CHAPTER FIVE

Neural Indices of Structured Sentence Representation: State of the Art

Ellen Lau
Department of Linguistics, University of Maryland, College Park, MD, United States
E-mail: ellenlau@umd.edu

Contents

Abstract

Natural language is characterized by structured, hierarchical relationships between morphemes, some of which can span an unbounded amount of intervening material. Determining how such relationships are neurally encoded in on-the-fly language comprehension is a fascinating challenge for cognitive neuroscience, and depends on foundational assumptions about both human parsing and the memory architecture. This chapter reviews current approaches towards discovering neurophysiological correlates to the encoding of structured sentence representations.

Consider the difference being given a list of things to pick up at the store which reads:

chocolate

eggs

Now consider being asked *Please go to the store and pick up some chocolate eggs*. In the former case, the two items *chocolate* and *eggs* become related in our mental representation at some level, by virtue of sharing membership in the temporary category 'things to pick up at the store right now'. But

ISSN 0079-7421
https://doi.org/10.1016/bs.plm.2018.08.004

in the latter case, the same two words become related in a different way. Here the listener understands this sequence of two words to correspond to a single phrase, which refers to a single referent in the discourse (a set of chocolate eggs), and where the conceptual properties of that referent follow from applying some of the conceptual properties labelled by *chocolate* to those labelled by *eggs*. This difference in how the same sequence is analyzed derives from the context and from the listener's knowledge of the grammar of English, which specifies that the head of a noun phrase follows noun or adjectival modifiers—in other words, *chocolate egg* means a type of egg that is made of chocolate, not a type of chocolate that is made out of eggs.

Natural language is characterized by structured, hierarchical relationships between words, including non-local relationships which can span across a potentially unbounded amount of intervening material. Although we've begun with a simple case, much more complex sets of relations between elements are routinely observed in natural utterances. For example, in the sentence *I like cold and creamy brownies with ice cream*, listeners relate the two descriptors of *cold* and *creamy* to *brownies* (even though they might be more plausible descriptors of *ice cream*), they also relate the locative phrase *with ice cream* to *brownies*, and furthermore the entire unit *cold and creamy brownies with ice cream* is related to the verb *like*. The hierarchical, non-linear phrasal relationships that form sentence structure are also famously illustrated by agreement phenomena; in subject-verb agreement in English, the form of the verb depends not on the number of the noun linearly preceding it, but on the number of the noun that 'heads' the phrase, hence *brownies with ice cream are delicious* and not *brownies with ice cream is delicious*. These are some of the many facts that motivate the descriptors 'structured' and 'hierarchical' for natural language. In other sentences a noun phrase may be linearly distant from the verb that it is related to, as in questions like *What kind of eggs did you ask your sister to buy?* In principle, the linear distance between the argument and the verb is unlimited, e.g. *What kind of eggs did you ask your sister to tell her daughter to buy?* This is the kind of fact that motivates the descriptor 'non-local'. The kind of relationships evidenced by these facts are often graphically depicted by linguists using 'tree diagrams' as in Figure 1, but it is important to recognize that while these diagrams can be used to illustrate different subtheories of linguistic relations, the format of the diagrams themselves is not a theory of mental representations (as is wrongly implied when newcomers to the field say that they want to use neuroimaging to ‘find out whether syntactic trees are really in the brain’). Rather, tree diagrams are just a convenient shorthand for illustrating the structured relations between words and phrases that *must* be represented by the brain in order to explain the sentences that people produce and understand.

This chapter reviews the current state-of-the-art in cognitive neuroscience towards discovering how the brain encodes this relational information in real time to allow successful language comprehension. This question is a challenging and important one, as there is little consensus about how short-term or working memory operations are neurally implemented in general, and relatively little is known about how relational information is encoded in short-term memory in particular. Therefore, much of the work reviewed here does not yet take the form of testing an explicit proposal for neural encoding of structured representations in language, but rather begins by asking a somewhat simpler question: what, if any, current neurophysiological measures track properties of these representations?

1. LINGUISTIC AND NON-LINGUISTIC RELATIONS

An important distinction to keep in mind throughout the following is between *linguistic* and *non-linguistic* representations. Even without language, humans and many other organisms seem able to represent and reason about concepts—for example, the ability of preverbal infants to associate different sensory and non-sensory properties with the same caregiver representation. It similarly seems that even in the absence of language, we can rapidly encode novel relations between existing conceptual representations—such as the fact that my mother possesses an apple, which I can immediately encode when I see her pick one up and put it in her pocket. We also know that in many cases humans can relatively easily retain these new conceptual relations in long-term memory, at least for the duration of a few minutes or hours.

In learning a language, we acquire ‘words’ (more technically, *morphemes*) that label some subset of these concepts. These linguistic objects come with their own set of language-specific restrictions about what kind of relations can be created between them. For example, even without knowing the words *chocolate* and *egg*, you could likely learn these conceptual categories from experience, and you could also learn that there is a prototypical way in which these concepts are combined in the world (e.g., that chocolate eggs *can* be any size but are most often smaller than ostrich eggs). But there is nothing about the concepts of *chocolate* and *egg* in themselves that causes the English phrase *chocolate egg* to refer to a kind of egg and *egg chocolate* to refer to a kind of chocolate—this is a fact about language, and not about the concepts. Therefore, sentence comprehension can be roughly viewed as a process of encoding *linguistic* relations between morphemes and phrases, which can then in turn drive the encoding of new relations between the concepts or discourse referents that the words label in long-term memory. 'Parsing' is the term that is usually used to refer to the process of determining the linguistic relations. The current chapter focuses on work whose aim was primarily to better understand the neural encoding of the linguistic relations, even though it is difficult in practice to clearly dissociate neural activity associated with encoding linguistic and non-linguistic relations.

Note that one reason to suppose that the encoding of linguistic relations might be qualitatively and mechanistically distinct from the conceptual relations is that there is no need to retain the linguistic relations after the sentence is over—we don't usually need to remember how a sentence was framed, we just need to remember the non-linguistic message that it conveyed. Indeed, classic behavioral work suggested that information about the form of the sentence became difficult to retrieve once the sentence was over, while the meaning persists over the long term (Sachs, 1967; Jarvella, 1971; although see; Murphy & Shapiro, 1994). As I will discuss in more detail below, this observation is very relevant for evaluating the likelihood that non-invasive measures of neural activity can track currently encoded linguistic relations at all. Retention of information in long-term memory should not rely on ongoing neural activity, if it is to be metabolically plausible, and therefore if linguistic relational information were carried forward in time through rapid encoding in long-term memory, we would have no hope of observing indices of those relations in subsequent neural activity. However, it is not so implausible to suppose that retention of information in short-term or working memory depends on ongoing neural activity—this is a popular if debated hypothesis in cognitive neuroscience—and therefore if linguistic relations were carried forward during the course of the sentence in working memory, there is a reasonable possibility that we could uncover indices of these relations in neural measures.

2. NEURAL CORRELATES OF LINGUISTIC RELATIONS: THREE CANDIDATES

One classic approach to psycholinguistic and neurolinguistics research on language has been to define some metric of structural complexity over sentences, and then to cast a wide net for any behavioral or neural dependent measure that correlates with this metric. In principle, this can be attempted without any explicit hypotheses about the parsing model—in other words, without any hypotheses about what operations occur in what sequence with what memory architecture such that the structural relations can be incrementally, accurately, and efficiently deduced from a serial input stream in real time. Although such a data-driven approach would be logically sound, it has an obvious weakness: it may well be that the processing model that links the input sentence to the output measure is complex enough that we will never be able to identify the true correlations between structure and neural activity without building in some basic hypotheses about it. For example, one strategy for parsing a sentence, known as 'bottom-up', is to record the incoming string until an end-of-sentence signal is encountered, and only then to compute the linguistic relations that support sentence meaning. An alternative parsing strategy, known as 'top-down', is to infer the relations incrementally by making well-informed guesses about what the rest of the string will contain. It should be clear from these extremes that the timecourse of neural activity associated with computing and representing linguistic relations could be wildly different depending on which parsing strategy humans actually use.

The rest of the chapter is therefore organized around three candidate types of neural correlates for linguistic relations (Fig. 1) that rely on simple (but not uncontroversial) hypotheses about the processing model. First, we can assume simple, 'word-by-word' incrementality, and then look for the neural correlates of the process of initially instantiating a local linguistic relation between two elements in the time-window immediately after the second element—for example, investigating the neural activity following *eggs* in the context of the phrase *chocolate eggs*. The second and third candidates discussed here are neural correlates of aspects of the structured memory representation—the processes involved in carrying forward these relational information over time—given particular hypotheses about the human memory architecture and its neural implementation. The second section

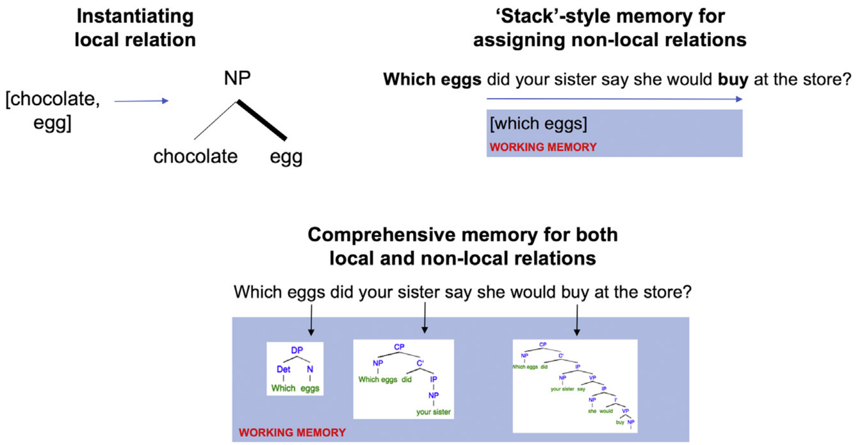
Instantiating local relation
[chocolate, egg]
NP
chocolate
egg
'Stack'-style memory for assigning non-local relations
Which eggs did your sister say she would buy at the store?
[which eggs]
WORKING MEMORY
Comprehensive memory for both local and non-local relations
Which eggs did your sister say she would buy at the store?
DP
Det
N
Which
eggs
CP
NP
C'
Which eggs
did
IP
NP
your sister
CP
NP
C'
Which eggs
did
IP
NP
VP
your sister
say
IP
NP
I'
she
would
VP
buy
NP
WORKING MEMORY

Figure 1 Three candidate types of neural correlates for linguistic relations

focuses on a classic conception of syntactic working memory that emphasizes the information which must be carried forward in cases of non-local relations in order to determine whether the input string could have been generated by the listener's grammar or not. The third section looks at a broader conception of memory for structure, which is at least the relational information which must be carried forward to ensure successful interpretation of the sentence, and at most a full structural analysis of the input, which might be useful for recovery if it turns out that part of the sentence was misanalysed.

2.1 Instantiating Local Linguistic Relations

The first aspect of the neural implementation of linguistic structure that I will consider is exactly the moment-to-moment neural processes that result in a unified syntactic, semantic, and conceptual representation for word sequences presented in a context that encourages linguistic structuring. While in subsequent sections I will examine how these structured representations may be carried forward in time over seconds or minutes, in this first section the question is, what kind of neural activity is associated with establishing these relational representations in the first place?

Perhaps the most straightforward approach to answering this question is simply to look for extra brain activity engaged at the moment when linguistic combination happens, relative to when it does not. In a highly influential paper Bemis and Pylkkänen (2011), recorded MEG responses while participants read simple two word phrases (*red boat*), looking for evidence of activity at *boat* reflecting the instantiation of syntactic or semantic relations between the two words. They compared this response to several controls designed not to elicit the instantiation of syntactic or semantic relations: *xkq boat*, in which the first stimulus was a consonant string that was not associated with a word or concept and so could not be linguistically related to the second, and *cup boat*, where the task instructions were designed to encourage processing of the sequence as an unrelated list of two nouns rather than a phrasal compound. In MEG source localization analyses, Bemis and Pylkkänen showed an increased response to the noun in the phrasal context by around 250 ms in left anterior temporal cortex, an effect that they soon replicated in other experiments (Bemis & Pylkkänen, 2012, 2013a, 2013b). In some of these studies Bemis and Pylkkänen also observed increased activity for the phrasal condition in ventromedial prefrontal cortex (Bemis & Pylkkänen, 2011) and angular gyrus (Bemis & Pylkkänen, 2012), although effects in these other regions have been slightly later and less reliable across studies. We have replicated these basic observations in EEG, showing that the same paradigm elicits differential responses within 180–400 ms after a noun is presented in a phrasal vs. non-phrasal context, in the form of more negative potentials for the noun in a phrasal context (Neufeld et al., 2016). We also observed an even earlier combinatorial effect, which we tentatively attributed to predictive structure-building operations.

This approach towards identifying neural correlates of combinatoric relations has thus seemed to bear fruit. However, further work suggests that at least one of these responses is better attributed to non-linguistic, conceptual combination processes rather than linguistic ones. Initially, Bemis and Pylkkänen argued that the increased anterior temporal activity indexed the computation of linguistic (specifically, syntactic) relations between the adjective and the noun, because the response was relatively early and was not observed for non-linguistic combination such as stimuli using images (Bemis & Pylkkanen, 2013a, 2013b). However, later studies cast doubt on this interpretation because they showed that not all cases of linguistic combination elicited an increased anterior temporal response; rather, there was a strong dependence on the conceptual properties of the nouns being composed (Westerlund & Pylkkänen, 2014; Zhang & Pylkkänen, 2015). For example, while *red boat* elicited an increased response relative to *xkq boat*, *blue canoe* did not elicit an increased response relative to *xkq canoe*, and while *tomato dish* elicited an increased response relative to *qptg dish*, *vegetable dish* did not elicit an increased response relative to *qptg dish*. Based on this pattern, Pylkkänen et al. suggested that the increased anterior temporal response was specifically tied to increases in semantic 'specificity'. This can be understood as restriction in the size of the set of objects that are denoted by the phrase (e.g., the set of tomato dishes is much smaller than the set of dishes, while the set of vegetable dishes is still reasonably large in comparison) or in terms of diagnosticity (in the phrase, *red boat*, the color adjective is critical for distinguishing the denoted item from other items in the broader boat category, whereas canoes already have a number of features that distinguish them from other boats). This hypothesis can be related to a broader literature on semantic memory and conceptual combination that is beyond the scope of the current chapter (e.g. Patterson, Nestor, & Rogers, 2007; Baron, Thompson-Schill, Weber, & Osherson, 2010; Boylan, Trueswell, & Thompson-Schill, 2017). However, several puzzles remain for the semantic specificity account. First, it is not yet clear how exactly to operationalize specificity or diagnosticity such that *red boat* > *boat* in this property but *vegetable dish* = *dish* (Zhang & Pylkkänen note that it is easier to quantify relative specificity for noun modifiers than adjectives). Second, the linking hypothesis between specific interpretation processes and the increase in anterior temporal activity at ~200 ms remains somewhat unclear—in other words, why exactly is there greater or more synchronized neural activity in anterior temporal cortex at 200 ms for phrases that denote more specific concepts?

In an EEG replication of the original Bemis and Pylkkänen (2011) work, we noted that the timing and distribution of the 'red boat' effect bear some resemblance to standard N400 effects in EEG (Neufeld et al., 2016); this also fits with prior localization work showing that some anterior temporal regions contribute to N400 effects (Van Petten & Luka, 2006; Lau et al., 2008, 2014; Lau, Weber, Gramfort, Hämäläinen, & Kuperberg, 2014). If this resemblance indeed reflects a common source, we could draw on current theories about N400 effects to inform the linking hypothesis for the 'red boat' effect. One family of theories suggests that increased N400 responses reflect the increased activation of conceptual and lexical memory networks (e.g. Kutas & Federmeier, 2000; Lau et al., 2008; Federmeier & Laszlo, 2009), and that this network activation is reduced, resulting in smaller N400 responses, when the prior context predicts or primes the subsequent input. In the context of this framework, the 'red boat' effect might reflect increased attention to or retrieval of conceptual features associated with the individual words, in the service of deriving inferences about the intended message or of retrieving exemplars of the entity or event described from episodic memory. For example, out of context, seeing the word *boat* may trigger retrieval of some minimal conceptual information to the effect of 'vehicle for transport in water', but in the context of *I started to deflate my boat*, *boat* may additionally trigger retrieval of episodic memories about inflatable boats I have seen, which may afford inferences about the likely appearance, approximate size, and uses of the boat. It is important to note that, like the specificity hypothesis above, this explanation would attribute the proximate cause of the 'red boat' effect to a non-combinatorial computation (activation of stored conceptual features) that is indirectly modulated by combinatorial contexts. Similarly, while fMRI studies consistently observe increased activity in anterior temporal cortex for structured sentences over lists (e.g. Mazoyer et al., 1993; Friederici, Meyer, & von Cramon, 2000; Vandenberghe, Nobre, & Price, 2002; Rogalsky & Hickok, 2008), this may not indicate that anterior temporal cortex instantiates the combinatorial relations itself, but rather that sentence structure increases the activation and retrieval of individual conceptual features (see Wilson et al., 2014; for more discussion).

Recent work in fMRI has made use of similar paradigms, where two-word and three-word phrases are compared to unstructured lists (Graves, Binder, Desai, Conant, & Seidenberg, 2010; Zaccarella & Friederici 2015a,b; Matchin, Hammerly, & Lau, 2017; Schell, Zaccarella, & Friederici, 2017), and have reported increased activity for phrases across a variety of brain regions associated with language processing such as anterior and posterior inferior frontal cortex, posterior superior temporal sulcus, anterior temporal cortex, and angular gyrus. While an existing fMRI literature had often reported comparisons between sentences and word lists, these newer studies on short phrases are much better-positioned to identify activity associated with forming linguistic relations than most of the prior fMRI literature on syntactic processing. This is because the poor temporal resolution of fMRI means that activity from forming the relations on the fly cannot be distinguished from activity associated with other components of sentence processing such as working memory and discourse updating, or post-stimulus task performance. However, even though the studies of shorter phrases represent an improvement in this respect, the impact of post-stimulus task performance on neural responses is still an ongoing concern and may explain why results have varied across these studies.

Although current evidence suggests that the 'red boat' effect reflects conceptual processing that may not be strictly combinatorial, other neural response differences for two-word combination may well have a linguistic source. Solving a few remaining methodological challenges may help to bring broader acceptance to this approach, which remains a simple and powerful way of identifying neural indices of instantiating both conceptual and linguistic relations with time-sensitive EEG/MEG measures. First, there are often concerns about nuisance differences in the 'control', non-combinatorial condition. The control has often replaced the first word of the phrase with a consonant string (*xkq boat*), but does this impact the response to *boat* in unexpected ways, particularly given the participant's task of matching a subsequent image probe to the prior linguistic input? We should work to develop alternative control conditions that may be subject to fewer concerns. In one recent study (Lau, Neufeld, & Idsardi, in prep.), we used identical two-noun sequences in both conditions, but manipulated the relations applied to the input through the use of task instructions, where participants were either encouraged to read the sequence as a list (*pizza, flower*) or as a novel compound (*pizza flower*). Another possibility would be to use a wholly non-linguistic placeholder for the first part of the control trial, which in the experimental context, may seem more natural to participants than consonant strings (e.g. ### ### *boat* vs. ### *red boat*). Second, many of the prior studies used two-word phrases that, because of their morphology, could not begin a well-formed sentence of English (**Red boat is approaching in the distance*). It is not clear what effects, if any, this should have on the extent to which linguistic relations will be automatically applied to the input, but it might be desirable for future studies to emphasize two-word sequences that could have begun a well-formed sentence (e.g. *red boats*).

A very different approach to this problem is illustrated by the work of Jonathan Brennan and others, which I will term the 'naturalistic neuro-computational' approach (Brennan, 2016). Rather than conducting a categorical comparison between cases that require combination and cases that do not, Brennan creates a model of the number and kind of combinatorial operations that might be hypothesized to occur at each word of a naturalistic connected text or discourse. Neural activity is then recorded while participants read or listen to the input, and the neural correlates of combinatorial operations are discovered by looking for correlations between the neural activity time-locked to each word and the combinatorial operations hypothesized to occur at each word. For example, in Brennan and Pylkkänen (2017), the MEG source localization estimate for each word of the input is correlated to the number of parsing steps that would be invoked by a left-corner parsing strategy (a popular intermediate between fully top-down and fully bottom-up) to instantiate the syntactic relations between the input word and the prior material. They find that the number of parse steps correlates with left anterior temporal activity between 350 and 500 ms after a word is presented. In other work, Brennan and colleagues use a similar approach to compare different models of parsing strategies and grammar, by asking which model results in the best correlation with neural data (Brennan, Stabler, Van Wagenen, Luh, & Hale, 2016; see also Nelson et al. 2017 for a similar approach using ECoG data). More recently Brodbeck, Presacco, and Simon (2018), have introduced a novel linear kernel estimation approach for MEG data that makes it possible to directly compare estimated spatiotemporal response functions for acoustic features, lexical properties, and combinatorial operations such as semantic composition.

The benefits of the naturalistic neuro-computational approach over the two-word phrase paradigm discussed above is that neural responses are being measured as participants are engaged in what is presumably much more natural language comprehension processes—listening to a continuous narrative—than in more controlled experiments (Brennan, 2016). The challenges are of course exactly in this lack of control; there are many strong internal correlations between linguistic factors—such as function words being shorter and more frequent as well as playing particular kinds of roles in syntactic relations—and so analysis of this kind of data must fully model all of these nuisance variables or risk misattributing variability in the neural response to the wrong factor.

Finally, a new methodology introduced by Nai Ding, David Poeppel and colleagues may provide yet another way of extracting neural responses associated with combinatorial operations (Ding et al., 2016). In what I will term the **constituent-rate** paradigm, different phrases associated with exactly the same structural relations are presented at a constant and unbroken rate. For example, in their first experiments in Mandarin and English they used an unbroken sequence of four syllable sentences with an [adjective-noun]-[verb-noun] structure, where all syllables were exactly the same length and no extra pauses were inserted between sentences (e.g. *dry fur rubs skin sick ducks hate ponds* …). The critical innovation is that a time-frequency analysis is applied to the neural response across the whole sequence of sentences. From their discussion, Ding and colleagues appear to be particularly interested in investigating whether oscillatory activity is used to support short-term memory encoding of structure (as discussed further in the last section of this chapter). However, this method would also pick out activity associated with combinatorial operations that occur at particular moments in the sentence. For example, if a combinatorial operation were invoked whenever a noun is combined with a modifier or a verb, then this operation would happen after the presentation of each of the four nouns in the string *dry fur rubs skin sick ducks hate ponds*, such that if each word had a duration of 250 ms, the corresponding neural activity would show a peak at 2 Hz. Ding and colleagues indeed show reliable peaks in the neural time-frequency data at the sentence rate and at the NP-VP rate in MEG (Ding et al., 2016) and EEG (Ding et al., 2017). Although it may seem a bit convoluted to use a time-frequency measure if the underlying parameter of interest is the *evoked* response associated with instantiating relations, the advantage of the methodology proposed by Ding et al. is that it requires no overt response from participants and has a very fast run-time (15—20 min), while the stereotyped nature of the stimuli means that there is less variability of the kind associated with the naturalistic neuro-computational approach. This gives the Ding et al. methodology the potential to be used to study combinatorial operations in populations that may not be able to perform a task or sit through a long recording session, such as infants and patients, or during sleep (Makov et al., 2017).

Here I summarize the takeaways from this first section. Probably the most well-established generalization thus far about rapid neural correlates of combinatorial operations is the large amount of evidence accrued by Liina Pylkkänen's group that the second word of a two word phrase often elicits increased activity in left anterior temporal cortex at around 250 ms, relative to the same word in non-combinatorial contexts. However, their evidence indicates that this activity reflects computations associated with *conceptual* combination, rather than the instantiation of linguistic relations associated with syntactic phrases or logical forms, and this approach still awaits broader exploration and validation by the field. Other approaches have intriguing early results that might be taken to suggest correlates of linguistic combination, but similarly need a stronger base of evidence before clear conclusions can be drawn.

There are several promising possibilities for making more progress in this area. First, as noted above, advances in the design of two-word phrase experiments could increase their chances of finding responses associated with linguistic structure building. Second, more experiments could take a similar approach to exploring structure-building operations at a slightly higher level, by contrasting the response to lists of phrases with full sentences with time-sensitive neurophysiological measures (Matchin, Brodbeck, Hammerly, & Lau, submitted). There are a number of experimental contexts in which participants find unstructured lists to be a fairly natural mode of input, and researchers could be more fully exploiting these contexts to create non-combinatorial controls for which we have good linking hypotheses.

2.2 Classic Conceptions of Syntactic Working Memory

Gallistel and King (2011) emphasize the importance of distinguishing between something like procedural memory, or 'learning', and something like declarative memory, or 'memory'. Their point is that the neural wiring changes that facilitate computational *procedures* when they are frequently applied—for example when we get better at two-digit multiplication with practice—may be very different from the changes that allow us to stably preserve symbolic *representations* in memory—for example when we hold the number '8' in memory while adding together two other numbers. In the first section of the chapter, I explored research aimed at the neural correlates of the procedures for identifying and forming structured relations between two linguistic elements. The rest of this chapter examines the search for neural indices associated with carrying forward these relations in memory after they are formed.

This second section reviews the neuroimaging literature that was classically grouped under the heading of investigations of 'syntactic working memory'. However, to understand why a few phenomena received most of the focus in this literature, it is important first to recognize that there is a lot of relational information that the combinatorial systems under investigation may not actually *need* to carry forward in memory—just in the way that, if I'm computing a running sum that starts with '2 + 4 + 1 + … ' I can carry forward only the current sum '7', without needing to also retain the inputs and functions that gave me that sum. For example, imagine an utterance that begins *Your red boat is* …. In order to compute the intended meaning of this phrase, the listener must indeed instantiate particular syntactic and semantic relations between *red* and *boat*, and between *your* and the phrase *red boat*. But for the purpose of assigning the correct structural relations to the rest of the sentence, all that needs to be carried forward in memory is just the fact that the sentence began with a noun phrase. For this purpose it doesn't matter whether there was adjectival modification relations inside the noun phrase or not; e.g., there is no rule in English grammar that says something like 'intransitive verbs can follow simple noun phrases, but only transitive verbs can follow noun phrases that contain adjectives'. Similarly, for interpretive purposes, what was initially a novel combinatorial object with relations between the meanings of the three words, can be converted in the discourse model into something like a simple pointer to a long-term memory representation of the particular beloved red boat that the listener has owned for years, and this can be what is carried forward in memory rather than the initial relational object. Therefore, it is easy to imagine a system where many of the on-the-fly syntactic and semantic relations do not have to be carried forward in memory in the same way that the addition relation between 2 and 4 does not have to be carried forward in memory after the result is computed. And note that in this system, there might be no point in time where the brain would have a representation of the data structure depicted by a full syntactic tree diagram—it might be that by the time the verb phrase was being computed, the relations within the noun phrase would have been discarded.

Because of the uncertainty about whether all relational information is maintained in memory throughout the sentence, then, the classic neuroimaging literature on syntactic working memory has had a particular focus on the cases of non-local relations, where immediate conversion of relational information to non-relational representations appears less straightforward. Wh-dependencies such as *I wonder which bird Thomaz looked carefully at* are one such case. Syntactically and interpretively, there appears to be a relation between the leading wh-phrase *which bird* and the verb phrase *looked carefully at*. But since these phrases are not adjacent to each other, and since they can in fact be indefinitely far apart (*I wonder which bird you claimed Thomaz looked carefully at*), forming this relation intuitively seems to require carrying relational information forward in memory across the intervening material. Note that it is easy to detect ungrammaticality at the end of the sentence if the object of the verb phrase is missing (*Thomaz looked carefully at*); the fact that the sentence becomes good in the context of a wh-phrase indicates that its presence must be available in memory, and crucially not just its presence but its structural relations with other parts of the input (as illustrated by the ungrammaticality of **I heard a bird from which country Thomaz looked carefully at*). The parser also needs to retain information about the syntactic properties of the wh-phrase, and not just the fact that it is a wh-phrase in general (*To whom/who does that belong? / *Whom/*Who does that belong?*). Finally, much behavioral psycholinguistic work has established that comprehenders do not process such dependencies with a 'reactive' strategy where the wh-phrase is forgotten until it is noted that an argument is missing, but indeed use a 'proactive' strategy for computing the dependency that indicates that the presence of the wh-phrase is remembered throughout the subsequent material (Stowe, 1986; Traxler & Pickering, 1996; see Phillips & Wagers, 2007 for review). Therefore, wh-dependencies have been seen as a paradigmatic case of syntactic working memory demands.

How might this kind of syntactic working memory be neurally implemented? One classic idea is that it involves sustained neural activity across the duration of the linguistic dependency. Highly influential single-cell recording studies in primates (Fuster & Alexander, 1971; Funahashi, Bruce, & Goldman-Rakic, 1989) showed that during a working memory task where information needed to be maintained over a delay, certain prefrontal neurons persistently fired across the delay period while the task-relevant representation had to be maintained, and that this sustained firing appeared to be specific to which representation was being held in memory. More recently, researchers have argued that sustained neural activity is not critical for all short-term memory representations, but more selectively for those in the focus of attention (Lewis-Peacock, Drysdale, Oberauer, & Postle, 2012; LaRocque, Lewis-Peacock, Drysdale, Oberauer, & Postle, 2013; Rose et al., 2016). Either way, these domain-general models provide motivation for perhaps the most easily testable hypothesis about how syntactic dependencies are encoded across time, where a key prediction is increased neural activity across the duration of the dependency, compared to a control case without an extended dependency.

Consistent with this hypothesis, for decades there have been numerous ERP reports of a sustained anterior negativity (SAN) across the timecourse of wh-dependencies relative to control conditions (Kluender & Kutas, 1993; King & Kutas, 1995; Fiebach, Schlesewsky, & Friederici, 2002; Ueno & Kluender, 2003; Phillips, Kazanina, & Abada, 2005). Some of these studies have contrasted object relative clauses (*The reporter who the senator harshly attacked*) with subject relative clauses (*The reporter who harshly attacked the senator*), where the dependency between *the reporter* and the verb can be almost immediately satisfied in the subject relative clause, but is lengthened by the intervening noun phrase in the object relative clause (Fiebach et al., 2002; King & Kutas, 1995). Other studies have contrasted long-distance wh-dependencies (*The lieutenant knew which accomplice the detective hoped that the shrewd witness would recognize* …) with conditions that contain only local dependencies (*The lieutenant knew that the detective hoped that the shrewd witness would recognize* …) (Phillips et al., 2005). Because these differences sustain even through positions of the sentence where there is no possibility that the dependency could be completed, these effects are hard to explain as a series of targeted memory retrievals, but are exactly the kind of data that a sustained neural activity model of working memory would predict (Sprouse et al., in prep). A number of fMRI studies have also demonstrated increased hemodynamic response in left inferior frontal cortex for the processing of long-distance dependencies relative to controls (e.g. Fiebach, Schlesewsky, Lohmann, Von Cramon, & Friederici, 2005; Santi & Grodzinsky, 2007), although the lack of timecourse information makes it difficult to determine whether this increase reflects sustained differences through the dependency. These data have sometimes been taken to suggest that sustained activity in left inferior frontal cortex reflects necessary machinery for computing non-local structural dependencies, and forms part of the explanation for the uniqueness of human language relative to other species (Friederici, 2012; Fitch, 2014).

However, this longstanding account is also not without problems. On the theoretical side, the last several decades have seen a number of prominent models in which the memory architecture for sentence processing does not require active, sustained maintenance of prior material in a dedicated short-term memory buffer (McElree, Foraker, & Dyer, 2003; Lewis & Vasishth, 2005), although they do include an attentional component that could perhaps be invoked to account for the sustained negativities. On the empirical side there are also various puzzles. Fiebach et al. (2002) observe a SAN for object vs. subject relative clauses even though in German, the linear distance spanned by the wh-dependency is the same in both; their fMRI results associating left inferior frontal cortex with this activity (Fiebach et al. 2005) have the same property. This could mean that what is carried forward in memory is *predicted* syntactic positions; in the subject condition, only a verb need be predicted to host the wh-phrase as a subject, while in the object condition, an object position needs to be predicted in addition to the (transitive) verb itself. However Fiebach et al. (2002), also fail to observe a SAN for sentences with shorter dependencies (… *who on Tuesday the doctor called*) even though the SAN had a relatively rapid onset in sentences with longer dependencies (in … *who on Tuesday afternoon after the accident the doctor called* … the SAN onsets at the word *afternoon*). Phillips et al. (2005) also observe differences between short and long distance dependencies, where the negativity is more posteriorly distributed in the short distance case. In a more recent study Yano and Koizumi (2018), find that the sustained anterior negativity observed for object-scrambled dependencies in Japanese disappears when a more supportive discourse context is provided. In fMRI Matchin, Sprouse, and Hickok (2014), failed to replicate effects of wh-dependency distance in left inferior frontal cortex. These findings are unexpected if sustained neural activity is necessary for accurate computation of non-local syntactic relations.

One natural means of reconciling the apparently conflicting data in the literature is to postulate that wide-scale sustained neural activity in inferior frontal cortex is not *necessary* for computing non-local syntactic relations, but that the sustained effects that have been observed rather index some optional process that makes these computations faster or more accurate. For example, the sustained anterior negativity observed in ERP studies might reflect subvocal rehearsal in an articulatory loop, which could be optionally invoked by the comprehender in order to facilitate memory retrieval in the context of a non-local dependency. This optional process might be engaged more by certain individuals, and this would explain why previous studies have reported individual differences in the amplitude of the SAN based on performance in memory span tasks (Fiebach et al. 2002) and on comprehension questions (King & Kutas, 1995) (although note that the reported effects are somewhat inconsistent with each other, as King and Kutas observed larger SANs for 'good' comprehenders, and Fiebach et al. reported larger SANs for 'low-span' participants who tended to make more comprehension errors).

It could also be that these effects reflect some non-memory property that is often correlated with long distance dependencies. One possibility that to our knowledge has not yet been seriously pursued is implicit prosody. Prior work by Steinhauer and colleagues (Steinhauer & Friederici, 2001; Steinhauer, 2003) argues that prosodic contours elicit characteristic ERP responses and that these responses can be observed even for text, as readers impose these prosodic contours implicitly. The variability in the SAN could be partially accounted for if dependencies of different lengths result in different (implicit) prosodic boundaries, and if individuals vary in the extent to which they apply implicit prosody in reading.

Overall, I believe that the classic literature on sustained neural effects in non-local structural dependencies deserves renewed attention because of its relevance to current theories of the sentence processing memory architecture and corresponding theories of the neural encoding of non-local structural relations. I think the best starting place is more direct evaluations of whether sustained activity is necessary for immediate resolution of non-local structural relations. This can be approached with manipulations designed to modulate the need for optional facilitatory mechanisms, but probably the most critical component of this work is to include sensitive behavioral or neural measures of the extent to which participants in fact resolved the dependency immediately. The extensive behavioral psycholinguistic literature on the instantiation of non-local structural relations in real time suggests several possible means of assessing this, such as evaluating the degree to which responses indicate immediate surprise when a wh-dependency cannot be resolved at the first verb (Stowe, 1986; Traxler & Pickering, 1996). Another important direction is to investigate sustained effects in long-distance dependencies other than wh-dependencies, where different hypotheses about the underlying mechanism may diverge in their predictions (e.g. Matchin et al., 2014).

2.3 Broader Conceptions of Syntactic Memory

The classic literature on syntactic working memory focused on cases where retaining relational information in memory would be strictly necessary for determining whether or not the sentence is grammatical. However, more recent neurophysiological work has assumed that more than this minimal relational information is carried forward. This could be the case for several reasons.

First, the goal of comprehension is to derive an interpretation from an utterance, not to verify grammaticality, and this may require additional relational information to be carried forward. For example, in a 'right-branching' sentence like '*Thomaz thinks that you invaded the house of your new boss*', a top-down parser can evaluate the grammaticality of the sentence by maintaining nothing in memory at each position of the sentence except a single prediction for the next category. After *Thomaz* is related to a noun phrase position, memory needs to contain the prediction for the verb phrase that will complete the sentence grammatically. After *thinks* is related as the head of that verb phrase, memory needs to contain the prediction for the complement clause that will complete the sentence grammatically. After *that* is related to the complementizer position, memory needs to contain the prediction for the embedded clause that will complete the sentence grammatically, etc. In this way, without retaining any of the prior syntactic relations in memory, a 'grammatical' judgment can be accurately returned if all predictions are fulfilled, and an 'ungrammatical' judgment if they are not. However, the described procedure alone obviously will not be sufficient for computing the interpretation, and it is somewhat unintuitive (though logically possible) to hypothesize that an incomplete meaning like *Thomaz thinks that you invaded* … can be converted to a wholly non-relational format in the way I suggested above for a noun phrase like *your red boat*.

Second, retaining relational information from previous parts of the sentence may aid in recovery when it turns out that the current structural analysis was incorrect and needs to be revised. Although certain misanalyses are famously difficult to recover from (strong 'garden paths' like *The horse raced past the barn fell*), recovering from many other kinds of misanalyses is often successful and associated with only mild processing cost (weak garden paths like *I know your friend [… likes ice cream]*). This entails either that comprehenders represent multiple alternative structures in parallel, or that they retain information about the prior input form in memory. If the latter, recovery would likely be easier if this information were structured rather than unstructured (so that the parts of the structural analysis that were still correct could be re-used rather than re-computed altogether).

Pallier, Devauchelle, and Dehaene (2011) reported a seminal fMRI study that was relatively unique in testing a specific model of structure maintenance. In their model, inspired by Smolensky and Legendre (2006), structural relations are built incrementally (as in a top-down parsing strategy), but those relations must be actively maintained in memory (inducing metabolic cost) until the phrase or constituent that they contribute to is complete (reminiscent of a bottom-up parsing strategy). This can be illustrated with the right-branching sentence above; since the verb phrase headed by *thinks* isn't completed until *boss*, their model assumes that the relational information for *thinks* and each of the intervening words has to be actively maintained in memory. They further hypothesize that there is a simple function from the amount of relational information to be maintained and metabolic cost as indexed by fMRI. Therefore, this model predicts that neural activity associated with representing syntactic relations should increase monotonically across the duration of a right-branching sentence, but should return to baseline when the sentence or phrase is completed.

In order to test the predictions of this model, Pallier et al. used a **constituency size paradigm**, creating contrasts between 12-word sequences which varied according to their phrasal structure; some sequences were 12-word right-branching sentences such that no phrase was completed until word 12, while other sequences were two 6-word sentences, so that activity should reset after word 6, three 4-word phrases, so that activity should reset after word 4, etc. The results of Pallier et al.'s study were in line with the predictions of their model: the fMRI response increased monotonically with connected phrase length in left inferior frontal, temporal, and parietal regions. In order to try to distinguish whether these effects reflected the representation of syntactic vs. interpretive relations, they also included a 'jabberwocky' manipulation in which content words were replaced by meaningless pseudowords (*I tosseive that you should begapt the tropufal of your tew viroate*). For jabberwocky materials, manipulating the length of connected structure modulated activity in the left inferior frontal cortex and the left posterior superior temporal sulcus only, leading Pallier et al. to hypothesize that these two regions maintain an ongoing representation of the syntactic relations between elements of an incomplete phrase. In later work they report electrocorticography (ECoG) data from patients in further support of this model (Nelson et al., 2017).

Although these results are intriguing and important, the Pallier et al. active maintenance model conflicts with the recent theories of memory discussed above in which even short-term memory representations are encoded passively without the need for ongoing neural activity. Given this ongoing debate, it is important to consider and rule out alternative explanations for the Pallier et al. results. In Matchin et al. (2017), we discuss several aspects of the design that could motivate alternative accounts of the data: Pallier et al. did not control lexical content across conditions, so that e.g. differences in the proportion of function words across conditions could have contributed to differences in the neural response, and because participants did not know whether an upcoming trial would be a full sentence or a list of phrases, some of the effects observed might have reflected disengagement or surprise upon the realization that a connected structure could not be formed from the input. We addressed these concerns in a partial replication of the constituency-size paradigm, contrasting scrambled lists, two-word phrases, and six-word sentences. Although we failed to find differences between the lists and the two-word phrases as the Pallier et al. model predicted, this could be because of the brief duration across which relations must be maintained in forming two-word phrases. However, we also point out that constituency-size effects could index the computation of syntactic predictions, rather than the representation of preceding syntactic relations.

A more straightforward means of testing the Pallier et al. model or others like it is through the use of temporally sensitive neurophysiological measures such as EEG and MEG. The Pallier et al. model makes very clear predictions about how the neural response should evolve over time: neural activity should systematically increase with each new morpheme across the duration of an incomplete phrase. However, the poor temporal resolution of fMRI measures means that this prediction could be only weakly evaluated in their original study. In a recent EEG study (Lau & Liao, 2018) we used a noun phrase coordination paradigm that allowed a tight lexical and semantic match between conditions with different constituent sizes. Here we compared trials in which coordinated noun phrases were presented (*new books and shiny floors*) with trials in which the same two noun phrases were presented as an unstructured list (*new books ### shiny floors*). Because in the coordination condition the constituent spans all five words of the trial, the Pallier et al. model predicts that in this condition should be observed increased neural activity across the second noun phrase associated with the maintenance of the structural relations from the first noun phrase. In two EEG experiments we saw exactly this pattern, with a sustained anterior negativity observed throughout the second noun phrase for the coordination condition relative to the list condition. However, we did not observe analogous effects for jabberwocky phrases (*blargal cloffs and slithy toves*), raising the possibility that these sustained differences reflect interpretive processes rather than the maintenance of structural relations. Fedorenko et al. (2016) report somewhat consistent data from intracranial ECoG recordings in patients. They compared the amplitude of high-gamma activity for participants reading sentences and word lists, and showed that across many electrode sites, high-gamma amplitude increased steadily across sentences but not word lists. They also show no corresponding increase for structured but meaningless jabberwocky sentences, leading them to argue that the response is a correlate of the representation of sentence-level meaning and not syntactic structure. Also potentially related are results of an MEG time-frequency analysis by Bastiaansen, Magyari, and Hagoort (2010), which showed an increase in amplitude in the beta frequency band (13—18 Hz) across the course of well-formed sentences, which was not observed for scrambled sentences, and which appeared to be truncated in stimuli that began as well-structured sentences prior to a word category violation.

Ding et al. (2016) implicitly appeal to a different mechanism for encoding prior structure. Their **constituent rate** paradigm involves repeating simple structures like sentences with an [[Adj N] [V N]] structure at a constant rate and evaluating the extent to which increases in MEG or EEG power are observed at the phrasal or sentence frequencies. As they do observe such increases, they gesture to the idea that oscillatory synchronization may support the encoding of structural relations. Although this model is not yet fully worked out (see Martin & Doumas, 2017 for a preliminary proposal), it is worth noting that proposals in which synchronous oscillations encode syntactic relations (e.g. that the syntactic relation between *red* and *boat* is encoded by synchronization of their firing) over the course of the sentence are thus implicitly assuming that this relational information needs to be maintained over time, vs. being rapidly transformed to a non-relational format. However, as discussed earlier, it is as yet unclear whether the time-frequency effects observed in the constituent rate paradigm are in fact due to differences in oscillatory properties across time, rather than reflecting event-related responses associated with the initial encoding of a relation (Zhou, Melloni, Poeppel, & Ding, 2016; Frank & Yang, 2018).

To sum up the results reviewed in this section, both Pallier et al. (2011) and Ding et al. (2016) suggest intriguing proposals about how structural relations are carried forward in memory through sustained neural activity, but initial evidence for these proposals from EEG, MEG, and fMRI are mixed. Even if these proposals turn out to be incorrect, however, a critical contribution of this work is to draw attention to questions about memory for structure that have largely been missing from the classic literature on syntactic working memory, that is: beyond what is minimally needed to confirm that a sentence is grammatically well-formed, what kind of structural relations need to be maintained in memory for interpretative and reanalysis purposes? More neurophysiological work is needed using constituent size and constituent rate paradigms of these types, looking for sustained differences in activity associated with structural properties of the prior input, rather than or in addition to the brief event-related manipulations that currently represent the vast majority of electrophysiological research on syntactic processing.

3. CONCLUSION

In this chapter, I have reviewed a body of past and present work aimed at discovering how the brain initially encodes structured relations in linguistic input and how the brain carries forward this relational information across the course of a sentence when needed. We can now see promising avenues for gaining traction on this problem, with the introduction of creative new paradigms for investigating neural responses to structural relations. I would argue that the most prominent obstacles for real progress are (a) the relatively small set of researchers currently working on this problem and (b) the insufficient synergy between the sub-disciplines of syntax, semantics, psycholinguistics, cognitive neuroscience, and computational linguistics that I believe is needed to devise good paradigms and draw appropriate conclusions. With respect to (a), we need to create a broader awareness that studying the neural bases of linguistic structure speaks to big-picture questions about neural computation, and can go far beyond studies that just record neural responses to ungrammatical sentences. With respect to (b), we need to create better interdisciplinary training approaches so that researchers in this area don't waste time reinventing the wheel or pursuing dead-end paths, but rather accelerate progress by building upon a solid foundation of existing well-established generalizations on linguistic relations and human sentence processing.

REFERENCES

Baron, S. G., Thompson-Schill, S. L., Weber, M., & Osherson, D. (2010). An early stage of conceptual combination: Superimposition of constituent concepts in left anterolateral temporal lobe. *Cognitive Neuroscience, 1*(1), 44–51.

Bastiaansen, M., Magyari, L., & Hagoort, P. (2010). Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *Journal of Cognitive Neuroscience, 22*(7), 1333–1347.

Bemis, D. K., & Pylkkänen, L. (2011). Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *Journal of Neuroscience, 31*(8), 2801–2814.

Bemis, D. K., & Pylkkänen, L. (2012). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cerebral Cortex, 23*(8), 1859–1873.

Bemis, D. K., & Pylkkänen, L. (2013a). Flexible composition: MEG evidence for the deployment of basic combinatorial linguistic mechanisms in response to task demands. *PLoS One, 8*(9), e73949.

Bemis, D. K., & Pylkkanen, L. (2013b). Combination across domains: An MEG investigation into the relationship between mathematical, pictorial, and linguistic processing. *Frontiers in Psychology, 3*, 583.

Boylan, C., Trueswell, J. C., & Thompson-Schill, S. L. (2017). Relational vs. attributive interpretation of nominal compounds differentially engages angular gyrus and anterior temporal lobe. *Brain and Language, 169*, 8–21.

Brennan, J. (2016). Naturalistic sentence comprehension in the brain. *Language and Linguistics Compass, 10*(7), 299–313.

Brennan, J. R., & Pylkkänen, L. (2017). MEG evidence for incremental sentence composition in the anterior temporal lobe. *Cognitive Science, 41*, 1515–1531.

Brennan, J. R., Stabler, E. P., Van Wagenen, S. E., Luh, W. M., & Hale, J. T. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language, 157*, 81–94.

Brodbeck, C., Presacco, A., & Simon, J. Z. (2018). Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension. *NeuroImage, 172*, 162—174.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience, 19*(1), 158.

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., & Poeppel, D. (2017). Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Frontiers in human neuroscience, 11*, 481.

Federmeier, K. D., & Laszlo, S. (2009). Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. *Psychology of Learning and Motivation, 51*, 1—44.

Fedorenko, E., Scott, T. L., Brunner, P., Coon, W. G., Pritchett, B., Schalk, G., & Kanwisher, N. (2016). Neural correlate of the construction of sentence meaning. *Proceedings of the National Academy of Sciences of the United States of America, 113*(41), E6256—E6262.

Fiebach, C., Schlesewsky, M., & Friederici, A. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language, 47*(2), 250—272.

Fiebach, C. J., Schlesewsky, M., Lohmann, G., Von Cramon, D. Y., & Friederici, A. D. (2005). Revisiting the role of Broca's area in sentence processing: Syntactic integration versus syntactic working memory. *Human Brain Mapping, 24*(2), 79—91.

Fitch, W. T. (2014). Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews, 11*(3), 329—364.

Frank, S. L., & Yang, J. (2018). Lexical representation explains cortical entrainment during speech comprehension. *PLoS One, 13*(5), e0197304.

Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences, 16*(5), 262—268.

Friederici, A. D., Meyer, M., & von Cramon, D. Y. (2000). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language, 74*(2), 289—300.

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology, 61*(2), 331—349.

Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science, 173*(3997), 652—654.

Gallistel, C. R., & King, A. P. (2011). *Memory and the computational brain: Why cognitive science will transform neuroscience*. John Wiley & Sons.

Graves, W. W., Binder, J. R., Desai, R. H., Conant, L. L., & Seidenberg, M. S. (2010). Neural correlates of implicit and explicit combinatorial semantic processing. *NeuroImage, 53*(2), 638—646.

Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Memory and Language, 10*(4), 409.

King, J. W., & Kutas, M. (1995). Who did what and when? Using word-and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7*(3), 376—395.

Kluender, R., & Kutas, M. (1993). Subjacency as a processing phenomenon. *Language & Cognitive Processes, 8*(4), 573—633.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463—470.

LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2013). Decoding attended information in short-term memory: An EEG study. *Journal of Cognitive Neuroscience, 25*(1), 127—142.

Lau, E., & Liao, C. H. (2018). Linguistic structure across time: ERP responses to coordinated and uncoordinated noun phrases. *Language, Cognition and Neuroscience, 33*(5), 633–647.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience, 9*(12), 920.

Lau, E. F., Weber, K., Gramfort, A., Hämäläinen, M. S., & Kuperberg, G. R. (2014). Spatiotemporal signatures of lexical–semantic prediction. *Cerebral Cortex, 26*(4), 1377–1387.

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience, 24*(1), 61–79.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*(3), 375–419.

Makov, S., Sharon, O., Ding, N., Ben-Shachar, M., Nir, Y., & Golumbic, E. Z. (2017). Sleep disrupts high-level speech parsing despite significant basic auditory processing. *Journal of Neuroscience*, 0168-17.

Martin, A. E., & Doumas, L. A. (2017). A mechanism for the cortical computation of hierarchical linguistic structure. *Plos Biology, 15*(3), e2000663.

Matchin, W., Hammerly, C., & Lau, E. (2017). The role of the IFG and pSTS in syntactic prediction: Evidence from a parametric study of hierarchical structure in fMRI. *Cortex, 88*, 106–123.

Matchin, W., Sprouse, J., & Hickok, G. (2014). A structural distance effect for backward anaphora in Broca's area: An fMRI study. *Brain and Language, 138*, 1–11.

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., … Mehler, J. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience, 5*(4), 467–479.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 48*(1), 67–91.

Murphy, G. L., & Shapiro, A. M. (1994). Forgetting of verbatim information in discourse. *Memory & Cognition, 22*(1), 85–94.

Nelson, M. J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., … Dehaene, S. (2017). Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences, 114*(18), E3669–E3678.

Neufeld, C., Kramer, S. E., Lapinskaya, N., Heffner, C. C., Malko, A., & Lau, E. F. (2016). The electrophysiology of basic phrase building. *PLoS One, 11*(10), e0158446.

Pallier, C., Devauchelle, A.-D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences, 108*(6), 2522–2527.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8*(12), 976.

Phillips, C., Kazanina, N., & Abada, S. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research, 22*(3), 407–428.

Phillips, C., & Wagers, M. (2007). Relating structure and time in linguistics and psycholinguistics. In *Oxford handbook of psycholinguistics* (pp. 739–756).

Rogalsky, C., & Hickok, G. (2008). Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cerebral Cortex, 19*(4), 786–796.

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science, 354*(6316), 1136–1139.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Attention, Perception, & Psychophysics, 2*(9), 437–442.

Santi, A., & Grodzinsky, Y. (2007). Working memory and syntax interact in Broca's area. *NeuroImage, 37*(1), 8–17.

Schell, M., Zaccarella, E., & Friederici, A. D. (2017). Differential cortical contribution of syntax and semantics: An fMRI study on two-word phrasal processing. *Cortex, 96*, 105–120.

Smolensky, P., & Legendre, G. (2006). *The harmonic mind: From neural computation to optimality-theoretic grammar*. MIT Press.

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language, 86*(1), 142–164.

Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research, 30*(3), 267–295.

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language & Cognitive Processes, 1*(3), 227–245.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language, 35*(3), 454–475.

Ueno, M., & Kluender, R. (2003). Event-related brain indices of Japanese scrambling. *Brain and Language, 86*(2), 243–271.

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language, 97*(3), 279–293.

Vandenberghe, R., Nobre, A. C., & Price, C. J. (2002). The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience, 14*(4), 550–560.

Westerlund, M., & Pylkkänen, L. (2014). The role of the left anterior temporal lobe in semantic composition vs. semantic memory. *Neuropsychologia, 57*, 59–70.

Wilson, S. M., DeMarco, A. T., Henry, M. L., Gesierich, B., Babiak, M., Mandelli, M. L., Miller, B. L., & Gorno-Tempini, M. L. (2014). What role does the anterior temporal lobe play in sentence-level processing? Neural correlates of syntactic processing in semantic variant primary progressive aphasia. *Journal of Cognitive Neuroscience, 26*(5), 970–985.

Yano, M., & Koizumi, M. (2018). Processing of non-canonical word orders in (in) felicitous contexts: Evidence from event-related brain potentials. *Language, Cognition and Neuroscience*, 1–15.

Zaccarella, E., & Friederici, A. D. (2015a). Reflections of word processing in the insular cortex: a sub-regional parcellation based functional assessment. *Brain and language, 142*, 1–7.

Zaccarella, E., & Friederici, A. D. (2015b). Merge in the human brain: A sub-region based functional investigation in the left pars opercularis. *Frontiers in Psychology, 6*, 1818.

Zhang, L., & Pylkkänen, L. (2015). The interplay of composition and concept specificity in the left anterior temporal lobe: An MEG study. *NeuroImage, 111*, 228–240.

Zhou, H., Melloni, L., Poeppel, D., & Ding, N. (2016). Interpretations of frequency domain analyses of neural entrainment: Periodicity, fundamental frequency, and harmonics. *Frontiers in Human Neuroscience, 10*, 274.