15 Intonation Units: Prosodic Regularity in Spontaneous Speech as a Window onto Cognitive Dynamics

MAYA INBAR, EITAN GROSSMAN, AND AYELET N. LANDAU

15.1 Introduction

Our use of language in daily life is rich and complex. Speaking involves the whole body, is embedded in context, and nearly always involves multiple people. Of relevance to the topic of this book, spontaneous spoken language is produced and comprehended on the fly, unfolding continuously over time. These factors contribute to the ongoing emergence of linguistic structure, which is constantly developing to meet our communicational needs. This usage-based approach to the study of language has contributed greatly to the understanding of the cognitive organization of language (Bybee, 2023; Bybee and Beckner, 2010; Christiansen and Chater, 2008; Ellis, 2019; Hopper, 1987; Mithun, 2009; Mushin and Doehler, 2021; Tomasello, 1998, 2003). The neural mechanisms supporting the use of language are likely also influenced by these processes (Scott, 2019). However, the neural mechanisms of language are often – and perhaps usually – studied by relying on language that is abstracted away from these complexities. In this chapter, we adopt the usage-based perspective and consider the temporal unfolding of spontaneous spoken language and concurrent brain activity. We focus on the prosodic organization of speech into sequences of intonation units (Chafe, 1994; henceforth, IUs), and the ~1 Hz rhythms that they form (Inbar et al., 2020, 2023; Stehwien and Meyer, 2022). Speech prosody is the domain of language that has to do with how pitch, duration, loudness, and other voice qualities unfold over time. Speech prosody plays a central role in communication, and – among its many functions – is used for chunking continuous speech into sequences of prosodic phrases, IUs. Cross-linguistic research suggests that IUs are fundamental to human language and constitute a universal property of language (Chafe, 1994; Himmelmann, 2022; Himmelmann et al., 2018). In the following, we survey the linguistic literature on IUs, including their formal definition and functional roles in speech (Sections 15.2 and 15.3, respectively). We continue by discussing neuroscientific findings pertaining to IUs, and their relation to low-frequency brain rhythms (Section 15.4). In addition, we review algorithmic tools for IU identification, to facilitate further investigations of IUs and the neural processes associated with them (Box 15.1). We build mainly upon Wallace Chafe's conception of IUs, yet consider

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

the influence of similar conceptions of prosodic chunking in a variety of research traditions. We aim to provide a comprehensive survey of current research on IUs, and to introduce them within the context of rhythms of speech, language, and the brain.

15.2 Intonation Units: Form

15.2.1 Phonetic-Prosodic Definition of IUs

IUs are defined by a set of phonetic-prosodic characteristics whose identification is an auditory, perceptual matter (Chafe, 1994; Couper-Kuhlen, 1993; Couper-Kuhlen and Barth-Weingarten, 2011; Cruttenden, 1997; Du Bois et al., 1992; Halliday, 1967; Kibrik, 2019; Schuetze-Coburn et al., 1991; Shattuck-Hufnagel and Turk, 1996). To identify IUs, one must listen to changes in pitch, duration, and loudness in the unfolding speech. Beginning with pitch, an IU is "a stretch of speech uttered under a single coherent intonation contour" (Du Bois et al., 1992). Intonation refers to the modulations of pitch in speech, which are controlled by variation in the frequency of vocal fold vibrations. A coherent intonation contour is composed of a unified pitch movement, on the one hand, and a sharp change in pitch level relative to the preceding utterance, on the other. The shape of the pitch movement is often characterized by reference to its nucleus and final part, and it is critical for determining the meaning of an utterance. Fundamental frequency (F0) is the correlate of pitch in the acoustic domain. Although it is an imperfect correlate, it allows to visualize pitch movements and resets. Figure 15.1A presents the F0 time course of an example speech excerpt in which three IUs were identified.

Moving to durational characteristics, sequences of IUs include an acceleration–deceleration dynamic of syllable delivery rate. Unit-initial syllables are short and often contracted (a phenomenon termed *anacrusis*; Cruttenden, 1997), while utterance-final syllables are lengthened. Final lengthening is a widespread process in the languages of the world (for recent quantitative demonstrations, see Paschen et al., 2022; Seifart et al., 2021), and is also considered a highly reliable cue for identifying IU boundaries. The audio waveform and annotations in Figure 15.1A illustrate changes in syllable delivery rate in the example speech excerpt. However, it should be mentioned that the perception of duration is a subjective phenomenon that cannot be fully captured in timing measurements.

In terms of loudness, sequences of IUs tend to include an increase–decrease dynamic in loudness. This dynamic relies on articulatory factors. The degree of contact between articulators is strongest at the onset of a new utterance (Keating et al., 2003), and this generates resets in loudness at IU onsets. Additionally, because of the gradual decrease in articulation pressure throughout the IU and the compression of the vocal folds, the vibration of the vocal folds might become irregular and particularly slow near the end of the IU, resulting in a low, raspy voice known as creaky voice (Chafe, 1994).

Pauses in speech often accompany the previously mentioned characteristics. However, they are neither a necessary nor a sufficient characteristic of an IU boundary. IUs may succeed each other without any audible pause in between, and on the other hand, pauses may appear midway through an IU, without a reset in pitch level or a change in syllable delivery rate.

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

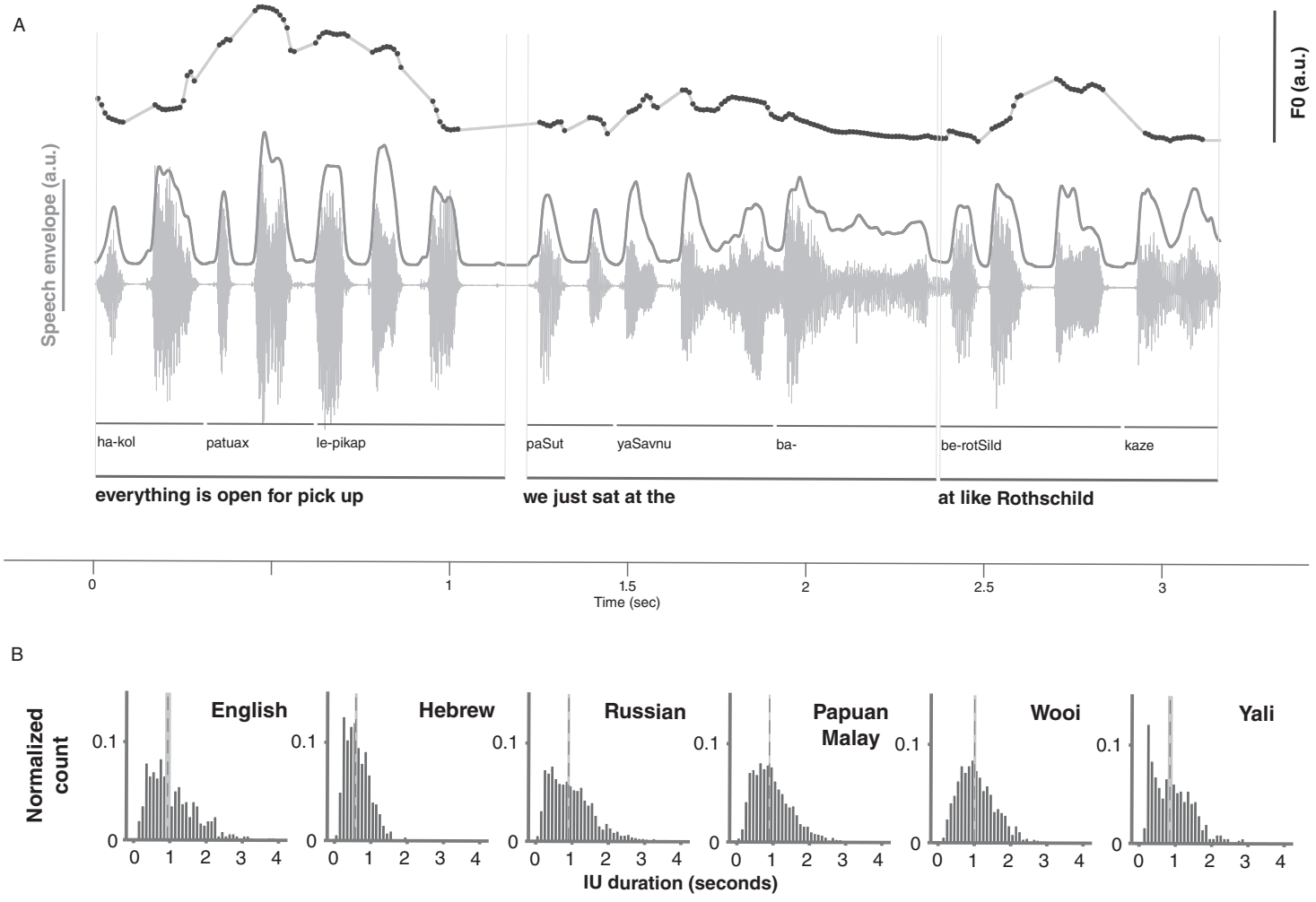


Figure 15.1 IUs: an example sequence and duration histograms.
(A) An example sequence of three IUs from a Hebrew phone call. From bottom to top: a literal translation of each unit; an X-SAMPA phonemic transcription of Hebrew orthographic words; the audio waveform; the speech envelope; the F0 time series. The highlighted portions in the F0 time series signify actual values measured from voiced segments. Thin lines in between provide an estimation via linear interpolation for voiceless segments and pauses. Speech recording available in Marmorstein et al. (2022).
(B) Probability distribution of IU durations in six languages, calculated in 50 ms bins and pooled across multiple speakers. Overlaid are the medians (dashed line, dark gray) and the bootstrapped 95% confidence intervals of the medians (light gray).
Adapted from Figure 2 in Inbar et al. (2020), licensed under CC BY.

A single IU typically features one accented syllable. Another common configuration is two accented syllables, one primary and the other secondary. Accented syllables are prominent in comparison to surrounding ones due to the pitch movement being centered on the syllable, and they are often louder and/or longer in duration. The accent is usually placed on a lexically stressed syllable, though this is not obligatory. An example exception is the case of contrastive accenting, as in: "I said fifTY not fifTEEN" (Couper-Kuhlen and Barth-Weingarten, 2011).[1] See Chapter 35 and Section 6 in general for a complementary view on the rhythm of stressed syllables in speech and its importance in language acquisition.

The phonetic-prosodic definition of IUs provided so far relies on perceptual phenomena, and thus attuned listeners can identify IUs even in unfamiliar languages (Himmelmann et al., 2018). The characteristics may be conventionalized and acquire meaning and become language-specific phonological categories. Trivial examples may be found in the association of a specific function with a certain configuration of the coherent pitch contour. For example, it is very common for languages to use particular prosodic contours for questioning. There may be more subtle conventionalized prosodic patterns, for example for strengthening one's own stance in interaction (Matalon, 2021).

Linguists have developed different systems for transcribing prosodic phenomena, reflecting a range of research traditions. Three widespread systems in current linguistic research are ToBI (Beckman et al., 2005; Pitrelli et al., 1994), GAT (Couper-Kuhlen and Barth-Weingarten, 2011; Selting et al., 1998), and DT (Du Bois et al., 1992), and they all converge on the phonetic-prosodic chunking of speech just described. ToBI (**To**nes and **B**reak **I**ndices) grew out of the autosegmental-metrical approach to the study of intonation (Ladd, 2008; Pierrehumbert, 1980). GAT (**G**esprächs**a**nalytisches **T**ranskriptionssystem, from German: discourse and conversation-analytic transcription system) was developed as an extension of the transcription style used in conversation analysis (Jefferson, 2004) for the representation of prosody. Similar to GAT, DT (**D**iscourse **T**ranscription) also emerged from the study of everyday conversation. These three systems developed in different research traditions for distinct uses and under varying sets of assumptions. However, they all include explicit marking of IUs, though they call them by different names: *intonation units* is the term used in DT, GAT uses the term *intonation phrases*, and ToBI (*intermediate*) *intonational phrases*. In the context of transcribing speech, it should be noted that IU boundary annotations tend to be relatively robust between transcribers (e.g., Himmelmann et al., 2018; Kibrik, 2019; Pitrelli et al., 1994).

When speech is segmented into IUs, the resulting segmentation reflects an assumption that prosodic phrase boundaries are binary in nature. As speech unfolds, a prosodic boundary either occurs or does not. This assumption may be too simplistic a conception of how speakers actually structure their speech prosodically (Barth-Weingarten, 2016; Barth-Weingarten and Ogden, 2021). Specifically, one can question the utility of a single prosodic cue versus several. Research on talk-in-interaction suggests that granularity may serve a function. For example, the prosodic boundaries between episodes in a narrative (Chafe, 1980) or frames in

[1] Note that the pitch movement just described should not be confused with the role of pitch in tone languages, in which pitch can distinguish lexical or grammatical meanings (Yip, 2006).

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

conversation (Goffmann, 1981) come in different strengths, depending on the amount of reorientation required to process the change.

15.2.2 Temporal Structure of IUs

Previous studies characterized the length of IUs using different measurement styles. One method measures the number of words per IU (Chafe, 1994; Himmelmann et al., 2018; Pawley and Syder, 2000). We present the counts aggregated from these studies in Table 15.1. However, as we have demonstrated elsewhere (see Box 2 in Inbar et al., 2020), what constitutes a word varies greatly across languages (Haspelmath, 2011), and so it is difficult to infer the temporal structure of IUs from word counts. Other studies measured the number of syllables per IU (see Table 15.1). Syllable structure also varies across languages, yet unlike the case for words, it is possible to find a common formal characterization for them: Syllables evolve around a vocalic nucleus (e.g., Blevins, 2006). Thus, it is possible to count and measure the durations of syllables in a cross-linguistically sensible way (Greenberg, 2022; Greenberg et al., 2003) and obtain an indirect estimation of IU length. Recent studies found that languages consistently differ in syllable production rate (Coupé et al., 2019; Pellegrino et al., 2011). Currently, it is unclear if this difference leads to differences in IU rates across languages, or if languages differ in the syllable counts per IU. Finally, it should be noted that an estimation of IU length does not capture information about the temporal structure of IU sequences.

In a recent study, we provided a quantitative characterization of the temporal structure of IU sequences (Figure 15.1B; see also Figure 2 in Inbar et al., 2020). We analyzed audio recordings of

Table 15.1 IU length in word and syllable counts

Language	Word count	Syllable count	IU type (see Section 15.3)	Reference
English	4.84		Substantive, e.g., "we had such a nice day"	Chafe (1994)
	1.36		Regulatory, e.g., "and" "well" "umm"	
Seneca (Iroquoian, east USA and Canada)	2		Any	
German	5.20		Any	Himmelmann et al. (2018)
Papuan Malay	4.08		Any	
Wooi (Austronesian, West Papua)	3.63		Any	
Yali (Papuan highland, West Papua)	3.37		Any	
Modern Hebrew	2.18		Any	Inbar et al. (2023)
English		7–10	Excluding IUs "consisting of a connective or an exclamation (e.g., Alors, 'Therefore', in French)"	Jun, 2005
French		7–10		
Japanese		7–10		
Korean		12–15		

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

spontaneous speech in six languages from different families and from different parts of the world. We used IUs as times of interest for analyzing periodicities in the speech envelope. The speech envelope (see Figure 15.1A) is a representation of amplitude fluctuations in speech, and it therefore presumably captures two of the cues for IU boundaries: (1) resets in loudness, and (2) the acceleration–deceleration dynamic of syllable delivery, because one of the sources for the amplitude peaks measured in the speech envelope is the succession of syllable nuclei (see Section 2 for further discussion of amplitude fluctuations in speech). We found that ~1 Hz fluctuations in the speech envelope aligned consistently with IU onset times, and consequently delineated the rate of the unfolding of IUs at a rate of approximately one per second. This finding held in each of the six languages that we studied, despite substantial differences between them in grammatical, areal, genealogical, and socio-cultural parameters. Stehwien and Meyer (2022) replicated these results in an additional language, finding that sequences of 2–5 IUs in German are periodic at 0.8–1.6 Hz. This study used a different statistical framework and radio news broadcasts rather than recordings of spontaneous speech, providing further validity and robustness to the finding. Together, these studies (Inbar et al., 2020; Stehwien and Meyer, 2022) corroborate early reports on the temporal structure of IU sequences of spoken American English (Chafe, 1987; Couper-Kuhlen, 1993) with quantitative evidence from a variety of languages.

This subsection described the temporal structure of sequences of IUs, namely, that one IU succeeds another at the rate of approximately one IU per second. Although the rhythm we characterize is not an isochronous beat, to which one can dance, tap, or nod (Cummins, 2009), nuanced research can demonstrate cases where the approximate rhythmic structure does, in fact, shape behavior. To conclude this subsection, we would like to present evidence that in local contexts, sequences of IUs, or rather sequences of accentuated syllables within IUs, do happen to give rise to rhythmicity in the sense of affording synchronization. Linguists have studied such rhythmic structures in talk-in-interaction (Auer et al., 1999; Couper-Kuhlen, 1993; Marmorstein and Matalon, 2022). In their analyses of stretches of spontaneous spoken language, they identify rhythmic structures that are composed of at least three accentuated syllables (e.g., in three IUs) that are evenly spaced in time. It is possible to tap along to these structures, and there is further evidence that interlocutors perceive this rhythmicity and act on it during interaction. Specifically,

Box 15.1 Methodological Advances: Algorithmic Tools for IU Identification

The identification of IUs has been for a long time an auditory, perceptual matter. The task of manual speech segmentation is laborious and time-consuming, and cannot keep up with the vast amounts of speech recordings. This limits the breadth of investigations pertaining to IUs, in terms of the languages studied and the amount of evidence in each. Automatic tools to speed up the manual process are much called for, yet the development of such tools relies on the unresolved relationship between acoustics and auditory perception. The reader is referred to Biron et al. (2021) for an overview of the existing approaches to boundary detection, as well as a discussion of their strengths and shortcomings.

Here we describe two applications that managed to provide segmentations of speech into IUs at an accuracy level above chance classification in large corpora of American

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

Box 15.1 (cont.)

English, Biron et al. (2021) and Suni et al. (2017). The two applications operationalize the cues for IU boundaries, finding signals and thresholds that correspond to the perception of a boundary. In both cases, the applications relied on the acceleration–deceleration dynamic of speech delivery rate. This dynamic is often considered the most salient characteristic of IU boundaries. In the application described by Biron et al. (2021), an instantaneous measure of speech rate was derived from time-stamped phone annotations and pauses. Speech rate calculated this way was found to drop by approximately twofold at the end of IUs in conversational American English, allowing for efficient automatic IU boundary detection. For the calculation of the instantaneous speech rate, this method requires phone timings, that is, time stamps at the level of individual speech sounds. Identifying and annotating phone timings is itself a laborious task, and is often achieved by automatic forced-alignment tools, with suboptimal results. Code streamlining the application presented in Biron et al. (2021) is unpublished but available.

The Wavelet Prosody Toolkit (Suni et al., 2017) is an application for the hierarchical representation and estimation of prosody. This algorithm derives F0 and intensity information from the speech audio signal, and duration information from time-stamped word annotations. Thus, the way that the acceleration–deceleration dynamic is captured differs from that of Biron et al. (2021). Speech rate is operationalized based on the word level rather than the phone level, and is additionally captured in the speech intensity signal, which gives an indirect measure of the syllable delivery rate. In addition, the application of Suni et al. relies on resets in loudness at IU boundaries, as captured in the speech intensity signal, as well as resets in pitch level at IU boundaries, as captured in the F0 signal. Together, these signals are subjected to an analysis that extracts from them a measure of discontinuity – the boundary strength score. Words that are associated with a strong boundary strength score are classified as ending an IU, thereby providing the segmentation. Suni et al. used a threshold for classification that optimized the performance in their training set relative to manual annotations, and more work is required to survey the stability of thresholds across speakers and languages. Note that this method requires word timings rather than phone timings. Thanks to ongoing developments in automatic speech recognition techniques, word annotations are becoming easier to obtain automatically and in ever more languages. In the context of language neuroscience, the acoustic boundary strength measure developed by Suni et al. (2017) has already been utilized in several recent studies (Anurova et al., 2022; Degano et al., 2023; Inbar et al., 2023; Vetchinnikova et al., 2023). These studies highlight the importance of using ecological stimuli in the study of language and its neural mechanisms, and demonstrate the gains in modeling the temporally unfolding prosodic structure. Algorithmic tools for IU identification facilitate the incorporation of IUs in experimental and quantitative research, and provide the stepping stone to study the construct of the IU itself on a large scale and in ever more languages.

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

when interlocutors take turns at talk, rhythmic structure can be constructed also across their turns, giving rise to *rhythmic integration* (Auer et al., 1999). When one interlocutor integrates their speech into a rhythm set by another at a point of speaker change, this promotes intact interactional exchange. However, when the speech is rhythmically off, interlocutors "engage in supplementary inferencing to account for the departure from expectation" (Auer et al., 1999, p. 70). In other words, the talk following a potential point of integration across turns provides evidence for the rhythmicity. The phenomenon of rhythmic integration highlights the importance of studying spontaneous speech in interactional contexts, for finding rhythm in language.

15.3 Intonation Units: Function

The prosodic phrasing described in Section 15.2 provides a systematic and temporally well-defined perceptual segmentation to ongoing naturalistic speech. Chafe (1987, 1994) termed the segmentation under discussion *intonation units*. We adopt this term rather than others because Chafe foregrounded the *role* of the prosodic segmentation of speech in communication. In the current section we describe the role of IUs in pacing the flow of information in the discourse, together with additional characterizations of their function. It is this aspect of IUs that has led Chafe and others working on language in day-to-day discourse to view IUs as a window onto dynamic cognitive processes in language (e.g., Chafe, 1994, 2018; Du Bois, 1987; Kibrik, 2019; Mithun, 2014; Park, 2002; Wahl, 2015).

In conversation, speakers build a common ground between them step by step. For example, when developing a narrative, speakers introduce the setting, participants, and the course of events in a gradual manner. Ideas that indeed call for introduction relative to the preceding discourse are produced one at a time, in separate IUs. This presumably derives from cognitive constraints on the amount of new information speakers can produce and comprehend. Such constraints may relate to attentional capacity and memory limitations (e.g., Christiansen and Chater, 2016; Simpson and del Prado Martín, 2015). Chafe (1994) termed this *the one-new-idea constraint*. IUs may include ideas that are not new – specifically, ideas that have been previously introduced in the discourse, or that have been evoked by an introduced schema (Du Bois, 1987). These may appear beside new ideas in a single IU, or in an IU of their own. However, IUs appear to express no more than one new idea at a time.

In discourse studies, new ideas are identified based on in-depth qualitative investigations of the discourse transcription. This type of investigation does not lend itself easily to analyses of large amounts of data, and so the field is lacking quantitative evidence for the one-new-idea-per-IU constraint across many languages and contexts. How might one operationalize newness of ideas in spontaneous discourse? In recent years it has become popular to estimate the amount of information that individual words convey using probabilistic language models (e.g., Armeni et al., 2017; Brodbeck et al., 2022; Broderick et al., 2018; Frank et al., 2015; Goldstein et al., 2022; Hale, 2001; Levy, 2008; Nelson et al., 2017; Shannon, 1948; Xu and Reitter, 2018). The field of natural language processing is booming with development in this area, offering constant improvement in the ability to capture context in its full complexity. With better models of context, the estimation of informativeness is expected to be more felicitous. Such informativeness estimates, when considered relative to the prosodic

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

segmentation of speech into IUs, may provide quantitative evidence for the one-new-idea constraint. Cross-linguistic research on phone and syllable rate suggests that language production exhibits a trade-off between production rate and informativeness, as operationalized by similar probabilistic models (Coupé et al., 2019; Pellegrino et al., 2011; Pimentel et al., 2021). This trade-off is hypothesized to stem from a communicative need to spread redundancy evenly across the speech signal (*smooth signal redundancy* and *uniform information density* hypotheses; Aylett and Turk, 2004; Jaeger, 2010). It is possible that prosodic chunking answers a similar need (as suggested in Turk, 2010), and that the one-new-idea constraint described in qualitative linguistic research could be understood within the broader principle of uniform information density.

Chafe proposed to distinguish between three types of IUs, namely, substantive, regulatory, and fragmentary IUs (for an illustration of the three types, see Chafe, 1994, p. 63). Starting with the latter, some IUs stop in the middle of production. These are termed *fragmentary* IUs, and are identified based on their truncated prosodic nature. IUs that are not truncated but rather produced in full may be subcategorized as *regulatory* or *substantive*. Regulatory IUs are those that have regulatory functions, such as organizing the discourse structure, addressing interpersonal matters between interlocutors, and expressing the speaker's mental processes. As noted by Chafe, regulatory IUs often coincide with the linguistic devices called *discourse markers* (Maschler and Schiffrin, 2015), and similarly pertain to the process of using language itself. Conversely, substantive IUs are those that pertain to the extra-lingual world, conveying ideas of events, states, or referents.

To characterize the function of IUs, one might consider their syntactic makeup. As just mentioned, substantive IUs convey ideas of events, states, or referents. Syntactically, ideas of events and states are typically produced in the form of a clause, with a predicate and all the complements and modifiers related to it. And indeed, the most common form of an IU is a clause (Chafe, 1994; see also Givón, 2018; Pawley and Syder, 2000). This has been demonstrated in a variety of studies in several languages, including English, Japanese, Mandarin Chinese, German, Russian, Sasak (Austronesian, Indonesia), Dolakhae Newar (Sino-Tibetan, Nepal), Wardaman (Yangmanic, Australia), and Korean (for reviews, see Croft, 1995, 2007; Kibrik, 2019; Park, 2002). IUs are nearly always full syntactic units of other types if not clauses (e.g., noun phrases, adverbs), containing a grammatical element with all its complements. Croft suggested that such units are "relatively highly integrated … stored as wholes in the mind and produced as single 'bursts' of articulation, that is, as a single intonation unit" (Croft, 2007, p. 5).

Finally, besides reflecting cognitive constraints on information flow, IUs are key units in the construction of interaction in an additional way. Transitions between interlocutors in conversation are smooth and swift, with little overlap and very short lags between turns (Levinson, 2016; Sacks et al., 1974; Schegloff, 2007). Interlocutors coordinate these transitions, and must rely on different cues for the task, that is, on different *projections* of upcoming completion. In interactional linguistics, projection means that an earlier part of a structure foreshadows its later trajectory, thus making its completion predictable (Auer, 2005; Couper-Kuhlen and Selting, 2017). Linguistic structures at various levels have recognizable trajectories, and so projections may be content-based, syntactic, phonological, prosodic, as in the case of IUs, and more. Prosodic projection via the formal characteristics of IUs – as described in Section 15.2 – has been shown to serve a crucial role in timing turn transitions (Bögels and Torreira, 2015,

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

2021; Ford and Thompson, 1996; Gravano and Hirschberg, 2011; Selting, 1996) and in temporal predictions in language processing more generally (Lamekina and Meyer, 2023).

15.4 Intonation Units: Cognitive Neuroscience

Spoken language is prosodically structured in sequences of IUs that succeed each other at a rate of approximately one unit per second. Usage-based linguistic research suggests that these phonetically defined units provide a window onto higher-level representations that are at play during language use. For example, as discussed in Section 15.3, IUs appear to pace new information in the discourse. They also allow speakers to anticipate turn changes in conversation. Against this backdrop, we may ask what cognitive functions and underlying neural mechanisms are associated with the processing of IUs. In this section we survey a study that found a neural response associated with IU closure. This response was found to relate to low-frequency brain rhythms as characterized using electrophysiological recordings. We discuss the neural response to IUs in light of previous neuroscientific studies on speech comprehension, including classical event-related potential (ERP) literature as well as literature on brain rhythms and cognition.

15.4.1 Neural Signatures

In a recent study, we measured the electroencephalography (EEG) of participants who listened to different speakers recounting an emotional life event in Hebrew (Genzer et al., 2022; Inbar et al., 2023). We analyzed the stimuli into IUs based on close auditory listening and categorized each word as either ending an IU or not. Additionally, we quantified an acoustically based measure of prosodic boundary strength at each word (see Box 15.1). We modeled the neural response at each EEG channel and time point relative to word offset. We found that the EEG response to words that end an IU differs from the response to words that do not end an IU, even when equating acoustic boundary strength. The EEG response at IU closure includes a negative deflection at right-anterior electrodes, starting shortly before the end of the final word in the IU and lasting circa 200 ms (Figure 15.2A). IU closure is further characterized by a centroparietal positive deflection between 150 and 500 ms after the final word in the IU (Figure 15.2B). In addition, we found that stronger acoustic boundaries elicit a larger anterior negativity between 100 and 400 ms (Figure 15.2C). Thus, when listening to spontaneous speech, the neural system is attuned to IUs. The response to IUs extends beyond a response to mere acoustics and captures processing at higher levels of representation that presumably unfolds in sequences of IUs.

Within the ERP literature, previous work has identified a component that resembles the response described in the previous paragraph, the closure positive shift (CPS) (Bögels et al., 2011; Steinhauer et al., 1999). The CPS was found in response to prosodic phrase boundaries in several languages other than Hebrew, including German, Dutch, English, Swedish, Japanese, Chinese, and Korean. See a detailed account of the CPS in Chapters 17 and 18. Prosodic boundaries in these studies rely on cues that are nearly identical to those defining IUs.

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

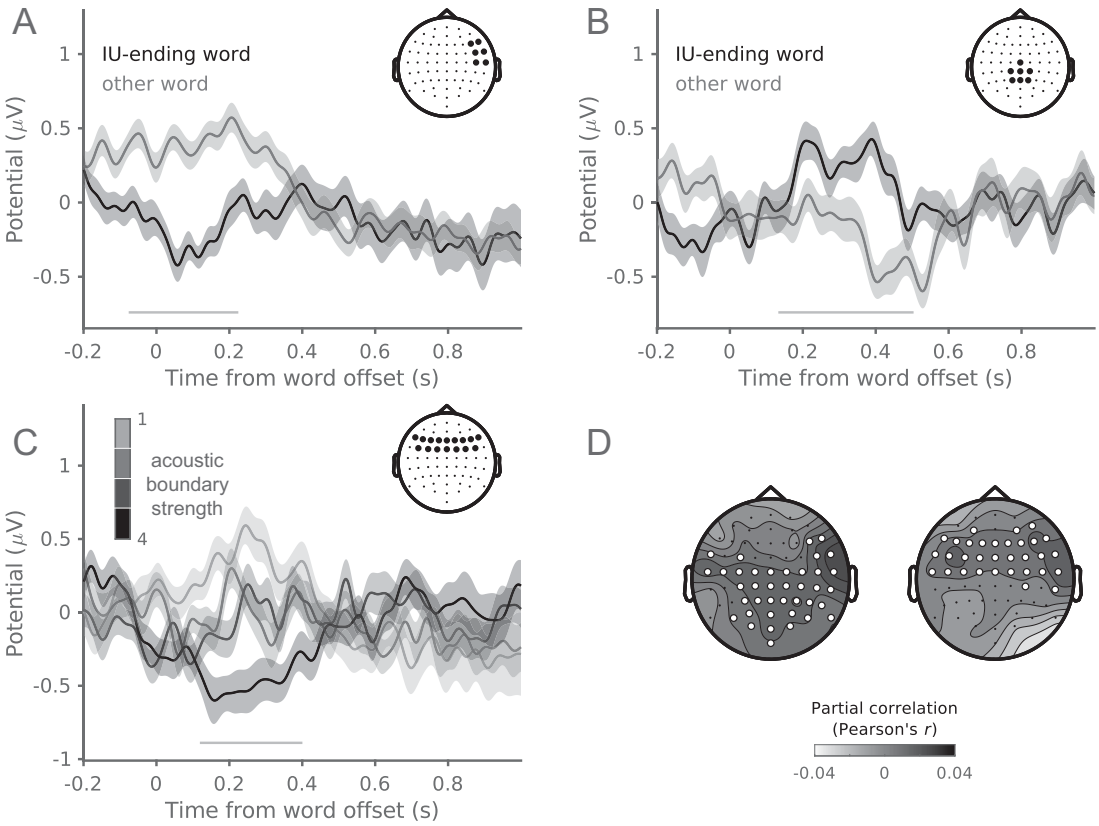


Figure 15.2 IUs: neural response.

(A) ERP traces in response to words that end an IU and other words with comparable acoustic boundary strength (dark and light gray, respectively), illustrating the right-anterior negative cluster. The traces show the grand average over the EEG electrodes highlighted in the inset topography. Shaded ribbons correspond to ±1 SEM. The horizontal bar marks time points of significant difference between conditions (corrected for multiple comparisons). **(B)** ERP traces in response to words that end an IU and other words with comparable acoustic boundary strength (dark and light gray, respectively), illustrating the centroparietal positive cluster. The traces show the grand average over the different sets of EEG electrodes highlighted in the inset topography. Shaded ribbons correspond to ±1 SEM. The horizontal bar marks time points of significant difference between conditions (corrected for multiple comparisons). **(C)** ERP traces in response to words that end an IU at different levels of acoustic boundary strength, illustrating the anterior negative cluster. Four different levels are presented, corresponding to quartiles of boundary strength scores within words that end an IU. The darker the gray, the stronger the acoustic boundary. The traces show the grand average over the EEG electrodes highlighted in the inset topography. Shaded ribbons correspond to ±1 SEM. The horizontal bar marks time points of significant difference between conditions (corrected for multiple comparisons). **(D)** Group average partial correlation maps of the unique predictive accuracy of IU closure (left) and acoustic boundary strength (right), when predicting left-out EEG data in the delta band. Highlighted electrodes denote electrodes in which the predictive accuracy was significantly higher than chance (corrected for multiple comparisons).

Adapted from Inbar et al. (2023), licensed under CC BY.

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

Importantly, however, these studies always used isolated, constructed sentences, and so, to the best of our knowledge, our analysis was the first to characterize a CPS-like response to spontaneous speech during naturalistic listening conditions. Furthermore, our modeling approach allowed to address a debate in the CPS literature considering the extent to which the CPS reflects a linguistic structuring of the input beyond the bottom-up response to acoustic boundary cues (Itzhak et al., 2010; Kerkhofs et al., 2007; Pannekamp et al., 2005; Pauker, 2013; Steinhauer, 2003). Our results suggested what appear to be two different components within the classical CPS. The early- and right-anterior negativity and the centroparietal positivity do not depend on acoustic boundary strength, while the late-anterior negativity is dependent on acoustic boundary strength. This neat result was made possible by combining linguistic theory that focuses on language use with a modeling approach that considers trial-level dynamics among multiple factors.

15.4.2 Brain Rhythms

Research in recent decades has highlighted the relation between temporal structure in speech and temporal structure in brain activity. Understanding this relationship and its role for language has been an influential research program and is one of the goals of the current book. For a detailed presentation of temporal structure in brain activity, see Chapter 3. This line of research suggests that neural activity tracks speech moment by moment (Giraud and Poeppel, 2012; Gross et al., 2013; Obleser and Kayser, 2019). Neural speech tracking refers to any significant relationship between brain activity and a speech stimulus. Neural speech tracking has been found in different timescales, among them, the theta band (4–8 Hz). This temporal scale is thought to reflect neural tracking of syllables, which across languages tend to have a similar temporal structure: four–eight syllables per second (Chandrasekaran et al., 2009; Ding et al., 2017; Greenberg et al., 2003; Pellegrino et al., 2011; cf. MacIntyre et al., 2022). Neural speech tracking has also been found in the delta band (1–2 Hz); however, in this band there was, until recently, less understanding as to what speech components are being tracked (see detailed discussions in Chapters 17 and 21). From a continuous perspective, since sequences of IUs in spontaneous speech give rise to rhythms of ~1 Hz, the time course of neural responses to IUs is expected to give rise to delta band neural speech tracking. In Inbar et al. (2023), we showed precisely this. IU-related activity captures previously described delta band neural activity (Figure 15.2D). Specifically, IU-related potentials were able to accurately predict unseen continuous EEG responses to our stimuli in the delta band. This finding was made possible by our analytic approach: We relied on regression ERPs and then used the stimulus parameters and the fitted potentials to predict continuous EEG signals.

Note that Meyer et al. (2016, 2020) hypothesized that there may be a link between low-frequency neural speech tracking and the CPS. As discussed briefly in Section 15.4.1, the CPS literature debated to what extent the component reflects a linguistic structuring of the input beyond the bottom-up response to acoustic boundary cues. This debate is also at the crux of the uncertainty relative to speech tracking in the delta band. Thanks to our modeling approach, in which the EEG was regressed against multiple predictors, low-frequency speech tracking could be probed relative to individual predictors while accounting for activity explained away by

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

other predictors. Acoustic boundary cues were found to uniquely predict delta band neural activity in one cluster (Figure 15.2D, right), whereas IU-closure status uniquely predicted delta band neural activity in a separate cluster (Figure 15.2D, left). We also studied the unique contribution of one specific higher-level linguistic representation to delta band neural activity, namely clause closure (see Figure 6 in Inbar et al., 2023). Clause closure also predicted unique delta band neural activity, yet to a lesser extent than IU closure and in a smaller cluster.

The current section has dealt exclusively with IUs in speech comprehension. We discussed IUs in speech comprehension as perceptual units relevant also to higher-level cognitive processes. One might expect that focusing on speech production will be no less illuminating regarding the cognitive processes and underlying neural mechanisms associated with IUs. This topic is beyond the scope of the current chapter, yet we would like to raise a question for future research. Speech production involves an ensemble of motor components, and some have suggested that speech timing might emerge from dynamical interactions between these motor components (Turk and Shattuck-Hufnagel, 2014). Are the motoric facets of IU production the sole contributors of their temporal structure? Conversely, perhaps we might identify additional, higher-level cognitive processes at play, such as the process of selecting what to say in the first place, involving both attention and memory (see Chapter 18).

15.5 Conclusion

This chapter provides a step forward in identifying a promising theoretical construct in the study of cognition, especially within the context of rhythms of language and the brain. We focus on the prosodic organization of spontaneous spoken language into intonation units (IUs). Prosodic chunking into IUs is fundamental to human language. We have presented the phonetic-prosodic definition of IUs and current research regarding their temporal structure. We have discussed the roles of IUs in pacing the flow of information in the unfolding discourse, as revealed in studies of everyday conversation. We have surveyed open questions in the neuroscientific literature that recently benefited from considering IUs in speech: questions pertaining to the nature of the CPS and delta band speech tracking. Finally, we have reviewed algorithmic tools for IU identification to facilitate further investigations relying on this construct. Our approach highlights the importance of studying language mechanisms from the perspective of their use, and applying advanced analytical methodologies for this task.

Box 15.2 Chapter Overview

Summary

Spontaneous spoken language is structured in time at the level of prosodic phrasing, a presumably universal property of language. Linguistic research suggests that this structure offers a window onto dynamic cognitive processes during communication. Neural speech tracking in the delta band is predictable based on the prosodic organization of speech into IUs.

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

Box 15.2 (cont.)

Implications

Considering language in its natural occurrence, that is, communication, has proven crucial for describing rhythm in language and the brain. Future research on the cognitive neuroscience of language will benefit from considering the complexities and richness of the way we use language.

Gains

The chapter introduced key points from linguistic research on IUs, and offered a usage-based approach to the neuroscientific study of language. The chapter further surveyed automatic applications for arriving at a prosodic segmentation of speech. These tools might facilitate investigations on prosodic phrasing and its interface with other domains of language and cognition in ecological settings.

REFERENCES

Anurova, I., Vetchinnikova, S., Dobrego, A., et al. S. (2022). Event-related responses reflect chunk boundaries in natural speech. *NeuroImage*, 255 (March), 119203. https://doi.org/10.1016/j.neuroimage.2022.119203

Armeni, K., Willems, R. M., and Frank, S. L. (2017). Probabilistic language models in cognitive neuroscience: Promises and pitfalls. *Neuroscience and Biobehavioral Reviews*, 83 (December 2016), 579–588. https://doi.org/10.1016/j.neubiorev.2017.09.001

Auer, P. (2005). Projection in interaction and projection in grammar. *Text – Interdisciplinary Journal for the Study of Discourse*, 25(1), 7–36. https://doi.org/10.1515/text.2005.25.1.7

Auer, P., Couper-Kuhlen, E., and Müller, F. (1999). *Language in Time: The Rhythm and Tempo of Spoken Interaction*. Oxford University Press.

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

Barth-Weingarten, D. (2016). *Intonation Units Revisited: Cesuras in Talk-in-Interaction*. John Benjamins Publishing Company.

Barth-Weingarten, D., and Ogden, R. (2021). Chunking spoken language: Introducing weak cesuras. *Open Linguistics*, 7(1), 531–548. https://doi.org/10.1515/opli-2020-0173

Beckman, M. E., Hirschberg, J., and Shattuck-Hufnagel, S. (2005). The original ToBi system and the evolution of the ToBi framework. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (Vol. 1) (pp. 9–54). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199249633.003.0002

Biron, T., Baum, D., Freche, D., et al. (2021). Automatic detection of prosodic boundaries in spontaneous speech. *PLoS ONE*, 16(5 May), 1–21. https://doi.org/10.1371/journal.pone.0250969

Blevins, J. (2006). Syllable: Typology. *Encyclopedia of Language & Linguistics*, 1986, 333–337. https://doi.org/10.1016/b0-08-044854-2/00047-x

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

Bögels, S., and Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, 52, 46–57. https://doi.org/10.1016/j.wocn.2015.04.004

Bögels, S., and Torreira, F. (2021). Turn-end estimation in conversational turn-taking: The roles of context and prosody. *Discourse Processes*, 58(10), 903–924. https://doi.org/10.1080/0163853X.2021.1986664

Bögels, S., Schriefers, H., Vonk, W., and Chwilla, D. J. (2011). Prosodic breaks in sentence processing investigated by event-related potentials. *Linguistics and Language Compass*, 5(7), 424–440. https://doi.org/10.1111/j.1749-818X.2011.00291.x

Brodbeck, C., Bhattasali, S., Cruz Heredia, A. A. L., et al. (2022). Parallel processing in speech perception with local and global representations of linguistic context. *ELife*, 11, 1–28. https://doi.org/10.7554/eLife.72056

Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., and Lalor, E. C. (2018). Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. *Current Biology*, 28(5), 803–809. https://doi.org/10.1016/j.cub.2018.01.080

Bybee, J. (2023). What is usage-based linguistics? In M. Díaz-Campos and S. Balasch (Eds.), *The Handbook of Usage-Based Linguistics* (pp. 7–29). Wiley. https://doi.org/10.1002/9781119839859.ch1

Bybee, J., and Beckner, C. (2010). Usage-based theory. In B. Heine and H. Narrog (Eds.), *Oxford Handbook of Linguistic Analysis* (pp. 827–856). Oxford Academic.

Chafe, W. L. (Ed.) (1980). *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production*. ABLEX.

Chafe, W. L. (1987). Cognitive constraints on information flow. In R. S. Tomlin (Ed.), *Coherence and Grounding in Discourse* (pp. 21–51). John Benjamins Publishing Company. https://benjamins.com/catalog/tsl.11.03cha

Chafe, W. L. (1994). *Discourse, Consciousness and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. University of Chicago Press.

Chafe, W. L. (2018). *Thought-Based Linguistics*. Cambridge University Press. https://doi.org/10.1017/9781108367493

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, 5(7), e1000436. https://doi.org/10.1371/journal.pcbi.1000436

Christiansen, M. H., and Chater, N. (2008). Language as shaped by the brain; the brain as shaped by development. *Behavioral and Brain Sciences*, 31(5), 489–558. https://doi.org/10.1017/S0140525X08005268

Christiansen, M. H., and Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39(2016), e62. https://doi.org/10.1017/S0140525X1500031X

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

Couper-Kuhlen, E. (1993). *English Speech Rhythm: Form and Function in Everyday Verbal Interaction*. John Benjamins Publishing Company.

Couper-Kuhlen, E., and Barth-Weingarten, D. (2011). A system for transcribing talk-in-interaction: GAT 2. English translation and adaptation of Selting, M. et al. (2009). Gesprächsforschung - Online-Zeitschrift Zur Verbalen Interaktion. www.gespraechsforschung-ozs.de/heft2011/px-gat2-englisch.pdf

Couper-Kuhlen, E., and Selting, M. (2017). Turn construction and turn taking. In E. Couper-Kuhlen and M. Selting (Eds.), *Interactional Linguistics* (pp. 31–111). Cambridge University Press. https://doi.org/10.1017/9781139507318.004

Croft, W. (1995). Intonation units and grammatical structure. *Linguistics*, 33(5), 839–882. https://doi.org/10.1515/ling.1995.33.5.839

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

Croft, W. (2007). Intonation units and grammatical structure in Wardaman and in cross-linguistic perspective. *Australian Journal of Linguistics*, 27(1), 1–39. https://doi.org/10.1080/07268600601172934

Cruttenden, A. (1997). *Intonation* (second edition). Cambridge University Press.

Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, 37(1), 16–28. https://doi.org/10.1016/j.wocn.2008.08.003

Degano, G., Donhauser, P. W., Gwilliams, L., Merlo, P., and Golestani, N. (2023). Speech prosody enhances the neural processing of syntax. *Communications Biology*, 7, 748. https://doi.org/10.1038/s42003-024-06444-7

Ding, N., Patel, A. D., Chen, L., et al. (2017). Temporal modulations in speech and music. *Neuroscience and Biobehavioral Reviews*, 81(Part B), 181–187. https://doi.org/10.1016/j.neubiorev.2017.02.011

Du Bois, J. W. (1987). The discourse basis of ergativity. *Language*, 63(4), 805–855.

Du Bois, J. W., Cumming, S., Schuetze-Coburn, S., and Paolino, D. (1992). Discourse transcription. In J. W. Du Bois, S. Cumming, S. Schuetze-Coburn, and D. Paolino (Eds.), *Santa Barbara Papers in Linguistics* (Vol. 4) (pp. 1–225). University of California.

Ellis, N. C. (2019). Essentials of a theory of language cognition. *Modern Language Journal*, 103, 39–60. https://doi.org/10.1111/modl.12532

Ford, C. E., and Thompson, S. A. (1996). Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff, and S. A. Thompson (Eds.), *Interaction and Grammar* (pp. 134–184). Cambridge University Press. https://doi.org/10.1017/CBO9780511620874.003

Frank, S. L., Otten, L. J., Galli, G., and Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, 140, 1–11. https://doi.org/10.1016/j.bandl.2014.10.006

Genzer, S., Ong, D. C., Zaki, J., and Perry, A. (2022). Mu rhythm suppression over sensorimotor regions is associated with greater empathic accuracy. *Social Cognitive and Affective Neuroscience*, 17(9), 788–801. https://doi.org/10.1093/scan/nsac011

Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15(4), 511–517. https://doi.org/10.1038/nn.3063

Givón, T. (2018). Toward a discourse definition of syntax: The communicative correlates of grammar. In *On Understanding Grammar: Revised Edition* (pp. 25–92). John Benjamins Publishing Company.

Goffman, E. (1981). *Forms of Talk*. University of Pennsylvania Press.

Goldstein, A., Zada, Z., Buchnik, E., et al. (2022). Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience*, 25(3), 369–380.

Gravano, A., and Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Computer Speech and Language*, 25(3), 601–634. https://doi.org/10.1016/j.csl.2010.10.003

Greenberg, S. (2022). Speech, rhythm, and the brain. *Acoustics Today*, 18(3), 22–31.

Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (2003). Temporal properties of spontaneous speech: A syllable-centric perspective. *Journal of Phonetics*, 31(3–4), 465–485. https://doi.org/10.1016/j.wocn.2003.09.005

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. https://doi.org/10.1371/journal.pbio.1001752

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies (NAACL '01). Association for Computational Linguistics, USA, pp. 1–8. https://doi.org/10.3115/1073336.1073357

Halliday, M. A. K. (1967). *Intonation and Grammar in British English*. De Gruyter Mouton.

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

Haspelmath, M. (2011). The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica*, 45(1), 31–80. https://doi.org/10.1515/flin.2011.002

Himmelmann, N. P. (2022). Prosodic phrasing and the emergence of phrase structure. *Linguistics*, 60(3), 715–743. https://doi.org/10.1515/ling-2020-0135

Himmelmann, N. P., Sandler, M., Strunk, J., and Unterladstetter, V. (2018). On the universality of intonational phrases: A cross-linguistic interrater study. *Phonology*, 35(2), 207–245. https://doi.org/10.1017/S0952675718000039

Hopper, P. J. (1987). Emergent grammar. *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, 13, 139–157.

Inbar, M., Grossman, E., and Landau, A. N. (2020). Sequences of intonation units form a ~ 1 Hz rhythm. *Scientific Reports*, 10(1), 1–9. https://doi.org/10.1038/s41598-020-72739-4

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

Itzhak, I., Pauker, E., Drury, J. E., Baum, S. R., and Steinhauer, K. (2010). Event-related potentials show online influence of lexical biases on prosodic processing. *NeuroReport*, 21(1), 8–13. https://doi.org/10.1097/WNR.0b013e328330251d

Jaeger, T. F. (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

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation Analysis: Studies from the First Generation* (Vol. 125) (pp. 13–31). John Benjamins Publishing Company. www.liso.ucsb.edu/liso_archives/Jefferson/Transcript.pdf

Jun, S. (2005). Prosodic typology. In S. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 430–458). Oxford University Press.

Keating, P., Cho, T., Fougeron, C., and Hsu, C.-S. (2003). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden, and R. Temple (Eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI* (pp. 145–163). Cambridge University Press.

Kerkhofs, R., Vonk, W., Schriefers, H., and 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

Kibrik, A. A. (2019). Upper Kuskokwim (Athabaskan, Alaska). *Anthropological Linguistics*, 61(2), 141–182. www.jstor.org/stable/10.2307/26974625

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

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

Levinson, S. C. (2016). Turn-taking in human communication: Origins and implications for language processing. *Trends in Cognitive Sciences*, 20(1), 6–14. https://doi.org/10.1016/j.tics.2015.10.010

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

MacIntyre, A. D., Cai, C. Q., and Scott, S. K. (2022). Pushing the envelope: Evaluating speech rhythm with different envelope extraction techniques. *Journal of the Acoustical Society of America*, 151(3), 2002–2026. https://doi.org/10.1121/10.0009844

Marmorstein, M., and Matalon, N. (2022). Responses within activities. *Interactional Linguistics*, 2(1), 42–78. https://doi.org/10.1075/il.21003.mar

Marmorstein, M., Matalon, N., Shor, L., et al. (2022). *HCSH: HUJI Corpus of Spoken Hebrew*. The Hebrew University of Jerusalem. www.huji-corpus.com

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

Maschler, Y., and Schiffrin, D. (2015). Discourse markers: Language, meaning, and context. In D. Tannen, H. E. Hamilton, and D. Schiffrin (Eds.), *The Handbook of Discourse Analysis* (second edition) (pp. 189–221). John Wiley & Sons, Inc.

Matalon, N. (2021). The camel humps prosodic pattern. In C. Mauri, I. Fiorentini, and E. Goria (Eds.), *Building Categories in Interaction: Linguistic Resources at Work* (pp. 155–186). John Benjamins Publishing Company. https://doi.org/10.1075/slcs.220.06mat

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

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

Mithun, M. (2009). Re(e)volving complexity: Adding intonation. In T. Givón and M. Shibatani (Eds.), *Syntactic Complexity: Diachrony, Acquisition, Neuro-cognition, Evolution* (pp. 53–80). John Benjamins Publishing Company.

Mithun, M. (2014). Syntactic and prosodic structures: Segmentation, integration, and in between. In T. Raso and H. Mello (Eds.), *Spoken Corpora and Linguistic Studies* (pp. 297–330). John Benjamins Publishing Company.

Mushin, I., and Doehler, S. P. (2021). Linguistic structures in social interaction. *Interactional Linguistics*, 1(1), 2–32. https://doi.org/10.1075/il.21008.mus

Nelson, M. J., El Karoui, I., Giber, K., et al. (2017). Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences*, 114(18), E3669–E3678. https://doi.org/10.1073/pnas.1701590114

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

Pannekamp, A., Toepel, U., Alter, K., Hahne, A., and 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

Park, J. S.-Y. (2002). Cognitive and interactional motivations for the intonation unit. *Studies in Language*, 26(3), 637–680. https://doi.org/10.1075/sl.26.3.07par

Paschen, L., Fuchs, S., and Seifart, F. (2022). Final lengthening and vowel length in 25 languages. *Journal of Phonetics*, 94, 101179. https://doi.org/10.1016/j.wocn.2022.101179

Pauker, E. (2013). How multiple prosodic boundaries of varying sizes influence syntactic parsing: Behavioral and ERP evidence. Unpublished PhD Dissertation, McGill University.

Pawley, A., and Syder, F. H. (2000). The one-clause-at-a-time hypothesis. In H. Riggenbach (Ed.), *Perspectives on Fluency* (pp. 163–199). University of Michigan Press.

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

Pierrehumbert, J. (1980). The phonology and phonetics of English intonation. *RELC Journal*, 15(1), 117–118. https://doi.org/10.1177/003368828401500113

Pimentel, T., Meister, C., Salesky, E., et al. (2021). A surprisal–duration trade-off across and within the world's languages. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 949–962. Online and Punta Cana, Dominican Republic. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.73

Pitrelli, J. F., Beckman, M. E., and Hirschberg, J. (1994). Evaluation of prosodic transcription labeling reliability in the ToBI framework. Proceedings of the Third International Conference on Spoken Language Processing, 2, pp. 123–126. https://doi.org/10.21437/ICSLP.1994-34

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696–735.

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press

Schegloff, E. A. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.004

Schuetze-Coburn, S., Shapley, M., and Weber, E. G. (1991). Units of intonation in discourse: A comparison of acoustic and auditory analyses. *Language and Speech*, 34(3), 207–234. https://doi.org/10.1177/002383099103400301

Scott, S. K. (2019). From speech and talkers to the social world: The neural processing of human spoken language. *Science*, 366(6461), 58–62. https://doi.org/10.1126/science.aax0288

Seifart, F., Strunk, J., Danielsen, S., et al. (2021). The extent and degree of utterance-final word lengthening in spontaneous speech from 10 languages. *Linguistics Vanguard*, 7(1), 1–14. https://doi.org/10.1515/lingvan-2019-0063

Selting, M. (1996). On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation. *Pragmatics*, 6(3), 357–388. https://doi.org/10.1075/prag.6.3.06sel

Selting, M., Auer, P., Barden, B., et al. (1998). Gesprächsanalytisches Transkriptionssystem (GAT). *Linguistische Berichte*, 173, 91–122.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(4), 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x

Shattuck-Hufnagel, S., and Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25(2), 193–246. https://doi.org/10.1007/BF01708572

Simpson, H. E., and del Prado Martín, F. M. (2015). Memory capacity limits in processing of natural connected speech: The psychological reality of intonation units. Proceedings of the 37th Annual Meeting of the Cognitive Science Society, CogSci 2015, pp. 2206–2211. Pasadena, USA.

Stehwien, S., and Meyer, L. (2022). Short-term periodicity of prosodic phrasing: Corpus-based evidence. Speech Prosody 2022, pp. 693–698. https://doi.org/10.31234/osf.io/c9sgb

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language*, 86(1), 142–164. https://doi.org/10.1016/S0093-934X(02)00542-4

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), 191–196. https://doi.org/10.1038/5757

Suni, A., Šimko, J., Aalto, D., and Vainio, M. (2017). Hierarchical representation and estimation of prosody using continuous wavelet transform. *Computer Speech and Language*, 45, 123–136. https://doi.org/10.1016/j.csl.2016.11.001

Tomasello, M. (Ed.) (1998). *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure Volume 1*. Lawrence Erlbaum Associates, Inc.

Tomasello, M. (Ed.) (2003). *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure Volume 2*. Lawrence Erlbaum Associates, Inc.

Turk, A. (2010). Does prosodic constituency signal relative predictability? A smooth signal redundancy hypothesis. *Laboratory Phonology*, 1(2), 227–262. https://doi.org/10.1515/labphon.2010.012

Turk, A., and Shattuck-Hufnagel, S. (2014). Timing in talking: What is it used for, and how is it controlled? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658), 20130395. https://doi.org/10.1098/rstb.2013.0395

Vetchinnikova, S., Konina, A., Williams, N., Mikušová, N., and Mauranen, A. (2023). Chunking up speech in real time: Linguistic predictors and cognitive constraints. *Language and Cognition*, 15(3), 453–479. https://doi.org/10.1017/langcog.2023.8

Wahl, A. (2015). Intonation unit boundaries and the storage of bigrams. *Review of Cognitive Linguistics*, 13(1), 191–219. https://doi.org/10.1075/rcl.13.1.08wah

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

Yip, M. (2006). Tone: Phonology. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (pp. 761–764). Elsevier. https://doi.org/10.1016/B0-08-044854-2/00054-7

https://doi.org/10.1017/9781009295888.019 Published online by Cambridge University Press