Electrophysiological Methods

Jon Sprouse
Diogo Almeida

1. Introduction

The potential value of electrophysiological measures like electroencephalography (EEG) and magnetoencephalography (MEG) for experimental syntax is easy to see: if one believes that cognition is mediated by electrical activity in the cortex, and if one believes that syntactic theories are ultimately theories of cognition, then any method that yields information about electrical activity in the cortex potentially provides information about syntactic theories. That said, we believe it is fair to say that, to date, EEG and MEG have played much larger roles in the construction and evaluation of theories of language processing (including sentence processing) than they have in the construction and evaluation of theories of grammar. One reason for this is that it is relatively difficult to construct a linking theory between syntactic theories and sentence processing theories that yields predictions that are dispositive of syntactic theories. This is because the relationship between syntactic theories and sentence processing theories is many-to-many. A second reason for this is specific to extracranial electrophysiological measures like scalp EEG and MEG: these methods only detect a subset of cortical activity. It is eminently possible that the subset of cortical activity that these methods can detect is not the subset of cortical activity that is relevant to syntactic theories. Despite these challenges, we believe that experimental syntacticians (who are so interested) should nonetheless consider exploring electrophysiology as potential data sources for syntactic theories. To be clear, when it comes to syntax, electrophysiology is a high-risk/high-reward method. But as long as it is undertaken after careful consideration, the risk may be worth it. To that end, we would like to use this chapter to provide a foundation for thinking about ways to incorporate EEG and MEG into the experimental syntax toolkit.

In this chapter we will focus (nearly) exclusively on extracranial electrophysiological techniques, namely scalp EEG, which measures electrical potentials on the scalp that are generated in the cortex, and MEG, which measures magnetic fields that are generated by electrical activity in the cortex. This choice is purely pragmatic – extracranial methods are far more likely to be available to experimental syntacticians than intracranial methods, which can only be used as part of a medical procedure performed by a licensed neurosurgeon (though we do discuss one intracranial study of syntax in section 6). In section 2, we begin by introducing the three content areas necessary to understand both the potential of EEG as a method in cognitive science and its specific challenges: the basics of electricity, the neurobiology underlying scalp potentials, and the math necessary to extract useful information from the EEG signal (with citations here and in the annotated bibliography for more information). In sections 3 and 4, we briefly review some of the ERP and time-frequency results (respectively) in the sentence processing literature, in order to provide a foundation for thinking about ways to leverage EEG results in service of syntactic theories. In section 5, we briefly discuss several recent studies that have attempted to directly link syntactic theories with sentence processing and electrophysiological measures. Our hope is that these brief discussions will provide a starting point for experimental syntacticians who are interested in pursuing their own studies. Section 6 concludes.

2. The three core content areas for understanding EEG

There are three core content areas that are necessary for understanding the potential of EEG and MEG as tools for experimental syntax: electricity, neurobiology, and wave mathematics. In this section we will provide a brief introduction to the core concepts of each, with pointers to more information for interested readers (see also the annotated bibliography, and of course, the two most prominent textbooks for EEG in cognitive science: Luck 2014 and Cohen 2014). We have three goals for this section: (i) to introduce readers to the concepts necessary to begin to work with EEG and MEG, (ii) to illustrate the promise that EEG and MEG hold for studying syntax as a component of cognition, and (iii) to explain the basis of the challenges that are unique to EEG and MEG in the study of syntax.

2.1 Electricity

EEG measures changes over time in electrical potential on the scalp. A typical EEG system consists of a set of electrodes embedded in a nylon cap or net (typically a power of two: 16, 32, 64, 128, or 256), and a specialized amplifier for recording the very small electrical potentials that the electrodes detect on the scalp. Electrical potential is the potential for electrical current (electrons) to flow between two locations. It is thus important to remember that the value that is recorded by the system at any given scalp location is ultimately the potential for electrons to flow between that location and a reference location. In an electrical circuit that reference location is the ground, but EEG amplifiers allow researchers to specify any reference location that they like by calculating the difference of the potential between one electrode and the ground and the potential between the chosen reference electrode and the ground. If the symbol A is the target electrode, G is ground, and R is the chosen reference electrode, then the potential recorded at A is given by this equation (A-G) - (R-G), which reduces to (A-R), thus underscoring the point that the potential recorded at electrode A is actually the potential for current to flow between A and R (and in the process this equation also removes electrical noise that may have been in the ground circuit). The unit of measure for electrical potential is the volt, with scalp EEG potentials typically in the microvolt range (i.e., one millionth of a volt, or 10^{-6}). In order to measure changes over time, EEG amplifiers must take discrete measurements, called samples, many times per second. Modern EEG amplifiers are able to take anywhere from 250 to 100,000 measurements per second (for a sampling rate of 250Hz to 100,000Hz), though the typical sampling rate for language experiments is 250Hz to 1000Hz. In human-made electrical systems, there are two types of electrical current: direct current (DC), wherein the electrons flow in one direction at all times, and alternating current (AC), wherein the electron flow alternates between two directions periodically. The biological electrical current of the brain cannot be as easily categorized as human-made systems; however, because cortical activity is oscillatory, it is often easiest to think of EEG as an AC signal. The fact that EEG is ultimately the measurement of an (AC-like) electrical signal has a number of practical implications for designing EEG experiments, such as the choice of reference location, the choice of sampling rate, and the effect of resistance/impedance on the measurements. These practical considerations are far beyond the scope of this chapter (see Luck 2014 for discussion and advice). The upshot is that it is critical for EEG researchers to invest some time in understanding the fundamentals of electrical signals and electrical circuits as these fundamentals do have consequences for EEG experimental design and analysis (see the annotated bibliography for resources).

2.2 Biology

The neurobiological source of the electrical activity that (scalp) EEG measures is relatively straightforward to state: it is the summed activity of the postsynaptic potentials of large populations of spatially-aligned pyramidal cells in the cerebral cortex. In this paragraph we will attempt to unpack this statement, albeit to a very cursory level of description. First, the cells that are generating the activity detected by scalp EEG are exclusively within the cerebral cortex. The cerebral cortex is the outer layer of the cerebrum, covering the gyri and sulci, typically 2mm to 3mm thick, and often called "gray matter" because of its coloring. The cortex is generally considered to be a critical component of cognition, but crucially, it is not the only component. Second, the cells that are generating this activity are pyramidal cells. These are a relatively common type of neuron in the cortex, but crucially, they are not the only type of neuron in the cortex. Third, (scalp) EEG can only detect electrical activity generated by large populations of pyramidal cells. This is because the relatively small potentials generated by these neurons dissipate quickly as they travel through the biological matter that intervenes between the cortex and the EEG electrode on the scalp. Fourth, these large populations of pyramidal cells must be spatially aligned. This is because each pyramidal cell is an electrical dipole: one end of the cell is positively charged and the other is negatively charged. If the cells are aligned, the electrical charges will sum, creating a larger electrical current that can reach the electrode on the scalp. If they are not aligned, the charges will not sum (as diametrically opposed dipoles will cancel each other), and the current will not be large enough to reach the electrode on the scalp. Fifth, the potentials that are measured by (scalp) EEG are post-synaptic potentials. Synapses are the junctions between two (or more) neurons, such that there is a presynaptic neuron and a postsynaptic neuron. Crucially, we can distinguish between the potential that is generated within the presynaptic neuron, called an action potential, and the potential that is (chemically) transmitted to the postsynaptic neuron, called the postsynaptic potential. Postsynaptic potentials can either be excitatory, which brings the postsynaptic neuron closer to generating an action potential, or inhibitory, which brings the postsynaptic neuron farther from generating an action potential. The structure and function of neurons and synapses is far beyond the scope of this chapter (see the annotated bibliography). The critical point here is that (scalp) EEG measures postsynaptic potentials, not action potentials. Finally, the pattern of activity that (scalp) EEG measures is the summation of all of the detectable activity emanating from the cortex. This is because the biological matter of the brain, skull, and scalp is conductive. Any given generator of EEG activity inside the cortex can potentially send some amount of signal to all of the electrodes on the scalp. If there are multiple such generators, which is likely given the complex nature of cognition, these signals will sum.

Though the preceding paragraph is only a relatively shallow review of the biology underlying EEG, even at this level of detail, it is a bit easier to see the source of both the potential benefits and challenges for the use of EEG in experimental syntax. The benefit of EEG is that it is a relatively direct measure of a signal that we think is relevant for cognition: postsynaptic potentials are the signals the messages that neurons send to each other. The challenges facing EEG come from the fact that those potentials are being recorded from the scalp. The summation of electrical signals through biological matter means that EEG is not an ideal tool for localizing the source of the EEG activity. The fact that only large populations of spatially aligned pyramidal cells can create the kind of activity that is detectable at the scalp means that EEG can only detect a small subset of cortical activity. This makes EEG a potentially risky method for cognitive theories, as it is possible that the activity that is most relevant will not be detected. It also means that scalp EEG is likely a poor tool for low-level neuroscience (a point made forcefully in Luck 2014), as it simply cannot detect any of the activities of neurons other than postsynaptic potentials (action potentials, ion gates, etc). Despite these challenges, there have been quite a number of results obtained using scalp EEG that appear relevant for cognitive theories of language processing (see section 3 below), therefore we believe it is valuable to at least try to leverage EEG in the domain of experimental syntax. That said, we do recommend that potential EEG researchers invest some time in understanding the neurobiology underlying EEG recordings in order to better appreciate the relationship of the signal to theories of the neurobiology of cognition.

2.3 Wave mathematics

There are two reasons that wave mathematics is fundamental to EEG analysis. The first is purely mathematical: EEG is a time-varying signal. As Fourier first demonstrated, any time-varying signal can be represented as the sum of some number of constituent sine waves of different frequencies. In other words, there are two equivalent representations for EEG signals: one in the time-amplitude domain, expressing the change in electrical potential over time, and one in the frequency domain, expressing the frequency and amplitude of the sine waves required to compose that signal. This equivalence makes available a number of advanced signal processing techniques from wave mathematics (the Fourier transform, the convolution theorem, etc). The second reason that wave mathematics is fundamental to EEG analysis is biological: the electrical activity generated by the cortex appears to be oscillatory in nature. This suggests that EEG signals may in fact be fundamentally composed of the sum of some number of neuronal oscillations, which in turn can be mathematically represented as sine waves, and analyzed using the tools of wave mathematics. Wave mathematics is a broad field, spanning trigonometry, calculus, and linear algebra. We provide some references to begin learning the most relevant concepts for EEG analysis in the annotated bibliography.

There are two fundamental analysis techniques in common use in the EEG literature: the event-related potential technique (ERP), and the time-frequency decomposition technique (TF). Here we will outline each technique, and then provide a brief discussion of their similarities and differences. We begin with the ERP technique, as it is by far the most common analysis technique for EEG experiments in the sentence processing literature. The first step of the ERP technique is to define each trial in the experiment as a time window of EEG activity around a critical stimulus (such as a word). These time windows are called epochs. The critical stimulus is designated time point 0 for convenience, with epochs typically ranging from 100ms or 200ms before the stimulus (-100ms or -200ms) to 1000ms or more after the stimulus. The second step of the ERP technique is to cut out these epochs from the continuous EEG recording, and then organize them according to experimental condition. The third step is to align all of the epochs within each condition by time point 0. The final step is to average (using the arithmetic mean) across all of the aligned epochs within each condition. The resulting averaged wave is called an ERP. The math underlying the ERP technique is simple (time-aligning and averaging), but powerful. The averaging procedure ensures that only features of the EEG signal that are time-locked to the stimulus (arising at the same time in each epoch) and phase-locked to the stimulus (peaks align with peaks, troughs with troughs, at each frequency) will survive. Any activity that is either not-time-locked or not-phase-locked or both will be diminished in the averaging. If this activity is randomly distributed in time and phase-locking it will approach zero as the number of epochs increases. This leads to the fundamental idea of the ERP technique: the signal that the technique returns is the time-locked and phase-locked activity; the noise that the technique discards is the not-time-locked and not-phase-locked activity. Time-locked and phase-locked activity is sometimes called *evoked* activity. The ERP technique will be appropriate for any theories that make predictions about evoked activity. (We have purposefully left out the processing steps that are necessary to eliminate other sources of noise from the EEG data, such as filtering, artifact detection, and baseline correction, so that we can focus on the underlying logic. See Luck 2014 for a comprehensive introduction to these steps.)

The TF technique is less common in the sentence processing literature than the ERP technique, but has been growing in popularity over the past 20 years or so. The first step of the TF technique is also to define epochs around a critical stimulus. The only difference is that the epochs in the TF technique often extend further back in time from the critical stimulus (typically -400 or -500ms) and sometimes extend further forward in time as well. This is because there is a direct relationship between the size of the epoch and the frequencies that can be reliably detected in the epoch (with lower frequencies requiring longer epochs to be reliably detected). The second step is also identical to the ERP technique: cut these epochs out of the continuous EEG recording. The third step is where the two methods diverge: perform time-frequency decomposition on each epoch independently. There are a number of techniques for time-frequency decomposition, such as Morlet wavelets, multitapers, and the short-time fast Fourier transform. All of these are beyond the scope of this chapter (but see Cohen 2014 for a comprehensive introduction). The critical idea is that each of these methods attempts to decompose the EEG in the epoch into a combination of sine waves of different frequencies, each with an amplitude (how much of the sine wave is present, sometimes reported as power, which is amplitude squared) and phase (the location in the cycle, reported as radians, as in angles in the unit circle) that varies over time. The fourth step is to align all of the epochs within each condition by time point 0. The final step is to average (using the arithmetic mean) across all of the aligned epochs, respecting the distinction between frequencies. This averaging procedure means that time-locked features of the epochs will be maintained, and not-time-locked features will be diminished. However, there is no phase-locking effect in this averaging, because power and phase have been separated into distinct quantities by the TF decomposition technique. Furthermore, because both amplitude and phase measures are always greater than or equal to 0 there is no way for these measures to cancel themselves out in an averaging procedure. This leads to the fundamental idea of the TF technique: the signal that the technique returns is time-locked activity at each frequency (with no commitment to phase-locking); the noise that the technique discards is the not-time-locked activity. Time-locked but not phase-locked activity is sometimes called *induced* activity. The TF technique by default returns the sum of induced and evoked activity, but can be modified in various ways to subtract out the evoked activity, leaving only the induced activity behind.

There are number of ways of thinking about the relationship between the ERP and TF techniques. One important similarity is that they both leverage time-locking to distinguish cognitive processes that are likely related to our experiment from all of the other processes that the brain might be deploying at any given moment. The primary difference between the two techniques centers on the role of phase-locking. This difference is not simply mathematical. The physiological events that give rise to phase-locked (evoked) activity and the physiological events that give rise to not-phase-locked (induced) activity are likely distinct. For example, one possible source of evoked activity is a phase-reset in the firing of a population of neurons; and one possible source of induced activity is a sustained oscillation of population of neurons. We say "one possible source" because the physiological source(s) of evoked activity is an open area of research (see Mazaheri and Jensen 2010 for a discussion of competing theories). The physiological source of results in the TF technique can never be stated with certainty from the EEG signal alone, as the mathematical decomposition methods will always return a representation that is composed of a series of sine waves for any time-varying signal, regardless of the source of the time-varying signal. For these reasons, the two techniques are complementary, and both should probably be in the EEG researcher's toolkit.

3. A brief review of common ERP effects during sentence processing

In this section we will review some of the well-established ERP effects in the linguistics and psycholinguistics literature. Our goal is twofold. First, any experimental syntactician interested in EEG must first become familiar with the work that has come before, in order to build on, and ultimately extend, that work. Therefore we wish to provide a starting point for building that knowledge. Second, one way to leverage EEG in service of syntactic theories is to use existing EEG effects, either ERP (this section) or TF (next section) to draw inferences about syntactic theories. This is not a simple task, as it requires first linking the EEG effects to underlying cognitive operations, and then linking those cognitive operation, via a sentence processing theory, to syntactic theories. We will discuss this challenge in more detail in section 6. In this section and the next, we wish to provide a foundation for the first step of this process, linking EEG effects to cognitive operations, by reviewing the literature on the functional interpretation of existing EEG effects. In this section we briefly review five ERPs that experimental syntacticians are likely to encounter in the sentence processing literature (for a broader review of ERP components, see Kappenman and Luck 2011: the Oxford Handbook of Event-Related Potential Components). For each we provide a brief review of the eliciting conditions of the ERP and the functional interpretation of the ERP.

3.1 The Early Left Anterior Negativity (ELAN)

The ELAN is a negative-going deflection that peaks in a relatively early processing window (between 100ms and 250ms post-stimulus onset), and is maximal over left anterior electrodes. The ELAN was first reported by Neville et al. 1991 to the transposition of a noun and a preposition in sentences like those in (1). Here and throughout, the critical words for the analysis will be in bold.

(1) a. grammatical control: The boys heard Joe's stories **about** Africa.
b. transposition: *The boys heard Joe's **about** stories Africa.

A similar effect was reported by Friederici et al. 1993 to German sentences like the one in (2):

(2) *Das Baby wurde im **gefürttert**
The baby was in-the fed.

The ELAN appears to be elicited by phrase structure violations, as in both of these cases, the critical word (in italics) cannot appear in that position. The ELAN has been elicited in a number of languages beyond English and German, including Mandarin Chinese (e.g., Ye et al. 2006), Dutch (e.g., Hagoort et al. 2003), French (e.g., Isel et al. 2007), Japanese (e.g., Mueller et al. 2005), and Spanish (e.g., Hinojosa et al. 2003). The ELAN is not affected by task (Hahne and Friederici 2002), by the probability of the violation in the experiment (Hahne and Friederici 1999), or by the frequency of a disambiguated structure (Ainsworth-Darnell, Shulman, and Boland 1998, Friederici et al. 1996). Taken as a whole, these results suggest that the ELAN is a very specific response to phrase structure violations, and not simply a response to difficult or unlikely structures.

Recent research on the ELAN has focused on the extremely early latency of the response. The 100-250ms post-stimulus window is remarkably early for syntactic analysis (and error diagnosis) given that estimates of lexical access often center around 200ms post-stimulus (Allopenna, Magnuson, and Tanenhaus 1998, van Petten, Coulson, Rubin, Plante, and Parks 1999). Four approaches have been offered to explain the early latency of the ELAN. Friederici (1995) adopts a parsing model in which the earliest stage considers only word category information (e.g., Frazier 1978), thus limiting the number of processes that need to be performed in the earliest time window. Lau et al. (2006) suggest that the early latency can be explained if the parser has predicted the properties of the critical word prior to encountering it, such that many of the syntactic features are in some sense "pre-parsed". Dikker et al. (2009) propose the "sensory ELAN hypothesis", in which the ELAN indexes a processing stage prior to lexical access that occurs in the sensory cortices (visual or auditory cortex). This pre-lexical processing is based purely on the form typicality of the words – i.e., the sensory cortices use the probability of certain phonetic forms to determine if the incoming string is most likely a noun, verb, etc. Finally, Steinhauer and Drury (2012) argue that at least some of the ELAN effects reported in the literature may be artifacts that arise when comparing two conditions that do not match in the word preceding the critical word. Though there is no consensus on the source of the ELAN, what is clear from this debate is that any adequate functional interpretation must take (i) the earliness of the response and (ii) the specificity of the response into consideration.

3.2 The Left Anterior Negativity (LAN)

While the LAN and the ELAN share many properties (i.e., they are both negative-going deflections that occur primarily over left anterior electrode sites), they differ along two critical dimensions. First, the LAN occurs in a slightly later time window, usually 300-500ms post-stimulus onset, which eliminates many of the complex timing questions associated with the ELAN. Second, the LAN has been elicited by a broad array of (morpho-)syntactic violations, such as agreement violations (Coulson et al. 1998, Gunter et al. 1997, Münte et al. 1997, Kaan 2002, Osterhout and Mobley 1995), case violations (Münte and Heinze 1994), phrase structure violations, (Friederici, Hahne, and Mecklinger 1996, Hagoort, Wassenaar, and Brown 2003) island constraint violations (Kluender and Kutas 1993b), and even garden-path sentences (Kaan and Swab 2003). The LAN has also been elicited during the processing of long-distance dependencies such as wh-movement, at both the displaced wh-word and the unambiguous cue for the gap location (Kluender and Kutas 1993a, Phillips, Kazanina, and Abada 2005), and during a memory period after processing grammatical semantically reversible sentences (Meltzer and Braun 2013).

One common functional interpretation of the LAN is as an index of morpho-syntactic agreement violations (e.g., Molinaro et al. 2011). However, there are two empirical concerns about the LAN that have led to (at least partially) competing interpretations. The first concern is that the LAN shows quite a bit of variability across experiments, in some cases not appearing at all for violations that are unambiguously morphosyntactic in nature. Tanner and van Hell (2014) argue that it is possible that the LAN is an illusion that could arise if the participants in a sample are really from two distinct populations: one that shows an N400 to violation and one that shows a P600 to the violation. As we will see when we review the N400 and P600 below, the timing and scalp distributions of N400s and P600s could potentially give rise to an illusory response with the timing and scalp distribution of a LAN if averaged together. The second concern is that the LAN also arises for conditions that do not obviously involve increased morpho-syntactic processing, but rather likely involve increasing working memory processing, like garden-path sentences, grammatical wh-dependencies, and semantically reversible sentences. This suggests that the morpho-syntactic processing interpretation of the LAN cannot be the whole story (see also Martín-Loeches et al. 2005 for some evidence that morpho-syntactic LANs and working memory LANs may have different scalp topographies).

3.3 The N400

The N400 is a negative-going deflection that is generally largest over centro-parietal electrode sites, and tends to occur 300-500ms post-stimulus onset (with a peak amplitude occurring at 400ms). The N400 was first found by Kutas and Hillyard (1980) when they presented participants with sentences that ended with unexpected words. They compared a baseline sentence with semantically congruent endings (3a) to sentences with semantically incongruent endings (3b) and sentences with endings that were incongruent due to the physical properties of the stimulus such as words written in all capital letters (3c):

(3)	a.	semantically congruent:	I spread the warm bread with **butter**.
	b.	semantically incongruent:	I spread the warm bread with **socks**.
	c.	physically incongruent:	I spread the warm bread with **BUTTER**.

Kutas and Hillyard (1980) observed a larger N400 for (3b) compared to (3a), and a larger P300 (also known as a P3b) to (3c) compared to (3a). This qualitative difference in the responses to (3b) versus (3a) suggests that the N400 is specifically related to semantic processes rather than general error detection. In the decades since its discovery, the N400 has been elicited by a broad array of linguistic and non-linguistic stimuli, with the common pattern being that they are all meaningful in some way: spoken words, written, words, signed words, pseudowords, acronyms, environmental sounds, faces, and gestures (see Kutas and Federmeier 2011 for a comprehensive review, and Lau 2008 for a review of the brain networks underlying the N400).

There are (at least) two dominant functional interpretations of the N400, though none appear to capture all of the N400 results in the literature. The first is that the N400 indexes the difficulty of semantic integration (Hagoort 2008, Osterhout and Holcomb 1992, Brown and Hagoort 1993). Under this view, increases in N400 amplitude reflect the increased difficulty of integrating incongruent, unexpected, or semantically unrelated words into the preceding context. The second view is that the N400 indexes processes related to the activation of semantic features in the mental lexicon. Under this view, decreases in N400 amplitude reflect the ease of activation (or pre-activation) for congruent, predicted, and semantically related words (Federmeier and Kutas 1999, Kutas and Federmeier 2000, Lau et al. 2009). The N400 is by far the most studied ERP effect in the language processing literature, and the pattern of results is both subtle and complex. Though it is tempting to set the N400 aside as a "semantic" effect, and therefore irrelevant to theories of syntax, the fact that the N400 touches upon issues like memory, predictability, and the mental lexicon means that it is a potentially valuable tool for probing theories of sentence processing.

3.4 The P600

The P600 (alternatively the "syntactic positive shift") is a positive-going deflection that is generally largest over centro-parietal electrode sites and tends to occur 500-800ms post-stimulus onset (although there is a good deal of variability in the latency in the ERP literature). Like the LAN, the P600 has been reported for a broad array of syntactic violations, in many cases co-occurring with a preceding LAN. P600's have been elicited to phrase structure violations (Hagoort, Brown, and Groothusen 1993, Friederici et al. 1993, Hahne and Friederici 1999, Friederici and Frisch 2000, Osterhout and Holcomb 1992), agreement violations (Hagoort, Brown, and Groothusen 1993, Kaan 2002), syntactic garden-paths (Friederici et al. 1996, Kaan and Swaab 2003, Osterhout, Holcomb, and Swinney 1994), and island violations (McKinnon 1996). P600's have also been elicited by the processing of grammatical sentences with particularly complex syntactic properties, such as ambiguous structures (Frisch, Schlesewsky, Saddy, and Alpermann 2002), wh-movement (Fiebach, Schlesewsky, and Friederici 2002, Kaan, Harris, Gibson, and Holcomb 2000, Phillips, Kazanina, and Abada 2005), and unexpected theta-role assignments (Kim and Osterhout 2005, Kuperberg, Sitnikova, Caplan, and Holcomb 2003, van Herten, Kolk, and Chwilla 2005, Kuperberg 2007, Bornkessel-Schlesewsky and Schlesewsky 2008, Stroud and Phillips 2012).

There are two central questions about the functional interpretation of the P600 in the literature. The first is whether there is a single functional interpretation that can cover the full range of P600 effects. Syntactic violations and ambiguous grammatical sentences could potentially be unified under an interpretation of the P600 as syntactic reanalysis (though questions remain as to how many distinct reanalysis operations there are). However, the fact that P600s arise in wh-dependencies (at the verb or preposition that selects the filler) is hard to capture under syntactic reanalysis, suggesting that perhaps the P600 is a family of responses with potentially distinct functional interpretations (see Gouvea et al. 2010 for a comparison of several types of P600s in a single experiment). The second question is whether the P600s that arise to ungrammatical sentences are specific to language or are a domain-general response to unexpected stimuli. One possibility is that these P600s are a temporally delayed version of the P300 (or P3b), which is a well-known domain-general response to unexpected stimuli (Coulson et al. 1998, Osterhout and Hagoort 1999, Sassenhagen et al. 2014).

3.5 Sustained Anterior Negativity (SAN)

Sustained anterior negativities are negative-going deflections that tend to appear over anterior electrode sites (though not exclusively), and tend to last for several words during the processing of wh-dependencies and relative-clause dependencies (King and Kutas 1995, Fiebach et al. 2002, Phillips et al. 2005). SANs have typically been interpreted as an index of working memory usage because (i) they appear during dependency processing, which almost certainly involves working memory, and (ii) similar anterior negativities have been reported for working memory tasks outside of sentence processing (e.g, Ruchkin et al. 1990). SANs have been less studied relative to some of the other ERPs that arise during sentence processing, partly because they appear to be related to a system outside of the grammar (working memory), and partly because the relationship between SANs and working memory theories is currently unclear. The original functional interpretation of SANs was as an index of working memory load due to maintenaing the filler in working memory (King and Kutas 1995), but more recent models of working memory in sentence processing have eliminated maintenance costs from the theory in favor of retrieval and interference costs (e.g., McElree 2003?, Lewis and Vasishth 2005). Nonetheless, SANs potentially provide an index for working memory effects (of some sort) during sentence processing, and therefore may be a useful tool for experimental syntacticians interested in dependencies.

4. A brief review of time-frequency effects during sentence processing

Similar to the previous section, our goal in this section is to provide a brief review of the time-frequency literature that experimental syntacticians interested in EEG can use as a starting point for new research. Time-frequency results are typically analyzed in frequency bands. These bands group together frequencies that tend to covary in various domains of cognition. The bands are named after Greek letters. The precise boundaries of the bands can vary by one or two Hertz from study to study, so here we simply provide an example of range boundaries, rather than a hard a fast definition: delta (1Hz-3Hz), theta (4Hz-7Hz), alpha (8Hz-12Hz), lower beta (13Hz-20Hz), upper beta (21Hz-30Hz), and gamma (>30Hz). Because the TF technique yields both power and phase information at each frequency or frequency band, at each electrode site, through time, there are a number of measures that can be derived, such as local changes in power (at one or more electrode sites), correlated fluctuations in power or phase in one frequency band across spatially distinct electrode sites (called *coherence*), and correlated fluctuations in power or phase across distinct frequency bands (called *cross-frequency coupling*). Compared to the ERP technique, time-frequency decomposition is a relatively new, and rapidly growing, segment of the sentence processing literature. There are far fewer established debates about the functional interpretation of TF results than there are about the functional interpretation of ERP results. Therefore in this section we review a targeted selection of results. This is not intended as a comprehensive review (for a more comprehensive review, see Bastiaansen et al. 2013). To that end, we have chosen to organize the results according to the types of linguistic manipulations in these studies, subdivided by the types of ERPs that they typically elicit: syntactic violations that lead to P600s, syntactic violations that lead to ELANs, semantic violations that lead to N400s, and dependencies that lead to SANs. Our hope is that this will allow readers to explore their own hypotheses about the functional interpretation of the various TF results (and maybe spur ideas for future research).

4.1 Syntactic violations that lead to P600s: increase in power in theta, decrease in alpha and beta

Bastiaansen, van Berkum, and Hagoort 2002 performed time-frequency decomposition on the EEG response to two types of syntactic violations in Dutch (relative to a grammatical control sentence): a gender agreement violation between an adjective and noun, and a number agreement violation between an adjective and noun. Examples are given in (4) below, where COM means common case, and NEU means neuter case..

(4)	a.	grammatical control:	Ik zag een donkere **wolk** aan de horizon I saw a dark.COM **cloud.COM** on the horizon
	b.	gender violation:	Ik zag een donker **wolk** aan de horizon I saw a dark.NEU **cloud.COM** on the horizon
	c.	number violation:	Ik zag enkele donkere **wolk** aan de horizon I saw several dark **cloud.SG** on the horizon

Both violations lead to a P600 response in the ERP domain (with relatively similar latency and scalp distribution). Both violations lead to an increase in power in the theta band 300-500ms post violation, with the gender violation showing a right-anterior scalp distribution, and the number violation showing a left-anterior scalp distribution. These results are potentially interesting in two ways. First, the latency of the time-frequency response (300-500ms) differs from the latency of the ERP response (500-800ms). Second, the scalp distributions of the time-frequency responses vary by violation, whereas the scalp distributions of the ERP responses do not. This result was one of the first demonstrating that time-frequency analysis can yield different information than ERP analysis (this is obviously true in principle, but Bastiaansen et al. demonstrated that it was also true in practice).

Davidson and Indefrey 2007 investigated the ERP and time-frequency responses to both number and phrase structure violations in English. Examples are given in (5) below.

(5)	a.	number violation:	The children **walks** to school
	b.	phrase structure violation:	Max's proof the **of** theorem

They found P600 responses to both violations in the ERP domain, as expected given the previous literature, and a decrease in power in both the alpha and beta frequency bands in the time-frequency domain. Crucially, they found a relationship between the ERP and time-frequency responses: participants who showed a larger P600 effect also showed a larger decrease in alpha and beta power. Davidson and Indefrey characterize this as an inverse relationship: an increase in the ERP correlates with a decrease in time-frequency power.

4.2 Syntactic violations that lead to ELANs: disruption in beta, decreases in power in alpha and gamma

Bastiaansen, Magyari, and Hagoort 2010 investigated the time-frequency response to word category violations compared to grammatical sentences and random re-orderings of the words in the sentence. Examples are given in (6) below.

(6)	a.	grammatical control:	Janneke kreeg de **zegen** bij de rivier. Janneke got the **blessing** at the river
	b.	word category violation:	Janneke kreeg de **zegenen** bij de rivier. Janneke got the **to-bless** at the river
	c.	random order:	De de Janneke zegen kreeg rivier bij The the janneke blessing got river at

Bastiaansen et al. used MEG for this particular study, so there is no ERP effect to report; that said, word category violations typically yield an ELAN effect in EEG studies. In the time-frequency domain, Bastiaansen et al. found a linear increase in power in the lower beta frequency band to grammatical sentences (i.e., the power in these two bands increased with each successive word in the sentence). They found that the word category violation disrupted this linear increase in beta (creating what could look like a decrease in beta, similar to the Davidson and Indefrey 2007 result), in addition to creating a decrease in power in the alpha band (again, similar to Davidson and Indefrey) and gamma bands. There was no linear increase in beta in response to the random order condition. They also report a linear increase in the theta band to all three conditions that does not appear to be disrupted by either the syntactic violation or the random ordering of words.

4.3 Semantic violations that lead to an N400: increase in power in theta and gamma

Hagoort, Hald, Bastiaansen, and Petersson 2004 report to the time-frequency response to two types of meaning-related violations in Dutch: violations of semantic congruency (e.g., trains cannot be sour), and violations of arbitrary facts about the world (e.g., trains in the Netherlands are not white). Though the stimuli were in Dutch, Hagoort et al. only repor the English transliterations:

(7)	a.	grammatical control:	The dutch trains are **yellow** and very crowded.
	b.	world violation:	The dutch trains are **white** and very crowded.
	c.	semantic violation:	The dutch trains are **sour** and very crowded.

Both of these violations yield N400s in the ERP domain, with a slightly larger N400 for the semantic violation than the world knowledge violation. In the time-frequency domain, both violations yield an increase in power in the theta band, with a slightly larger increase to the semantic violation. The world knowledge violation also yielded an increase in power in the gamma band. Though this result suggests a direct relationship between the size of the N400 effect and the size of the increase in power in the theta band, it appears as though Hagoort et al. report total power – that is, a power analysis that includes both phase-locked (i.e., ERP) and non-phase-locked activity. Thus the larger power increase for semantic violations could simply reflect the larger ERP effect. The Davidson and Indefrey 2007 study mentioned previously computed induced power only – that is, non-phase-locked power. Davidson and Indefrey also investigated violations of semantic congruency, and found an inverse relationship between the size of the N400 effect in the ERP domain and the size of the (induced only) theta band increase in the time-frequency domain: larger N400 effects lead to smaller (induced only) theta band increases.

Wang, Zhu, and Bastiaansen 2012 further elaborated the investigation of semantic congruency violations by partially crossing congruency and predictability, leading to three conditions: congruent and predictable, congruent and unpredictable, and incongruent (which is also unpredictable).

(8)	a.	congruent+predictable:	In the concert hall an **orchestra** played the second symphony of Beethoven.
	b.	congruent+unpredictable:	In the concert hall an **expert** played the second symphony of Beethoven.
	c.	incongruent:	In the concert hall a **finding** played the second symphony of Beethoven.

In the ERP domain, Wang et al. found the expected cline in N400 deflections: the incongruent condition leads to the largest N400 deflection, the congruent and unpredictable condition leads to a smaller N400 deflection, and the congruent and predictable condition leads to the smallest N400 deflection. In the time-frequency domain, Wang et al. found two effects. First, the congruent and predictable condition showed an increase in power in the gamma band relative to the two other conditions (which showed no difference relative to each other). Second, the incongruent (and unpredictable) condition showed an increase in power in the theta band relative to the two other conditions (which showed no difference relative to each other). Wang et al. take this to suggest that gamma activity may be related to predictability, since it appears to divide the conditions by predictability, whereas theta activity may be related to (semantic congruency) error detection, as it appears to divide the conditions by (semantic congruency) error.

4.4 Dependencies that lead to SANs: increased coherence in theta, beta, and gamma,

Weiss, Mueller, Schack, King, Kutas, and Rappelsberger 2005 investigated the time-frequency response to subject and object relative clauses, as in (9):

(9) a. subject RC: The fireman who __ speedily rescued the cop sued the city …

b. object RC: The fireman who the cop speedily rescued __ sued the city …

As mentioned in section 3.1, in the ERP domain, object relative clauses elicit a sustained anterior negativity relative to subject relative clauses (King and Kutas 1995). In the time-frequency domain, Weiss et al. found that object relative clauses showed increased coherence between anterior and posterior electrode sites in the gamma band during the relative clause (relative to subject relative clauses). They also found that object relative clauses showed increased coherence between anterior and posterior electrode sites in the theta and beta bands for several words after the gap location of the relative clause (relative to subject relative clauses).

5. Linking electrophysiology and syntax

Like any sentence-processing-related measure, linking syntactic theories and EEG/MEG responses requires specifying a linking hypothesis between syntactic theories and sentence processing theories. This is a recurring theme in this handbook. With that link in hand, it is, in principle, possible to look for differential predictions made by the combined syntactic and sentence processing theories in the EEG/MEG domain – ERPs, time-frequency responses, etc. This is no small order. In this section, we would like to briefly discuss a few of the studies that have explored different approaches to linking syntax and EEG/MEG. Our hope is that these examples will provide a starting point for experimental syntacticians who are beginning to think about their own experiments.

Bemis and Pylkkänen 2011 attempted to isolate syntactic and semantic combinatorial processes by recording MEG while presenting participants with two-word sequences that form a syntactically and semantically well-formed phrase, such as ‘red boat’ (presented one word at a time, visually), and comparing the activation to two item sequences that do not form a syntactic or semantic phrase, such as noun lists (‘cup, boat’), and sequences that include an unpronounceable consonant string (‘xkq boat’). The logic of this design is that these two-word phrases likely involve the fundamental processes of syntactic and semantic composition, while avoiding many of the other processes that arise during the processing of complete sentences. They found two potentially interesting patterns of activity: an increase in activity for the two-word condition 200-250ms after the onset of the second word that localizes to an area of the cortex that has been linked to syntactic processing in the past (left anterior temporal lobe), and an increase in activity for the two-word condition 300-500ms after the onset of the second word that localizes to an area of the cortex that has been linked to semantic processing in the past (ventromedial prefrontal cortex). These results suggest that minimal designs like this could be used to isolate fundamental syntactic and semantic processes, while side-stepping certain potential confounds (a research program that the Pylkkänen lab has been exploring in the domain of semantic processing). The are two familiar challenges to expanding on this approach: (i) the theoretical challenge of identifying fundamental syntactic processes beyond basic phrasal composition, and (ii) the methodological challenge of isolating those processes in concrete stimuli. Nonetheless, these results are an encouraging proof of concept for minimal designs.

Brennan and Pylkkänen 2016 recorded MEG while participants read a 1279-word story, one word at a time. They constructed a context-free grammar for prepositional phrases that could be used to syntactically analyze 224 words in the story (i.e., the prepositions, nouns, determiners, and adjectives inside of prepositional phrases). They combined that grammar with two parsers: a left-corner parser, which is a psycholinguistically plausible model for how humans process sentences, and a type of bottom-up parser, which is not generally considered a psycholinguistically plausible model for how humans process sentences. They then calculated the number of parser operations triggered by each of the 224 words according to each parser, and looked for correlations between the parser operation counts at each word, and source-localized MEG activity. The idea behind this analysis is that the number of parser operations can serve as a type of incremental complexity metric for each parser; if an area of cortex shows a pattern of activation that correlates with this complexity metric, it suggests that the area of cortex is involved in the syntactic processing predicted by the parser. They found no cortical areas that showed activity that significantly correlated with the bottom-up parser. For the left-corner parser, they found a significant correlation between the number of parse steps and activity in the left anterior temporal lobe 300-500ms after word onset. These results suggest that incremental complexity metrics derived from psycholinguistically plausible grammar plus parser combinations can be correlated with electrophysiological activity in a way that complexity metrics from sophisticated-yet-implausible parsers cannot. To expand on this approach, one could imagine (i) scaling up the grammatical analysis to cover more complex syntactic phenomena, and (ii) comparing the predictions made by two distinct grammars combined with the same parser.

Nelson et al. 2017 recorded intracranial EEG from 12 participants (who were undergoing a separate medical procedure requiring intracranial EEG and several days of waiting) while they read sentences between 3 and 10 words long, one word at a time. Crucially, these sentences contained a subject noun phrase that varied in length, such as ‘Ten students’ and ‘Ten sad students of Bill Gates’. Nelson et al. measured broadband high gamma activity (typically 60Hz to 200Hz), which can be interpreted as an index of the intensity of the activity of the neurons that are local to a given intracranial electrode. They report a number of findings, including correlations between incremental complexity metrics for top-down and left-corner parsers (combined with a context-free grammar that covers the sentences), and the spatial distributions of the various results. Their primary finding is that there is an increase in high gamma activity as the constituent length of the subject noun phrase increases, with a decrease in activity at both potential and actual constituent boundaries. They interpret this pattern as a potential correlate of phrase-structure building (e.g., minimalist *merge*). These result are particularly encouraging for future intracranial EEG studies; unfortunately, they may be less encouraging for extracranial EEG studies given that high gamma is typically not detectable on the scalp (because higher frequencies tend to have lower amplitude, and therefore are more likely to be attenuated by the biological tissue between the cortex and the scalp). Nonetheless, these results suggest that there are (potential) electrophysiological correlates of syntactic structure building that are detectable with some form of current EEG technology, which in turn suggests that it may be worthwhile for experimental syntacticians to look for similar correlates in frequency ranges that are detectable on the scalp.

Ding et al. 2016 used a technique known as a steady state response to demonstrate that speakers construct units of syntactic structure that are larger than syllables and words, such as phrases and sentences. The idea of a steady state response is to present stimuli at a constant rate to induce a response in the brain at that presentation rate. Ding et al. auditorily synthesized monosyllabic words in Mandarin, and presented them to native speakers at a rate of 4Hz (250ms per word) while recording MEG. Crucially, the words were arranged into four-word sentences, each consisting of two two-word phrases (e.g., "New plans gave hope."), with no breaks between the sentences. Ding et al. looked at the frequency-response induced by this design. They found statistically significant increases in power at 1Hz, 2Hz, and 4Hz. The 4Hz activity is not surprising – the physical stimuli were presented at 4Hz. The 1Hz and 2Hz activity is a different story. The 1Hz activity appears to reflect the construction of complete sentences (4 words presented at 250ms per word yields one complete sentence per second). The 2Hz activity appears to reflect the construction of the two phrases that constitute each sentence. Ding et al. provide causal support for this interpretation by showing that unstructured word lists generate activity only at 4Hz, and that the Mandarin sentences only generate 4Hz activity in English speakers who do not speak Mandarin. They also show that the frequency responses can be modulated by manipulating the size of the phrases in the sentences (i.e., a three-word verb phrase or three-word noun phrase). These results suggest that the 1Hz and 2Hz activity increases were not driven by the 4Hz presentation rate (i.e., they were not harmonics of the presentation rate), but rather were driven by syntactic processing of the stimuli. This in turn suggests that steady state designs may be another potential tool for experimental syntacticians to explore. The primary challenge to expanding the use of steady state designs is to figure out how to induce more complex or subtle syntactic phenomena at a constant rate.

Hale, Dyer, Kuncoro, and Brennan 2018 demonstrate a slightly different path forward for linking syntactic analyses and electrophysiological responses. Hale et al. introduce a number of new approaches to linking EEG and sentence processing that are beyond the scope of this chapter, from leveraging recurrent neural network grammars (e.g., Dyer et al. 2016), to leveraging beam search as a sort of parsing algorithm (e.g., Stern et al. 2017). For our purposes, we want to focus on their use of the information-theoretic measure *surprisal* as an incremental measure of complexity during sentence processing (i.e., an incremental complexity metric). Surprisal is a measure of the unexpectedness of word given the previous words in the sentence. Surprisal is defined in such a way that it is larger for unexpected words, and smaller for expected words. (Surprisal is mathematically defined as the logarithm of the reciprocal of the probability of the appearance of that word as the next word in the string, which can be equivalently calculated as the negative log of the transitional probability of the word; see Hale 2016 for a detailed review). Surprisal is calculated directly from a probabilistic grammar, without any explicit reference to a mechanistic parsing theory like top-down or left-corner parsers. We can thus make a distinction between information-theoretical metrics like surprisal and automata-theoretic (or memory-theoretic) metrics like stack size or number of parsing operations. It is an open question to what extent the two types of metrics provide different windows into the properties of human sentence processing. To explore the use of information-theoretic metrics, Hale et al. recorded EEG while participants passively listened to a spoken presentation of the first chapter of Alice in Wonderland. They calculated various information-theoretic metrics for each word in the presentation, including surprisal, and used those metrics to search for activity in the (time-amplitude domain of the) EEG that correlated with those metrics. They identified two significant effects: a positivity over frontal electrode sites around 250ms after word onset, and a positivity over central electrode sites around 600ms after word onset. It is interesting to note that these two effects appear similar to two well-known ERPs in the literature – the P2 (not discussed here, but plausibly linked to predictability in the lexical access literature), and the P600 (as discussed in section 3.4). These results suggest that information-theoretical complexity measures provide another potential tool for connecting to explicit syntactic analyses to electrophysiological responses during sentence processing.

7. Conclusion

It is our hope that this chapter provides a relatively useful introduction to both the prospects and challenges of using electrophysiological measures like EEG and MEG to study syntax. We believe that there is quite a bit of potential for new work in this area, both in principle, because syntactic theories are ultimately intended to be theories of cognition, and in practice, because new methods and linking theories are constantly being developed to link syntax, sentence processing, and electrophysiology. That said, this line of research is not without risk, both because of the biological facts of EEG and MEG (i.e., they only detect a subset of cortical activity), and because of the size of the methodological challenge (linking syntactic and sentence processing theories is no small task). There are also a number of practical challenges, such as the time requirements for data collection (often several months for one experiment), and the time requirements to learn and deploy the complex data analysis techniques discussed in previous sections (again, often several months). As such, we recommend that experimental syntacticians adopt electrophysiological methods only after careful consideration of all of these factors. It is a classic instance of a high-risk/high-reward method, at least in the context of syntactic theory. That said, for those who decide that the they are willing to assume those risks, we believe that it is a potentially exciting and valuable tool for experimental syntax.

References

Ainsworth-Darnell, Kim, Shulman, Harvey G., & Boland, Julie E. 1998. Dissociating brain responses to syntactic and semantic anomalies: Evidence from event-related potentials. Journal of Memory and Language 38: 112–130.

Allopenna, Paul, James S. Magnuson, and Michael K. Tanenhaus. 1998. Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. Journal of Memory and Language 38: 419-439.

Bastiaansen, Marcel C. M., Jos J. A. van Berkum, and Peter Hagoort. 2002. Syntactic processing modulates the theta rhythm of the human EEG. NeuroImage 17: 1479–1492.

Bastiaansen, Marcel C. M., Lilla Magyari, and Peter Hagoort. 2010. Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. Journal of Cognitive Neuroscience 22: 1333–1347.

Bastiaansen, Marcel C. M., Ali Mazaheri, and Ole Jensen. 2013. Beyond ERPs: Oscillatory neuronal dynamics. In the Oxford Handbook of Event-Related Potential Components. Edited by Steven J. Luck and Emily S. Kappenman. 31-50. Oxford University Press.

Bemis, Douglas, and Liina Pylkkänen. 2011. Simple Composition: A Magnetoencephalography Investigation into the Comprehension of Minimal Linguistic Phrases. The Journal of Neuroscience 31: 2801-2814.

Bornkessel-Schlesewsky, Ina, and Matthias Schlesewsky. 2008. An alternative perspective on “semantic P600” effects in language comprehension. Brain Research Reviews 59: 55–73.

Brennan, Jonathan R., and Liina Pylkkänen 2016. MEG Evidence for Incremental Sentence Composition in the Anterior Temporal Lobe. Cognitive Science 41: 1515-1531.

Brown, Colin, and Peter Hagoort. 1993. The processing nature of the n400: Evidence fiom masked priming. Journal of cognitive neuroscience 5: 34–44.

Cohen, Mike X. 2014. Analyzing neural time series data. Cambridge, MA: MIT Press.

Coulson, Seana, Jonathan King, and Marta Kutas. 1998. Expect the unexpected: Event- related brain response to morphosyntactic violations. Language and cognitive processes 13: 21–58.

Davidson, Doug J., and Peter Indefrey. 2007. An inverse relation between event-related and time-frequency violation responses in sentence processing. Brain Research 1158: 81–92.

Dikker, Suzanne, Rabagliati, Hugh, & Pylkkänen, Liina. 2009. Sensitivity to syntax in the visual cortex. Cognition 110: 293–321.

Ding, Nai, Lucia Melloni, Hang Zhang, Xing Tian, and David Poeppel. 2016. Cortical tracking of hierarchical linguistic structures in connected speech. Nature Neuroscience 19: 158-164.

Dyer, Chris, Adhiguna Kuncoro, Miguel Ballesteros, and Noah A. Smith. 2016. Recurrent neural network grammars. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 199–209, San Diego, California.

Federmeier, Karen D., and Marta Kutas. 1999. A rose by any other name: Long-term memory structure and sentence processing. Journal of Memory and Language 41: 469–495.

Fiebach, Christian J., Matthias Schlesewsky, and Angela D. Friederici. 2002. Separating syntactic memory costs and syntactic integration costs during parsing: the processing of german wh-questions. Journal of Memory and Language 47: 250–272.

Frazier, Lyn 1978. On comprehending sentences: syntactic parsing strategies. University of Connecticut. PhD Dissertation.

Friederici, Angela D. 1995. The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. Brain and language 50: 259-281.

Friederici, Angela D., and Stefan Frisch. 2000. Verb argument structure processing: The role of verb-specific and argument-specific information. Journal of Memory and language 43: 476-507.

Friederici, Angela D., Anja Hahne, and Axel Mecklinger. 1996. Temporal structure of syntactic parsing: early and late event-related brain potential effects. Journal of Experimental Psychology: Learning, Memory, and Cognition 22: 1219-1248.

Friederici, Angela D., Erdmut Pfeifer, and Anja Hahne. 1993. Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. Cognitive Brain Research 1: 183-192.

Frisch, Stefan, Matthias Schlesewsky, Douglas Saddy, and Annegret Alpermann. 2002. The P600 as an indicator of syntactic ambiguity. Cognition 85: B83–B92.

Gouvea, Ana, Colin Phillips, Nina Kazanina, and David Poeppel. 2010. The linguistic processes underlying the P600. Language and Cognitive Processes 25: 149-188.

Gunter, Thomas G., Laurie A. Stowe, and Gusbertus Mulder. 1997. When syntax meets semantics. Psychophysiology 34: 660–676.

Hagoort, Peter. 2008. The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. Philosophical Transactions of the Royal Society B: Biological Sciences 363: 1055-1069.

Hagoort, Peter, Colin Brown, and Jolanda Groothusen. 1993. The syntactic positive shift (SPS) as an ERP measure of syntactic processing. Language and cognitive processes 8: 439-483.

Hagoort, Peter, Lea Hald, Marcel C. M. Bastiaansen, and Karl Magnus Petersson. 2004. Integration of word meaning and world knowledge in language comprehension. Science 304: 438–441.

Hagoort, Peter, Marlies Wassenaar, and Colin M. Brown. 2003. Syntax-related ERP-effects in Dutch. Cognitive Brain Research 16: 38-50.

Hahne, Anja, & Friederici, Angela D. 1999. Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. Journal of Cognitive Neuroscience 11: 194–205.

Hahne, Anja, & Friederici, Angela D. 2002. Differential task effects on semantic and syntactic processes as revealed by ERPs. Cognitive Brain Research 13: 339–356.

Hale, John. 2016. Information-theoretical complexity metrics. Language and Linguistics Compass 10: 397–412.

Hale, John, Chris Dyer, Adhiguna Kuncoro, and Jonathan R. Brennan. 2018. Finding Syntax in Human Encephalography with Beam Search. arXiv:1806.04127.

Hinojosa, José A., Martin-Loeches, Manuel, Casado, Pilar, Muñoz, Francisco, & Rubia, Francisco J. 2003. Similarities and differences between phrase structure and morphosyntactic violations in Spanish: An event-related potentials study. Language and Cognitive Processes 18: 113–142.

Isel, Frédéric, Hahne, Anja, Maess, Burkhard, & Friederici, Angela D. 2007. Neurodynamics of sentence interpretation: ERP evidence from French. Biological Psychology 74: 337– 346.

Kaan, Edith. 2002. Investigating the effects of distance and number interference in processing subject-verb dependencies: An ERP study. Journal of Psycholinguistic Research 31:165– 193.

Kaan, Edith, Anthony Harris, Edward Gibson, and Phillip J. Holcomb. 2000. The P600 as an index of syntactic integration difficulty. Language and Cognitive Processes 15: 159–201.

Kaan, Edith, and Tamara Y. Swaab. 2002. The brain circuitry of syntactic comprehension. Trends in Cognitive Sciences 6: 350–356.

Kaan, Edith, and Tamara Y. Swaab. 2003. Electrophysiological evidence for serial sentence processing: a comparison between non-preferred and ungrammatical continuations. Cognitive Brain Research 17: 621–635.

Kim, Albert, and Lee Osterhout. 2005. The independence of combinatory semantic processing: Evidence from event-related potentials. Journal of Memory and Language 52: 205–225

King, Jonathan, and Marta Kutas. 1995. Who did what and when? Using word- and clause level ERPs to monitor working memory usage in reading. Journal of Cognitive Neuroscience 7: 376–395.

Kluender, Robert, and Marta Kutas. 1993a. Bridging the gap: Evidence from erps on the processing of unbounded dependencies. Journal of cognitive neuroscience 5: 196–214.

Kluender, Robert, and Marta Kutas. 1993b. Subjacency as a processing phenomenon. Language and cognitive processes 8: 573–633.

Kuperberg, Gina R. 2007. Neural mechanisms of language comprehension: Challenges to syntax. Brain Research 1146: 23–49.

Kuperberg, Gina R., Tatiana Sitnikova, David Caplan, and Phillip J. Holcomb. 2003. Electrophysiological distinctions in processing conceptual relationships within simple sentences. Cognitive Brain Research 17: 117–129.

Kutas, Marta, and Karen D. Federmeier. 2000. Electrophysiology reveals semantic memory use in language comprehension. Trends in Cognitive Sciences 4: 463–470

Kutas, Marta, and Kara D. Federmeier. 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annual review of psychology 62: 621-647.

Kutas, Marta, and Steven Hillyard. 1980. Event-related brain potentials to semantically inappropriate and surprisingly large words. Biological Psychology 11: 99–116.

Lau, Ellen, Diogo Almeida, Paul C Hines, and David Poeppel. 2009. A lexical basis for n400 context effects: Evidence from meg. Brain and Language 111: 161–172.

Lau, Ellen, Clare Stroud, Silke Plesch, and Colin Phillips. 2006. The role of structural prediction in rapid syntactic analysis. Brain and language 98: 74-88.

Lewis, Richard L., and ShravanVasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. Cognitive Science 29: 375–419.

Luck, Steven J. 2014. An introduction to the event-related potential technique. Cambridge, MA: MIT Press.

Martín-Loeches, Manuel , Francisco Muñoz, Pilar Casado, A. Melcón, Carlos Fernández-Frías. 2005. Are the anterior negativities to grammatical violations indexing working memory? Psychophysiology 42: 508-519.

Mazaheri, Ali, and Ole Jensen. 2010. Rhythmic Pulsing: Linking Ongoing Brain Activity with Evoked Responses. Frontiers in Human Neuroscience 4: 177.

McElree Brian, Stephani Foraker, and Lisbeth Dyer. 2003. Memory structures that subserve sentence comprehension. Journal of Memory and Language 48: 67–91.

McKinnon, Richard. 1996. Constraints on movement phenomena in sentence processing: Evidence from event-related brain potentials. Language and cognitive processes 11: 495–524.

Meltzer, Jed A., and Allen R. Braun. 2013. P600-like positivity and Left Anterior Negativity responses are elicited by semantic reversibility in nonanomalous sentences. Journal of Neurolinguistics 26: 129-148.

Molinaro, Nicola, Horacio A. Barber, and Manuel Carreiras. 2011. Grammatical agreement processing in reading: ERP findings and future directions. Cortex 47: 908-930.

Mueller, Jutta L., Hahne, Anja, Fujii, Yugo, & Friederici, Angela D. 2005. Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. Journal of Cognitive Neuroscience 17: 1229–1244.

Münte, Thomas F., and Hans Jochen Heinze. 1994. ERP negativities during syntactic processing of written words. In Cognitive electrophysiology, 211-238. Boston, MA: Birkhäuser.

Münte, Thomas F., Mike Matzke, and Sönke Johannes. 1997. Brain activity associated with syntactic incongruencies in words and pseudo-words. Journal of cognitive neuroscience 9: 318–329.

Nelson, Matthew J., Imen El Karoui, Kristof Giber, Xiaofang Yang, Laurent Cohen, Hilda Koopman, Sydney S. Cash, Lionel Naccache, John T. Hale, Christoph Pallier, Stanislas Dehaene. 2017. Neurophysiological dynamics of phrase-structure building during sentence processing. Proceedings of the National Academy of Sciences 114: E3669-E3678.

Neville, Helen, Janet L. Nicol, Andrew Barss, Kenneth I. Forster, and Merrill F. Garrett. 1991. Syntactically based sentence processing classes: Evidence from event-related brain potentials. Journal of Cognitive Neuroscience 3: 151-165.

Osterhout, Lee, and Peter Hagoort. 1999. A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 debate. Language and Cognitive Processes 14: 1-14.

Osterhout, Lee, and Phillip J. Holcomb. 1992. Event-related brain potentials elicited by syntactic anomaly. Journal of memory and language 31: 785–806

Osterhout, Lee, Phillip J. Holcomb, and Daniel Swinney. 1994. Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. Journal of Experimental Psychology: Learning, Memory, and Cognition 20: 786–803.

Osterhout, Lee, and Linda A. Mobley. 1995. Event-related brain potentials elicited by failure to agree. Journal of Memory and Language 34: 739–773

Phillips, Colin, Nina Kazanina, and Shani H. Abada. 2005. ERP effects of the processing of syntactic long-distance dependencies. Cognitive Brain Research 22: 407–428

Ruchkin, Daniel S., Ray Johnson, Howard Canoune, Walter Ritter. 1990. Short-term memory storage and retention: an event related brain potential study. Electroencephalography and Clinical Neurophysiology 76: 419–439.

Sassenhagen, Jona, Matthias Schlesewsky, and Ina Bornkessel-Schlesewsky. 2014. The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. Brain and Language 137: 29-39

Steinhauer, Karten, and John E. Drury. 2012. On the early left-anterior negativity (ELAN) in syntax studies. Brain and Language 120: 135-162.

Stern, Mitchell, Daniel Fried, and Dan Klein. 2017. Effective inference for generative neural parsing. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 1695–1700, Copenhagen, Denmark.

Stroud, Clare, and Colin Phillips. 2012. Examining the evidence for an independent semantic analyzer: An ERP study in Spanish. Brain and Language 120: 108-126.

Tanner, Darren, and Janet G. Van Hell. 2014. ERPs reveal individual differences in morphosyntactic processing. Neuropsychologia 56: 289-301.

van Herten, Marieke, Herman H. J. Kolk, and Dorothee J. Chwilla. 2005. An erp study of P600 effects elicited by semantic anomalies. Cognitive Brain Research 22: 241–255.

Van Petten, Cyma, Coulson, Seana, Rubin, Susan, Plante, Elena, & Parks, Marjorie. 1999. Time course of word identification and semantic integration in spoken language. Journal of Experimental Psychology: Learning Memory & Cognition 25: 394–417.

Wang, Lin, Zude Zhu, and Marcel C. M. Bastiaansen. 2012. Integration or predictability? A further specification of the functional role of gamma oscillations in language comprehension. Frontiers in Psychology 3: 1–12.

Weiss, Sabine, Horst M. Mueller, Baerbel Schack, Jonathan W. King, Marta Kutas, and Peter Rappelsberger. 2005. Increased neuronal communication accompanying sentence comprehension. International Journal of Psychophysiology 57: 129–141.

Ye, Zheng, Luo, Yue-Jia, Friederici, Angela D., & Zhou, Xiaolin. 2006. Semantic and syntactic processing in Chinese sentence comprehension: Evidence from event-related potentials. Brain Research 1071: 186–196.

Annotated bibliography for learning electrophysiological methods

Steve Luck's textbook is the place to start for anyone interested in learning EEG. It provides a complete introduction to all of the fundamentals of EEG and the most common analysis technique, the event-related potential technique. The newest edition also includes a number of online chapters that delve into more advanced topics and analysis techniques.

Luck, Steven J. 2014. An introduction to the event-related potential technique. Cambridge, MA: MIT Press.

Mike X Cohen's textbook is the textbook to read for anyone interested in learning time-frequency decomposition. It can, in principle, be read on its own, but it will likely be most useful to readers who are already familiar with the basics of EEG and the ERP technique.

Cohen, Mike X. 2014. Analyzing neural time series data. Cambridge, MA: MIT Press.

EEG data analysis requires familiarity with one or more programming languages. At present, there are far more resources for EEG data analysis written in Matlab than any other language. EEGLAB is a free toolbox for Matlab that provides a complete solution to EEG analysis. It has three particular strengths: (i) users can write their own toolboxes for EEGLAB to extend its functionality, (ii) it has both a scripting option and a graphical user interface, and (iii) it has the most well-developed set of ICA (independent components analysis) tools.

Delorme, Arnaud, and Scott Makeig. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. Journal of Neuroscience Methods 134: 9-21

Download: https://sccn.ucsd.edu/eeglab/index.php

ERPLAB is a free toolbox for EEGLAB developed by the Luck lab. It provides a complete solution to using the ERP technique (and also implements a number of methodological recommendations found in Luck 2014). As a toolbox for EEGLAB, it provides both a scripting and graphical user interface option.

Lopez-Calderon, Javier, and Steven J. Luck. 2014. ERPLAB: an open-source toolbox for the analysis of event-related potentials. Frontiers in human neuroscience 8: 213.

Download: https://erpinfo.org/erplab/

Fieldtrip is a free toolbox for Matlab that provides a complete solution to EEG analysis. Though it can be used for standard ERP analysis, Fieldtrip's strength lies in advanced analysis techniques, like time-frequency analysis. We typically use ERPLAB for ERP analysis and Fieldtrip for time-frequency analysis.

Oostenveld, Robert, Pascal Fries, Eric Maris, and Jan-Mathijs Schoffelen. 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational intelligence and neuroscience 2011: 1.

Download: http://www.fieldtriptoolbox.org/

For readers who wish to learn more about Matlab programming, Mike X Cohen has written a Matlab textbook that is specifically tailored to EEG data analysis.

> Cohen, Mike X. 2017. Matlab for brain and cognitive scientists. Cambridge, MA: MIT Press.

The statistical analysis of EEG data is a particularly complex topic given the extreme multiple comparisons problem posed by multiple electrodes and high sampling rates. Mass univariate permutation tests provide a good solution to this problem. The Fieldtrip toolbox implements its own mass univariate permutation tests. For EEGLAB, there are two plugins. The Mass Univariate Toolbox implements permutation tests for one condition and two-condition experimental designs (*t*-tests). The Factorial Mass Univariate Toolbox implements permutation tests for factorial designs.

> Groppe, David M., Thomas P. Urbach, and Marta Kutas. 2011. Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. Psychophysiology 48: 1711-1725.
>
> Download: https://openwetware.org/wiki/Mass_Univariate_ERP_Toolbox
>
> Fields, Eric C., and Gina R. Kuperberg. 2018. Having Your Cake and Eating It Too: Flexibility and Power with Mass Univariate Statistics for ERP Data. PsyArXiv. doi:10.31234/osf.io/qfkgc.
>
> Download: https://github.com/ericcfields/FMUT/wiki

EEG data analysis requires relatively complex mathematics. In most cases, the software solutions discussed above will perform the math without requiring user intervention. For readers interested in a deeper understanding of the math, here is a list of free, open-source math textbooks. The important concepts will likely be found in trigonometry and linear algebra (sine/cosine, dot product, Fourier transform, convolution, complex numbers, etc).

A list of open-source textbooks: https://aimath.org/textbooks/approved-textbooks/