18 Cognitive and Neural Constraints on Timing and Rhythm in Language

LENA HENKE, ELEONORA BEIER, MIKAEL ROLL, AND LARS MEYER

18.1 Introduction

Language comprehension is guided by external acoustic cues but may also have an internal, implicit realization, influencing processing from the inside out – that is, endogenously. In particular, there are a number of endogenous constraints that influence the timing and pace of language processing. Some of these constraints may reflect cognitive and neural processing limitations, such as properties of working memory or even the electrophysiology of the brain as such. These context-invariant endogenous constraints interact with the current context of linguistic representations that are built during incremental language processing.

In this chapter, we first review behavioural evidence from psycholinguistics, describing the complex interaction of syntactic, prosodic, and information structures during the integration of words into larger multi-word units across substantial temporal intervals. In the second part, we summarize evidence from event-related brain potentials (ERPs) supporting the idea that temporal processing windows in the order of multiple seconds shape language production and comprehension endogenously. Lastly, we suggest that the repetitive pace of these integration windows may trigger a certain extent of rhythmicity in language behaviour. We discuss recent evidence that temporal electrophysiological constraints are driven by underlying endogenous oscillatory activity that acts as a pacemaker on language processing.

18.2 Timing Constraints on Linguistic Processing in Behaviour

The timing of linguistic processing is driven by the incremental computation of linguistic representations, including prosodic, syntactic, and information structures. As language comprehension unfolds over time, the online computation of these structures (e.g., "where" to place a boundary within a syntactic tree) may be influenced by contextual temporal constraints (e.g., "when" to expect a syntactic boundary given the preceding acoustic and abstract context). Here, we first focus on the temporal relationship between (implicit) prosody and syntax, and

https://doi.org/10.1017/9781009295888.022 Published online by Cambridge University Press

how they relate to the context-invariant constraints of working memory capacity. We then describe how syntax and prosody further interact with information structure to place contextual constraints on the timing of language processing.

There is a close relationship between syntactic and prosodic structure (for reviews, see Dahan, 2015; Dahan and Ferreira, 2019; for further discussion of the prosody–syntax interface and its interaction with informativity, see Chapter 19). While not strictly one to one, syntactic phrase boundaries often co-occur with prosodic boundaries in spoken language (Clifton et al., 2002; Ferreira, 1993; Selkirk, 2011). Thus, prosody is an important cue to syntax: Listeners rely on the perception of prosodic boundaries to infer syntactic structure, particularly in cases of syntactic ambiguity. For example, in the sentence *The friend of the moviestar who was sitting on the balcony was under investigation*, the relative clause (*who was sitting on the balcony*) can either be attached "high" in the syntactic tree (to the first noun, *the friend*) or "low" (to the second noun, *the moviestar*; see Figure 18.1). During incremental processing of the sentence, comprehenders must choose what the relative clause attaches to, thus arriving at different semantic interpretations of the sentence (whether it was *the friend* or *the moviestar* that was *on the balcony*). Studies suggest that this syntactic choice depends on cross-linguistic attachment preferences (e.g., a preference for low attachment in English, as opposed to high attachment in Spanish; Carreiras and Clifton, 1999) as well as constraints in working memory capacity, further discussed below (Swets et al., 2007; Traxler, 2009). Importantly, prosodic boundaries following either the first or second noun have been shown to influence syntactic attachment choices, generally leading to more high attachment when the boundary follows the second noun (Carlson et al., 2001; Clifton et al., 2002; Frazier et al., 2006).[1] However, syntactic choices are not simply driven by the presence or absence of a prosodic boundary or its acoustic realization (e.g., the absolute duration of a pause or pitch height of a boundary tone) but rather by the global prosodic structure over the entire sentence (e.g., the duration of a prosodic pause or height of a boundary tone relative to previous pauses and intonation patterns; Clifton et al., 2002). This suggests that during spoken language comprehension, listeners incrementally build a representation of a prosodic structure[2], which then influences the computation of a syntactic structure (parsing) by constraining the temporal location of syntactic boundaries. Thus, it is the relationship between the acoustic characteristics of prosodic boundaries over a long time window (the entire sentence), rather than the individual duration and pitch of each boundary, that is thought to play a critical role in syntactic processing.

[1] Similarly, Chapter 19 reviews evidence for prosodic disambiguation of temporarily ambiguous garden-path sentences, with faster speech tempo and shorter durational patterns signaling more embedded syntactic structures.

[2] Building a representation of prosodic structure itself depends on several timing constraints such as the principles of metrical and intonational phonology (Ladd, 2008; Liberman and Prince, 1977), lexical stress, syntactic and information structural constraints (Beier and Ferreira, 2018; Calhoun, 2010), as well as unplanned disfluencies (e.g., pauses due to hesitations; Ferreira, 2007). To reconcile these many constraints in the interpretation of prosodic structure (and to use it as a cue to syntax), listeners are thought to engage in Bayesian probabilistic reasoning, inferring the speaker's intended meaning by weighing the acoustic evidence (e.g., the duration of a pause) by its possible explanations (e.g., phonological or syntactic constraints, as opposed to an unplanned disfluency; following the rational speaker hypothesis; Dahan and Ferreira, 2019; Frazier et al., 2006).

https://doi.org/10.1017/9781009295888.022 Published online by Cambridge University Press

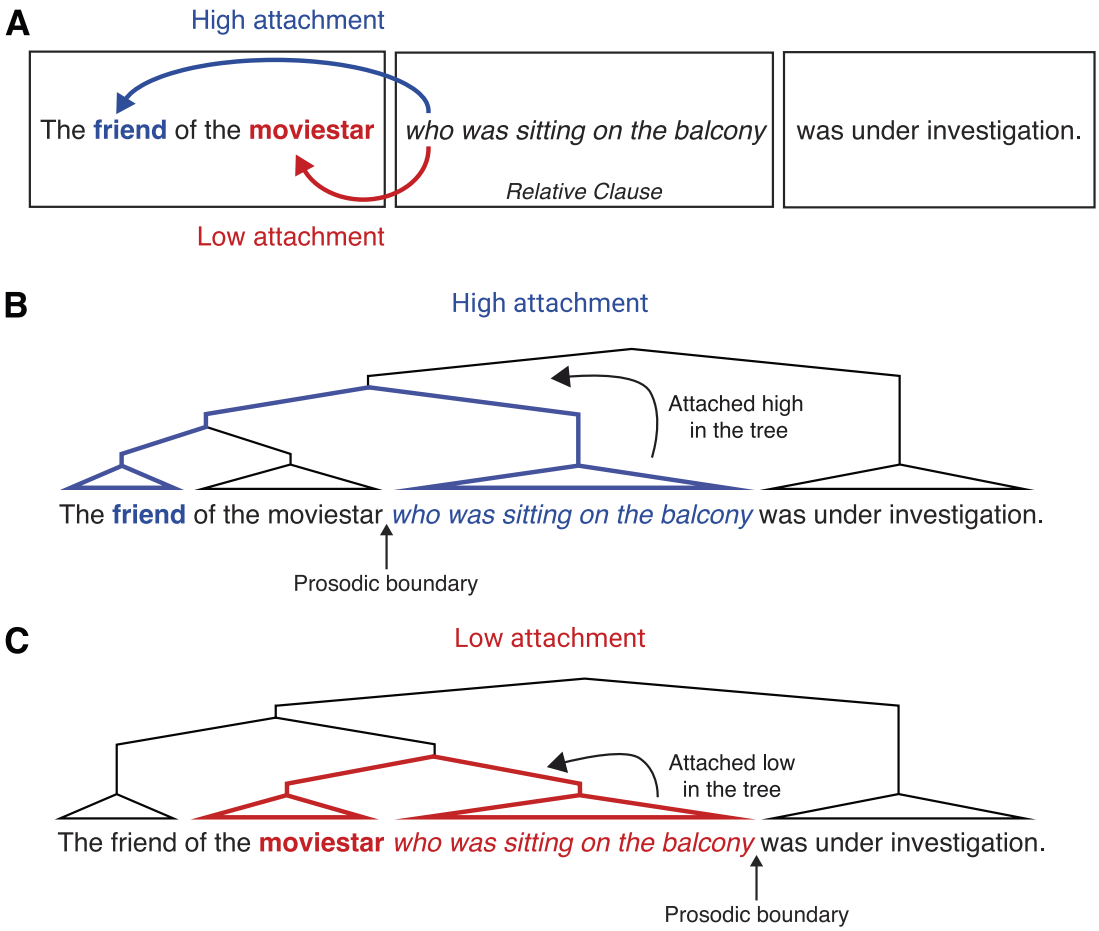


Figure 18.1 Relative clause attachment ambiguities.
(A) Illustration of the attachment ambiguity highlighting the two possible attachment sites for the relative clause. The relative clause may be attached high in the syntactic tree (bold) combining it with *the friend* (B) when there is a (implicit) prosodic boundary after *the moviestar*. Alternatively, it may be attached low in the syntactic tree (bold) such that it is *the moviestar* who *was sitting on the balcony* (C), when the boundary only occurs later in the sentence.

The influence of prosodic constraints on syntactic parsing is not limited to spoken language. Even in silent reading, comprehenders appear to make syntactic decisions based on an internal, implicit realization of prosodic structure (Fodor, 2002). For example, Swets et al. (2007) appeal to the concept of *implicit prosody* in order to explain their findings on the relationship between syntactic attachment preferences and working memory capacity. It was previously suggested that in cases of syntactic attachment ambiguities (as in the example above), comprehenders with low working memory span prefer low attachment, since the second noun is more recent and accessible than the first noun, which would need to be retrieved from working memory (the recency principle or late closure; Frazier, 1979; Traxler, 2007). However, Swets et al. found the

https://doi.org/10.1017/9781009295888.022 Published online by Cambridge University Press

opposite pattern: Readers with low working memory span were more likely to attach high compared to readers with high spans. In a following study, they tested whether implicit prosody mediates this chunking strategy. Readers with high spans may tend to chunk a larger amount of information together, combining the second noun with the relative clause in a single prosodic phrase; on the other hand, readers with low spans may chunk less information together, adding an implicit prosodic boundary between the second noun and the start of the relative clause. These two prosodic patterns would be consistent with a preference to attach low or high, respectively (see Figure 18.1B, C). Swets et al. (2007) provide evidence for this interpretation by presenting their sentences in chunks, forcing all readers to insert a boundary following the second noun, resulting in high attachment regardless of readers' working memory span. Later similar findings support this idea (Hemforth et al., 2015; Jun and Bishop, 2015; Traxler, 2009), and a large literature suggests that silent reading involves the computation of an implicit prosodic structure (for a review, see Breen, 2014). Thus, even in the absence of overt articulation, comprehenders are thought to compute an internal prosodic structure constrained by context-invariant individual differences in working memory capacity. In particular, working memory capacity may determine the amount of chunked information and therefore the duration of prosodic phrases.

While Swets et al.'s findings contradict initial assumptions about the role of working memory in syntactic ambiguity resolution, they support an indirect effect of individual differences in working memory capacity mediated by prosodic phrasing. This effect is consistent with the early "sausage-machine" model of parsing (Frazier and Fodor, 1978), which argues that syntactic parsing occurs over two stages: an initial stage that groups words into chunks of limited size (the Preliminary Phrase Packager), and a secondary stage that builds a full syntactic analysis based on these chunks (the Sentence Structure Supervisor). In this model, the size of the initial chunks was posited to be limited to approximately six words by working memory limitations; it follows that individual differences in working memory capacity would lead to differences in the size of these initial chunks.[3] In their discussion, Swets et al. (2007) speculate that the size of the chunks produced through this first stage of parsing (constrained by working memory) may be responsible for the introduction of implicit prosodic boundaries and, subsequently, syntactic attachment choices. We further elaborate on the context-invariant constraint of working memory capacity on the size and duration of syntactic phrases in the next sections.

One additional contextual constraint placed by prosody and syntax relates to their interaction with another important aspect of language: information structure. Information structure is a term from linguistic pragmatics that describes how sentences organize given information (= what the listener already knows) and new information (= what the listener does not know yet), within the context of a broader discourse or conversation. In many languages, sentences often present given information first and new information last. This *given-before-new* strategy reflects greater accessibility of given information during production, and is thought to facilitate comprehension by helping the listener retrieve concepts from long-term memory before appending new information (Bock and Irwin, 1980; Christianson and Ferreira, 2005;

[3] In subsequent work (Fodor, 1998), it was proposed that the duration / length constraint operates on the level of prosody rather than syntax.

https://doi.org/10.1017/9781009295888.022 Published online by Cambridge University Press

Ferreira and Yoshita, 2003; Haviland and Clark, 1974). However, speakers can deviate from this default information structure, cueing listeners towards focused, important information through prosodic prominence (e.g., *JOHN ate the sandwich*) or the use of specific syntactic structures (e.g., *It was John who ate the sandwich*; Carlson et al., 2009; but see Vallduví, 2016, for debates on terminology). There are cross-linguistic differences in whether speakers focus important information primarily prosodically (e.g., English and Dutch; Breen et al., 2010; Nooteboom and Kruyt, 1987) or syntactically (e.g., Spanish and Italian; Samek-Lodovici, 2005; Zubizarreta, 1998), also leading listeners to differently weigh each of these focus cues to infer information structure across languages (Calhoun et al., 2021; Kember et al., 2019; Yan and Calhoun, 2020).

Inferring the upcoming information structure of a sentence as it unfolds places temporal constraints on all levels of language processing. Listeners are thought to predict the timing of important information so that attention and processing resources can be pre-allocated to it (Beier and Ferreira, 2022; Cutler and Fodor, 1979; Kristensen et al., 2013). In addition, information structure has been shown to interact with other key aspects of language comprehension, including semantic (Wang et al., 2009, 2011) and syntactic processing (Wang et al., 2012). Focused words are read for longer (Lowder and Gordon, 2015) and are remembered better (Birch and Garnsey, 1995), while comprehenders may be more likely to engage in "good-enough" processing for already given information (Ferreira and Lowder, 2016). However, more research is needed to establish how listeners use syntactic and prosodic focus cues to incrementally infer upcoming information structure and how this interacts with language comprehension over time.

In summary, as language unfolds over time, processing is constrained by the incremental computation of syntactic, prosodic, and information structures and the temporal relationship between them. During spoken language comprehension, acoustic features are weighed to infer a sentence's prosodic structure, which then guides syntactic choices. This occurs even in silent reading, where context-invariant individual differences in working memory capacity constrain the duration of (implicit) prosodic and syntactic phrases. Prosody and syntax further serve as cues to upcoming information structure, guiding the comprehender's attention towards the timing of important information and interacting with all levels of linguistic processing.

18.3 Time-Limited Integration Windows: Evidence from ERPs

In the previous section, we discussed early psycholinguistic work that suggested that phrases or constituents are constrained to a length of around six words (Frazier and Fodor, 1978). However, this constraint may not reflect a limitation in the number of words but rather a temporal constraint that limits the duration of linguistic structures within our processing capacity. For instance, studies of prosodic phrasing provide evidence that the duration of phrases plays an important role in the determination of boundaries (for a detailed discussion on prosodic phrasing as a cue for cognitive dynamics, see Chapter 15). Accordingly, prolonged phrase durations lead to a different overt prosodic phrasing and increase the likelihood for

https://doi.org/10.1017/9781009295888.022 Published online by Cambridge University Press

implicit prosodic boundaries (Hirose, 2003; Hwang and Schafer, 2009; Webman-Shafran and Fodor, 2016). Hence, temporal windows may shape the formal side of language production and perception.

The proposed six-word limit – assuming an average speech rate of approximately 150 words per minute (Tauroza and Allison, 1990) – results in multi-word units of about 2.4 seconds. Likewise, speech production data suggests that spoken language tends to be delivered in chunks of around 2.6 seconds (Vollrath et al., 1992). Notably, this slow time constant corresponds to the maximum duration of word forms held in working memory. Baddeley et al. (1975) argued that phonological working memory traces decay after 2–3 seconds. Sachs (1967, 1974) found that after more than 3 seconds of intervening speech, listeners had great difficulties identifying form-related (grammatical) but not semantic changes. She proposed that the form and syntactic details of sentences are clear and accessible in working memory for only a few seconds. During that short time window, however, the conveyed message is recoded into a more durable semantic representation.

Roll et al. (2012) hypothesized that intonational phrases tend to be uttered within time windows of 2–3 seconds to be parseable while the word forms are still active in working memory. A more specific time constant was obtained by letting speakers read sentences that could be phrased in different ways. Speakers increased their speech rate for longer phrases, seemingly trying to fit them within a time window of 2.7 seconds. This duration was taken as the ideal duration of intonational phrases. To determine whether prosodic phrasing automatically proceeds in windows of 2.7 seconds during perception while controlling for acoustic factors, silent reading has been investigated using ERPs. In auditory processing, intonational phrase boundaries yield a positive-going shift in listeners' ERPs called the *closure positive shift* (CPS; Steinhauer et al., 1999). A CPS has even been found in silent reading at commas, where readers close phrase boundaries (Steinhauer and Friederici, 2001). Roll et al. (2012) presented participants with sentences word by word to see if they would spontaneously integrate or chunk them into 2.7-second phrases. Sentences could be chunked in three different possible phrasings. Phrases could correspond to one, two, or three clauses. The sentence material was presented at different rates so that either one, two, or all three clauses of each sentence would fall within a time window of 2.7 seconds. In this way, a CPS was observed whenever a possible phrase boundary corresponded to a 2.7-second duration from a clause onset (see Figure 18.2). Due to the varying presentation rate, the phrase endings that elicited a CPS thus contained one, two, or three clauses. An additional control condition with an extra word in each clause allowed the researchers to investigate what happened when clauses exceeded the 2.7-second time window. For these 3.4-second-long clauses, there was a difference for participants with high and low working memory spans. Whereas the high-span group showed no difference between longer and shorter clauses, the low-span group exhibited a left parietal negativity for the longer compared to the shorter clauses, suggesting that stimuli exceeding the common time window were more taxing on working memory.

Implicit prosodic phrasing based on temporal processing windows of 2.7 seconds has also been found to influence syntactic parsing, in line with the behavioural evidence of working memory constraints on syntactic processing described above (e.g., Swets et al., 2007). When the time-induced prosodic phrasing pattern mismatches the expected syntactic parsing of a

https://doi.org/10.1017/9781009295888.022 Published online by Cambridge University Press

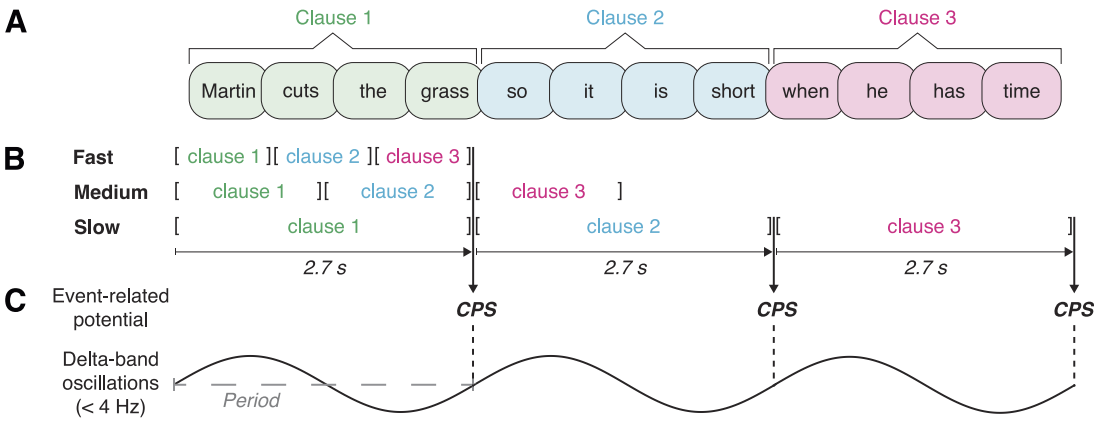


Figure 18.2 Presentation rate manipulations and their neural responses.
(A) Example sentence consisting of three clauses. (B) Sentences were presented word by word in three different presentation rates; following the experimental manipulation, one, two, or all three clauses of the sentence could fall into a temporal window of 2.7 seconds (adjusted from Roll et al., 2012). (C) ERP (i.e., CPS) at those clause boundaries that coincide with a duration of 2.7 seconds; regularity of the evoked response may reflect the period of low-frequency neural oscillatory activity.

sentence, an ERP related to syntactic reanalysis (the so-called P600; Hagoort et al., 1993) has been found, both in silent reading (Schremm et al., 2015) and speech processing (Henke and Meyer, 2021). Further evidence for a limited time window for the recoding of formal language features into semantics comes from an experiment on grammatical agreement (Roll et al., 2013). Two (dis)agreeing words, such as *We . . . run(s)*, were separated by four different time intervals based on increases in presentation rate. The temporal distances were below, at, or above 2.7 seconds. At the two faster rates, resulting in time intervals below 2.7 seconds, disagreeing elements produced the typical left anterior negativity (LAN) found for grammatical disagreement. This was interpreted as the processing of grammatical matching based on the form of the words. At a distance over 2.7 seconds, the negativity had a right-lateralized distribution, hinting at some other cognitive process being involved in matching the two elements, possibly semantically based. At 2.7 seconds, there was no effect of agreement, but participants' working memory span correlated with the negativity in amplitude at the electrodes displaying the LAN. In favour of the idea that agreement processing at distances over 2.7 seconds rather involves semantic representations, a sharp divide in response times has been found for acceptability judgments of grammatical agreement with distances above or below 2.7 seconds. Yet, for semantic matching, response times instead slowly increased with presentation rate (Schremm et al., 2016).

Overall, time-driven processing constraints of approximately 2–3 seconds may shape the integration of words into larger multi-word groups. In particular, they may guide the insertion of implicit prosodic boundaries during reading, which allows the parsing of words into a syntactic structure before word traces fade from working memory. Semantic processing of language seems to be differently affected by this temporal constraint.

https://doi.org/10.1017/9781009295888.022 Published online by Cambridge University Press

18.4 Neural Oscillations as Time Constraints?

Behavioural and ERP research both indicate a temporal constraint on the integration of words during language processing. Recently, it was hypothesized that this temporal limitation of processing windows discovered by the above behavioural and ERP work may stem from an underlying periodic electrophysiological substrate (Henke and Meyer, 2021; Meyer et al., 2016), so-called neural oscillations (cf. Chapter 3). Indeed, this work suggests that endogenous oscillatory activity may time-limit the duration of multi-word chunks, and it was also found to constrain acoustically driven chunking (Rimmele et al., 2021). Cycles of low-frequency delta-band oscillations (<4 Hz) have a period of up to several seconds and are thus long enough to accommodate the integration of words into larger multi-word units. In that way, delta-band oscillations may function as an endogenous pacemaker for language processing (Meyer et al., 2020a, 2020b).

Delta-band oscillations have been associated with the tracking of prosody (Bourguignon et al., 2013; Gross et al., 2013) but were also found to align with linguistic structure that is not apparent in the acoustic signal (Ding et al., 2016; Meyer et al., 2016). Using isochronous stimuli, neural activity within the delta frequency band was found to track the rates of phrases and sentences, although those could not be inferred from the acoustic input (Ding et al., 2016). Notably, tracking vanished when the phrasal rate fell out of the delta-band range (i.e., was faster; Lo, 2021). Thus, delta-band oscillations may have an endogenous tempo playing an active role in language processing.

In natural language, it has proven difficult to investigate an endogenous function of delta-band oscillations because speech contains physical rhythms; even subtle acoustic cues may mask endogenous activity. Therefore, it is necessary to consider a different sensory domain that avoids this confound. In particular, text input does not provide any temporal structure on the chunk level; although punctuation may guide the structuring of the text input, punctuation marks occur in space and not at specific time intervals. Thus, rhythmic sampling during reading would be endogenously imposed. Accordingly, Henke et al. (2023) examined simultaneous recordings of eye movements and electroencephalography (EEG) during naturalistic reading and observed that fixation durations fluctuate rhythmically at ~1 Hz in coherence with delta-band activity over the occipital cortex.[4] This suggests that neural activity may impose its preferred processing pace onto reading, independent of physical rhythms in the stimulus. Complementary work on self-paced reading has shown that the behavioural periodicity of reading times (here, ~2 Hz) is temporally aligned to chunk boundaries – as defined by a computational formalization based on dependency annotations and part-of-speech tags – as well as sentence boundaries (Lo et al., 2023). This alignment suggests that the observed periodicity is indeed related to the formation of multi-word chunks during language processing and may be behaviourally relevant to minimize memory demands. These findings on reading may also relate to psycholinguistic work on implicit prosodic phrasing (discussed in detail

[4] Activity was observed above the occipital cortex responsible for processing visual information. Yet, the limited spatial resolution of the EEG cannot rule out that this constraint may also stem from motor areas (for an example from speech processing, see Assaneo and Poeppel, 2018).

https://doi.org/10.1017/9781009295888.022 Published online by Cambridge University Press

above). Implicit prosody as a chunking strategy may in fact be a reflection of underlying neural processing windows. Yet, as opposed to previously proposed integration windows of 2.7 seconds, the observed frequencies of 1–2 Hz in reading suggest rather short time windows of 0.5–1 second. These shorter and longer time windows could relate to the multi-scale nature of prosody that comprises both intermediate as well as full prosodic phrases (Stehwien and Meyer, 2022). In particular, multiple short intermediate phrases may be integrated into longer prosodic phrases. Alternatively, the size of the integration windows may differ across modalities; for instance, the here reported results suggest shorter windows for visual as compared to auditory language processing. More work is needed to unravel the different time windows and their underlying neural implementation.

Returning to the longer integration windows and supporting the idea of an endogenous constraint of electrophysiological processing, the phase of delta-band oscillations was found to predict the termination of multi-word chunks (Meyer et al., 2016) and, more specifically, to enforce chunking after 2.7 seconds (Henke and Meyer, 2021). These studies employed ambiguous sentences similar to those used in the behavioural psycholinguistic studies described above. The results suggest that chunking, and as a consequence attachment, is affected by the phase of delta-band oscillations. Reconsidering the behavioural results from this perspective, it may be that the period of these oscillations subserves (implicit) prosodic phrasing (Glushko et al., 2022) and underlies the different chunking strategies of the high and low working memory span groups described above (i.e., the high-span group integrates more information together, while the low-span group terminates a chunk earlier; Swets et al., 2007). Note here that the duration of the chunking window for delta-band oscillations also aligns with findings on the CPS (e.g., Roll et al., 2012) described above. Based on this, it was suggested that the CPS may consist of underlying neural oscillations within the delta band (Meyer et al., 2016); the increased amplitude that is observed as ERP may arise from phase-locking of oscillatory activity across trials (Sauseng et al., 2007). Yet, to date, this remains to be shown (see Inbar et al., 2023, for the first evidence consistent with a relationship between the CPS, prosodic boundaries, and delta-band tracking during naturalistic speech listening).

As mentioned previously, delta-band oscillations may also be exogenously guided by prosody (see Chapter 15 for further discussion on exogenous versus endogenous influences within the delta band), aligning their phase to slow modulations in the acoustic speech signal (e.g., Bourguignon et al., 2013; Ding and Simon, 2014; Gross et al., 2013). Strikingly, oscillations were also found to inherit the exogenous stimulus frequency, so that the aligned oscillatory rhythm persists even after stimulus offset – a phenomenon called *entrainment* (Kösem et al., 2018; Luo and Poeppel, 2007; Obleser and Kayser, 2019). Entrainment thus results from an exogenous (acoustic) cue, which then affects subsequent processing from the inside out. For instance, presenting participants acoustically with repetitive prosodic contours – whose duration fell within the delta-band range – influenced the perception of subsequently presented sentences that either matched or did not match the duration of the prosodic contours (Lamekina and Meyer, 2023). When their duration matched, processing of the visually presented sentences was facilitated as indicated by decreased response times in behaviour. This indicates that the prosodic contours lead to a temporal prediction of the upcoming sentence duration, possibly driven by entrained delta-band oscillations. Notably, the facilitatory effect was numerically strongest for prosodic contours of 2.5

https://doi.org/10.1017/9781009295888.022 Published online by Cambridge University Press

seconds as compared to shorter (1.5 seconds) or longer contours (3.5 seconds). Supporting the role of the delta band as a pacemaker for processing, entrainment may be strongest when the exogenous stimulus matches the preferred endogenous pace.

While we propose that the period of oscillatory cycles at delta-band frequency imposes a temporal constraint on chunk-level integration, the course of each individual cycle may yet pose another constraint on processing: Excitability of the underlying cells changes along the course of the cycle and hence processing abilities are not uniform. On the neuronal level, there are excitable phases, where processing is facilitated, and unexcitable phases that hinder processing (Lakatos et al., 2005). A promising avenue for future research is to assess whether these cycles of excitability are aligned to the expected informativeness of the incoming language input, following the sentence's information structure. For example, prosodic and syntactic cues to upcoming information structure may phase-reset ongoing delta-band oscillations so that the more excitable phase of the oscillations will coincide with the expected timing of more important, new information, while the less excitable phase will occur during less important, given information. Further research is needed to address this and similar questions regarding the relationship between oscillatory cycles of excitability and higher-order language processing.

In conclusion, delta-band oscillations may underlie the previously proposed time constraint and serve as an endogenous pacemaker for language processing. In particular, their preferred wavelength may constrain the formation of multi-word chunks, and even alignment to an external stimulus may be best when the stimulation duration matches their preferred endogenous pace. Additionally, excitability gradients along the cycle may yet pose another timing constraint on the processing of the information structure of the input.

18.5 Discussion

Converging findings from behaviour, ERPs, and oscillatory research point towards temporal constraints on linguistic processing. These constraints can either be context-dependent, resulting from the available linguistic and acoustic information, or context-invariant, representing endogenous processing limitations.

The psycholinguistic literature indicates that (implicit) prosody, syntax, and information structure interact in guiding the formation of multi-word chunks. During speech comprehension, acoustic cues of the prosodic structure guide syntactic parsing through the alignment of prosodic and syntactic boundaries. Even in silent reading, comprehenders use implicit prosodic structure as a chunking strategy to support syntax; this is in line with ERP evidence on the CPS during silent reading. Prosody and syntax further constrain linguistic processing by acting as cues to the information structure of a sentence, guiding attention and processing resources towards the timing of important information.

Additionally, studies on (implicit) prosodic phrasing provide evidence for a specific temporal limit on language production and comprehension. In particular, a context-independent temporal constraint of approximately 2–3 seconds shapes the integration of words into larger multi-word chunks. These temporal integration windows may reflect cognitive and neural processing limitations. At this point, there is no evidence that can disentangle the nature of

https://doi.org/10.1017/9781009295888.022 Published online by Cambridge University Press

these limitations; that is, whether they may result from the decay function of working memory or low-frequency delta-band oscillations. However, these accounts may not be mutually exclusive: Oscillatory cycles could potentially be the electrophysiological unit that corresponds to the working memory buffer that binds multiple to-be-integrated words during language comprehension. Indeed, temporal processing constraints have also been related to inter-individual differences in working memory capacity (Roll et al., 2012, 2013). Likewise, from an oscillatory perspective, a static cycle duration will find little support in terms of ecological validity for language processing. Hence, oscillatory cycles will likely be flexible within a certain frequency range and may vary across individuals. Furthermore, it remains open if this temporal constraint applies in the same way to different levels of processing. While temporal constraints directly seem to affect prosodic and syntactic processing, effects at higher levels of processing such as semantics or information structure may only be indirectly induced by lower-level constraints on chunking.

Contrary to the idea of a temporal constraint discussed here, it may also reflect an information constraint. Cross-linguistic research has found similar information rates for syllables across languages (Coupé et al., 2019; Pellegrino et al., 2011), suggesting a close relationship between the speech production rate and the informativeness. This information rate may be optimally tuned to our processing capacity. In this way, there may be a limit to the amount of information that can be processed within an integration window, which in turn may determine chunking behaviour. Limitations on information processing would also be consistent with the idea that speakers prefer utterances that convey information uniformly across speech signals (Entropy Rate Constancy/Uniform Information Density Hypothesis; Genzel and Charniak, 2002; Jaeger, 2010; Xu and Reitter, 2018).

Future studies may address these questions by combining measures from cognitive neuroscience with psycholinguistic paradigms to tease apart the specific neural processes underlying language comprehension (Beier et al., 2021; Meyer et al., 2020b).

Box 18.1 Chapter Overview

Summary

We present evidence for temporal constraints on (implicit) prosody, syntax, and information structure. These constraints may link to cognitive and neural (e.g., oscillatory) processing limitations and interact with the buildup of structure during comprehension. We suggest that these constraints shape linguistic structure at an optimal pace for brain processing.

Implications

Cognitive, neural, and linguistic timing constraints influence language processing, either contextually or context-invariantly. These constraints appear to interact with the physical timing of the speech signal in shaping the timing of processing. We thus recommend to take them into account when designing experiments and interpreting findings on speech/language processing.

https://doi.org/10.1017/9781009295888.022 Published online by Cambridge University Press

Box 18.1 (cont.)

Gains

We review evidence for context-dependent and context-invariant timing influences on language processing. This may help to acknowledge that behavioural and neural rhythms during language processing – and possible speech and language as such – are shaped both exogenously by speech timing and endogenously by cognitive, neural, and linguistic constraints.

REFERENCES

Assaneo, M. F., and Poeppel, D. (2018). The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Science Advances*, 4(2), eaao3842. https://doi.org/10.1126/sciadv.aao3842

Baddeley, A. D., Thomson, N., and Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 575–589. https://doi.org/10.1016/S0022-5371(75)80045-4

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

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

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

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

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

Bourguignon, M., De Tiège, X., De Beeck, M. O., et al. (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.

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

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

Calhoun, S., Wollum, E., and Kruse Va'ai, E. (2021). Prosodic prominence and focus: Expectation affects interpretation in Samoan and English. *Language and Speech*, 64(2), 346–380. https://doi.org/10.1177/0023830919890362

https://doi.org/10.1017/9781009295888.022 Published online by Cambridge University Press

Carlson, K., Clifton, C., and 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

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

Carreiras, M., and Clifton, C. (1999). Another word on parsing relative clauses: Eyetracking evidence from Spanish and English. *Memory & Cognition*, 27(5), 826–833. https://doi.org/10.3758/BF03198535

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

Clifton, C., Carlson, K., and Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech*, 45(2), 87–114. https://doi.org/10.1177/00238309020450020101

Coupé, C., Oh, Y., Dediu, D., and Pellegrino, F. (2019). Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances*, 5(9), eaaw2594. https://doi.org/10.1126/sciadv.aaw2594

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

Dahan, D. (2015). Prosody and language comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(5), 441–452. https://doi.org/10.1002/wcs.1355

Dahan, D., and Ferreira, F. (2019). Spoken language comprehension. In P. Hagoort (Ed.), *Human Language: From Genes and Brains to Behavior* (pp. 21–33). MIT Press.

Ding, N., and Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, 8, 1–7. https://doi.org/10.3389/fnhum.2014.00311

Ding, N., Melloni, L., Zhang, H., Tian, X., and 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(2), 233. https://doi.org/10.1037/0033-295X.100.2.233

Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes*, 22(8), 1151–1177. https://doi.org/10.1080/01690960701461293

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

Ferreira, V. S., and Yoshita, H. (2003). Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research*, 32(6), 669–692. https://doi.org/10.1023/A:1026146332132

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 the First International Conference on Speech Prosody, pp. 83–90. Aix-en-Provence, France.

Frazier, L. (1979). On comprehending sentences: Syntactic parsing strategies. Unpublished doctoral dissertation, University of Connecticut. www.proquest.com/openview/b6e1df35c2d27107ce8b0e0427ffa5e2/1.pdf?pq-origsite = gscholar&cbl = 18750&diss = y

Frazier, L., and Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6(4), 291–325. https://doi.org/10.1016/0010-0277(78)90002-1

Frazier, L., Carlson, K., and Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, 10(6), 244–249. https://doi.org/10.1016/j.tics.2006.04.002

Genzel, D., and Charniak, E. (2002). Entropy rate constancy in text. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 199–206. https://doi.org/10.3115/1073083.1073117

https://doi.org/10.1017/9781009295888.022 Published online by Cambridge University Press

Glushko, A., Poeppel, D., and Steinhauer, K. (2022). Overt and implicit prosody contribute to neurophysiological responses previously attributed to grammatical processing. *Scientific Reports*, 12(1), 1–18. https://doi.org/10.1038/s41598-022-18162-3

Gross, J., Hoogenboom, N., Thut, G., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, 11(12), e1001752.

Hagoort, P., Brown, C., and Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8(4), 439–483. https://doi.org/10.1080/01690969308407585

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

Hemforth, B., Fernandez, S., Clifton, C., et al. (2015). Relative clause attachment in German, English, Spanish and French: Effects of position and length. *Lingua*, 166, 43–64. https://doi.org/10.1016/j.lingua.2015.08.010

Henke, L., Lewis, A. G., and Meyer, L. (2023). Fast and slow rhythms of naturalistic reading revealed by combined eye-tracking and electroencephalography. *Journal of Neuroscience*, 43(24), 4461–4469. https://doi.org/10.1523/JNEUROSCI.1849-22.2023

Henke, L., and 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

Hirose, Y. (2003). Recycling prosodic boundaries. *Journal of Psycholinguistic Research*, 32(2), 167–195. https://doi.org/10.1023/A:1022448308035

Hwang, H., and Schafer, A. J. (2009). Constituent length affects prosody and processing for a dative NP ambiguity in Korean. *Journal of Psycholinguistic Research*, 38(2), 151–175. https://doi.org/10.1007/s10936-008-9091-1

Inbar, M., Genzer, S., Perry, A., Grossman, E., and Landau, A. N. (2023). Intonation units in spontaneous speech evoke a neural response. *Journal of Neuroscience*, 43(48), 8189–8200. https://doi.org/10.1523/JNEUROSCI.0235-23.2023

Jaeger, F. T. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23–62. https://doi.org/10.1016/j.cogpsych.2010.02.002

Jun, S.-A., and Bishop, J. (2015). Priming implicit prosody: Prosodic boundaries and individual differences. *Language and Speech*, 58(4), 459–473. https://doi.org/10.1177/0023830914563368

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

Kösem, A., Bosker, H. R., Takashima, A., et al. (2018). Neural entrainment determines the words we hear. *Current Biology*, 28(18), 2867–2875. https://doi.org/10.1016/j.cub.2018.07.023

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

Ladd, D. R. (2008). *Intonational Phonology* (second edition). Cambridge University Press.

Lakatos, P., Shah, A. S., Knuth, K. H., et al. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology*, 94(3), 1904–1911. https://doi.org/10.1152/jn.00263.2005

Lamekina, Y., and Meyer, L. (2023). Entrainment to speech prosody influences subsequent sentence comprehension. *Language, Cognition and Neuroscience*, 38(3), 263–276. https://doi.org/10.1080/23273798.2022.2107689

Liberman, M., and Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8(2), 249–336. www.jstor.org/stable/4177987

Lo, C.-W. (2021). Testing low-frequency neural activity in sentence understanding, thesis. https://doi.org/10.7302/2952

https://doi.org/10.1017/9781009295888.022 Published online by Cambridge University Press

Lo, C.-W., Anderson, M., Henke, L., and Meyer, L. (2023). Periodic fluctuations in reading times reflect multi-word-chunking. *Scientific Reports*, 13(1), 1. https://doi.org/10.1038/s41598-023-45536-y

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

Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, 54(6), 1001–1010. https://doi.org/10.1016/j.neuron.2007.06.004

Meyer, L., Sun, Y., and Martin, A. E. (2020a). "Entraining" to speech, generating language? *Language, Cognition and Neuroscience*, 35(9), 1138–1148. https://doi.org/10.1080/23273798.2020.1827155

Meyer, L., Sun, Y., and Martin, A. E. (2020b). 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

Meyer, L., Henry, M. J., Gaston, P., Schmuck, N., and Friederici, A. D. (2016). Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cerebral Cortex*, 27(9), 4293–4302. https://doi.org/10.1093/cercor/bhw228

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

Obleser, J., and Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in Cognitive Sciences*, 23(11), 913–926. https://doi.org/10.1016/j.tics.2019.08.004

Pellegrino, F., Coupé, C., and Marsico, E. (2011). Across-language perspective on speech information rate. *Language*, 87(3), 539–558. https://doi.org/10.1353/lan.2011.0057

Rimmele, J. M., Poeppel, D., and Ghitza, O. (2021). Acoustically driven cortical δ oscillations underpin prosodic chunking. *eNeuro*, 8(4), ENEURO.0562-20.2021. https://doi.org/10.1523/ENEURO.0562-20.2021

Roll, M., Gosselke, S., Lindgren, M., and Horne, M. (2013). Time-driven effects on processing grammatical agreement. *Frontiers in Psychology*, 4, 1–8. https://doi.org/10.3389/fpsyg.2013.01004

Roll, M., Lindgren, M., Alter, K., and 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

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 2(9), 437–442. https://doi.org/10.3758/BF03208784

Sachs, J. S. (1974). Memory in reading and listening to discourse. *Memory & Cognition*, 2(1), 95–100. https://doi.org/10.3758/BF03197498

Samek-Lodovici, V. (2005). Prosody–syntax interaction in the expression of focus. *Natural Language & Linguistic Theory*, 23(3), 687–755. https://doi.org/10.1007/sl

Sauseng, P., Klimesch, W., Gruber, W. R., et al. (2007). Are event-related potential components generated by phase resetting of brain oscillations? A critical discussion. *Neuroscience*, 146(4), 1435–1444. https://doi.org/10.1016/j.neuroscience.2007.03.014

Schremm, A., Horne, M., and 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

Schremm, A., Horne, M., and Roll, M. (2016). Time-driven effects on processing relative clauses. *Journal of Psycholinguistic Research*, 45(5), 1033–1044. https://doi.org/10.1007/s10936-015-9391-1

Selkirk, E. (2011). The syntax–phonology interface. In J. Goldsmith, J. Riggle, and A. C. L. Yu (Eds.), *The Handbook of Phonological Theory* (pp. 435–484). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781444343069.ch14

Stehwien, S., and Meyer, L. (2022). Short-term periodicity of prosodic phrasing. Speech Prosody 2022.

Steinhauer, K., and 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

https://doi.org/10.1017/9781009295888.022 Published online by Cambridge University Press

readers. *Journal of Psycholinguistic Research*, 30(3), 267–295. https://doi.org/10.1023/A:1010443001646

Steinhauer, K., Alter, K., and Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, 2(2), 2. https://doi.org/10.1038/5757

Swets, B., Desmet, T., Hambrick, D. Z., and Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, 136(1), 64–81. https://doi.org/10.1037/0096-3445.136.1.64

Tauroza, S., and Allison, D. (1990). Speech rates in British English. *Applied Linguistics*, 11(1), 90–105. https://doi.org/10.1093/applin/11.1.90

Traxler, M. J. (2007). Working memory contributions to relative clause attachment processing: A hierarchical linear modeling analysis. *Memory & Cognition*, 35(5), 1107–1121. https://doi.org/10.3758/BF03193482

Traxler, M. J. (2009). A hierarchical linear modeling analysis of working memory and implicit prosody in the resolution of adjunct attachment ambiguity. *Journal of Psycholinguistic Research*, 38(5), 491–509. https://doi.org/10.1007/s10936-009-9102-x

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

Vollrath, M., Kazenwadel, J., and Krüger, H. P. (1992). A universal constant in temporal segmentation of human speech: A reply to Schleidt and Feldhütter (1989). *Naturwissenschaften*, 79(10), 479–480. https://doi.org/10.1007/BF01139205

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

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

Wang, L., Bastiaansen, M., Yang, Y., and Hagoort, P. (2012). Information structure influences depth of syntactic processing: Event-related potential evidence for the Chomsky illusion. *PLoS ONE*, 7(10), 1–9. https://doi.org/10.1371/journal.pone.0047917

Webman-Shafran, R., and Fodor, J. D. (2016). Phrase length and prosody in on-line ambiguity resolution. *Journal of Psycholinguistic Research*, 45(3), 447–474. https://doi.org/10.1007/s10936-015-9358-2

Xu, Y., and Reitter, D. (2018). Information density converges in dialogue: Towards an information-theoretic model. *Cognition*, 170, 147–163. https://doi.org/10.1016/j.cognition.2017.09.018

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

Zubizarreta, M. L. (1998). *Prosody, Focus, and Word Order*. MIT Press. https://cir.nii.ac.jp/crid/1130282272784575488

https://doi.org/10.1017/9781009295888.022 Published online by Cambridge University Press