.

RTs below 100 ms or above 1500 ms were excluded, and the power analysis was performed on invRTs following Brysbaert and Stevens (2018). Their script was additionally modified to iteratively perform the power analysis on 100 random subsets of 36 items. The required sample size was determined by varying the number of participants using the *extend()* function from the *simr* package. For more details, the script we used for the power analysis is available on OSF (https://osf.io/ru4y8/).

[4] While this led to an effect size of focus position that is smaller than the minimally interesting effect size of 10 ms, this was necessary in order not to inflate the size of all other effects. As the main effect of focus position is secondary to other effects and not theoretically interesting, it was not prioritized for the purposes of the power analysis.

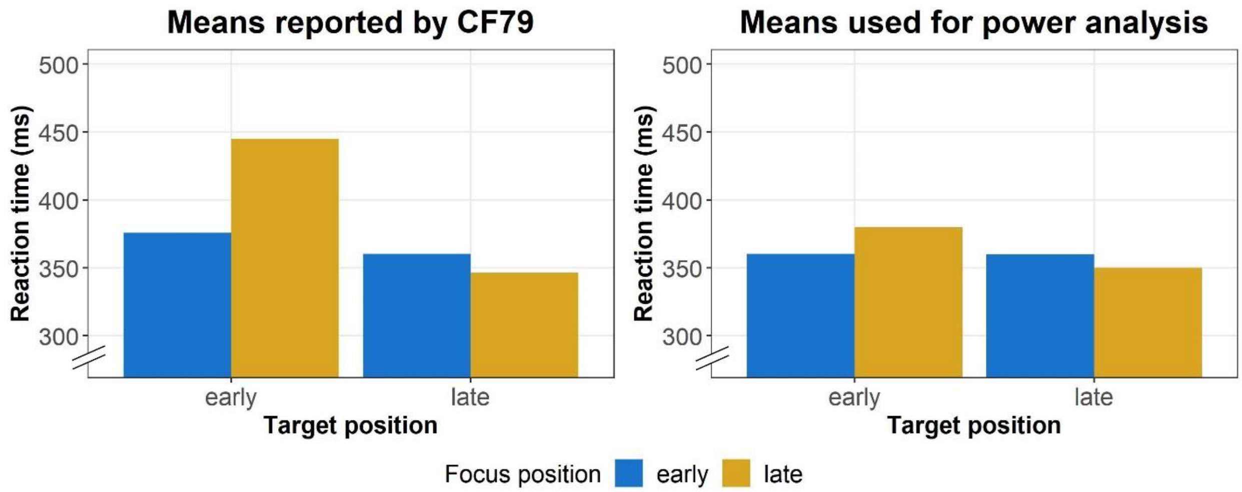


Figure 4.1. Pattern of means across conditions, as found by CF79 and as used in our power analysis. Note that the y axis is truncated at 300 ms to permit better visualization of mean differences.

Power analysis results

The power simulation for the model described above, using the entire Perea et al. (2015) dataset, found 100% power to detect the effects of interest at an alpha level of 0.02. As the Perea et al. (2015) dataset includes data from 120 stimuli and 40 participants, this indicates that 1200 observations per cell in our 2x2 design should be sufficient to reach adequate power. However, as described above, we also performed the analysis iteratively on 100 random subsets of 36 stimuli, using the *extend()* function to test power for different numbers of participants. With this method, we found that our originally proposed sample of 120 participants should lead to above 90% power using an alpha level of 0.02. However, as this would lead to 1080 observations per cell, which is fewer than the 1600 observations per cell recommended by Brysbaert and Stevens (2018), we opted to increase the target sample size to 180. Using 36 stimuli, this would give us 1620 observations per cell which should provide adequate power to detect our predicted effects.

Results

Reaction times

Reaction time results are visualized through Figure 4.2 and summary descriptive statistics are provided in Table 4.2. Interestingly, RT averages were overall higher in our sample (*M* = 650 ms) than in the original CF79 data (*M* = 381 ms). This could be due to the fact that our experiment was conducted online rather than in person, and that the hardware participants used to perform the experiment at home (e.g., personal laptops) may have been less responsive than the laboratory equipment employed by CF79. On the other hand, RT averages were more comparable to the mean RTs reported in a recent replication of Cutler (1976), which was conducted in person (*M* = 567 ms; Rysling et al., 2020).

The size of the interaction in our data (-82.31 ms) was comparable to the original (-85 ms in CF79), while the main effects in our data (Target position: 22.93 ms; Focus position: -8.58 ms) were smaller, though in the same direction as those reported by CF79 (Target position: 58.5 ms; Focus position: -26.5 ms).

Table 4.2.
Descriptive statistics of reaction times data.

Focus Position	Target Position	*Mean*	*Median*	*SD*	*SE*	*1st Quantile*	*3rd Quantile*
Early	Early	636.42	603	207.99	5.05	499	733
Early	Late	654.65	630	193.98	4.83	525.75	750.25
Late	Early	686.15	656	213.82	5.27	539.25	788
Late	Late	622.07	601	188.23	4.68	503.75	710

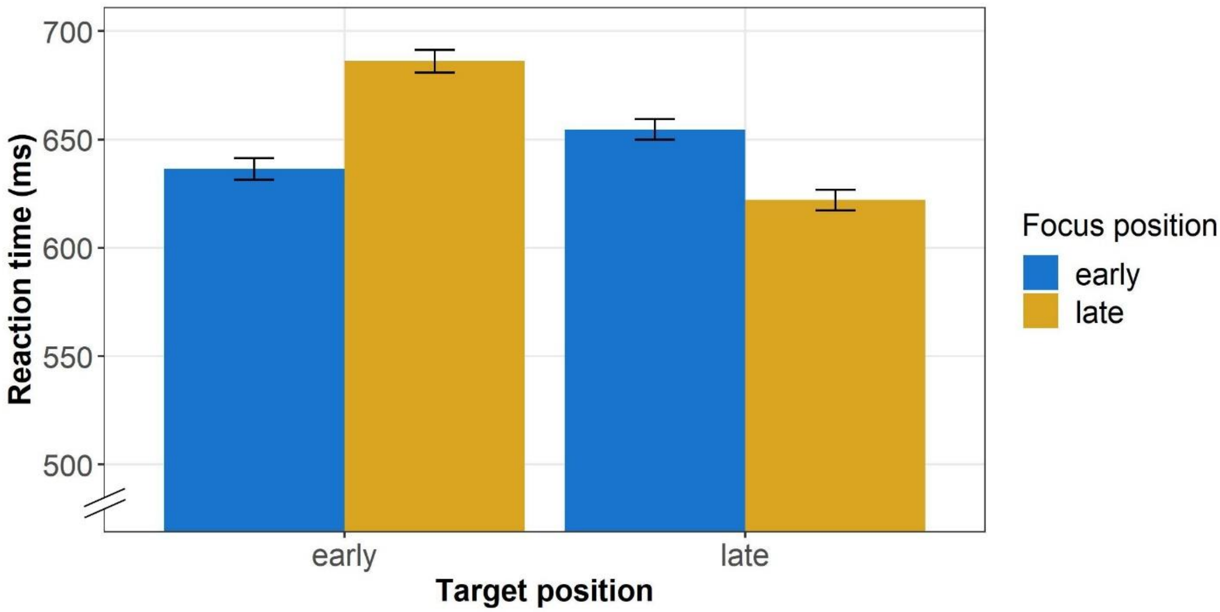


Figure 4.2. Mean reaction times by Focus and Target position. Error bars represent Standard Error of the Mean. Note that the y axis is truncated at 500 ms to permit better visualization of mean differences.

Mixed effects model

Transformed RTs (invRTs) were analyzed through a mixed effects model, as described under *Analysis plan*. The model found a significant effect of Target position, with faster RTs to targets in later positions than earlier positions, supporting prediction a ($\beta = 0.03$, $SE = 0.006$, $t = 4.82$, $p < 0.0001$). There was a marginally significant effect of Focus position, with faster RTs when the preceding question focused on the earlier part of the sentence than the later part of the sentence, following prediction b ($\beta = -0.01$, $SE = 0.006$, $t = -1.96$, $p = 0.0503$). Critically, there was a significant interaction between Target and Focus position, supporting prediction c and the main finding of CF79 ($\beta = -0.05$, $SE = 0.006$, $t = -9.94$, $p < 0.0001$). The significance of the interaction was further tested through model comparison, where a likelihood ratio test found the model including the interaction to be a significantly better fit than the model without the interaction ($\chi^2(1) = 98.18$, $p < 0.0001$). Pairwise comparisons found that RTs to the early target were faster in response to the early focus question compared to the late focus question ($\beta = -0.14$, $SE = 0.02$, $t = -8.49$, $p < 0.0001$), while RTs to the late target were faster in response to the late focus question compared to the early focus question ($\beta = -0.09$, $SE = 0.02$, $t = -5.6$, $p < 0.0001$), supporting prediction d. In summary, the mixed effects model replicated the original main effects of Target position and Focus position (though the latter was marginally significant, but in the expected direction) and, most importantly, replicated the interaction that represents the main finding of CF79.

Original ANOVA

We also performed the original analyses reported in CF79, analyzing non-transformed RTs using repeated measures ANOVAs. The by-subject ANOVA found a significant main effect of Target position ($F1(1, 199) = 26.90$, $p < 0.001$), a significant main effect of Focus position ($F1(1, 199) = 6.23$, $p < 0.05$) and a significant interaction ($F1(1, 199) = 92.68$, $p < 0.001$). In the by-item ANOVA, the main effects of Target and Focus position were not significant, while there was a significant interaction ($F2(1, 35) = 98.31$, $p < 0.001$). Similarly, the combined *min F'* analysis found that the main effects of Target and Focus position were not significant, while it found a significant interaction (*min F'*(1, 124) = 47.71, p < 0.0001). Thus, using the original analysis reported by CF79, only the critical interaction was replicated, while the two main effects were not. This is likely due to the use of two separate ANOVAs combined through the *min F'* statistic and the analysis of non-transformed RTs, both of which have been shown to lead to less power compared with mixed effects models on transformed RTs (Baayen et al., 2008; Stevens & Brysbaert, 2016; Brysbaert & Stevens, 2018; Ratcliff, 1993).

Exploratory analysis: is it worth transforming RTs?

We performed an exploratory analysis to test whether inverse-transformed, log-transformed, or non-transformed RTs would lead to a better model fit of our data.
One limitation of our pre-registered analysis plan is that, compared to the original CF79 analyses, we made two changes: the use of mixed effects models rather than *min F'*, and the transformation or raw RTs into invRTs. Therefore, the differences in significance between these two analyses could be due to either of these changes. While the benefits of mixed effects models over separate by-subject and by-items repeated measures ANOVAs are now largely uncontroversial in psycholinguistics (Baayen et al., 2008), there is considerable variability in the pre-processing steps and choice of transformation used for RT data (Baayen & Milin, 2010; Ratcliff, 1993; Whelan, 2008). In particular, the distribution of RT data (and its corresponding transformation) may vary depending on the task (Baayen & Milin, 2010).

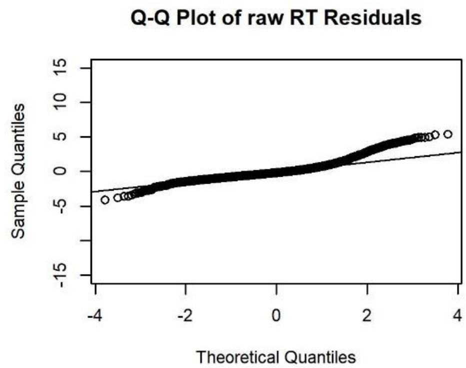
In this exploratory analysis, we addressed whether the inverse transformation of RT data was beneficial by comparing the fit of the mixed effect model on invRT to the fit of the same model on raw RTs. We also tested whether the inverse transformation improved model fit compared to log-transformed RTs (logRTs), another common transformation of RT data.

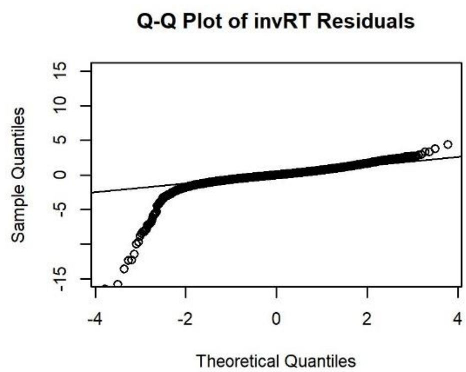
The same model as described under *Analysis plan* was performed on raw and on log-transformed RTs. Both models found a significant effect of Target position (raw RTs: $\beta = 12.83$, $SE = 2.02$, $t = 6.34$, $p < 0.0001$; logRTs: $\beta = -0.02$, $SE = 0.003$, $t = 5.9$, $p < 0.0001$), a significant effect of Focus position (raw RTs: $\beta = -4.58$, $SE = 2.02$, $t = -2.27$, $p < 0.05$; logRTs: $\beta = -0.007$, $SE = 0.003$, $t = -2.36$, $p < 0.05$), and a significant interaction (raw RTs: $\beta = -20.56$, $SE = 2.02$, $t = -10.18$, $p < 0.0001$; logRTs: $\beta = -0.03$, $SE = 0.003$, $t = -11.03$, $p < 0.0001$). Therefore, in terms of significance, these two alternative models found the same pattern as the pre-registered model on invRTs (with the difference that the effect of Focus position, marginally significant with invRTs, was fully significant with raw and logRTs).

The overall model fit was compared by visualizing the normality of residuals through Quantile-Quantile (Q-Q) plots, and by comparing R^2 across models, following Baayen and Milin (2010). Inspection of the residuals through Q-Q plots (Figure 4.3) reveals that the inverse transformation actually led to larger deviations from normality in the residuals compared to the analysis of raw RTs. On the other hand, log-transformed RTs led to smaller deviations in the residuals than invRTs, indicating that a log transformation may be more appropriate for RTs acquired through the phoneme monitoring task. To corroborate the visual inspection of the residuals and to determine whether the inverse transformation was ineffective, we turned to R^2 as a quantitative measure of model fit.

The conditional R^2 (the proportion of variance explained by both fixed and random effects) was compared across models using the R package *performance* (Lüdecke et al., 2021). This metric reveals that the model on logRTs explained the most variance (R^2 = 0.4), followed by the model on invRTs (R^2 = 0.36), and finally by the model on raw RTs (R^2 = 0.35). Thus, based on this measure of model performance, we conclude that a log transformation would have been preferable to an inverse transformation, but we also see that the inverse transformation was still (slightly) beneficial compared to the analysis of raw RTs. These findings support the prevailing current recommendation to transform RTs (Baayen & Milin, 2010) and they suggest that a log transformation is most appropriate for phoneme monitoring data.

Nonetheless, it should be noted that all three models failed to fully capture variance and led to non-normal residuals, while also supporting the same overall conclusions; thus, the decision whether to transform the RTs and the choice of transformation did not seem to greatly influence the outcome of the mixed effects model used for this dataset. The data and models used for this study are available on OSF (https://osf.io/ru4y8/) to enable further investigations on the effectiveness of other pre-processing steps, such as outlier exclusion.





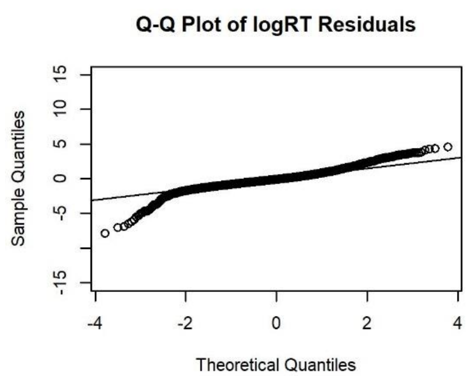


Figure 4.3. Q-Q plots illustrating the normality of residuals for the same mixed effects model on raw, inverse-transformed, and log-transformed RTs. Residuals were standardized to allow for comparison. Better model fit is indicated by residuals falling along the line.

Memory test

Descriptive statistics for the memory test results can be found in Table 4.3, and the results are visualized in Figure 4.4. Across conditions, participants correctly identified 58% of the target words, which is above chance level (25%) but lower than the average performance reported by CF79 (70%). The logistic mixed effects model found that participants were more likely to remember words that had contained the target phoneme compared to words that did not contain the target phoneme ($\beta = 0.29$, $SE = 0.03$, $z = 11.65$, $p < 0.0001$). On the other hand, participants were equally likely to remember words regardless of whether they had been focused through the preceding question ($\beta = 0.02$, $SE = 0.03$, $z = 11.65$, $p = 0.47$). Interestingly, the results of the memory test did not replicate the original post-test results reported by CF79, who found a significant effect of whether the word had been focused and no effect of whether the word had contained the target phoneme.

Table 4.3.
Descriptive statistics of memory test data.

WasFocused	WasTarget	*Mean*	*SD*	*SE*
Focused	Non-target	0.51	0.50	0.01
Focused	Target	0.65	0.48	0.01
De-focused	Non-target	0.51	0.50	0.01
De-focused	Target	0.63	0.48	0.01

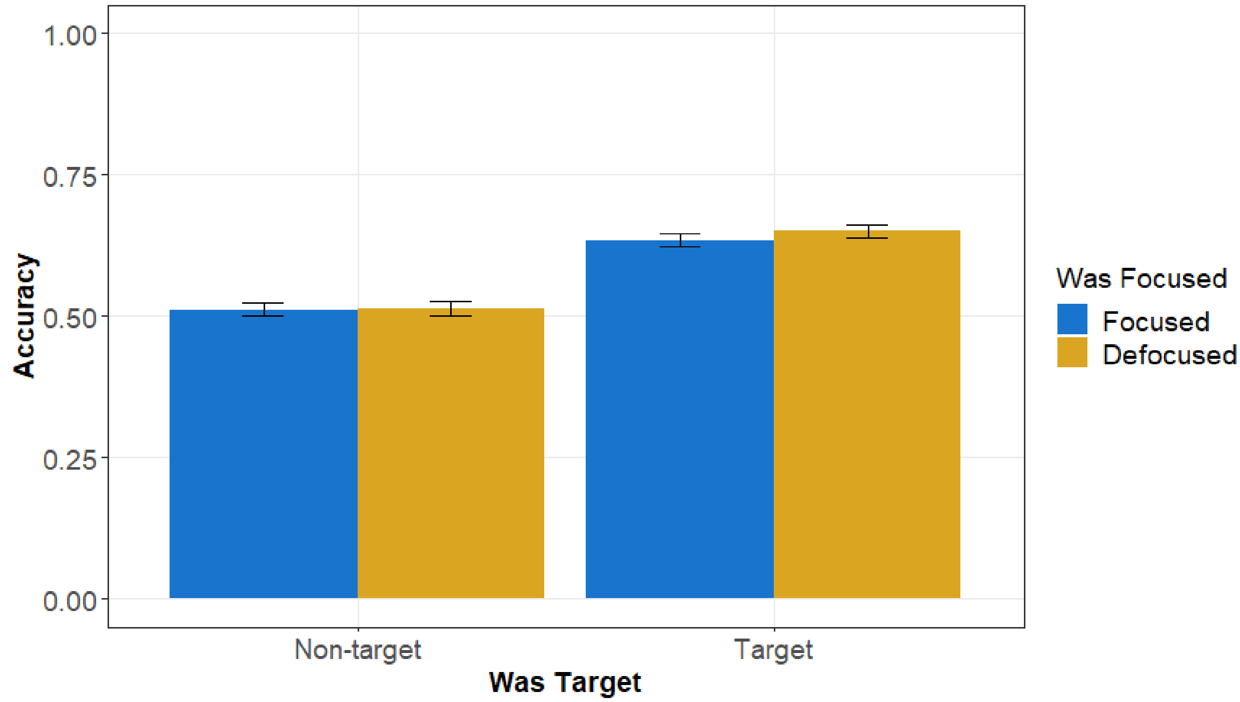


Figure 4.4. Average memory test accuracy as a function of whether the word had contained the target phoneme (WasTarget) and whether the word had been focused (WasFocused). Error bars represent Standard Error of the Mean.

Discussion

In this study, we attempted to replicate the findings of a foundational experiment by Cutler and Fodor (1979). The original study measured the effect of discourse cues of information structure on reaction times (RTs) using a phoneme monitoring task. They found that RTs were faster for target words that had been focused by the preceding question, as opposed to de-focused, suggesting that discourse cues guide the listener's attention to the timing of focused, important information.

We successfully replicated the critical interaction between target position and focus position, which represents the main finding of CF79. The less theoretically relevant main effects of target and focus position were also replicated, though their effect sizes were smaller than those reported by CF79. While the overall pattern of means was replicated, RTs in our sample were on average almost twice as long as those reported by CF79. There are several potential causes. As this replication was conducted online rather than in person, participants may have been overall less attentive (despite the measures we took to ensure data quality) and their hardware may have been less responsive than the laboratory equipment used by CF79. Additionally, as RTs were time-locked to the onset of the target phonemes, there may have been slight differences in the way the target phoneme onset was determined in our stimuli. Nonetheless, average RTs in our study were similar to those found by Rysling et al. (2020), a more recent replication of the phoneme monitoring experiment conducted by Cutler (1976).

Lastly, we did not replicate the findings on the post-experiment memory test; in our data, participants were more likely to remember words that had contained the target phoneme while there was no effect of whether a word had been focused during the phoneme monitoring task; this is contrary to CF79, who found an effect of focus and no effect of having contained the target phoneme. While the effect of focus on memory reported by CF79 might have been a false positive, it is also possible that this effect was not observed in our data due to unavoidable differences in the experimental setups and the participant samples. First, as our experiment was conducted online, we had less control over how much time participants took to finish the experiment and, therefore, the amount of time between the phoneme monitoring task and memory test. Participants were instructed to complete the experiment in one sitting, but it is possible that some participants took a break in between the two tasks. Secondly, as the longer RTs found in our sample suggest, participants may have been overall less attentive than those in the original study, leading to overall lower recollection of the target words. On the other hand, we found that words containing the target phoneme were better recalled in our data, an effect that was not significant in CF79. Aside from differences in experimental setup and participants, what may have contributed to these contrasting findings is the fact that, in this replication, we constructed new distractor items and new word alternatives for the memory test, as these materials were not available in the original article. We did our best to follow the description of these materials provided by CF79, but it is possible that there were slight changes in the way the distractors and the memory test alternatives were designed across studies, which may have led to different effects on recall. As the memory test was not the primary focus of this replication, we defer any further explanation of these effects to future replications and investigations specifically designed to explore the memory processes involved in phoneme monitoring.

A second goal of this study was to compare the results of analyses following modern statistical practices to the results of outdated analyses closer to those used by CF79. This comparison allowed us to further assess the robustness of the original findings while also providing insights into the usefulness of updated statistical standards. The original CF79 study analyzed non-transformed (raw) RTs using two separate repeated measures ANOVAs (by-subject and by-item) combined into a single *min F'* statistic (Clark, 1973). Following the advances in statistical practices since the 1970s, we improved on this analysis in two ways. First, because RTs typically present a right-skewed distribution, breaking the assumptions of linear models, RTs underwent an inverse transformation (invRTs). Second, rather than with two separate ANOVAs, RTs were analyzed with a single mixed effects model, which captures by-subject and by-item variability more effectively (Baayen et al., 2008).

We reproduced the original analyses by analyzing raw RTs with two ANOVAs and a combined *min F'*. Using this analysis, only the interaction was replicated, while the two main effects of target and focus position were not significant. This supports the claim that the *min F'* statistic and the analysis of raw RTs both have lesser power than the use of mixed effects models on transformed RTs (Baayen et al., 2008; Stevens & Brysbaert, 2016; Brysbaert & Stevens, 2018; Ratcliff, 1993). We also performed an exploratory analysis to test whether the inverse transformation contributed to a better fitting model. By running the same mixed effects model on raw RTs, invRTs, and log-transformed RTs (logRT; another common transformation of RTs), we found that for our dataset, logRTs led to a better fitting model than invRTs. Therefore, we suggest that the phoneme monitoring task may lead to a distribution of RTs that more closely follows the log-normal distribution, as opposed to the inverse-normal distribution, and recommend that further investigations using this task use log-transformed RTs. Nonetheless, all three mixed effects models led to the same pattern of results, suggesting that the choice of transformation did not greatly impact the model outcome.

In conclusion, we found that the effect of discourse cues of information structure on phoneme monitoring RTs is robust, as reflected through significant interactions across all the analyses we conducted. This finding lends support to CF79's original claim that discourse, similarly to prosody, is a cue that listener can use to infer the timing of upcoming focused information. Thanks to these cues, listeners may then pre-allocate attention to those points in time, facilitating processing.

This replication sets a stronger foundation for this claim and for the many diverging lines of research that have followed it. For the most part, these lines of research have investigated different types of cues (e.g., discourse vs. prosody vs. syntax) independently, leading to inconsistent results. Thus, we hope that this replication will revive interest in the broader picture proposed by CF79 and spur more research into how these cues interact. In particular, further research is needed to assess whether all cues to information structure influence comprehension through the same domain-general mechanisms (e.g., guiding attention) or whether they have independent effects at different stages of processing. Additionally, more research is needed to reconcile contradictory findings regarding the interaction of information structure cues with other aspects of language comprehension, such as semantic prediction, depth of processing, and good-enough language processing (F. Ferreira & Lowder, 2016; Rothermich et al., 2012; Wang et al., 2011). Overall, we hope that this replication will inspire new avenues for research into information structure, as well as promote the replication of other classic psycholinguistics findings.

References

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition*, *6*(2), 81–96. https://doi.org/10.1017/s1366728903001056

Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, *74*, 187–195. https://doi.org/10.1016/j.jesp.2017.09.004

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baayen, R. H., & Milin, P. (2010). Analyzing Reaction Times. *International Journal of Psychological Research*, *3*(2), 12–28.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting Linear Mixed-Effects Models using lme4* (arXiv:1406.5823). arXiv. http://arxiv.org/abs/1406.5823

Beier, E. J., & Ferreira, F. (2022). Replication of Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. Cognition, 7(1), 49–59. *Journal of Memory and Language*, *126*, 104339. https://doi.org/10.1016/j.jml.2022.104339

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7), 1044–1098. https://doi.org/10.1080/01690965.2010.504378

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 1–20.

Calhoun, S. (2010). How does informativeness affect prosodic prominence? *Language and Cognitive Processes*, *25*(7–9), 1099–1140. https://doi.org/10.1080/01690965.2010.491682

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. https://doi.org/10.1016/S0022-5371(73)80014-3

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics*, *20*(1), 55–60.

Cutler, A., Butterfield, S., & Williams, J. N. (1987). The perceptual integrity of syllabic onsets. *Journal of Memory and Language*, *26*(4), 406–418. https://doi.org/10.1016/0749-596X(87)90099-4

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, *7*(1), 49–59. https://doi.org/10.1016/0010-0277(79)90010-6

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech*, *20*, 1–10. https://doi.org/10.1177/002383097702000101

Cutler, A., & McQueen, J. M. (2014). How prosody is both mandatory and optional. In J. Caspers, Y. Chen, W. Heeren, N. Pacilly, N. Schiller, & E. van Zanten (Eds.), *Above and Beyond the Segments* (pp. 71–82). John Benjamins.

Dienes, Z. (2021). Obtaining Evidence for No Effect. *Collabra: Psychology*, *7*(1), 28202. https://doi.org/10.1525/collabra.28202

Fernández, L. M., & Vadillo, M. A. (2020). Flexibility in reaction time analysis: Many roads to a false positive? *Royal Society Open Science*, *7*, 190831.

Ferreira, F., & Lowder, M. W. (2016). Prediction, information structure, and good-enough language processing. *Psychology of Learning and Motivation*, *65*, 217–247. https://doi.org/10.1016/bs.plm.2016.04.002

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication Bias in Psychology: A Diagnosis Based on the Correlation between Effect Size and Sample Size. *PLOS ONE*, *9*(9), e105825. https://doi.org/10.1371/journal.pone.0105825

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Lehiste, I. (1970). *Suprasegmentals* (pp. viii, 194). MIT Press.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). *Emmeans: Estimated marginal means, aka least-squares means. R package version 1.1.2.*

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. Oxford University Press.

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, *79*(6), 487–509. https://doi.org/10.1037/h0033467

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*, *53*(4), 1551–1562. https://doi.org/10.3758/s13428-020-01514-0

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Perea, M., Vergara-Martínez, M., & Gomez, P. (2015). Resolving the locus of cAsE aLtErNaTiOn effects in visual word recognition: Evidence from masked priming. *Cognition*, *142*, 39–43. https://doi.org/10.1016/j.cognition.2015.05.007

Pitt, M., & Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology. Human Perception and Performance*, *16*(3), 564–573. https://doi.org/10.1037/0096-1523.16.3.564

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532.

Roettger, T. B., Mahrt, T., & Cole, J. (2019). Mapping prosody onto meaning – the case of information structure in American English. *Language, Cognition and Neuroscience*, *34(7)*, 841-860. https://doi.org/10.1080/23273798.2019.1587482

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia*, *50*(2), 232–244. https://doi.org/10.1016/j.neuropsychologia.2011.10.025

Rysling, A., Bishop, J., Clifton, C., & Yacovone, A. (2020). Preceding syllables are necessary for the accent advantage effect. *The Journal of the Acoustical Society of America*, *148*, EL285. https://doi.org/10.1121/10.0001780

Shields, J. L., McHugh, A., & Martin, J. G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology*, *102*(2), 250–255. https://doi.org/10.1037/h0035855

Slote, J., & Strand, J. F. (2016). Conducting spoken word recognition research online: Validation and a new timing method. *Behavior Research Methods*, *48*(2), 553–566. https://doi.org/10.3758/s13428-015-0599-7

Stevens, M., & Brysbaert, M. (2016). *When do we have enough power in language research? Evidence from priming studies* [Unpublished manuscript. Ghent University.]. Retrieved from http://crr. ugent. be/papers/When% 20do% 20we% 20have% 20enough% 20power% 20in% 20language% 20research. pdf.

Wang, L., Bastiaansen, M., Yang, Y., & Hagoort, P. (2011). The influence of information structure on the depth of semantic processing: How focus and pitch accent determine the size of the N400 effect. *Neuropsychologia*, *49*(5), 813–820. https://doi.org/10.1016/j.neuropsychologia.2010.12.035

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*, 475–482.

CHAPTER 5: GENERAL DISCUSSION

In this dissertation, I have reviewed the theoretical background and psycholinguistic evidence for the idea that the information structure (IS) of sentences, or the organization of new and given information, plays an integral role in language comprehension. Several cognitive mechanisms have been posited to be influenced by IS, but to date there is no clear picture of how these different mechanisms interact with each other and with different aspects of IS. In Chapter 1, I have grouped the mechanisms proposed by previous literature under three main accounts. Under the More Resources account, comprehenders use the discourse context, syntactic and prosodic focus cues to pre-allocate attention and processing resources towards the points in time when the new, important information is expected to occur, facilitating processing of focused information. Under the Deeper Processing account, focused information is processed more deeply than given information. While these two accounts are theoretically compatible, studies have led to divergent findings with respect to whether focused words receive facilitated or deeper processing. Lastly, under the Altered Predictions account, IS is likely to interact with semantic predictions about upcoming words, but it is unclear what this interaction entails.

I presented three experiments aimed at disentangling these mechanisms. Experiment 1 (Chapter 2) consists of an eye-tracking reading study assessing the interaction of syntactic focus with semantic prediction. This experiment was designed to tease apart the More Resources account (predicting a smaller semantic predictability effect for focused words, thanks to facilitated processing) from the Deeper Processing account (predicting a larger semantic predictability effect for focused words, thanks to deeper processing). The results of this experiment are most consistent with the More Resources account, as the effect of predictability was smallest for focused words and largest for de-focused words; however, as described in the

Chapter 2 discussion and below, these findings do not completely preclude the influence of Deeper Processing and Altered Predictions.

Experiment 2 (Chapter 3) is an online replication of Experiment 1 using self-paced reading. This experiment found that the effect of semantic predictability persists throughout the discourse and that syntactically focused words are read for longer. While it did not clearly replicate the interaction pattern found in Experiment 1 (other than a marginally significant interaction in the same direction in the first spillover region), in the Chapter 3 discussion I review several reasons why self-paced reading may not have been an optimal choice of task. For example, syntactic manipulations and semantic predictability appear to influence self-paced reading times at different points in the sentence (e.g., the target word vs. the spillover regions), so this task may not be suitable to measuring their interaction.

Lastly, Experiment 3 (Chapter 4) presents an online replication of Cutler and Fodor's (1979) foundational study, which established the More Resources account and had a large role in shaping subsequent literature. This experiment has been published as a registered report in a special issue of the *Journal of Memory and Language* (Beier & Ferreira, 2022). Using an online version of the phoneme monitoring task, we successfully replicated the original findings, showing faster RTs to words focused by the preceding discourse question.

Taken together, these experiments suggest that part of the reason for the literature’s inconsistent findings is that all proposed mechanisms may be at play under different circumstances, and as reflected by different measures. The More Resources account was supported by the successful replication of Cutler and Fodor (1979) and by the results of Experiment 1, showing that the effect of semantic predictability was reduced by syntactic focus and enhanced by a de-focusing construction. This result is in line with Rothermich et al.'s (2012) findings of a smaller N400 effect for sentences where the location of prosodically prominent words could be predicted thanks to greater rhythmic regularity. Under this view, cues to focus (whether prosodic or syntactic) facilitate the processing of focused words; thus, it is easier to integrate unpredictable words when they are focused and harder to do so when they are de-focused.

However, the results are also consistent with some predictions of the Deeper Processing account. Focused words were read for longer in the self-paced reading task of Experiment 2, similarly to previous eye-tracking findings (Lowder & Gordon, 2015), which were interpreted as a sign of deeper processing and greater memory encoding. Importantly, in Experiment 1 the smaller effect of semantic predictability for focused words was driven by longer reading times for predictable words, while the larger effect for de-focused words was driven by longer reading times for unpredictable words. This suggests that the processing of predicted vs. unpredicted words is differently affected by IS: predicted words are processed more deeply when in focus, while unpredicted words are harder to integrate when de-focused due to fewer processing resources. Lastly, these experiments did not rule out the possibility that the strength or content of semantic predictions themselves may have changed as a function of syntactic focus, as assumed under the Altered Predictions account.

It is clear that while the experiments presented in this dissertation are a first step towards a better understanding of the cognitive mechanisms involved by IS, they do not provide a final answer as to their exact interplay. Two experiments aimed at further delineating these processes are described in the following section.

Future directions

There are many open questions about IS left unanswered by current literature. Does focus facilitate processing at all levels of comprehension, or only specific components? Do both predicted and unpredicted words receive deeper processing? Does IS influence semantic predictions by affecting pre-activation, integration or both? Does focus have different effects on different measures, from reaction times to eye-tracking reading times to the size of the N400? Do syntactic and prosodic cues to focus have independent effects at different levels of processing, are they used interchangeably, or cumulatively? How exactly does IS direct attention to focused information, and what do the "processing resources" consist of? When do listeners interpret focus cues as presentational vs. contrastive, and what effect does this have on comprehension? In the following, I describe two experiments currently in development that will address some of these questions.

Exploring the Altered Predictions account

In my original dissertation plan, Experiment 1 was followed by an auditory study using the visual world paradigm (Huettig et al., 2011) to characterize the effects of syntactic focus on semantic predictions prior to the target word. While the health and safety regulations imposed during the Covid-19 pandemic precluded us from completing this in-person experiment and prompted us to switch to online methods, all materials for this study have been developed (including two stimulus norming studies) and data collection is set to begin later this year.

The goal of this study will be to discern the specific effect of syntactic focus on semantic prediction, exploring the claims of the Altered Predictions account. Experiments 1 and 2 only measured their interaction for reading times at and after the target word. As I have argued above and in Chapter 1, patterns similar to those expected under the More Resources and Deeper Processing accounts could also arise from a change in the strength or content of the semantic predictions themselves, before encountering the target word. Thus, in this visual world study, I will assess whether syntactic focusing and de-focusing constructions influence which objects are fixated before the target word.

The stimuli consist of 27 stories from Experiment 1, those with target words that could be easily pictured in the visual world paradigm (e.g., objects). An additional 24 stimuli were developed and cloze normed following the same criteria to arrive at a total of 51 stories. As in Experiment 1, each story comprises an introductory sentence, a context sentence, and a target sentence that can lead the target word to either be focused, neutral, or de-focused.

Each item was recorded in all three focus conditions, but only in the predictable version for one of the two possible target words (e.g., only the "lunch-sandwich" version). This is because the other target word (e.g., "omelet"; unpredictable but plausible) will be used as a distractor item in the visual display. The items were recorded by a female native English speaker naive to the hypotheses of the study, who was instructed to use a similar prosody across focus conditions (e.g., avoiding a pitch accent on the target word when syntactically focused) in order to prevent confounds between syntactic and prosodic focus. As eye-movements will be measured prior to the onset of the target word, it is important that the same amount of time passes between the end of the context word ("lunch") and the target word ("sandwich") for all three focus constructions, to avoid confounds. A custom Praat script was used to calculate the average duration of this construction across focus conditions for each item. The construction duration was then equated to each item's average.

To prevent participants from assuming that all items are entirely predictable, an additional 51 filler items were developed. Of these, 34 contain an unpredictable (but plausible) word. Seventeen unpredictable fillers conform to the sentence structure of the experimental items, while 17 unpredictable and 17 predictable fillers present different structural patterns. This will provide some structural variety while it will also prevent participants from noticing that the experimental items present a unique structure and are always predictable. Filler items were recorded by the same speaker as the experimental items.

The visual display for each stimulus will consist of four color images representing the Predictable Target (e.g., a sandwich), an Unpredictable (but plausible) Distractor (e.g., an omelet), a Context Distractor related to the broader context but locally implausible (e.g., a notebook, related to studying but unrelated to making lunch and not an edible item), and an Unrelated Distractor (e.g., a plant). The purpose of the Unrelated Distractor is to provide a baseline, as an item that should not be fixated. On the other hand, the purpose of the Context Distractor is to provide an alternative possible item for participants to fixate during the de-focusing construction. This is because the de-focusing condition involves a subordinate clause (e.g., "After he made a sandwich, he…"). During this construction participants may engage in prediction for what might happen after the target word (e.g., they may predict that John will return to studying, and therefore fixate the notebook). Thus, this broad-context distractor is critical for two reasons: (1) so that participants are not restricted to items related to words in the subordinate clause, which they should not be focusing on; (2) to test whether participants are engaging in prediction of what will be presented after the target during the subordinate clause, which might explain why they would not be engaging in as much prediction of the de-focused target item.

All images have been collected by research assistants from publicly available repositories (e.g., Konklab image sets retrieved from https://konklab.fas.harvard.edu/#; Konkle & Oliva, 2012), from previous studies (Lowder & Ferreira, 2016) and from internet search. An image norming study was conducted online through Qualtrics in order to ensure that the selected images were perceived as either related or unrelated to the story context, following our design. In the norming study, participants were presented with each item's introductory and context sentences (thus, the target sentence containing the target word was omitted). They were shown one of the four images associated with the story and asked to rate how well it fit the theme of the story on a scale from 1 (Not at all) to 7 (It fits the story extremely well). Participants saw each item only once, and conditions were counterbalanced across four lists.

Predictable Targets were rated on average as 6.14 (median: 7), Unpredictable Distractors as 3.21 (median: 3), Context Distractors as 4.82 (median: 5), and Unrelated Distractors as 1.3 (median: 1). Filler items were also included in the norming study in order to ensure the images presented similar characteristics to those for the experimental items. For filler items, Predictable Targets were rated on average as 5.3 (median: 6), Unpredictable Distractors as 3.55 (median: 3), Context Distractors as 5.11 (median: 6), and Unrelated Distractors as 1.3 (median: 1). Thus, the image norming study confirmed that, in accordance with our design, Predictable Targets were perceived as the best fit for the context, while Unpredictable Distractors as significantly less fitting. Context Distractors were perceived as well fitting (higher than Unpredictable Distractors, lower than Predictable Targets), although the cloze norms confirm they could not be plausible replacements for the target words. Lastly, Unrelated Distractors were shown to be ill-fitting to the context.

In the experiment, each trial will begin with a fixation cross at the center of the screen, followed by the appearance of the visual display. After 3000 ms, participants will hear the recorded story through headphones. We will measure fixations towards each image during three areas of interest: from the onset of the context word (e.g., lunch) to the onset of the target sentence (Area 1); from the onset of the target sentence to the onset of the target word (Area 2); and from the onset of the target word until 1500 ms have elapsed (Area 3).

If syntactic focus influences semantic predictions prior to the target word, we expect to find the majority of the differences between focus conditions during Area 2. One possible pattern of results is that the focusing construction will lead to stronger pre-activation of the predictable target word, thus leading to a greater percentage and/or faster switch of fixations towards the Predictable Target image compared to the neutral condition; the de-focusing construction may have the opposite effect, and instead lead to more fixations towards the Context Distractor (as a potential prediction for what may happen after). An alternate possible pattern of results is that the focusing construction will lead listeners to expect the unexpected, possibly due to a contrastive focus interpretation. Thus, during the focusing construction participants may be more likely to fixate the Unpredictable Distractor compared to the neutral condition.

Overall, this study will extend this dissertation by testing the claims of the Altered Predictions account. By clarifying the exact influence of syntactic focus on semantic pre-activation, this experiment will allow for easier interpretation of the results of Experiment 1, as well as previous inconsistent findings with respect to the interaction of IS with the N400 semantic prediction effect.

Neural mechanisms of temporal attention

Another area of investigation I plan to explore in my future studies regards the specific mechanisms through which attention is pre-allocated in time during language comprehension. Under the More Resources account, IS directs attention towards focused information to facilitate processing. Studies have measured fMRI and ERP correlates of attention at or after the presentation of focused target words (Chen et al., 2014; Kristensen et al., 2013; Yang et al., 2019), or have assumed the involvement of attention from secondary measures (e.g., RTs or the N400 amplitude; Cutler & Fodor, 1979; Rothermich et al., 2012). An integral component of the More Resources account, however, is that attention should be deployed in advance, resulting from a temporal prediction for when the new information will occur. To date, there is no clear evidence that IS cues engage such temporal predictions to guide attention.

Non-linguistic studies of perceptual processing show that attention can be pre-allocated through a range of temporal prediction mechanisms (Nobre & Van Ede, 2018). When events occur at regularly timed intervals, the brain switches from a continuous vigilance mode to a rhythmic mode, whereby attention is dynamically pre-allocated to the most likely temporal occurrence of future events (Breska & Deouell, 2017; Rimmele et al., 2018). These *rhythm-based predictions* rely on neural entrainment in the brain, such that neural oscillations synchronize with the periodic stimulus (Lakatos et al., 2008), thereby facilitating the processing of stimuli occurring at expected times. When events do not occur at periodic intervals, their timing can often be predicted through non-rhythmic temporal regularities (Nobre & Van Ede, 2018; Rimmele et al., 2018). For example, a *cue* event may indicate that some other event will occur after a set interval of time (e.g., a yellow traffic light is a cue that a red light will appear soon). This association of cues with determinate time intervals in memory allows for *memory-*

based temporal predictions that engage neural mechanisms that are temporally and neuroanatomically distinct from neural entrainment (Breska & Deouell, 2017; Breska & Ivry, 2018).

As reviewed in Chapter 1, there is considerable debate as to whether rhythm-based predictions underlie comprehension (also see Beier et al., 2021; Beier & Ferreira, 2018), and studies have yet to provide evidence for the involvement of memory-based temporal predictions in language. My proposed experiment will address these gaps by characterizing the role of rhythm-based and memory-based temporal predictions in guiding attention during spoken language comprehension. I predict that the regular speech rhythm induced by the alternation of stressed and unstressed syllables will engage rhythm-based predictions, while discourse cues (preceding questions) will guide attention through memory-based predictions.

The stimuli will consist of 320 question-answer pairs. Similarly to Cutler and Fodor (1979) and Wang et al. (2011), the preceding question will lead to either early or late focus in the answer (e.g., "Which cook bought the meat?" / "What kind of meat did the cook buy?"). Similarly to Rothermich et al. (2012), there will be two versions of the answer with varying degrees of rhythmic regularity. In the regular rhythm version, stressed and unstressed syllables will alternate periodically (e.g., "The <u>cook</u> who <u>worked</u> at <u>night</u> bought <u>beef</u> for <u>work</u>", stresses underlined). In the irregular rhythm version, a stress clash and lapse will render the metrical structure infelicitous (e.g., "The <u>cook</u> who <u>worked</u> <u>night</u> shifts bought <u>beef</u> for <u>work</u>"). A target word (e.g., "beef") will therefore be focused either through the discourse or through the predictable rhythmic prosody. Thus, this experiment will combine discourse and metrical prosody cues to IS to investigate their respective effects on temporal predictions and attention.

Participants will listen to all question-answer pairs with conditions counterbalanced across four lists, while EEG will be recorded. I will rely on several EEG measures to index attention throughout the sentence. These include the suppression of alpha oscillations (8-13 Hz), an established correlate of the deployment of attention (Bengson et al., 2014; Boudewyn et al., 2015; Liu et al., 2016); and the phase coherence of delta oscillations (0.5-4 Hz), which is associated with temporally predicted events (Breska & Deouell, 2017). I will also develop computational models to assess whether the increased rhythmic regularity induces neural entrainment, based on prior work (Kojima et al., 2021; Oganian & Chang, 2019). I will compare the collected EEG data to the predictions of an oscillatory entrainment model and those of an evoked response model, addressing the current debate on whether evidence for cortical tracking of speech is related to neural entrainment for non-linguistic stimuli.

In a planned follow-up study, I will also manipulate the semantic predictability of the target word (e.g., “beef”) by varying the preceding discourse question (e.g., “What kind of vegetable did the cook buy?” which renders “beef” incongruent), in a similar design to Wang et al. (2011). In this way, I will be able to directly compare the effects of discourse and rhythmic prosody on the N400 effect in a single paradigm, reconciling the findings of Wang et al. and Rothermich et al.

In summary, my next step towards a full understanding of the mechanisms involved by IS will be to investigate the neurocognitive processes of temporal prediction that underlie the pre-allocation of attention towards focused words, as assumed under the More Resources account.

Relating to linguistic theories: do we have IS backwards?

In this dissertation I have focused primarily on the cognitive processes thought to be involved by IS. Therefore, I have followed the definitions commonly used in psycholinguistics, which distinguishes between presentational focus (any new information) and contrastive focus, in opposition to given information already known to both interlocutors. Across different cognitive accounts, the implicit assumption is that speakers purposely mark new information through prosodic and syntactic cues, with the goal of facilitating comprehension in the listener. Listeners then rely on these cues rationally to infer the timing of new information (Roettger & Franke, 2019).

However, as briefly reviewed in Chapter 1, not all linguistic theories agree on this definition of IS. In particular, a recently published linguistic perspective (Kratzer & Selkirk, 2020) puts into question the assumptions held by psycholinguists, and it has the potential to completely turn the field on its head.

Kratzer and Selkirk (2020) argue against the idea, common to psycholinguistics and many linguistic theories, that any new or contrastive information is focused against a “baseline” of given information. For example, Selkirk's (1984, 1995) older view was that there is a morphosyntactic marker for Focus, which influences the prosodic realization of sentences such that both new and contrastive information receive a pitch accent in English. In their novel account, Kratzer and Selkirk claim that “newness” is the baseline in communication, and it therefore does not need to be marked through focus. Rather, IS is determined by two distinct morphosyntactic features. Given information is marked through [G], which signals a match with a previously mentioned referent, and it is characterized through a lack of prosodic prominence in English. On the other hand, what they refer to as FoCus-marked information ([FoC]) always invokes a contrast with potential alternatives in the discourse, and in English it aims for the highest prosodic prominence. As an example, the traditional view of Focus-marking would lead to the following:

(1) What color is your new car?

My new car is [red]$_{F}$.

(2) Is your new car green?

My new car is [red]$_{F}$.

Here, "red" receives the same type of Focus marking (in English, a pitch accent) regardless of whether it is merely new information or whether it marks a contrast with "green". Given information, on the other hand, is the default and it is not marked. Kratzer and Selkirk's account would instead lead to the following:

(3) What color is your new car?

[My new car]$_{G}$ is red.

(4) Is your new car green?

[My new car]$_{G}$ is [red]$_{FoC}$.

Here, purely new information (as in (3)) is the default and it is not Focus-marked. Rather, the given information is marked with [G] and the contrastive information with [FoC].

The fact that non-contrastive new information is also associated with prosodic prominence in English, in their view, merely results from default prosody; for example, new information is typically presented at the end of sentences, which also happens to receive the highest prosodic prominence through the default Nuclear Stress Rule.

Further, Kratzer and Selkirk argue for the necessity of morphosyntactic markers for givenness and contrast by providing several cross-linguistic examples of how these markers influence language production. In some views (e.g., Calhoun, 2010), IS can be derived purely through the metrical organization of stresses/accents in prosody; whether a word constitutes new, given or contrastive information can be inferred through deviations from the default prosody alone. Kratzer and Selkirk challenge this interpretation through examples of languages, such as Gungbe and Wolof, where contrast is only marked through syntax (e.g., a focus particle) and not prosody. They further argue for a distinction between presentational focus (the default) and contrastive focus through examples from languages such as Italian, where the two lead to different syntactic structures. Overall, the range of phenomena from intonation/prosody, pragmatics and semantics associated with IS are merely reflections of the [FoC] and [G] morphosyntactic features, which determine the prosodic and syntactic realization of sentences across languages. In their own words, "there is no such thing as information structure that deserves this name" (p. 42).

Kratzer and Selkirk's account questions the assumptions of psycholinguistic studies of IS, and in particular the idea that listeners use prosodic and syntactic focus cues to infer the timing of new information among a baseline of given information. Instead, listeners may assume that all information they will receive is new by default – after all, communicating new information is one of the primary goals of language. Rather than marking newness, speakers may use prosody and syntax to signal a deviation from this assumption. They may mark given information to signal that it is not the answer to the QUD, but rather a recapitulation of already mentioned concepts to facilitate the progression of discourse. And they may mark contrastive information to signal that a previously understood idea was incorrect and needs to be revised. Thus, listeners use cues to IS not to infer the timing of new information, but rather the timing of given content (to be “ignored”) and contrastive focus (leading to a revision of the common ground).

This alternative view poses consequences for the cognitive mechanisms associated with IS. For example, under the More Resources account, focused new information receives facilitated processing thanks to the pre-allocation of attention; under the Deeper Processing account, focused new information receives deeper processing. If we no longer assume that new information has special status, then the level and ease of processing of new information becomes the baseline. Rather than pre-allocating attention and processing resources *towards* new information, listeners deploy attention and processing resources *away* from Given-marked constituents. Thus, temporal predictions would be used not to infer the timing of new information, but the timing of the given information, to be shallowly processed.

My proposed experiment on the attentional mechanisms involved by IS may reveal whether this is the case. For example, my original prediction (based on the mainstream psycholinguistic view) is that listeners will rely on the discourse questions to predict the timing of new information in the answer, through memory-based temporal predictions. However, if listeners are in fact predicting the timing of given information, we may measure the opposite pattern, where memory-based temporal predictions follow the location of given information. Moreover, this alternative view suggests that rhythmic regularity and discourse cues to IS may influence processing in fundamentally different ways. While discourse may *inhibit* the processing of given information, rhythmic prosody may still *facilitate* the processing of prosodically prominent words falling on metrically strong positions. Rhythm-based temporal predictions through neural entrainment have been shown to facilitate processing of temporally expected events outside of language (Lakatos et al., 2008; Nobre & Van Ede, 2018), so if this mechanisms is engaged by speech rhythm, I would expect it to have a similar effect. Thus, this proposed experiment may reveal whether discourse cues and prosodic rhythm are interchangeable as previously assumed (Cutler & Fodor, 1979) or whether they have complementary and opposite effects on processing (inhibiting given information, facilitating prosodically prominent information), which would help explain the previous inconsistent findings regarding the effects of these two cues on the N400.

The only type of focus to draw attention beyond the baseline of newness would then be contrastive focus. Contrastively-marked information requires more processing resources than merely new information: not only does it append new data to the listener's store of knowledge, it also revises an incorrect representation previously encoded in long-term memory. The good-enough language processing framework (F. Ferreira & Patson, 2007) reveals that in the case of locally ambiguous sentence structures, comprehenders often fail to revise the first incorrect interpretation (F. Ferreira et al., 2001; F. Ferreira & Henderson, 1991). This is also the case for non-ambiguous sentences that depart from the canonical organization of thematic roles (e.g., passive sentences) or that convey implausible ideas contradicting a frequent schema (e.g., a man biting a dog; Chantavarin, 2021; Ferreira, 2003). Thus, speakers may use contrastive focus to ensure that listeners will not process their utterances in a good-enough manner, and instead revise the incorrect representation.

This shift in perspective may help explain the previous inconsistent findings with respect to the effects of focus, as "presentational focus" (which in this view is actually the default expectation of newness) may have fundamentally different effects on processing than contrastive focus. This is in line with studies showing a dissociation between the effects of focus and newness on eye-tracking reading times (Chen et al., 2012) and ERPs (Chen et al., 2014).

Because prosodic cues associated with newness and contrastive focus are often hard to distinguish (Breen et al., 2010; Roettger et al., 2019), some experiments that claim to manipulate presentational focus may in fact lead listeners to perceive a contrast. Further, some linguistic perspectives argue that syntactic cleft focus constructions (such as it- and wh-clefts) always imply a contrast (Rochemont, 1986).

Following this logic, the IS of the stimuli used in Experiments 1 and 2 may be as follows:

(1) What he made was a [tasty sandwich]$_{\text{FoC}}$…

(2) He made a tasty sandwich…

(3) After he made a [tasty sandwich]$_{\text{G}}$…

where the focusing construction induced the expectation of contrastive focus. My planned visual world study may reveal whether this is the case by measuring fixations towards less predictable but plausible alternatives.

Overall, Kratzer and Selkirk's view challenges the cognitive accounts of IS and may prompt a reinterpretation of previous findings in a new light. An important takeaway is that psycholinguistic studies need to be careful about how they manipulate newness, givenness and contrast; a more specific definition of "focus" may be required in order not to confound these concepts. In the current literature, presentational and contrastive focus have been studied mostly independently. However, it may be necessary to manipulate both in the same paradigm in order elucidate their respective effects on processing, and determine which IS cues are interpreted as presentational vs. contrastive.

Conclusions: towards an integrated view of information structure

In this dissertation I have reviewed a broad collection of phenomena falling under the umbrella term *information structure*. These phenomena range from the simple distinction of *new* and *given* information in discourse, to various aspects of prosody and intonation (e.g., metrical structure, pitch accents) to various syntactic devices (e.g., clefts, preposing, ellipsis, focus particles like "only"…). Literature on this topic arises from disparate fields including psycholinguistics and cognitive neuroscience, pragmatics, formal semantics and linguistics. Yet across this large body of research, there is no consensus on the very definition of IS.

How is IS represented in the mind? Some perspectives treat IS as a semantic/pragmatic representation separate from syntax. Some view IS as arising from prosody alone. Kratzer and Selkirk (2020) argue that "there is no such thing as information structure" (p. 42) but rather IS-related phenomena emerge only from morphosyntactic markers visible to syntactic operators, which influence both prosody and syntactic structure.

My view is that at its core, IS is a cognitive representation of both interlocutors' state of knowledge, which is separately built in each interlocutor's mind. It takes into account what has been previously mentioned in the discourse (the current conversation or a conversation held in the past), plausible inferences based on the discourse, world knowledge assumed to be shared, the current availability of a concept in memory, and ideas/prejudices about the interlocutor (e.g., a little girl and an old man may be expected to know and believe different things). This complex representation is formed prior to the linearization of the spoken output, and it is continuously updated over the course of the conversation, determining both the content and shape of each utterance. It determines the content of what is said, as the speaker chooses which concepts and ideas to convey based on what they think the listener already knows. And it determines the syntactic structure and prosody used to convey this content. Both have a default structure (e.g., canonical word order and default prosody guided by the Nuclear Stress Rule), which typically corresponds to the given-before-new order. The speaker may choose to deviate from these defaults in order to place emphasis on a specific concept (e.g., to indicate a contrast) or to de-emphasize it (e.g., to indicate that something refers to what was previously mentioned).

At the same time, a separate IS is formed in the mind of the listener, combining what the listener already knows, what the listener thinks the speaker knows, and what the listener thinks the speaker knows the listener knows… and so on. Thus, IS not only involves each interlocutor's state of knowledge, but also a recursive simulation of what each person knows their interlocutor knows. Based on this simulation, IS influences comprehension in a rational manner. Listeners expect the speaker's utterance to conform to default syntax and prosody, and infer that any deviation from these defaults is meant to signal givenness or contrast. These expectations and inferences continuously guide all aspects of comprehension, from attention to depth of processing to semantic prediction.

This dissertation represents a first step towards an integrated understanding of this complex interaction of mental representations and cognitive processes. More research is clearly needed in order to fully grasp what IS is, how it is represented, and how it drives both language production and comprehension. What I hope to have demonstrated, however, is that IS is at the very core of language: there cannot be a complete understanding of production and comprehension that does not take into account the mental representation of each interlocutor's state of knowledge.

References

Beier, E. J., Chantavarin, S., Rehrig, G., Ferreira, F., & Miller, L. M. (2021). Cortical tracking of speech: Toward collaboration between the fields of signal and sentence processing. *Journal of Cognitive Neuroscience*, *33*(4), 574–593. https://doi.org/10.1162/jocn_a_01676

Beier, E. J., & Ferreira, F. (2018). The temporal prediction of stress in speech and its relation to musical beat perception. *Frontiers in Psychology*, *9*, 431. https://doi.org/10.3389/fpsyg.2018.00431

Beier, E. J., & Ferreira, F. (2022). Replication of Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. Cognition, 7(1), 49–59. *Journal of Memory and Language*, *126*, 104339. https://doi.org/10.1016/j.jml.2022.104339

Bengson, J. J., Kelley, T. A., Zhang, X., Wang, J.-L., & Mangun, G. R. (2014). Spontaneous Neural Fluctuations Predict Decisions to Attend. *Journal of Cognitive Neuroscience*, *26*(11), 2578–2584. https://doi.org/10.1162/jocn

Boudewyn, M. A., Long, D. L., Traxler, M. J., Lesh, T. A., Dave, S., Mangun, G. R., Carter, C. S., & Swaab, T. Y. (2015). Sensitivity to referential ambiguity in discourse: The role of attention , working memory, and verbal ability. *Journal of Cognitive Neuroscience*, *27*(12), 2309–2323. https://doi.org/10.1162/jocn

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7), 1044–1098. https://doi.org/10.1080/01690965.2010.504378

Breska, A., & Deouell, L. Y. (2017). Neural mechanisms of rhythm-based temporal prediction: Delta phase-locking reflects temporal predictability but not rhythmic entrainment. *PLoS Biology*, *15*(2), 1–30. https://doi.org/10.1371/journal.pbio.2001665

Breska, A., & Ivry, R. B. (2018). Double dissociation of single-interval and rhythmic temporal prediction in cerebellar degeneration and Parkinson's disease. *Proceedings of the National Academy of Sciences*, *115*(48), 12283–12288. https://doi.org/10.1073/pnas.1810596115

Calhoun, S. (2010). The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language*, *86*(1), 1–42.

Chantavarin, S. (2021). *Examining The Competing Interpretations Hypothesis in the Comprehension of Noncanonical Structures* [Doctoral dissertation]. University of California, Davis.

Chen, L., Li, X., & Yang, Y. (2012). Focus, newness and their combination: Processing of information structure in discourse. *PLoS ONE*, *7*(8). https://doi.org/10.1371/journal.pone.0042533

Chen, L., Wang, L., & Yang, Y. (2014). Distinguish between focus and newness: An ERP study. *Journal of Neurolinguistics*, *31*, 28–41. https://doi.org/10.1016/j.jneuroling.2014.06.002

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, *7*(1), 49–59. https://doi.org/10.1016/0010-0277(79)90010-6

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203. https://doi.org/10.1016/S0010-0285(03)00005-7

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of reanalysis. *Journal of Psycholinguistic Research*, *30*, 3–20.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, *30*(6), 725–745.

Ferreira, F., & Patson, N. D. (2007). The 'Good Enough' Approach to Language Comprehension. *Language and Linguistics Compass, 1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Kojima, K., Oganian, Y., Cai, C., Findlay, A., & Chang, E. (2021). Low-frequency neural tracking of speech amplitude envelope reflects the convolution of evoked responses to acoustic edges, not oscillatory entrainment. *BioRxiv*.

Konkle, T., & Oliva, A. (2012). A Real-World Size Organization of Object Responses in Occipitotemporal Cortex. *Neuron*, *74*(6), 1114–1124. https://doi.org/10.1016/j.neuron.2012.04.036

Kratzer, A., & Selkirk, E. (2020). Deconstructing information structure. *Glossa: A Journal of General Linguistics*, *5*(1), 1–53. https://doi.org/10.5334/gjgl.968

Kristensen, L. B., Wang, L., Petersson, K. M., & Hagoort, P. (2013). The interface between language and attention: Prosodic focus marking recruits a general attention network in spoken language comprehension. *Cerebral Cortex*, *23*(8), 1836–1848. https://doi.org/10.1093/cercor/bhs164

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, *320*(5872), 110–113. https://doi.org/10.1126/science.1154735

Liu, Y., Bengson, J., Huang, H., Mangun, G. R., & Ding, M. (2016). Top-down modulation of neural activity in anticipatory visual attention: Control mechanisms revealed by simultaneous EEG-fMRI. *Cerebral Cortex*, *26*, 517–529. https://doi.org/10.1093/cercor/bhu204

Lowder, M. W., & Ferreira, F. (2016). Prediction in the processing of repair disfluencies: Evidence from the visual-world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(9), 1400–1416.

Lowder, M. W., & Gordon, P. C. (2015). Focus takes time: Structural effects on reading. *Psychonomic Bulletin & Review*, *22*(6), 1733–1738. https://doi.org/10.3758/s13423-015-0843-2

Nobre, A. C., & Van Ede, F. (2018). Anticipated moments: Temporal structure in attention. *Nature Reviews Neuroscience, 19*(1), 34–48. https://doi.org/10.1038/nrn.2017.141

Oganian, Y., & Chang, E. F. (2019). A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Science Advances*, *5(11),* eaay6279.

Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018). Proactive sensing of periodic and aperiodic auditory patterns. *Trends in Cognitive Sciences*, *22*(10), 870–882. https://doi.org/10.1016/j.tics.2018.08.003

Rochemont, M. S. (1986). *Focus in Generative Grammar*. Benjamins.

Roettger, T. B., & Franke, M. (2019). Evidential Strength of Intonational Cues and Rational Adaptation to (Un-)Reliable Intonation. *Cognitive Science*, *43*(7), e12745. https://doi.org/10.1111/cogs.12745

Roettger, T. B., Mahrt, T., & Cole, J. (2019). Mapping prosody onto meaning – the case of information structure in American English. *Language, Cognition and Neuroscience*, *34(7)*, 841-860. https://doi.org/10.1080/23273798.2019.1587482

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia*, *50*(2), 232–244. https://doi.org/10.1016/j.neuropsychologia.2011.10.025

Selkirk, E. (1984). *Phonology and syntax: The relationship between sound and structure*. The MIT Press. https://dl.acm.org/doi/abs/10.5555/6521

Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. In J. A. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 550–569). Blackwell.

Wang, L., Bastiaansen, M., Yang, Y., & Hagoort, P. (2011). The influence of information structure on the depth of semantic processing: How focus and pitch accent determine the size of the N400 effect. *Neuropsychologia*, *49*(5), 813–820. https://doi.org/10.1016/j.neuropsychologia.2010.12.035

Yang, C. L., Zhang, H., Duan, H., & Pan, H. (2019). Linguistic focus promotes the ease of discourse integration processes in reading comprehension: Evidence from event-related potentials. *Frontiers in Psychology*, *9*, 1–14. https://doi.org/10.3389/fpsyg.2018.02718