

Prosody vs. syntax, or prosody and syntax?

Evaluating accounts of delta-band tracking

Coopmans, Cas W.^{1,2*}, Martin, Andrea E.^{1,2}

1 Donders Institute for Brain, Cognition and Behaviour, Radboud University,
Nijmegen, The Netherlands

2 Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

*Corresponding author:

cas.coopmans@mpi.nl

Abstract

Recent studies have shown that neural activity tracks the syntactic structure of phrases and sentences in connected speech. This work has sparked intense debate, with some researchers aiming to account for the effect in terms of the overt or imposed prosodic properties of the speech signal. In this chapter, we present four types of arguments against attempts to explain putatively syntactic tracking effects in prosodic terms. The most important limitation of such prosodic accounts is that they are architecturally incomplete, as prosodic information does not arise in speech autonomously. Prosodic and syntactic structure are interrelated, so prosodic cues are informative about the intended syntactic analysis, and syntactic information can be used to aid speech perception. Rather than trying to attribute neural tracking effects exclusively to one linguistic component, we consider it more fruitful to think about ways in which the interaction between the components drives the neural signal.

Keywords

speech processing, language processing, neural tracking, syntactic structure, prosodic phrasing, brain rhythms

1. Introduction

Spoken language is a complex and intricately structured physical signal. Its rich variation stems from a myriad of sources, which are not easily decomposable into discrete variables. What is more, many of these sources of information are related, yielding correlations in the speech signal that are difficult to disentangle experimentally. For spoken language processing, however, this redundancy might actually prove to be a valuable property, because it enables the brain to leverage one feature to draw inferences about another feature, thereby facilitating comprehension. In other words, what seems like an annoying property from an experimental perspective might turn out to be an integral condition for perceptual inference.

To give a key example, the structures of syntax and prosody are related, due to which syntactic and prosodic variation are correlated in the speech signal (Cooper & Paccia-Cooper, 1980; Ferreira, 1993; Gee & Grosjean, 1983; Nespor & Vogel, 1986; Selkirk, 1978, 1984; Shattuck-Hufnagel & Turk, 1996).¹ This means that syntax and prosody probabilistically cue one another in spoken language comprehension (Cole, 2015; Cutler et al., 1997; Martin, 2016, 2020; Wagner & Watson, 2010). From the perceiver's perspective, the redundancy is quite useful, because it supports syntactic inference when the intended syntactic analysis of a spoken utterance is underdetermined. This scenario is essentially the norm for pre-linguistic infants, who strongly rely on prosodic information because their syntactic competence is yet not sufficient to fully parse the utterances they hear. But the situation is also common in adults, who use prosody to disambiguate sentences that are structurally ambiguous. From the experimental perspective of a cognitive neuroscientist, however, this mixing of information sources is very challenging, because it makes it difficult to determine their unique contributions to the neural signal. For instance, as modulations of syntax and prosody in spoken language are temporally correlated, syntactic and prosodic accounts of empirical findings in the electrophysiological literature are largely compatible with the same set of results: like syntactic effects, neural signatures of prosodic processing frequently show up in the delta band (0.5-4 Hz; Boucher et al.,

¹ Prosody is a multifaceted phenomenon with many diverse functions. In this chapter, we only focus on linguistic prosody (in particular, the structure of prosodic constituency) and its relationship to syntactic structure. This is because the literature on neural speech tracking is mostly concerned with this aspect of suprasegmental variation. It does mean that we do not touch upon the role of emotional prosody in spoken language processing. However, as the prosodic realization of emotional states adds another perceptually informative cue to the speech signal, it should also be incorporated in models of language in the brain.

2019; Bourguignon et al., 2013; Ghitza, 2017; Glushko et al., 2022; Henke & Meyer, 2021; Inbar et al., 2023; Meyer et al., 2017; Rimmele et al., 2021). It is therefore often unclear to what extent seemingly syntactic effects are actually due to prosodic properties of the stimulus that are (often unavoidably) correlated with the syntactic manipulations, or vice versa.

As a case in point, a seminal study by Ding and colleagues showed that neural activity ‘tracks’ the hierarchical structure of phrases and sentences in connected speech (Ding et al., 2016). This effect has been replicated with different manipulations of syntactic structure, in different languages, and in different experimental situations (e.g., Blanco-Elorietta et al., 2020; Burroughs et al., 2021; Martin & Dumas, 2017; Makov et al., 2017), suggesting that the signal in question is indeed a neurophysiological marker of syntactic processing. On the other hand, several authors have instead argued that these neural tracking effects do not reflect syntactic structure. One type of account attributes the effect to the overt or imposed prosodic properties of the speech signal (e.g., Boucher et al., 2019; Glushko et al., 2022; Inbar et al., 2023). Such prosodic accounts of syntactic tracking effects are seemingly successful for two related reasons. They are descriptively accurate because syntax and prosody are temporally correlated, both exhibiting periodicities in the delta range. The electrophysiological effects they elicit during spoken language processing therefore have similar spectro-temporal properties, so syntactic effects are often amenable to a prosodic explanation. A deeper, more explanatory reason for the success of prosodic accounts is architectural: syntax and prosody are ontologically related (Cooper & Paccia-Cooper, 1980; Féry, 2016; Ferreira, 1993; Gee & Grosjean, 1983; Nespor & Vogel, 1986; Selkirk, 1978, 1984; Shattuck-Hufnagel & Turk, 1996). It is due to their interdependence that syntax and prosody are temporally correlated and largely compatible with the same electrophysiological results; temporal correlations are based on ontological relations. Insofar as prosodic accounts aim to explain syntactic tracking effects as being purely prosodic, they will therefore be architecturally incomplete, as prosodic information does not arise in speech autonomously.

Accordingly, we will show that prosodic accounts of neural tracking effects often explicitly or implicitly incorporate a role for syntax. But this does not completely invalidate those accounts. Quite the opposite, it is actually a good state of affairs, as we think that the quest for sources of information that uniquely or exclusively explain neural effects misses what we perceive to be the goal of the neurobiology of language,

that is, to explain how language comprehension in the brain actually works. In naturally-produced spoken language, neither prosody nor syntax is ever presented in isolation, so in order to understand how the brain makes sense of language, it should be explained how the interaction between different sources of (linguistic) information, and their functional interdependence, drives the neural signal. In this chapter, we present four types of arguments against prosodic accounts of syntactic tracking effects, and in favor of studying syntax and prosody together when investigating the neural foundations of speech tracking. The chapter is structured as follows: In Section 2, we introduce the central concept of neural tracking, and we explain which inferences are (not) licensed from such neural data (*logical* argument). Section 2.1 then discusses empirical studies which show that the brain tracks syntactic structure in connected speech. In Section 2.2, we review studies that have used similar experimental paradigms to investigate tracking of prosodic structure. We critically evaluate whether a prosodic account of these findings is empirically accurate (*empirical* argument). In Section 3, we discuss the tight relationship between syntax and prosody (*ontological* argument) from the perspectives of cognitive architecture (Section 3.1) and processing (Section 3.2). The final Section 4 provides a summary of our main arguments and recommendations for future work (*strategic* argument).

2. Neural tracking

Neural tracking refers to (pseudo-)rhythmic modulations in the neural signal as a consequence of (pseudo-)rhythmic repetition of energy or information in the stimulus. This process has been argued to result from endogenous oscillations synchronizing with or entraining to repeated properties of the input ('intrinsic synchronicity' in Meyer et al., 2020) and/or from the repetition of responses evoked by the input (Frank & Yang, 2018; Oganian et al., 2023). We use the term 'tracking' descriptively, which means that we do not take a stance here on the neurobiological origin of this effect. Rather, we are concerned with the cognitive implications of observing tracking effects: if the brain represents variable (or type) X , the repeated presentation of token x at frequency f will lead to an increase in power in the frequency spectrum of the neural response at f , showing that the brain recognized x . In such a case, we say that the brain tracks X .

As such, empirical observations of neural tracking immediately present us with an inferential problem: if two relevant variables are temporally correlated in the input,

their repeated presentation will elicit a neural response whose spectral properties align with the timescales of both variables. Logic dictates that, without additional information, we cannot determine the representational source of this tracking effect. A crucial corollary is that neural tracking effects cannot be explained exclusively in terms of either one of these variables. More generally, as the mapping between neural signals and cognitive events is rarely (if ever) one-to-one (Mehler et al., 1984; Westlin et al., 2022), the interpretation of neural data is inherently uncertain. Yet, due to the use of ingenious experimental designs, it has been possible to show that electrophysiological brain activity tracks both syntactic and prosodic structure.

2.1 Tracking syntactic structure

The syntactic structure of phrases and sentences is hierarchically organized (see Figure 1A). Speech, on the other hand, is essentially linear in its physicality, which means that syntax is not visible in the acoustic signal in the way lower-level linguistic information is. To understand spoken language, the brain must infer, or internally construct, syntactic structure using knowledge of grammar.

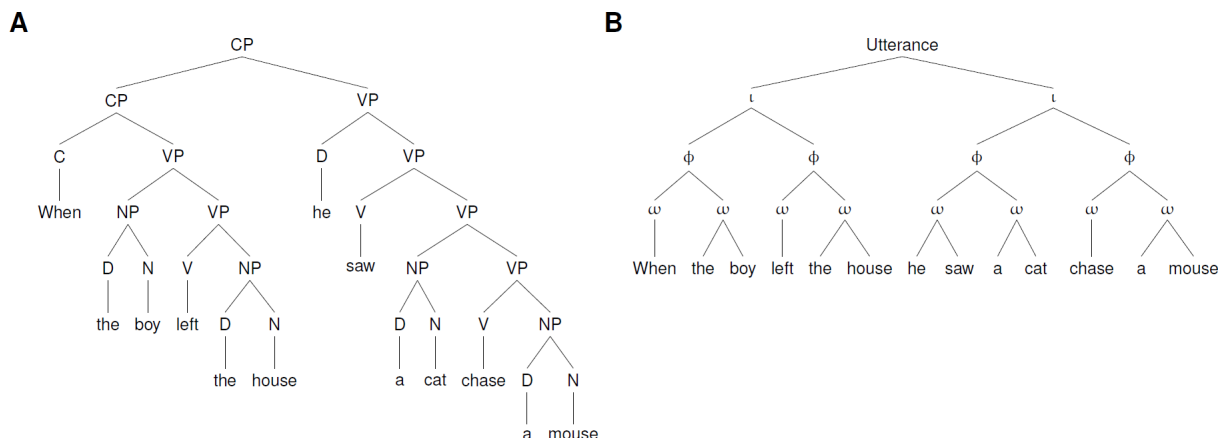


Figure 1. A syntactic hierarchy **(A)** and a prosodic hierarchy **(B)** for the sentence "When the boy left the house, he saw a cat chase a mouse". The prosodic structure is based on a subjectively natural pronunciation of the sentence. It uses the labels from Selkirk (2011), where ι = intonational phrase, ϕ = phonological phrase, ω = prosodic word. As our aim is to illustrate differences in the general organization of prosodic and syntactic hierarchy (i.e., (a)symmetry of and (non)correspondence between prosodic

and syntactic constituents), the syntactic structure in (B) is simplified. CP = complementizer phrase, VP = verb phrase, NP = noun phrase, D = determiner.

Two main paradigms have been used to study neural tracking of syntactic structure. One paradigm relies on strict control of the presentation rate of linguistic structure, whereby this structure is frequency tagged. The idea behind this approach to tracking is that when a unit of information is repeatedly presented at a specific frequency, the neural response to that type of information becomes synchronized with its presentation rate. In a magnetoencephalography (MEG) study by Ding et al. (2016), participants listened to connected speech sequences of monosyllabic words, which were isochronously presented at 4 Hz (i.e., each word lasted 250 ms). Within these sequences, two adjacent words could repeatedly be grouped into 500-ms phrases, and two adjacent phrases could repeatedly be grouped into 1000-ms sentences, such as in [s [NP new plans] [VP gave hope]]. It was found that electrophysiological brain activity concurrently tracked the time courses of words, phrases and sentences, showing peaks in the neural power spectrum at 4 Hz (words), 2 Hz (phrases), and 1 Hz (sentences). Because all words were synthesized independently and presented isochronously, the speech sequences contained no overt prosody. As a result, only the words were clearly defined by acoustic boundaries and therefore physically discernible in the speech signal. Phrases and sentences, instead, were dissociated from prosodic cues and thus had to be internally constructed using knowledge of grammar (Burroughs et al., 2021; Ding et al., 2016; Martin & Doumas, 2017). Indeed, neural activity emerged only at those timescales that correspond to linguistic structures: when the sequences contained words that could be grouped into two-word phrases at most, such as [NP dry fur] [NP new plans], brain recordings revealed a tracking effect at the 2-Hz phrase rate only. No effect emerged at the 1-Hz sentence rate, supposedly because there are no naturally sensible groupings present at 1 Hz. Moreover, neural tracking of syntactic structure does not appear during sleep (Makov et al., 2017), nor when the listener does not understand the language (Ding et al., 2016) or when the speech is embedded in noise (Blanco-Elorietta et al., 2020), suggesting that syntactic tracking effects are linked to actual understanding (i.e., comprehension of the structures).

Because overt prosodic cues are commonly removed from the speech input in frequency-tagging studies, these effects do not easily lend themselves to a prosodic explanation. In another relevant paradigm, the same question is addressed using

naturalistic stimuli, in which prosody is overtly present but matched across conditions. Kaufeld et al. (2020) investigated whether neural speech tracking is modulated by the syntactic and semantic content of the linguistic input. In an electroencephalography (EEG) study, they presented participants with naturally spoken stimuli in three conditions: regular sentences, pseudoword sentences, and unstructured word lists. Tracking was quantified through phase coupling between the speech envelope and neural activity. The presentation rate of phrases, which was derived from manual annotations of the stimuli, served as the frequency band of interest (within the delta band). At this phrasal timescale, speech-brain coupling was stronger for regular sentences than for both pseudoword sentences and word lists (see also Coopmans et al., 2022a). Importantly, this effect was replicated in more direct measures of syntactic tracking, that is, when coupling was computed between the EEG signal and abstract annotations that reflect syntactic structure but contain no acoustic information. These findings show that the brain tracks the timescale of syntactic phrases more strongly when this timescale contains meaningful information and is therefore relevant for language comprehension (Coopmans et al., 2022a; Kaufeld et al., 2020; Keitel et al., 2018; ten Oever et al., 2022). Thus, like the tracking effects of frequency-tagging studies, these coupling-based tracking effects are linked to actual understanding of the structures.

Because these effects have been attributed to the tracking of syntactic structure (Blanco-Elorietta et al., 2020; Burroughs et al., 2021; Coopmans et al., 2022a; Ding et al., 2016; Martin & Doumas, 2017, 2019; Kaufeld et al., 2020; Makov et al., 2017; ten Oever et al., 2022), we will conveniently refer to them as ‘syntactic tracking effects’. Note that this is not meant to imply that the effects are necessarily or exclusively reflective of syntactic structure. One should be skeptical of such claims of exclusivity. The brain tracks anything it recognizes, including transitional probabilities (Henin et al., 2021), rule-based chunks (Jin et al., 2020), and semantic properties of words (Frank & Yang, 2018). If these variables occur at the frequency of phrases (as they often do, even in natural language), their tracking effects will show up at the phrase rate. As already mentioned in Section 2, this logically means that not one of these factors exclusively explains tracking effects. Nevertheless, the idea that syntactic tracking effects can be explained in terms of prosody is quite prominent, because syntax and prosody are temporally correlated. Due to their similar periodicity, they are largely compatible with the same electrophysiological results and are therefore experimentally

and analytically difficult to tease apart. In this chapter, we evaluate, both empirically and conceptually, how well prosodic explanations of syntactic tracking effects fare. In the next section (Section 2.2), we will discuss empirical studies linking electrophysiological signals to prosodic processing. To foreshadow our conclusion, none of these studies demonstrates purely prosodic effects. Rather, in each case do we find modulatory effects of syntactic structure. This is expected, given the strong relationship between syntax and prosody in natural language, which will be discussed in Section 3.

2.2 Tracking prosodic structure

The prosodic structure of spoken utterances is hierarchically organized. As illustrated in Figure 1B, an utterance contains one or more intonational phrases (denoted ι), which are composed of phonological phrases (ϕ ; also called intermediate phrases), which are in turn composed of prosodic words (ω ; Beckman & Pierrehumbert, 1986; Nespor & Vogel, 1986; Selkirk, 1984, 2011; Shattuck-Hufnagel & Turk, 1996).² Prosodic phrasing is phonetically marked in various ways. The boundaries between ι -phrases, for instance, are cued by a silent pause, a change in pitch or fundamental frequency and lengthening of segments preceding the boundary (Beckman & Pierrehumbert, 1986; Cooper & Paccia-Cooper, 1980; Ferreira, 1993; Nespor & Vogel, 1986; Price et al., 1991; Selkirk, 1984; Wightman et al., 1992). In addition to these phonetic markers of prosodic structure, people process prosody even if it is not overtly realized in the stimulus (e.g., in reading). This is known as implicit prosody, a subvocally activated prosodic representation that is projected onto the stimulus and can affect syntactic processing in much the same way as overt prosody does (Breen, 2014; Fodor, 1998, 2002).

Prosodic structure has rhythmic properties, with phrasal boundaries periodically occurring roughly once per second in naturally produced speech (Inbar et al., 2020; Stehwien & Meyer, 2022). These low-frequency prosodic regularities provide a possible source of information for the brain to exploit. Indeed, electrophysiological

² While there is general agreement about the hierarchical nature of prosodic structure, theories differ in the proposed number of levels in the hierarchy, the labels assigned to the constituent types, and their definition (for an excellent review, see Shattuck-Hufnagel & Turk, 1996). For the purpose of illustration, we use the levels described in Selkirk (2011). And because we are concerned with the (temporal) correspondence between prosodic and syntactic phrasing, we restrict ourselves to phonological units above the level of the word, thus excluding feet and syllables.

studies have shown that sentential prosody is tracked by slow neural activity in the delta range. While this overlaps quite strongly with the reported timescale of syntactic processing, we will show in our discussion of these results in the current section that none of them unequivocally shows that syntactic tracking effects can be explained in terms of prosodic processing.

In spoken language comprehension, the boundaries of multi-word chunks are accompanied by a slow event-related potential (ERP) component in the EEG signal. This so-called Closure Positive Shift (CPS) is elicited by $\bar{\iota}$ -boundaries that are phonetically marked through a pitch change and pre-boundary lengthening (Steinhauer et al., 1999). Given the low-frequency rhythmicity of $\bar{\iota}$ -phrases, it is expected that the neural response to repeated phrase closures has a low frequency as well. Indeed, a closure-related CPS is periodically elicited at regular 2- to 3-second intervals (Roll et al., 2012; Schremm et al., 2015), and CPS effects elicited by $\bar{\iota}$ -boundaries contribute to delta-band speech tracking (Inbar et al., 2023). As these CPS effects are observed even when overt prosodic cues are absent, it has been suggested that there are endogenous, time-driven constraints on the grouping of words into multi-word chunks, and that these constraints operate at a frequency corresponding to the (lower) delta band (Henke & Meyer, 2021; Meyer et al., 2017; Roll et al., 2012; Vetchinnikova et al., 2023).

In this context, it is relevant to mention a set of seemingly discrepant findings from the ERP literature. On the one hand, it has been shown in adult listeners that a CPS can be elicited by prosodically modulated stimulus sequences from which all syntactic cues are removed (e.g., hummed sentences, sentence melodies; Gilbert et al., 2015; Pannekamp et al., 2005; Steinhauer & Friederici, 2001). This shows that prosodic structure alone is sufficient to elicit a CPS. On the other hand, the CPS in spoken sentence processing is modulated by whether contextually induced syntactic expectations do or do not support a prosodic boundary (Kerkhofs et al., 2007), even to the extent that a CPS can be elicited by syntactic phrase boundaries that are not marked by prosodic boundary cues (Itzhak et al., 2010). Moreover, infant studies show that prosodically-induced CPS effects are dependent on syntactic development, as they are only observed in children who have acquired knowledge of phrase structure (Männel & Friederici, 2011). In contradiction to the earlier conclusion, these findings indicate that prosodic boundary cues are neither necessary nor sufficient to elicit a CPS. Addressing the apparent discrepancy, the results can be reconciled by the idea

that, due to the tight interplay between syntax and prosody, the presence of one type of information in the input automatically activates the other. If prosodic modulations trigger syntactic representations, and syntactic cues activate prosodic structure, prosodic phrasing can be initiated even when acoustic or syntactic markers of phrase closure are not explicitly present. Either type of information is sufficient to elicit a closure-related CPS, provided that the listener is linguistically proficient – the interactive mapping between syntax and prosody requires understanding of speech structure based on syntactic knowledge (Männel & Friederici, 2011). Summing up this brief discussion of the ERP literature, even for a brain response that might initially seem to be a pure reflection of (implicit) prosodic processing, the role of syntax cannot be ignored.

A similar argument can be made about the results of a recent frequency-tagging study. Specifically, Glushko et al. (2022) claim that the syntactic tracking effects observed by Ding et al. (2016) predominantly reflect prosodic rather than syntactic processing. In Ding et al. (2016), the 2-Hz peak corresponded to the phrasal timescale and was therefore interpreted as reflecting syntactic tracking. The prosodic account put forward by Glushko and colleagues instead holds that the 2-Hz peak for sentences with a 2+2 syntactic structure, like [NP new plans] [VP gave hope], reflects an implicit grouping of the sentence into “new plans / gave hope” (the / indicates a prosodic break), which contains two equally sized ϕ -phrases. This prosodic analysis is possible despite the absence of prosodic cues in isochronous speech because people are known to project implicit prosodic structure onto the input (e.g., following a same-size-sister constraint, see Section 3.1). In an EEG experiment, Glushko and colleagues used sentences with a 1+3 syntactic structure, such as [NP John] [VP likes big trees], which can be analyzed via a 2+2 prosodic grouping as well. In line with their hypothesis, these isochronously presented sentences elicit a 2-Hz peak in the neural power spectrum despite the absence of a major syntactic boundary in the middle of the sentence (i.e., between “likes” and “big”). This delta effect initially suggests that people prosodically analyzed these sentences in a way that is not suggested by their syntactic structure. However, it is not obvious that the listeners’ implicit prosodic analysis is purely prosodic, or whether it is also sensitive to the syntactic structure of the sentence. Notice that the 2+2 prosodic grouping of a sentence with a 1+3 structure retains some of its syntactic constituency. In “John likes big trees”, for instance, the 2+2 grouping yields “John likes / big trees”. While the elements to the left of the

prosodic break (“John likes”) do not form a syntactic constituent, the elements to the right of the break (“big trees”) do, making this prosodic grouping syntactically not entirely unacceptable. In other words, the prosodic units are not identical to the syntactic constituents, but the prosodic boundary does correspond to a relevant syntactic boundary. It is easy to come up with sequentially similar sentences for which this is not the case, such as [NP the cute kid] [VP laughs]. Applying a 2+2 prosodic grouping to this 3+1 structure yields “the cute / kid laughs”. This grouping intuitively sounds much less natural, plausibly because neither of the two resulting prosodic phrases corresponds to a syntactic phrase. Thus, while it initially seems that implicit groupings are induced by mechanisms other than syntactic processing, they might still be sensitive to syntax, in the sense that where people place implicit prosodic boundaries is affected by the syntactic structure underlying the sentence.

Studies with naturally produced speech also show that prosodic modulations affect low-frequency neural activity. For instance, Bourguignon et al. (2013) found similar levels of speech-brain coupling in the delta band for listeners’ native speech, non-native speech, and hummed speech. As these effects are independent of people’s ability to process the meaning and syntactic structure of the input, they were explained in terms of the similar prosodic rhythmicity of the different stimulus types. Relatedly, Boucher et al. (2019) reject the idea that delta activity tracks abstract, non-sensory information. Instead, they argue that delta oscillations underlie sensory chunking via entrainment to articulated sounds. In their EEG study, participants listened to sentences, meaningless syllable strings and tone sequences, all of which were prosodically matched by having similar patterns of timing, pitch and energy. Delta entrainment was considerably reduced for tone sequences compared to both types of speech stimuli, arguably because tone sequences do not contain articulated sounds. No difference was found between sentences and syllable strings, suggesting that the syntactic content of sentences did not affect delta entrainment. To explain these results, Boucher et al. (2019) argue that delta oscillations entrain to temporal groups marked by articulated sounds, which are equally present in spoken sentences and syllable strings. This account, however, cannot explain existing work that does show a relationship between delta-band activity and the high-level linguistic content of speech. First, both Kaufeld et al. (2020) and Coopmans et al. (2022a) find stronger speech-brain coupling at the delta rate for sentences than for prosodically matched stimuli that lack syntactic structure or meaning. Second, Keitel et al. (2018) reported stronger

delta-band tracking for comprehended than for uncomprehended spoken sentences. And third, recent EEG studies by Meyer and colleagues show that delta phase and power reflect people's syntactic parsing choices independent of delta entrainment to speech prosody (Henke & Meyer, 2021; Meyer et al., 2017). If delta tracking only reflects the sensory chunking of speech into temporal groups, it is unclear why it is stronger for regular sentences than for prosodically matched pseudoword sentences, why it correlates with behavioral measures of comprehension, and why it is predictive of people's syntactic analysis of a sentence.

In sum, it has been shown that delta-band neural activity tracks overt and implicit prosody during speech comprehension. Our empirical evaluation reveals both that these prosodic tracking effects are modulated by syntax and that prosodic explanations either do not cover the full range of empirical results or implicitly incorporate a notion of syntactic structure. In the rest of this chapter, we will argue that this is expected, because syntax and prosody in natural language are strongly tied together (Section 3). For this reason, we should not aim to explain syntactic tracking effects as being prosodic in nature, but rather, face the challenging task of determining how prosody and syntax complement and constrain one another, both in the stimulus and in the neural signal. We end by discussing such an interactive approach to syntax and prosody in the brain (Section 4).

3. The relationship between syntactic and prosodic constituency

3.1 Structural alignments and discrepancies

Prosody is often characterized in terms of the *boundaries* marking the edges of prosodic constituents and the relative *prominence* (or accent) assigned to a designated element within these constituents (Cole, 2015; Shattuck-Hufnagel & Turk, 1996; Wagner & Watson, 2010). Both of these play an important role in language processing, but they signal different aspects of the meanings of sentences: while prosodic prominence tends to mark focus and discourse status, prosodic boundaries are commonly related to the boundaries of syntactic structure. Concerning the latter, Selkirk (1978) observed that a particular intonational contour is required for certain syntactic constituents, such as preposed adverbials (e.g., “*In Pakistan // Tuesday is a holiday*”), non-restrictive relative clauses (e.g., “*Tuesday // which is a weekday // is a*

holiday”), and parenthetical expressions (e.g., “Tuesday is // *Jane said* // a holiday”), each of which forms a separate i-phrase (delineated by //). Here, the boundaries of prosodic constituents directly coincide with the boundaries marking major syntactic phrases. In most cases, however, intonational phrasing is not directly informed by syntax but rather based on prosodic structure, which is derived from syntactic constituency but not identical to it (Ferreira, 1993; Féry, 2016; Gee & Grosjean, 1983; Nespor & Vogel, 1986; Selkirk, 1984).

Recall that segments at the end of prosodic phrases are often lengthened and followed by a pause (Cooper & Paccia-Cooper, 1980; Ferreira, 1993; Nespor & Vogel, 1986; Price et al., 1991; Wightman et al., 1992). Structurally ambiguous phrases, like “old men and women”, can be disambiguated by these prosodic cues. That is, the word “men” and the subsequent pause are longer in “old men / and women”, where they occur at the end of a noun phrase, than in “old / men and women”, where they occur phrase-medially. As prosodic phrase boundaries tend to align with syntactic phrase boundaries, acoustic cues such as pre-boundary lengthening provide information about syntactic structure. However, these cues are not diagnostic, because pre-boundary lengthening can vary in the absence of syntactic differences. For instance, compared to longer subjects, shorter subjects are less likely to be accompanied by boundary-related phonetic cues in production (pre-boundary lengthening and subsequent pause; Gee & Grosjean, 1983; Watson & Gibson, 2004) and they do not elicit a CPS in comprehension (Hwang & Steinhauer, 2011). Rather than directly reflecting syntax, it appears that the magnitude of pre-boundary lengthening effects is related to the perceived strength of prosodic boundaries (Ferreira, 1993; Price et al., 1991; Wightman et al., 1992). Thus, syntactic structure affects phonetic variation indirectly, through intermediate prosodic structure.

It has repeatedly been observed that prosodic structure reflects syntactic structure in important ways, but is not isomorphic to it (Ferreira, 1993; Féry, 2016; Gee & Grosjean, 1983; Nespor & Vogel, 1986; Selkirk, 1984). Several arguments have been presented for a prosodic structure representation that is distinct from syntactic structure (Selkirk, 2011; Shattuck-Hufnagel & Turk, 1996). First, the two have different formal properties. Compared to syntactic hierarchies, which are deeply embedded and fundamentally asymmetrical, prosodic structure is symmetrical and rather flat (Gee & Grosjean, 1983; Nespor & Vogel, 1986; Selkirk, 1978, 1984). Second, prosodic constituents may systematically deviate from syntactic constituents. For instance, the

sentences “John is eager to please” and “John is easy to please” are superficially similar and, when produced naturally, receive the same analysis in terms of prosodic constituency. However, their underlying syntactic structures are very different: “John” is the subject of “please” in the first sentence, but its object in the second sentence. Moreover, prosodically well-formed sequences may be syntactically unacceptable, as in the ungrammatical “John is easy to please Mary”. Note again that the prosodically similar “John is eager to please Mary” is fully acceptable, showing that syntactic deviance need not be reflected in prosodic deviance. Third, and conversely, there is considerable variability in the prosodic realization of one and the same syntactic structure. Non-syntactic factors such as speech rate, semantic coherence, discourse status, and independent phonological well-formedness constraints all affect prosodic structuring (Ferreira, 1993; Frazier et al., 2004; Gee & Grosjean, 1983; Nespor & Vogel, 1986; Selkirk, 1984, 2011; Watson & Gibson, 2004).

One such non-syntactic factor that has received considerable attention is the same-size-sister constraint, which reflects speakers’ tendency to place boundaries at locations such that the prosodic phrases on both sides of the boundary are roughly the same weight and length (Cooper & Paccia-Cooper, 1980; Fodor, 1998; Gee & Grosjean, 1983). This preference for balance results in symmetry in the prosodic hierarchy (see Figure 1B), contrasting with the asymmetrical structure of syntactic hierarchy (Figure 1A). A well-known case is presented by recursively embedded clauses, such as “this is the cat that chased the rat that ate the cheese”. The syntactic structure of this sentence is asymmetrical and deeply nested, as shown by the number of closing brackets at the end of [this [is [the cat [that chased [the rat [that ate [the cheese]]]]]]]. When produced naturally, however, the intonational bracketing is rather flat, roughly corresponding to “this is the cat // that chased the rat // that ate the cheese”, with all *i*-phrases being approximately the same size, and with none of the *i*-breaks corresponding to clause boundaries (Chomsky & Halle, 1968). That prosodic boundaries might occur at places that are not major syntactic boundaries is also seen in Figure 1B, where the subject and the finite verb in the main clause form a prosodic constituent (the ω -word “he saw”) to the exclusion of the object (“a cat chase a mouse”), which is not consistent with the syntactic constituency analysis (Figure 1A).

To sum up, syntax and prosody are systematically related, but their structures are not isomorphic. According to Féry (2016), the relation of syntax to prosody is one of homomorphism. Homomorphic maps are structure-preserving, but in contrast to

isomorphisms they need not be one-to-one, which means that the inverse relation is not necessarily structure-preserving. In other words, the structure of syntax is preserved in prosodic structure, but because of the smaller number of constituents in the prosodic structure, not all syntactic details are retained in the map.³ Syntactic structure can therefore not uniquely be derived from prosodic patterns. This would explain why prosodic boundaries nearly always index syntactic boundaries, while many syntactic boundaries are not prosodically marked.

3.2 Syntax-prosody interactions in language processing

Given that the boundaries of prosodic structure tend to align with syntactic phrase boundaries, prosodic information can be used as an informative cue during syntactic processing, and conversely, knowledge of syntax can be used to make inferences about prosodic structure (for reviews, see Cutler et al., 1997; Wagner & Watson, 2010). Concerning the former, behavioral studies with syntactically ambiguous sentences show that people use prosodic boundary cues to resolve temporary syntactic ambiguities in comprehension and production (Kjelgaard & Speer, 1999; Marslen-Wilson et al., 1992; Millotte et al., 2007; Schafer et al., 2000; Speer et al., 1996). For instance, in a constrained production experiment by Schafer et al. (2000), participants had to produce sentences with early vs. late closure ambiguities, as in the examples below. In the late closure analysis of the subordinate clause, “the square” is the direct object of “moves” (as in (2)), whereas it is the subject of the main clause when the subordinate clause is closed early (as in (1)). Speakers consistently placed an i-boundary at the subordinate clause boundary, whose location varied depending on the meaning the speakers wanted to convey.

- 1.) When that triangle moves // the square will ...
- 2.) When that triangle moves the square // it

³ This is a description of the formal relation between the two systems, which, for the purpose of scientific explanation, relies on the assumption that prosodic structure is affected by syntax only. As such, the map might be described as a surjective homomorphism, in which certain syntactic properties are omitted from prosodic structure because the latter is less detailed. As discussed, however, the prosodic structuring of an utterance is also influenced by non-syntactic principles, such as phonological well-formedness constraints. When these interact with constraints on the syntax-prosody correspondence (in the optimality-theoretic sense; e.g., Féry, 2016, Selkirk, 2011), the prosodic structure corresponding to an utterance can deviate from its syntactic structure (e.g., the prosodic boundaries do not index syntactic boundaries). In all, then, the mapping between syntax and prosody will be descriptively many-to-many.

In a subsequent comprehension experiment, it was found that listeners could use the speakers' prosodic phrasing to constrain their syntactic analysis (Schafer et al., 2000). The same results are reported for globally ambiguous sentences, which show that speakers and listeners typically avoid attaching two syntactic constituents that are separated by a prosodic break. In a sentence like "when you learn gradually you worry more", people use prosodic boundaries to disambiguate the intended reading (here, left- vs. right-attachment of "gradually"), both in speaking and in listening (Kraljic & Brennan, 2005; Price et al., 1991; Snedeker & Trueswell, 2003).

Similar effects of prosodic phrasing on the interpretation of locally ambiguous sentences are reported in the ERP literature. These effects reflect not only the processing of the prosodic boundary itself (see Section 2.2), but also its downstream consequences for sentence interpretation. Consider the sentences in (3) and (4), which are identical up to the embedded verb (Bögels et al., 2010). Depending on the argument structure of this verb, *de soldaat* "the soldier" is either the object in the embedded sentence (in (4), when the embedded verb is the transitive *vermoorden* "to kill") or the direct object of the main verb *beval* "ordered" (in (3), when the embedded verb is the intransitive *vuren* "to fire"). If the sentence is presented without overt prosodic boundaries and is truncated before the embedded verb, people have a preference for the intransitive reading.

3.) De commandant beval // de soldaat te vuren en ...

The commander ordered // the soldier to fire and ...

4.) De commandant beval // de soldaat te vermoorden en...

The commander ordered // to kill the soldier and...

When these sentences are presented with the intonational phrasing indicated, the i-boundary after the verb *beval* "ordered" elicits a CPS in the ERP signal, indexing the closure of a prosodic phrase. Moreover, because people are biased against integrating information across a prosodic boundary, the presence of this i-boundary creates the expectation that the upcoming words will form a constituent. This means that *de soldaat* "the soldier" is initially not interpreted as belonging to *beval* "ordered", but rather as the object of the upcoming verb, which is therefore expected to be transitive (as in (4)). Indeed, the disambiguating intransitive verb in (3) elicited an increased

N400, reflecting processing difficulty associated with a verb whose argument structure violates a prosody-induced syntactic expectation (Bögels et al., 2010; Friederici et al., 2007; Steinhauer et al., 1999). Corroborating the behavioral data, these ERP results show that prosodic information can determine people's syntactic analyses of locally ambiguous sentences by completely reversing their default parsing preferences (see also Henke & Meyer, 2021).

It is important to note that prosody can only disambiguate two interpretations of the same sentence if their constituent boundaries are located at different places (Nespor & Vogel, 1986). An ambiguous sentence like “flying planes can be dangerous” is difficult to disambiguate prosodically because the words that cause the ambiguity – “flying planes” – form a syntactic constituent in both meanings. Even for structurally ambiguous sentences, however, a local prosodic boundary is not an unambiguous cue to syntactic closure or attachment. Rather, what matters is the global prosodic representation, in which the informativeness of a prosodic boundary is evaluated relative to other prosodic and phonological cues, including the strength of other boundaries and the lengths of the constituents it flanks. That is, a ϕ -boundary is more decisive syntactically when it is the only prosodic cue than when it is preceded by a phonologically stronger ι -boundary (Carlson et al., 2001; Clifton et al., 2002). And prosodic boundaries are perceived to be more informative about the intended structure of a sentence when they flank short constituents than when they flank long constituents (Clifton et al., 2006). Other things being equal, the probability of a prosodic boundary increases when the constituents surrounding it are longer (Ferreira, 1993; Frazier et al., 2004; Gee & Grosjean, 1983; Hwang & Steinhauer, 2011; Watson & Gibson, 2004). Thus, when the surrounding constituents are short, the prosodic boundary is not justified by constituent length, so listeners assume it to be driven by syntactic structure (Clifton et al., 2006). In accordance with the non-isomorphism between prosodic and syntactic structure, prosodic phrasing cues syntactic decomposition probabilistically, not diagnostically.

Concerning the effect of syntax on prosodic parsing, it has been shown that syntactic constituency guides the perception of prosody. Cole et al. (2010) asked untrained listeners to prosodically transcribe naturalistic speech and found that syntactic context made an independent contribution to the perception of prosodic boundaries. Similarly, in a prosodic boundary detection task with spoken sentences, Buxó-Lugo and Watson (2016) found that the probability of boundary marking was

higher for syntactically licensed locations than for non-licensed locations, even when the acoustic properties of the boundaries were experimentally controlled. These two studies show that prosodic boundaries are perceptually more salient when they coincide with the boundaries of syntactic constituents. This is in line with acquisition studies showing that children are sensitive to the acoustic correlates of syntactic structure. Very young infants prefer speech in which artificially inserted pauses coincide with syntactic phrase boundaries compared to speech in which the pauses are inserted in the middle of phrases (Hirsh-Pasek et al., 1987; Jusczyk et al., 1992; Kemler Nelson et al., 1989). What is more, infants use their sensitivity to the prosodic marking of syntactic units to recognize these units in natural speech. After being familiarized with phonological word sequences that were spoken both as a prosodically well-formed syntactic constituent and as a syntactic non-constituent (i.e., the prosodic structuring indexed syntactic constituency), infants prefer listening to passages in which the sequences constitute a constituent over passages in which the sequences cross a phrase boundary and form a non-constituent (Soderstrom et al., 2003). Prosodic well-formedness thus facilitates the identification and recognition of syntactic units in continuous speech.

There might be a role for delta tracking in modulating this relationship between syntactic and prosodic grouping. Infants' preference for coinciding syntactic and prosodic boundaries holds more strongly for child-directed speech than for adult-directed speech (Kemler Nelson et al., 1989). A possible reason is that child-directed speech is characterized by strong delta-band regularities (i.e., enhanced amplitude modulation and rhythmic regularity; Leong et al., 2017), which support the perceptual inference of higher-level linguistic structure. Indeed, it has been observed in young infants that speech-brain coupling at the prosodic stress rate is higher for child-directed speech than for adult-directed speech (Menn et al., 2022). Given that prosodic boundaries very often coincide with syntactic boundaries, the infant brain might be able to bootstrap its syntactic acquisition by neurally tracking the prosodic delta-band modulations that are prominent in child-directed speech (Jusczyk et al., 1992; Männel & Friederici, 2011; Soderstrom et al., 2003).

4. Integrating prosodic and syntactic accounts of neural tracking effects

In the preceding sections, we have given several reasons for why we should not attempt to explain syntactic tracking effects in prosodic terms (or vice versa). Our arguments are logical (i.e., due to the absence of one-to-one mappings, neural signals do not uniquely index cognitive events; see Section 2), empirical (i.e., the neural correlates of prosodic processing are modulated by syntax; see Section 2.2), and ontological (i.e., prosodic and syntactic constituency are ontologically related, Section 3). In this last section, we will offer a fourth argument, which is more strategic. That is, because neither prosody nor syntax ever appears in isolation in natural language, it makes sense if neurobiological investigations of language attempt to determine how the interaction between prosody and syntax, both in the stimulus and in the brain, facilitates processing. This aligns naturally with what we perceive to be the goal of the neurobiology of language, that is, to explain how the interplay between different sources of linguistic information ultimately yields comprehension via neurobiological mechanisms.

By presenting this strategic argument in favor of an integrative approach, we do not mean to say that a disjunctive approach, in which it is investigated whether and to what extent single cognitive factors modulate neural processing, should not be pursued. On the contrary, this approach is required to determine whether the brain cares about a given cognitive feature in the first place, so it naturally precedes the integrative approach. However, when its results are interpreted, it is important to keep in mind that in natural language processing the brain never encounters that feature in isolation. It should therefore be acknowledged that single-component explanations are only part of the story, and that a conjunctive understanding is ultimately desired. Moreover, when researchers try to isolate the contribution of one factor (e.g., prosody, syntax) as part of the disjunctive approach, it is important that they take adequate caution in their experimental designs by establishing proper baselines for neural responses that are amenable to a high-level linguistic explanation. This is particularly important when one aims to dissociate the neural contributions of two factors that are highly correlated. It can be done by including control conditions that only differ from the experimental condition in the variable of interest (e.g., natural sentences vs. prosodically matched pseudoword sentences; Coopmans et al., 2022a; Kaufeld et al., 2020), or by including prosodic control predictors in encoding models of language-related brain activity (Slaats et al., 2023). Properly accounting for prosodic variance

will be helpful in detecting and understanding the remaining effects whose variance in neural dynamics is more directly attributable to syntactic processing (i.e., ‘true’ syntactic tracking effects).

In line with this disjunctive approach, many recent studies have tried to dissociate different linguistic factors in order to account for the linguistic properties underlying tracking effects. This approach has been very effective and successful, but such functional decompositions are not the end result when the ultimate aim is to explain how language is represented in the brain. While syntax and prosody comprise formally distinct systems (Beckmann & Pierrehumbert, 1986; Féry, 2016; Ferreira, 1993; Nespor & Vogel, 1986; Selkirk, 1978, 1984, 2011; Shattuck-Hufnagel & Turk, 1996), it is unlikely that the brain strictly distinguishes them during natural language comprehension, precisely because they are so strongly tied together. Thus, beyond trying to attribute tracking effects exclusively to one linguistic component, or looking at the remaining variance to be explained by that one component, we consider it critical to also think about ways in which the interaction between the components (linguistic or otherwise) drives the neural signal. When it comes to syntax and prosody, this type of interactive approach is common both in theoretical linguistics (e.g., Féry, 2016; Nespor & Vogel, 1983; Selkirk, 1984, 2011) and psycholinguistics (e.g., Bögels et al., 2010; Buxó-Lugo & Watson, 2016; Coopmans et al., 2022b; Friederici et al., 2007; Hwang & Steinhauer, 2011; Itzhak et al., 2010; Kerkhofs et al., 2007; Kjelgaard & Speer, 1999; Steinhauer et al., 1999), but the literature on neural speech tracking remains predominantly focused on single-component explanations.

With the advanced analysis techniques that have become available in recent years, it is now possible to study of the neural basis of language comprehension in naturalistic contexts (e.g., audiobook listening), where the interaction between, and co-occurrence of, syntactic and prosodic information is natural and commonplace. In a recent MEG decoding study, Degano et al. (2023) utilized this opportunity and found that the neural encoding of syntactic information in natural speech is boosted by prosody. This important result indicates that the alignment between syntax and prosody yields a syntactic representational gain in the neural signal (Degano et al., 2023), and more generally, it shows the promise of the interactive approach. The time is ripe for cognitive neuroscientists to investigate how the brain interprets and constructs spoken language by (de)composing the mutually constraining rhythms of syntactic and prosodic structures. But while doing so, it is important to keep in mind that neural

tracking of a given variable, be it physical or linguistic, may only be the tip of the iceberg in terms of the infrastructure for language comprehension in the brain.

Acknowledgements

We are grateful to Hatice Zora and two anonymous reviewers for valuable comments on an earlier draft of this chapter. Andrea E. Martin was supported by an Independent Max Planck Research Group and a Lise Meitner Research Group “Language and Computation in Neural Systems”, by NWO Vidi grant 016.Vidi.188.029 to AEM, and by Big Question 5 (to Prof. dr. Roshan Cools & Dr. Andrea E. Martin) of the Language in Interaction Consortium funded by NWO Gravitation Grant 024.001.006 to Prof. dr. Peter Hagoort. Cas W. Coopmans was supported by NWO Vidi grant 016.Vidi.188.029 to AEM.

References

- Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, 3, 255–309.
- Blanco-Elorrieta, E., Ding, N., Pytkänen, L., & Poeppel, D. (2020). Understanding requires tracking: Noise and knowledge interact in bilingual comprehension. *Journal of Cognitive Neuroscience*, 32(10), 1975–1983.
https://doi.org/10.1162/jocn_a_01610
- Bögels, S., Schriefers, H., Vonk, W., Chwilla, D. J., & Kerkhofs, R. (2010). The Interplay between Prosody and Syntax in Sentence Processing: The Case of Subject- and Object-control Verbs. *Journal of Cognitive Neuroscience*, 22(5), 1036–1053. <https://doi.org/10.1162/jocn.2009.21269>
- Boucher, V. J., Gilbert, A. C., & Jemel, B. (2019). The Role of Low-frequency Neural Oscillations in Speech Processing: Revisiting Delta Entrainment. *Journal of Cognitive Neuroscience*, 31(8), 1205–1215.
https://doi.org/10.1162/jocn_a_01410
- Bourguignon, M., Tiège, X. D., Beeck, M. O. de, Ligot, N., Paquier, P., Bogaert, P. V., Goldman, S., Hari, R., & Jousmäki, V. (2013). The pace of prosodic phrasing couples the listener's cortex to the reader's voice. *Human Brain Mapping*, 34(2), 314–326. <https://doi.org/10.1002/hbm.21442>
- Breen, M. (2014). Empirical Investigations of the Role of Implicit Prosody in Sentence Processing. *Language and Linguistics Compass*, 8(2), 37–50.
<https://doi.org/10.1111/lnc3.12061>
- Burroughs, A., Kazanina, N., & Houghton, C. (2021). Grammatical category and the neural processing of phrases. *Scientific Reports*, 11(1).
<https://doi.org/10.1038/s41598-021-81901-5>
- Buxó-Lugo, A., & Watson, D. G. (2016). Evidence for the influence of syntax on prosodic parsing. *Journal of Memory and Language*, 90, 1–13.
<https://doi.org/10.1016/j.jml.2016.03.001>
- Carlson, K., Clifton, C., & Frazier, L. (2001). Prosodic Boundaries in Adjunct Attachment. *Journal of Memory and Language*, 45(1), 58–81.
<https://doi.org/10.1006/jmla.2000.2762>
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row.

- Clifton, C., Carlson, K., & Frazier, L. (2002). Informative Prosodic Boundaries. *Language and Speech*, 45(2), 87–114.
<https://doi.org/10.1177/00238309020450020101>
- Clifton, C., Carlson, K., & Frazier, L. (2006). Tracking the what and why of speakers' choices: Prosodic boundaries and the length of constituents. *Psychonomic Bulletin & Review*, 13(5), 854–861. <https://doi.org/10.3758/BF03194009>
- Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience*, 30(1–2), 1–31. <https://doi.org/10.1080/23273798.2014.963130>
- Cole, J., Mo, Y., & Baek, S. (2010). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes*, 25(7–9), 1141–1177.
<https://doi.org/10.1080/01690960903525507>
- Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and Speech*. Harvard University Press.
- Coopmans, C. W., de Hoop, H., Hagoort, P., & Martin, A. E. (2022a). Effects of structure and meaning on cortical tracking of linguistic units in naturalistic speech. *Neurobiology of Language*, 3(3), 386–412.
https://doi.org/10.1162/nol_a_00070
- Coopmans, C. W., Struiksma, M. E., Coopmans, P. H. A., & Chen, A. (2022b). Processing of Grammatical Agreement in the Face of Variation in Lexical Stress: A Mismatch Negativity Study. *Language and Speech*, 66(1), 202–213.
<https://doi.org/10.1177/00238309221098116>
- Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the Comprehension of Spoken Language: A Literature Review. *Language and Speech*, 40(2), 141–201. <https://doi.org/10.1177/002383099704000203>
- Degano, G., Donhauser, P. W., Gwilliams, L., Merlo, P., & Golestani, N. (2023). *Speech prosody enhances the neural processing of syntax*. bioRxiv.
<https://doi.org/10.1101/2023.07.03.547482>
- Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, 19(1), 158–164. <https://doi.org/10.1038/nn.4186>
- Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, 100, 233–253. <https://doi.org/10.1037/0033-295X.100.2.233>

- Féry, C. (Ed.). (2016). Intonation and Syntax: The Higher-Level Prosodic Constituents. In *Intonation and Prosodic Structure* (pp. 59–93). Cambridge University Press. <https://doi.org/10.1017/9781139022064.005>
- Fodor, J. D. (1998). Learning To Parse? *Journal of Psycholinguistic Research*, 27(2), 285–319. <https://doi.org/10.1023/A:1023258301588>
- Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. *Proceedings of Speech Prosody*, 83–90.
- Frank, S. L., & Yang, J. (2018). Lexical representation explains cortical entrainment during speech comprehension. *PLOS ONE*, 13(5), e0197304. <https://doi.org/10.1371/journal.pone.0197304>
- Frazier, L., Clifton, C., & Carlson, K. (2004). Don't break, or do: Prosodic boundary preferences. *Lingua*, 114(1), 3–27. [https://doi.org/10.1016/S0024-3841\(03\)00044-5](https://doi.org/10.1016/S0024-3841(03)00044-5)
- Friederici, A. D., von Cramon, D. Y., & Kotz, S. A. (2007). Role of the Corpus Callosum in Speech Comprehension: Interfacing Syntax and Prosody. *Neuron*, 53(1), 135–145. <https://doi.org/10.1016/j.neuron.2006.11.020>
- Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15(4), 411–458. [https://doi.org/10.1016/0010-0285\(83\)90014-2](https://doi.org/10.1016/0010-0285(83)90014-2)
- Ghitza, O. (2017). Acoustic-driven delta rhythms as prosodic markers. *Language, Cognition and Neuroscience*, 32(5), 545–561. <https://doi.org/10.1080/23273798.2016.1232419>
- Gilbert, A. C., Boucher, V. J., & Jemel, B. (2015). The perceptual chunking of speech: A demonstration using ERPs. *Brain Research*, 1603, 101–113. <https://doi.org/10.1016/j.brainres.2015.01.032>
- Glushko, A., Poeppel, D., & Steinhauer, K. (2022). Overt and implicit prosody contribute to neurophysiological responses previously attributed to grammatical processing. *Scientific Reports*, 12(1), Article 1. <https://doi.org/10.1038/s41598-022-18162-3>
- Henin, S., Turk-Browne, N. B., Friedman, D., Liu, A., Dugan, P., Flinker, A., Doyle, W., Devinsky, O., & Melloni, L. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. *Science Advances*, 7(8), eabc4530. <https://doi.org/10.1126/sciadv.abc4530>

- Henke, L., & Meyer, L. (2021). Endogenous oscillations time-constrain linguistic segmentation: Cycling the garden path. *Cerebral Cortex*, 31(9), 4289–4299. <https://doi.org/10.1093/cercor/bhab086>
- Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, 26(3), 269–286. [https://doi.org/10.1016/S0010-0277\(87\)80002-1](https://doi.org/10.1016/S0010-0277(87)80002-1)
- Hwang, H., & Steinhauer, K. (2011). Phrase Length Matters: The Interplay between Implicit Prosody and Syntax in Korean “Garden Path” Sentences. *Journal of Cognitive Neuroscience*, 23(11), 3555–3575. https://doi.org/10.1162/jocn_a_00001
- Inbar, M., Genzer, S., Perry, A., Grossman, E., & Landau, A. N. (2023). Intonation Units in spontaneous speech evoke a neural response. *The Journal of Neuroscience*. <https://doi.org/10.1523/JNEUROSCI.0235-23.2023>
- Inbar, M., Grossman, E., & Landau, A. N. (2020). Sequences of Intonation Units form a ~ 1 Hz rhythm. *Scientific Reports*, 10(1), Article 1. <https://doi.org/10.1038/s41598-020-72739-4>
- Itzhak, I., Pauker, E., Drury, J. E., Baum, S. R., & Steinhauer, K. (2010). Event-related potentials show online influence of lexical biases on prosodic processing. *NeuroReport*, 21(1), 8. <https://doi.org/10.1097/WNR.0b013e328330251d>
- Jin, P., Lu, Y., & Ding, N. (2020). Low-frequency neural activity reflects rule-based chunking during speech listening. *ELife*, 9, e55613. <https://doi.org/10.7554/eLife.55613>
- Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L. J., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, 24(2), 252–293. [https://doi.org/10.1016/0010-0285\(92\)90009-Q](https://doi.org/10.1016/0010-0285(92)90009-Q)
- Kaufeld, G., Bosker, H. R., ten Oever, S., Alday, P. M., Meyer, A. S., & Martin, A. E. (2020). Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *The Journal of Neuroscience*, 40(49), 9467–9475. <https://doi.org/10.1523/JNEUROSCI.0302-20.2020>
- Keitel, A., Gross, J., & Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLOS Biology*, 16(3), e2004473. <https://doi.org/10.1371/journal.pbio.2004473>

- Kerkhofs, R., Vonk, W., Schriefers, H., & Chwilla, D. J. (2007). Discourse, Syntax, and Prosody: The Brain Reveals an Immediate Interaction. *Journal of Cognitive Neuroscience*, 19(9), 1421–1434.
<https://doi.org/10.1162/jocn.2007.19.9.1421>
- Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic Facilitation and Interference in the Resolution of Temporary Syntactic Closure Ambiguity. *Journal of Memory and Language*, 40(2), 153–194. <https://doi.org/10.1006/jmla.1998.2620>
- Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology*, 50(2), 194–231.
<https://doi.org/10.1016/j.cogpsych.2004.08.002>
- Leong, V., Kalashnikova, M., Burnham, D., & Goswami, U. (2017). The Temporal Modulation Structure of Infant-Directed Speech. *Open Mind*, 1(2), 78–90.
https://doi.org/10.1162/OPMI_a_00008
- 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*, 37(32), 7772–7781.
<https://doi.org/10.1523/JNEUROSCI.0168-17.2017>
- Männel, C., & Friederici, A. D. (2011). Intonational phrase structure processing at different stages of syntax acquisition: ERP studies in 2-, 3-, and 6-year-old children. *Developmental Science*, 14(4), 786–798.
<https://doi.org/10.1111/j.1467-7687.2010.01025.x>
- Martin, A. E. (2016). Language processing as cue integration: Grounding the psychology of language in perception and neurophysiology. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00120>
- Martin, A. E. (2020). A compositional neural architecture for language. *Journal of Cognitive Neuroscience*, 32(8), 1407–1427.
https://doi.org/10.1162/jocn_a_01552
- Martin, A. E., & Dumas, L. A. A. (2017). A mechanism for the cortical computation of hierarchical linguistic structure. *PLOS Biology*, 15(3), e2000663.
<https://doi.org/10.1371/journal.pbio.2000663>
- Martin, A. E., & Dumas, L. A. A. (2019). Predicate learning in neural systems: Using oscillations to discover latent structure. *Current Opinion in Behavioral Sciences*, 29, 77–83. <https://doi.org/10.1016/j.cobeha.2019.04.008>

- Mehler, J., Morton, J., & Jusczyk, P. W. (1984). On reducing language to biology. *Cognitive Neuropsychology*, 1(1), 83–116.
<https://doi.org/10.1080/02643298408252017>
- Menn, K. H., Michel, C., Meyer, L., Hoehl, S., & Männel, C. (2022). Natural infant-directed speech facilitates neural tracking of prosody. *NeuroImage*, 251, 118991. <https://doi.org/10.1016/j.neuroimage.2022.118991>
- Meyer, L., Henry, M. J., Gaston, P., Schmuck, N., & Friederici, A. D. (2017). Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cerebral Cortex*, 27(9), 4293–4302.
<https://doi.org/10.1093/cercor/bhw228>
- Meyer, L., Sun, Y., & Martin, A. E. (2020). Synchronous, but not entrained: Exogenous and endogenous cortical rhythms of speech and language processing. *Language, Cognition and Neuroscience*, 35(9), 1089–1099.
<https://doi.org/10.1080/23273798.2019.1693050>
- Millotte, S., Wales, R., & Christophe, A. (2007). Phrasal prosody disambiguates syntax. *Language and Cognitive Processes*, 22(6), 898–909.
<https://doi.org/10.1080/01690960701205286>
- Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, 16(1), 55–68. <https://doi.org/10.1017/S030500090001343X>
- Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Foris.
- Oganian, Y., Kojima, K., Breska, A., Cai, C., Findlay, A., Chang, E., & Nagarajan, S. S. (2023). Phase Alignment of Low-Frequency Neural Activity to the Amplitude Envelope of Speech Reflects Evoked Responses to Acoustic Edges, Not Oscillatory Entrainment. *Journal of Neuroscience*, 43(21), 3909–3921.
<https://doi.org/10.1523/JNEUROSCI.1663-22.2023>
- Pannekamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A. D. (2005). Prosody-driven Sentence Processing: An Event-related Brain Potential Study. *Journal of Cognitive Neuroscience*, 17(3), 407–421.
<https://doi.org/10.1162/0898929053279450>
- Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America*, 90(6), 2956–2970. <https://doi.org/10.1121/1.401770>

- Rimmele, J. M., Poeppel, D., & Ghitza, O. (2021). Acoustically driven cortical delta oscillations underpin prosodic chunking. *ENeuro*, 8(4), 1–15.
<https://doi.org/10.1523/ENEURO.0562-20.2021>
- Roll, M., Lindgren, M., Alter, K., & Horne, M. (2012). Time-driven effects on parsing during reading. *Brain and Language*, 121(3), 267–272.
<https://doi.org/10.1016/j.bandl.2012.03.002>
- Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational Disambiguation in Sentence Production and Comprehension. *Journal of Psycholinguistic Research*, 29(2), 169–182.
<https://doi.org/10.1023/A:1005192911512>
- Schremm, A., Horne, M., & Roll, M. (2015). Brain responses to syntax constrained by time-driven implicit prosodic phrases. *Journal of Neurolinguistics*, 35, 68–84.
<https://doi.org/10.1016/j.jneuroling.2015.03.002>
- Selkirk, E. (2011). The syntax-phonology interface. In J. A. Goldschmit, J. J. Riggle, & A. C. L. Yu (Eds.), *The handbook of phonological theory* (pp. 435–484). Wiley-Blackwell.
- Selkirk, E. O. (1978). On Prosodic Structure and Its Relation to Syntactic Structure. *Indiana University Linguistics Club*.
- Selkirk, E. O. (1984). *Phonology and syntax: The relationship between sound and structure*. MIT press.
- Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25(2), 193–247. <https://doi.org/10.1007/BF01708572>
- Slaats, S., Weissbart, H., Schoffelen, J.-M., Meyer, A. S., & Martin, A. E. (2023). Delta-Band Neural Responses to Individual Words Are Modulated by Sentence Processing. *Journal of Neuroscience*, 43(26), 4867–4883.
<https://doi.org/10.1523/JNEUROSCI.0964-22.2023>
- Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, 48(1), 103–130. [https://doi.org/10.1016/S0749-596X\(02\)00519-3](https://doi.org/10.1016/S0749-596X(02)00519-3)
- Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, 49(2), 249–267. [https://doi.org/10.1016/S0749-596X\(03\)00024-X](https://doi.org/10.1016/S0749-596X(03)00024-X)

- Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, 25(2), 249–271.
<https://doi.org/10.1007/BF01708573>
- Stehwien, S., & Meyer, L. (2022). Short-Term Periodicity of Prosodic Phrasing: Corpus-based Evidence. *Speech Prosody*, 693–698.
- Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, 2(2), Article 2. <https://doi.org/10.1038/5757>
- 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. <https://doi.org/10.1023/A:1010443001646>
- ten Oever, S., Carta, S., Kaufeld, G., & Martin, A. E. (2022). Neural tracking of phrases in spoken language comprehension is automatic and task-dependent. *ELife*, 11, e77468. <https://doi.org/10.7554/eLife.77468>
- Vetchinnikova, S., Konina, A., Williams, N., Mikušová, N., & Muraenen, A. (2023). Chunking up speech in real time: Linguistic predictors and cognitive constraints. *Language and Cognition*, 1–27.
<https://doi.org/10.1017/langcog.2023.8>
- Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, 25(7–9), 905–945.
<https://doi.org/10.1080/01690961003589492>
- Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19(6), 713–755. <https://doi.org/10.1080/01690960444000070>
- Westlin, C., Theriault, J. E., Katsumi, Y., Nieto-Castanon, A., Kucyi, A., Ruf, S. F., Brown, S. M., Pavel, M., Erdogmus, D., Brooks, D. H., Quigley, K. S., Whitfield-Gabrieli, S., & Barrett, L. F. (2023). Improving the study of brain-behavior relationships by revisiting basic assumptions. *Trends in Cognitive Sciences*, 27(3). <https://doi.org/10.1016/j.tics.2022.12.015>
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal*

Book chapter to be published in the forthcoming *Rhythms of Speech and Language: Culture, Cognition, and the Brain* edited by Lars Meyer and Antje Strauss (Cambridge University Press).

of the Acoustical Society of America, 91(3), 1707–1717.

<https://doi.org/10.1121/1.402450>