



**Article type: Advanced Review**

## Prosody and Language Comprehension

**Delphine Dahan\***

This review provides a summary of the most recent advances on the study of how prosody is used during language comprehension. Prosody is characterized as an abstract structure composed of discrete tonal elements aligned with the segmental composition of the sentence organized in constituents of increasing size, and this structure is influenced by the phonological, syntactic, and informational structures of the sentence. Here, we discuss evidence that listeners are affected by prosody when establishing those linguistic structures. Prosody has been shown to influence the segmentation of the utterance into syllables and words, and, in some cases, whether a syllable or word is judged to be present or not. The literature on how prosody informs the structural relationship between words and phrases is also discussed, contrasting views that assume a direct (albeit probabilistic) link between syntax and prosody with those that posit a complex interface between syntax and prosodic structure. Finally, the role of prosody in conveying important aspects pertaining to the sentence's information structure (i.e., which parts of the sentence's meaning are highlighted and brought forward to the discourse, which ones are presupposed and left in the background, which attitudes are being conveyed about the concepts or propositional content) has long been recognized. Current research focuses on which prosodic elements contribute to marking which the dimensions (or semantic primitives) of the information structure.

### How to cite this article:

*WIREs Cogn Sci* 2015. doi: 10.1002/wcs.1355

### INTRODUCTION

Prosody, in the sense I will be using throughout this review, refers to a structure defined by phrasing and prominence. Prosody groups syllables and words together and renders some elements more prominent than others. While some of the phrasing is realized by intonation and rhythm modulations, others are revealed through phenomena that affect the segments of the utterance. Prosody plays an important role in the establishment of the phonological, lexical, syntactic, and semantic representation of the sentence, as well as in retrieving some aspects of the information structure of the sentence (e.g., what is presupposed or what is not) and the pragmatic implications that a listeners may infer regarding the speaker's intentions. For example, people share the intuition that the

segmentally identical phrases “crisis turnip” and “cry sister nip” can be distinguished from each other based upon the prosodic rendering that the speaker adopts.<sup>1</sup> Likewise, the sentences “Raoul murdered the man with a gun” and “I asked the teacher who left” have more than one literal meaning and speakers may be able to reveal which they intend through the sentence's prosody.<sup>2,3,4</sup> The prosody of a sentence can sometimes affect a listener's conclusion about the speaker's beliefs, attitude, or intention.<sup>5,6</sup> For example, the sentence “It looks like a zebra” with an accent on the final word is characterized by a canonical declarative prosodic contour and is perceived as supporting an affirmative interpretation (“It looks like a ZEBRA. [and indeed, it is one]”); on the other hand, the same sentence with a so-called ‘contrastive’ accent on “looks” and a rising intonation supports a

\*Correspondence to: [dahan@psych.upenn.edu](mailto:dahan@psych.upenn.edu)

Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

negative or contradictory interpretation (“It LOOKS like a zebra [but it actually isn’t one]”).

Thus, the prosody of a sentence is intimately connected to the sentence’s multiple linguistic structures. This review provides a summary of some recent advances made in characterizing how prosody contributes to language comprehension. It is limited to studies conducted on the English language because these continue to make up the bulk of the research on the topic. In pursuing such investigation, psycholinguists generate hypotheses about the nature of the information listeners extract from the utterance and how the information is utilized in the establishment of linguistic structures. This review is as much a review of the most exciting findings on this topic as it is about the nature of the hypotheses scholars have considered and the assumptions they entertain along the way.

## **PROSODY AS AN ABSTRACT STRUCTURE**

The study of prosody has a long history. Until the mid 1970s, little consensus had been reached on what determines a speaker’s choice of intonation and rhythm. Since then, a widely accepted framework has emerged. According to the ‘autosegmental-metrical’ theory, the variations of pitch, loudness, and rhythm of an utterance are the phonetic realization of an abstract structure composed of phonological objects (see Ref. [7-9] for reviews). This prosodic structure consists of a succession of intonational elements or targets that are associated with particular landmarks in the segmental description of the utterance. How these targets are linked with elements in the segmental string is determined by a ‘metrical grid’ (i.e., an abstract representation that specifies the organization of syllables into constituents of increasing size, with the smallest constituents nested into constituents of larger sizes) and the level of prominence that each syllable can receive. The intonation of an utterance, or its tune, consists of two kinds of discrete elements. Pitch accents are tonal objects that are associated with specific syllables and that mark them as prominent. Pitch accents are viewed as properties of words. Edge tones, on

the other hand, are pitch movements that mark the right edge of constituents. The domains of edge tones are the constituents they mark. Both kinds of tonal elements are further specified by their level, ‘high’ (H) or ‘low’ (L), as opposed to configurations, such as ‘rise’ or ‘fall’. Because the F0 values that these elements can take are variable (a necessary consequence of assuming two levels), tonal elements are defined as local F0 ‘peaks’ or ‘valleys’ whose specific phonetic values depend on factors considered orthogonal to their phonological category, such as the degree of emphasis or the position in the utterance. Finally, the discrete phonetic elements are assumed to be connected with one another via interpolation or other simple principles. Thus, within the auto-segmental-metrical framework, a complete theory must specify the inventory of tonal elements and a series of principles or rules regarding how a sequence of these abstract, discrete elements translates into a continuous F0 contour.

Unfortunately, the field has not reached a consensus regarding either of these aspects. There is disagreement on how to decide whether a portion of an utterance’s F0 contour is associated with a tone or is merely a transition, how to decide whether a pitch accent contains a single target or two (i.e., a bi-tonal pitch accent), and whether the pitch range of the F0 contour plays a role in this decision. For instance, there is much debate on whether the F0 range that characterizes the onset of a rise in a pitch accent (i.e., medium or low in the speaker’s pitch range) can be taken to specify a phonological distinction between two accent types, L+H\* vs. H\*, or whether the pitch range marks gradient, non-phonological variations.<sup>10,11</sup> The classification of tonal element as H or L is also contentious. Although H pitch accents have been loosely described as local F0 peaks associated with accented syllables, and L pitch accents as local F0 valleys, there is no consensus on defining this locality.<sup>12,13</sup> Finally, it still is an open question whether all intonational categories are expressed through discrete phonological elements or if some may be conveyed by the continuous variation of a physical dimension (such as the alignment of a pitch accent with the accented syllable’s vowel).<sup>14,15</sup>

An assumption shared by all phonological approaches to intonation and prosody is that discrete tonal elements convey meaning, an assumption that Ladd<sup>8</sup> dubbed ‘the linguist’s theory of intonational meaning’: Prosodic elements, alone or combined with other elements to form a tune and in conjunction with the other linguistic structures of the sentence and a rich pragmatic system, are assumed to give rise to meaning as we experience it.<sup>16-20</sup> However, the specific meanings of tones, as well as tunes, are debated. It is useful to keep in mind that tonal elements may resemble segments in this respect: While phonemes are elements of meaning because phonemic contrasts capture (potential) contrasts in meaning, they themselves have no independent meaning (the /p/ in ‘peach’, in contrast to the /b/ in ‘beach’, convey nothing that differentiates the semantic features of a peach compared to those of a beach).

Because of the assumption that tonal elements are discrete categories that convey discrete meaning, an approach to establishing the inventory of such categories has consisted of showing that the interpretation of a sentence changes as a function of its prosodic characteristics. However, a contrast in meaning (such as an intonation that conveys uncertainty or incredulity) is not synonymous with a phonological distinction because such distinction can emerge from contrasting two values located on a continuous dimension.<sup>10,21</sup> Phonologists often find it more productive to focus on the phonetic forms that an utterance can take and tease apart phonetic from phonological variations.<sup>10,12,13,22-24</sup>

Much debate remains despite active research on these questions. Nonetheless, many scholars have adopted the framework offered by Pierrehumbert<sup>25</sup> and its later instantiation,<sup>26</sup> or the transcription system known as ‘ToBI’ (an acronym for ‘Tones and Boundary Indices’),<sup>27</sup> which substantially departs from Pierrehumbert’s original theory. To those who are not well versed into the phonological literature, the popularity of these theories may be interpreted as evidence for a consensus while in fact, as described above, many of their assumptions have not received firm empirical support or have been challenged.

This brief summary of the state of research on prosody and intonational phonology highlights the difficulties in describing how the various linguistic structures of a sentence affect the prosodic structure that speakers assign to their utterance, and how listeners interpret the continuously varying dimensions of speech into elements that, once integrated with other linguistic elements of the sentence, give rise to an interpretation. The prosodic structure of a sentence is a representation that is independent of, albeit related to, the sentence’s other linguistic structures. It is via these relationships that prosody contributes to the meaning of the sentence. However, the role of prosody has been conceptualized, and consequently investigated, differently for different linguistic structures. These disparities, I argue, reflect the division of labor that scholars have (largely implicitly) assumed between prosody and other markers of a sentence’s linguistic structures. For the syntax of a sentence, the establishment of the structure is viewed as depending on the linguistic elements present in the sentence and the grammar, which dictates both how these elements combine into phrases and how these phrases relate to one another; when the sentence is spoken, the prosodic phrasing may help determine which of those relationships is intended. Thus, prosody is conceptualized as complementing syntactic knowledge. By contrast, the information structure of a sentence, i.e., the way by which the sentence relates to the context in terms of the discourse status of its content and the attentional and intentional states of the discourse participants, relies heavily on the pattern of prominence (and perhaps phrasing) that the speaker chooses to produce. Other devices are available, such as pronouns or syntactic constructions, but the role of prosody in marking the information structure in English is substantial. Here, the contribution of prosody to the sentence’s information structure is viewed as leading, as opposed to supporting, and prosodic elements are often conceived as conveying meaning that bears on information structure directly. As far as the phonology of a sentence is concerned, the situation is a little more complex because the structure encompasses many aspects; regarding the establishment of the segmental content of a

sentence, one from which morpho-syntactic elements emerge, prosody is generally viewed as having a supportive role.<sup>28</sup>

These often implicit assumptions have had implications on the nature of the questions that scholars have addressed: When examining the role of prosody in establishing the phonology and syntax of a sentence, investigators have asked whether or not prosody contributes to it, and if it does, how to characterize the mechanism by which the contribution takes place. Regarding the role of prosody in establishing the information structure of a sentence, on the other hand, scholars have taken for granted that prosody plays a role; instead, research has revolved around the way prosody conveys the dimensions of information structure and whether there is a one-to-one mapping between elements of the two structures. Here, I discuss the current state of knowledge on the role of prosody in establishing each of those structures.

#### **THE ROLE OF PROSODY IN DERIVING THE PHONOLOGICAL REPRESENTATION OF AN UTTERANCE**

Speech comprehension requires listeners to make some hypotheses about the linguistic elements present in the utterance they are trying to interpret. The significant role that prosody plays in this process is becoming increasingly clear. It has long been recognized that many temporal cues to phonemic contrasts (such as the length of the delay between the burst and the onset of vocalic vibrations, an acoustic correlate to the voicing contrast amongst stop consonants) are interpreted relative to global temporal aspects, such as the syllable duration or speaking rate.<sup>29</sup> The phenomenon could be viewed as confined to a hypothetical early stage of processing, one in which the acoustic signal is translated into a linear string of phonemic elements. Such phonemic representation, one could argue, constitutes the sole basis for establishing the lexical and morpho-syntactic composition of the utterance. This view is still commonly held, despite substantial evidence for its shortcomings.

The assumption of a phonemic string devoid of any phonetic marking of syllable, word, and phrase boundaries denies a role for these cues in assisting with segmenting and parsing the segmental string into linguistic units. The absence of systematic cues to word boundaries, as pointed out by many empirical papers published in the 1970s and 1980s, may have contributed to the prevalence of this view.<sup>30-32</sup> However, a large body of work has since documented the multiple ways in which the prosodic structure of an utterance is marked in the phonetic characteristics of the segments themselves (see Ref. [33] for a review). If prosody plays such a fundamental role in explaining fine-grained systematic variation in the production of individual speech sounds, one may expect listeners to take advantage of the marking of prosodic constituents in the process of isolating linguistic constituents. Indeed, while prosodic and lexical constituents are not isomorphic, their relationship is quite systematic: The edges of prosodic words (and of the prosodic phrases higher on the prosodic hierarchy) are always aligned with word boundaries. For instance, if an edge tone is aligned with the syllable “ham”, this syllable must be word final (and, perhaps, correspond to the monosyllabic “ham”) rather than be parsed as the first syllable of a polysyllabic word such as “hamster”. Thus, prosodic phrasing could facilitate the segmentation of an utterance into word-size units by favoring a parsing of the input that respects the prosodic-lexical alignment and disfavors one that violates it. This hypothesis has now received ample empirical support.<sup>34-37</sup> In addition to local cues, segmentation may be assisted by some stress-based patterning or rhythm, what Dilley and colleagues call ‘distal’ prosody.<sup>38-40</sup> People appear to build an expectation on how syllables-to-come will be grouped based on the pattern of tonal alternation that precedes them and this expectation affects how they perceive the current grouping, above and beyond the local acoustic-phonetic cues that may cue grouping.

A number of accounts of the mechanism by which the prosodic structure of an utterance affects segmentation have been proposed. According to one approach, the acoustic-phonetic characteristics that arise from the

prosodic structure at word junctures, such as aspiration, glottalization, lengthening, etc., are integrated with other cues, such as phonotactic, lexical, and syntactic knowledge, in order to generate the segmentation that best fit the set of cues.<sup>28,41</sup> For instance, the lengthening of a syllable, given the context, is interpreted as a cue to a word boundary. This approach posits that the phonetics of an utterance directly mark word boundaries.

An alternative account proposes that phonetic cues to the prosodic structure affect the perceived lexical composition of the utterance *indirectly*, via the establishment of a prosodic structure. Such structure is built in parallel to the phonemic and lexical analysis of the utterance.<sup>34,36</sup> In this approach, specific phonetic characteristics, such as syllable lengthening or articulatory strengthening, cause listeners to posit elements of prosodic structure. Such prosodic structure would then serve as a source of support for some parsing of the segmental string over others. For instance, the posited presence of an intonational-phrase boundary would favor lexical hypotheses whose edges are aligned with the boundary and disfavor those that straddle it.

The proposal of a prosodic structure being built in parallel to a segmental representation of the utterance implies some independence in the construction of these two representations, in spite of the fact that the utterance's phonetics merges the two. But recent research has revealed some interdependence between the two. In their seminal study, Dille and Pitt<sup>42</sup> showed that whether or not people perceived the presence of a short function word in a phrase depended on the perceived duration of the fragment of speech comprising the syllable that preceded the potential function-word site and the function word itself, if produced. If the fragment was perceived as too long to be a single syllable given the speech rate, a function word was posited. Conversely, if the fragment was perceived as too short to also include a function word, its presence was largely undetected. Duration perception could have been induced from an estimate of long-range speaking rate, but the effect was also sensitive to a very local perturbation.<sup>43</sup> In the light of this work, one must conclude that the segmental representation

of an utterance is only partially determined by the spectral cues of those segments. Equally important are the temporal characteristics of the utterance to which listeners try to attribute lexical content. It is not clear how this finding can emerge from a model in which the prosodic and phonemic structures of an utterance are independently assessed.

Bayesian accounts of speech perception, however, are in principle compatible with this result and, more generally, with an approach in which the phonological structure of an utterance depends on the establishment of an abstract prosodic structure. In the Bayesian framework, the goal of the listener can be conceptualized as explaining away the data (the phonetics of the utterance) by selecting what is most likely to have generated them. Listeners evaluate how well a given phonological (both segmental and prosodic) representation accounts for the phonetics of the utterance based on the likelihood that this representation would yield the utterance. The key aspect is that the analysis of the utterance in terms of its segmental and prosodic composition is done simultaneously and in consideration of the influence that one imposes on the other. In this approach, the distinction between bottom-up and top-down processing is blurred because some form of loop takes place between the data and linguistic knowledge: The data give rise to a set of hypotheses which, in turn and based on a generative mechanism, can be used to evaluate how well they predict the data. Furthermore, expectations against which the data are evaluated can be adjusted to incorporate contextual information, a powerful mechanism to account for the fact that the same phonetic cue can give rise to different percepts across contexts.<sup>44,45</sup>

The notion of explaining away the data based on a hypothesized prosodic structure has also been proposed to account for the role of prosody in establishing the syntactic structure of a sentence, a topic the next section is devoted to.

## **THE CONTRIBUTION OF PROSODY TO SYNTACTIC STRUCTURE**

The meaning of a sentence depends not only on the identity of its linguistic elements but

also on how these relate to one another. Words have syntactic properties and the grammar of the language imposes constraints and restrictions on the way words relate to one another. However, the existence of constituent-structure ambiguities, i.e., strings of lexical elements that can syntactically relate to one another in multiple ways, suggests a role for prosody in mitigating ambiguity in the structure associated with a string of linguistic elements (see Ref. [46] for an extended review of such ambiguities).

The large majority of this work has focused on the role of prosodic phrasing and the location of constituent boundaries. Many accounts that have been proposed to explain the role of prosody in establishing a sentence's syntactic structure are limited to the role of prosodic constituency. Prosodic constituency is closely related but not isomorphic to syntactic constituency. A theory of how speakers determine prosodic phrasing of a sentence given its syntax must account for both the alignment and the discrepancy between the two structures. A number of such accounts have been proposed.<sup>20,47-52</sup> However, research on how listeners make use of the prosodic characteristics of the utterance in order to derive its syntactic structure has not always built on these insights.

The accounts that have been proposed can be divided into two broadly defined approaches.<sup>53</sup> According to one such approach, listeners derive syntactic-constituent boundaries directly from the phonetics that result from prosodic-phrase boundaries. There is little to no attempt to relate the interpretation of prosodic phrasing to the factors that affected its production. Furthermore, these acoustic cues are interpreted locally. This view is embodied in the 'Anti-attachment' hypothesis proposed by Watson and Gibson.<sup>52</sup> According to this hypothesis, the presence of the prosodic boundary is interpreted as "a strong cue not to attach the upcoming word to the last potential attachment site before the boundary" (p. 285). This strategy is valid, the authors argue, because it reflects the dependency between intonational-phrase boundaries and syntactic constituency. A similar view is held by Snedeker and colleagues.<sup>3,54</sup> This class of accounts is rooted in cue-integration approaches, which dominate research on visual and auditory perception.<sup>55</sup> Acoustic

correlates associated with prosodic boundaries are viewed as probabilistic cues to syntactic structure, which can be combined with other cues such as the syntactic properties of the linguistic elements. The role of these local cues is assumed to be stronger when the contribution of other morpho-syntactic cues is small and therefore, is best observed in the lexically ambiguous sentences usually used in this research.

This position has been criticized on theoretical as well as empirical grounds. First, Ladd<sup>8</sup> warns against assuming that an abstract structure such as syntax can be directly encoded in the phonetics of an utterance because the relationship between syntax and phonetics is mediated by a prosodic structure, which itself reflects influences of other aspects of the sentence. Looking for the marking of syntax in the phonetics is akin to assuming that the semantic contrast between a beach and a peach resides in the delay between the burst of the initial stop consonant and the onset of the vocal-fold vibrations. Second, there is good empirical evidence, much of it reported by Clifton, Carlson, and Frazier, that prosodic boundaries and their phonetic correlates are not interpreted locally but in relation with other aspects of the utterance's prosodic structure. For example, the effect that a boundary at a point of structural ambiguity has on attachment at that site depends on the presence and/or size of other boundaries in the utterance.<sup>56</sup> Thus, using sentences such as "Susie learned that Bill telephoned when John visited", the tendency to attach the clause "When John visited" to the most recent verb (resulting in an interpretation in which "Bill telephoned after John visited") decreases with the presence of an intermediate phrase boundary before "when John visited", but significantly more so when this break is the only one in the sentence compared to when a break is also present before "that Bill telephoned".

When phrased in terms of cue-based approach, one in which prosodic breaks function as cues to syntactic constituency, the phenomenon just described is difficult to account for because the prosodic break inserted before the relative clause "that Bill telephoned" is not predicted to have any impact on deciding between the two syntactic constituencies:

Indeed, the attachment of the relative clause remains the same in each of the interpretations considered (but see Ref. [54] for a diverging opinion).

Clifton and colleagues initially explained their findings by proposing that listeners interpret a prosodic boundary with respect to the presence and size of preceding boundaries at syntactically relevant sites. However in more recent work,<sup>53,57</sup> these authors have moved away from a (relative) cue-based approach. Instead, they advocate for a view dubbed the ‘rational speaker hypothesis’, in which listeners rely on what they know to be the linguistic and non-linguistic constraints on the speaker’s production of prosodic structures to derive the underlying causes of the observed prosodic structure. This approach assumes an intimate dependency between the production of prosodic structure and its interpretation. This view is also grounded in a Bayesian approach of perception and interpretation in which observers/listeners infer the underlying structures or hypotheses based on how well the hypotheses predict the observed data. Importantly here, observers fit multiple and interdependent hypotheses simultaneously. Thus, listeners explain away the prosodic structure of the utterance by attributing it to the linguistic structures and other non-linguistic factors known to affect it (i.e., the phonological, syntactic, and information structures). Because attribution to all of these factors is done simultaneously, the explanation for a given prosodic event in terms of, e.g., syntactic constituency, is not expected to remain the same if its presence can be explained by the influence of other factors, e.g., phonological constituent size. For instance, while a prosodic-phrase boundary before the phrase “last night” in “Susie learned that Bill telephoned last night” increases the tendency to interpret the last phrase as attaching to the main clause, i.e., “Susie learned”, the presence of the same boundary before the longer phrase “last night after the meeting” has no such effect.<sup>57</sup> This is because, the authors argue, the prosodic-phrase boundary is attributed to the phonological size of the phrase and not to its structural relationship with the immediately preceding attachment site.

Contrary to the cue-based approach to the role of prosodic phrasing on the retrieval of syntax, the rational speaker hypothesis recognizes the role that the prosodic structure per se plays. The structure can be thought of as an independent representation over which linguistic and non-linguistic forces exert their influence simultaneously. Furthermore, this approach acknowledges that the question of how listeners use prosody to derive structural information cannot be addressed independently of the role of prosody to the retrieval of phonological and semantic information. Indeed, listeners are considering all linguistic structures simultaneously when ‘explaining’ the prosodic structure of the utterance. Under this approach, the key question is how listeners determine which structure(s) to attribute the phonetic evidence to. This is an area of research that holds much promise (see Ref. [58] for a laudable step in this direction).

## **THE CONTRIBUTION OF PROSODY TO INFORMATION STRUCTURE**

The information structure of an utterance captures the properties by which the utterance relates to the context in terms of the discourse status of its content and the attentional and intentional states of the discourse participants. This structure is intimately connected to what is called common ground, i.e., “a set of propositions that a given conversational participant supposes to be mutually agreed upon for the purposes of the conversation” (Ref. [20], p. 9). Information structure comprises a number of dimensions or features, which can be expressed via the use of grammatical devices, including prosody.<sup>59</sup>

As pointed out earlier, investigation on the contribution of prosody to a sentence’s information structure departs from research on the contribution of prosody to phonology or syntax. Here, a role for prosody in conveying aspects of the sentence’s information structure is taken for granted; research focuses on how prosodic elements may reveal this structure.

The role of prosody, and accent placement in particular, in relating a sentence to its discourse context can be readily demonstrated. For instance, the utterance “I

would love some COFFEE” (with an accent on “coffee”) is ill-fitted to the preceding question “Would you like some coffee?”<sup>60</sup> The role of prosody is also evident in the sentence “I asked the teacher who left”, whose interpretation changes whether the *wh*-word receives a pitch accent or not,<sup>4</sup> or in phrases such as “the propeller of the plane that the mechanic was carefully examining”, in which the attachment of the relative clause to the first or second noun in the complex noun phrase is modulated by whether the first or second noun received a pitch accent.<sup>61,62</sup> In the case of the last example, the impact of the pitch-accent location on the attachment site of the relative clause has been explained by resorting to the notion of salience: A relative clause that acts as a restrictive modifier of the noun phrase it relates to is more likely to modify a salient, as opposed to non-salient, constituent.<sup>63</sup> This salience may in turn be related to the notion of ‘predictability’, a construct that some have proposed to account for the speaker’s choice to render a word more or less acoustically salient.<sup>64-67</sup> Proponents of the predictability account of prominence, however, rarely connect their work to the information-structure literature. It is therefore unclear whether word predictability is a concept that encompasses what linguists have described as information structure, or if it is a dimension that captures variability left unexplained by information structure.

A difficulty in navigating the literature on prosody and information structure resides in the terminology, often confusing, that scholars have used to characterize the dimensions of information structure. Here, I follow Kruijff-Korbayova and Steedman<sup>68</sup> in characterizing the structure along two independent dimensions. The first one differentiates the part of the utterance that relates it to the current discourse and the interlocutors’ common ground (i.e., theme) from the part of the utterance that advances the discourse (rheme). The second dimension is one that highlights a contrast between the actual content of a part of the utterance and those alternatives that the context (explicitly or implicitly) makes available. This is referred to as ‘contrast’ or ‘kontrast’. The notion of contrast is embedded in Rooth’s<sup>69</sup> Alternative Semantics, which proposes that

contrast introduces a presupposition of alternatives to the element in contrast, these alternatives being present in or accommodated by the context. An entity or proposition that is not in contrast is said to be in background.

A clear distinction between these two dimensions has not always been achieved. Particularly confusing is the use of the term ‘focus’, which is sometimes used to refer to contrast, as in the case of the so-called ‘narrow focus,’ and sometimes to rheme in the case of ‘broad focus.’ There is a debate whether there is a semantic or phonological distinction between ‘contrastive focus’, where the alternative set is known and bounded, and ‘non-contrastive focus,’ where it is unknown and/or unbounded.<sup>20,70</sup> Finally, the ‘given / new’ distinction, one that is frequently made in the literature, is difficult to align with the two dimensions just mentioned. The term ‘given’ is usually used to refer to the part of the sentence that has been explicitly mentioned in the current discourse and that is readily accessible. Personal pronouns are grammatical devices used to make reference to such given, highly accessible entities. Thus, given constituents are equivalent to what we refer here as themes, and new constituents, rhemes. When a personal pronoun carries a pitch accent, its referent is set in contrast with contextually-defined alternatives. Unfortunately, the terms ‘given’ and ‘new’ have also been used to refer to those words conveying contrast or no contrast.<sup>68</sup>

How may pitch-accent placement mark the contrast/no-contrast and theme/rheme dimensions? In order to understand the relationship between prosody and information structure, researchers have often induced these discourse distinctions in a sentence by placing it in the context of a short narrative or a question. For example, a context question of the form “What’s happening?” is assumed to be followed by an ‘out-of-the-blue’, all-rheme, utterance. More specific questions introduced by a *wh*-word, on the other hand, create a distinction between the part of the utterance that answers the question and therefore adds information to common ground (the rheme) and the part of the utterance that is already part of common ground because it is stated in the question (the theme). For instance, the question “Who fried an



omelet?” preceding the utterance “Damon fried an omelet” assigns the role of theme to the phrase “fried an omelet” and that of rheme to “Damon”.<sup>71</sup>

A few aspects of English intonation and information structure are uncontroversial: In an all-rheme utterance, all non-pronominal noun phrases must carry pitch accents, regardless of their position in the sentence; verbs, on the other hand, may or may not carry a pitch accent. When a sentence includes a theme that was mentioned in the immediately preceding sentence and that is therefore salient in the current discourse (i.e., given), whether or not the theme carries an accent or not depends on its position relative to the rheme: It is always deaccented if it occurs after the rheme, but optionally accented if before it.<sup>70</sup> Themes can also carry accents if they are in a contrastive relationship with alternatives that the context has introduced. Rheme constituents are generally accented, although which of their words can or must carry an accent is debated (see below).

Thus, the rheme / theme distinction cannot be reduced to the presence vs. absence of accent. Some have proposed that accent placement conveys contrast alone and is orthogonal to the theme/rheme distinction because both themes and rhemes can carry an accent.<sup>20,72,73</sup> For others, information structure makes a three-way distinction: rhemes (new), themes (given), and contrastive constituents (‘contrastive focus’). Given constituents are generally deaccented (and optionally accented if they precede new constituents), while both new and contrastive focus constituents (or rheme and theme marked by contrast, in our terminology) carry an accent.

There is excellent evidence that the absence of a pitch accent on a noun phrase (sometimes called ‘deaccenting’) is a signal that listeners immediately integrate with (partial) segmental information to select the referent of the noun in the set of entities that are highly salient in the discourse (‘given’).<sup>74-78</sup> Conversely, the presence of a pitch accent on a noun is immediately detected. However, the information status of the accented constituent that people consider varies as a function of what entities matching the segmental information on the noun are available in the discourse and their

respective prominence: The referent of an accented noun phrase is not considered to be the most prominent entity, but it can be taken to be new (rheme) or a contrastive theme.

If, as the evidence indicates, the presence of a pitch accent does not convey the distinction between rheme and (contrastive) theme, how might it be prosodically encoded? Opinions on this question vary and, unfortunately the empirical data do not adjudicate the debate. Some have claimed that the distinction between theme and rheme is conveyed by the type of pitch accent.<sup>19,20,79</sup> There is a general agreement that H\* accents mark rhemes, i.e., elements that convey information to be added to common ground. According to Pierrehumbert and Hirschberg,<sup>19</sup> L+H\* accents specifically encodes the contrast between the element that receives the accent and others in a set or in a scale relationship with that element (see Ref. [80] for a formal definition of scale). Such an accent would be used when the speaker wants to highlight the contrast between a thematic element and its alternatives.

Steedman<sup>20</sup> also assumes a distinction between H\* and L+H\* that pertains to the information structure, but makes a very different claim. According to Steedman, any accent conveys a contrast between the accented element and its alternatives as introduced by or inferred from the context. H\* convey contrast with a set of alternatives on the rheme; by contrast, L+H\* marks contrast on a theme. Unfortunately, Steedman’s and Pierrehumbert and Hirschberg’s claims are difficult to differentiate empirically.

The contrastive aspect of the L+H\* accent on themes has been specifically evaluated.<sup>80-82</sup> For example, Ito and Speer<sup>80</sup> gave participants successive instructions to pick up objects from a large array organized in terms of object types varying in their colors (e.g., “Hang the red angel. After that, hang the BLUE angel”). They monitored people’s eye gaze to the object array upon hearing the accented prenominal color adjective, and showed that when the adjective received an L+H\* accent, people anticipated an object of the same type as the one they had just selected more often than they did when the color adjective received an H\* accent. The L+H\* pitch accent, Ito and Speer argue, was interpreted as evoking a set

composed of the group of objects represented by the most salient entity, i.e., the one most recently encountered, with the color adjective conveying the contrastive dimension. An H\* accent on the color adjective, however, did not lead people to preferentially anticipate an object from a different set from the most recent one. Rather, the H\* accent on the adjective appeared to be associated with both contrastive and non-contrastive (new) interpretations, the conclusion that Watson, Tanenhaus, and Gunlogson<sup>82</sup> reached in their own assessment of the meaning of these accents. The significance of these results, of course, depends on a phonological distinction between the two types of accents. As discussed above, this is a highly debated matter.

Embracing a radically divergent position, some scholars contend that differences in the information status of accented constituents are not conveyed by accent types, but by their relative prominence.<sup>70,72,73</sup> According to this view, and contrary to tonal distinctions such as the presence or absence of pitch accent or the type of pitch accent, relative prominence is encoded in the phonological representation of the utterance specified in the metrical grid, and can capture gradient distinctions, which in turn are phonetically realized by the constituent duration or in the height of the pitch peak. However, while Calhoun<sup>73</sup> argues for greater prominence (and higher F0 peak) for rhemes, Katz and Selkirk<sup>70</sup> propose that (contrastive) themes are more prominent (and of longer duration) than rhemes. Thus, while the proposal that information structure may be conveyed by gradient, as opposed to categorical, distinctions is intriguing, further work is warranted.

These previous remarks have assumed that theme or rheme constituents consist of single nouns. In this case, the placement of an accent is straightforward. However, constituents are likely to comprise multiple content words. How do speakers decide which content words must or may carry an accent? Conversely, how do listeners know the scope of an accented constituent if the grammar allows some content words of an accented constituent to forgo accentuation? Selkirk,<sup>48,83</sup> Gussenhoven,<sup>84</sup> and Beckman<sup>85</sup> have proposed syntactic constraints on the relationship between accent placement and theme / rheme constituency (sometimes

called ‘focus structure’). Among other things, these theories differ on whether surface structure (i.e., the linear organization of the phrases in the sentence) can constrain the mapping of the focus structure onto accent placement. As far as rhemes are concerned, an accent on an argument can project to its head, making accent on the head optional.<sup>76</sup> This projection doesn’t apply as readily to adjuncts (although see Ref. [77]). Furthermore, the position of adjuncts relative to their heads affects whether they may carry an accent.<sup>77</sup> Similar positional constraints apply to themes. A consequence of the projection of an accent to its constituent is that it becomes difficult to distinguish an accent marking only the word it falls onto from one marking a whole constituent. Some have proposed that the type of accent or other phonetic cues may disambiguate between these two interpretations, although evidence supporting the claim is equivocal.<sup>71,75</sup>

A role for prosodic phrasing in the marking of a sentence’s focus structure has been proposed by Steedman.<sup>20</sup> In his theory, prosodic phrasing, and not accent placement in interaction with syntactic constraints, marks the scope of the theme and rheme constituents. Empirical studies looking at the role of prosodic phrasing on information structure have been limited (but see Ref. 87). This is unfortunate because of the well-established role of prosodic phrasing on the marking of a sentence’s syntactic structure. Because theme and rheme constituents are not necessarily isomorphic with syntactic constituents as traditionally defined, the prosodic phrasing of a sentence may be under the influence of multiple linguistic structures simultaneously. Echoing the preceding sections, one may characterize listeners’ task as one in which the constraints that these structures can exert on prosodic phrasing must be considered simultaneously in order to explain the utterance’s prosodic structure.

## CONCLUSION

The prosodic structure that speakers choose to adopt to convey a sentence’s meaning is under the influence of a multitude of factors. These factors cut across the traditional levels of linguistic structures. Listeners use prosody in conjunction with the morpho-syntactic elements

of the sentence in order to establish these structures. Recent investigation on how this process is achieved has revealed the importance of considering the influence of these factors simultaneously because of their interdependence on the prosodic structure of a sentence and, therefore, on the resulting phonetics. Thus, attempts to find reliable phonetic marking of words boundaries or syntactic constituency may not be successful strategies because the phonetics associated with the prosodic structure depends on other factors, each one having its own influence alone or in conjunction with others. The presence or absence of, e.g., a prosodic boundary cannot be reliably interpreted as local marking of a syntactic-constituent edge.

A promising approach has framed the issue listeners face as a data-explanation problem: Given the options that a speaker faces in choosing a prosodic structure to render the linguistic structures associated with the meaning the speaker wishes to convey, which of these structures (the hypotheses) may have given rise to the utterance? Segmentally-related information, along with pitch, intensity, and rhythm, serve as clues to simultaneously infer these structures, thereby narrowing down the number of hypotheses. This view characterizes perception and interpretation as 'reverse engineering' solutions where production and perception of prosody form two sides of the same coin.

## FURTHER READINGS

Speer S, Blodgett A. Prosody. In: Traxler M, Gernsbacher MA, eds. *Handbook of psycholinguistics (2<sup>nd</sup> edition)*. San Diego, CA: Academic Press; 2006, 505-537.

Carlson K. How prosody influences sentence comprehension. *Language and Linguistics Compass* 2009, 3:1188-1200

## REFERENCES

1. Dilley LC, Mattys SL, Vinke L. Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language* 2010, 63:274-294.
2. Schafer AJ, Speer SR, Warren P. Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In: Trueswell JC, Tanenhaus MK (eds.) *Approaches to studying world-situated language use*. Cambridge, The MIT Press, 2005, 209-225.
3. Snedeker J, Trueswell J. Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language* 2003, 48:103-130.
4. Schafer A, Carlson K, Clifton Jr C, & Frazier L. Focus and the interpretation of pitch accent: Disambiguating embedded questions. *Language and Speech* 2000, 43:75-105.
5. Ward G, Hirschberg J. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 1985, 61:747-776.
6. Kurumada C, Brown M, Bibyk S, Pontillo DF, Tanenhaus MK. Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition* 2014, 133:335-342.
7. Goldsmith JA. *Autosegmental and metrical phonology*. Oxford, UK: Blackwell; 1990.
8. Ladd DR. *Intonational phonology*. Cambridge: Cambridge University Press; 2008.
9. Shattuck-Hufnagel S, Turk AE. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 1996, 25:193-247.
10. Ladd DR, Morton R. The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics* 1997, 25:313-342.
11. Ladd DR, Schepman A. "Sagging transitions" between high pitch accents in English: Experimental evidence. *Journal of Phonetics* 2003, 31:81-112.
12. Dilley LC, Brown M. Effects of pitch range variation of F0 extrema in an imitation task. *Journal of Phonetics* 2007, 35:523-551.
13. Dilley LC. Pitch range variation in English tonal contrasts: continuous or categorical? *Phonetica* 2010, 67:63-81.
14. Gussenhoven C. Discreteness and gradience in intonational contrasts. *Language and Speech* 1999, 42:283-305.
15. Gussenhoven C. Experimental approaches to establishing discreteness of intonational contrasts. In: Sudhoff S, Lenertová D, Meyer R, Pappert S., Augurzky P, Mleinek I, Richter N, Schliesser J, eds *Methods in Empirical Prosody Research*. Berlin, Walter de Gruyter, 2006, 321-334.
16. Gussenhoven C. *On the grammar and semantics of sentence accents*. Dordrecht: Foris; 1984.

17. Bolinger D. *Intonation and its parts*. Palo Alto, CA: Stanford University Press; 1986.
18. Bolinger D. *Intonation and its uses*. Palo Alto, CA: Stanford University Press; 1989.
19. Pierrehumbert J, Hirschberg J. The meaning of intonational contours in the interpretation of discourse. In: Cohen PR, Morgan J, Pollack ME, eds. *Intentions in communication*. Cambridge, MA: The M.I.T. Press; 1990, 271-311.
20. Steedman M. The surface-compositional semantics of English intonation. *Language* 2014, 90:2-57.
21. Hirschberg J, Ward G. The influence of pitch range, duration, amplitude, and spectral features on the interpretation of L\*+H LH%. *Journal of Phonetics* 1992, 20:241-251.
22. Pierrehumbert, JB, Steele SA. Categories of tonal alignment in English. *Phonetica*, 1989, 46:181-196.
23. Dilley LC, Heffner CC. The role of F0 alignment in distinguishing intonation categories. *Journal of Speech Sciences* 2013, 3:3-67.
24. Braun B, Kochanski G, Grabe E, Rosner BS. Evidence for attractors in English intonation. *Journal of the Acoustical Society of America* 2006, 119: 4006-4015
25. Pierrehumbert JB. *The Phonology and Phonetics of English Intonation*. PhD Thesis, MIT; 1980.
26. Beckman ME, Pierrehumbert JB. Intonational structure in Japanese and English. *Phonology* 1986, 3:255-309.
27. Beckman ME, Hirschberg J, Shattuck-Hufnagel S. The original ToBI system and the evolution of the ToBI framework. In: Jun SA, ed *Prosodic Models and Transcription: Towards Prosodic Typology*. Oxford, Oxford University Press, 2005, 9–54.
28. Mattys SL, Melhorn JF. Sentential, lexical, and acoustic effects on the perception of word boundaries. *Journal of the Acoustical Society of America* 2007, 122:554-567.
29. Miller JL. Mapping from acoustic signal to phonetic category: Internal category structure, context effects and speeded categorization. *Language and Cognitive Processes*, 2001, 16:683-690.
30. Lehiste I. An acoustic–phonetic study of internal open juncture. *Phonetica* 1960 5:5-54.
31. Nakatani LH, Dukes KD. Locus of segmental cues for word juncture. *The Journal of the Acoustical Society of America* 1977, 62:714-719.
32. Cole RA, Jakimik J, Cooper WE. Segmenting speech into words. *The Journal of the Acoustical Society of America* 1980, 67: 1323-1332.
33. Cho T. Laboratory phonology. In Kula, NC, Botma ED, Botma B, Nasukawa K. (eds.), *The Continuum companion to phonology*. London: Continuum International, 2011, 343-368.
34. Salverda AP, Dahan D, McQueen JM. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 2003, 90: 51-89.
35. Christophe A, Peperkamp S, Pallier C, Block E, Mehler J. Phonological phrase boundaries constrain lexical access I. adult data. *Journal of Memory and Language* 2004, 51:523-547.
36. Cho T, McQueen JM, Cox EA. Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics* 2007, 35:210-243.
37. Salverda AP, Dahan D, Tanenhaus MK, Crosswhite K, Masharov M, McDonough J. Effects of prosodically modulated sub-phonetic variation on lexical competition. *Cognition* 2007, 105: 466-476.
38. Dilley LC, McAuley JD. Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language* 2008, 59:294-311.
39. Brown M, Salverda AP, Dilley LC, Tanenhaus MK. Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin & Review* 2011, 18:1189-1196.
40. Heffner CC, Dilley LC, McAuley JD, Pitt MA. When cues combine: How distal and proximal acoustic cues are integrated in word segmentation. *Language and Cognitive Processes* 2013, 28:1275-1302.
41. Mattys SL, White L, Melhorn JF. Integration of multiple segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General* 2005, 134:477-500.
42. Dilley LC, Pitt MA. Altering context speech rate can cause words to appear or disappear. *Psychological Science* 2010, 21:1664-1670.
43. Baese-Berk MM, Heffner CC, Dilley LC, Pitt MA, Morrill TH, McAuley JD. Long-term temporal tracking of speech rate affects spoken-word recognition. *Psychological Science* 2014, 25:1546-1553.
44. McMurray B, Jongman A. What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review* 2011, 118:219-246.
45. Kleinschmidt DF, Jaeger TF. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review* 2015, 122:148-203.
46. Price PJ, Ostendorf M, Shattuck-Hufnagel S, Fong C. The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America* 1991, 90:2956-2970.
47. Selkirk E. The syntax-phonology interface. In Goldsmith J, Riggle J, Yu A, eds. *The Handbook of Phonological Theory*. Oxford: Blackwell; 2011, 435-483.
48. Selkirk EO. *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: The MIT Press; 1984

49. Steedman M. Structure and intonation. *Language* 1991, 67:260-296.
50. Steedman M. Information structure and the syntax-phonology interface. *Linguistic Inquiry* 2000, 31:649-689.
51. Watson D, Gibson E. The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes* 2004, 19:713-755.
52. Watson D, Gibson E. Intonational phrasing and constituency in language production and comprehension. *Studia Linguistica* 2005, 59:279-300.
53. Frazier L, Carlson K, Clifton Jr C. Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences* 2006, 10:244-249.
54. Snedeker J, Casserly E. Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. *Language and Cognitive Processes* 2010, 25:1234-1264.
55. Massaro DW. *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: The MIT Press; 1998.
56. Clifton Jr C, Carlson K, Frazier L. Informative prosodic boundaries. *Language and Speech* 2002, 45:87-114.
57. Clifton Jr C, Carlson K, Frazier L. Tracking the what and why of speakers' choices: Prosodic boundaries and the length of constituents. *Psychonomic Bulletin & Review* 2006, 13:854-861.
58. Brown M, Salverda AP, Gunlogson C, Tanenhaus M K. Interpreting prosodic cues in discourse context. *Language, Cognition and Neuroscience* 2015, 30:149-166.
59. Féry C, Krifka M. Information structure. In: van Sterkenburg P, ed. *Unity and diversity of languages*. Amsterdam: John Benjamins; 2008, 123-136.
60. Wagner M, Watson DG. Experimental and theoretical advances in prosody: a review. *Language and Cognitive Processes* 2010, 25:905-945.
61. Speer SR, Crowder RG, Thomas LM. Prosodic structure and sentence recognition. *Journal of Memory and Language* 1993, 32:336-358.
62. Schafer A, Carter J, Clifton Jr C, Frazier L. Focus in relative clause construal. *Language and Cognitive Processes* 1996, 11:135-163.
63. Lee EK, Watson DG. Effects of pitch accents in attachment ambiguity resolution. *Language and Cognitive Processes* 2011, 26:262-297.
64. Aylett M, Turk A. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 2004, 47:31-56.
65. Watson DG, Arnold JE, Tanenhaus MK. Tic tac toe: Effects of predictability and importance on acoustic prominence in language production. *Cognition* 2008, 106:1548-1557.
66. Bell A, Brenier JM, Gregory M, Girand C, Jurafsky D. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 2009, 60:92-111
67. Arnold JE, Kaiser E, Kahn JM, Kim LK. Information structure: linguistic, cognitive, and processing approaches. *WIRE Cognitive Science*, 2013, 4:403-413.
68. Kruijff-Korbayová I, Steedman M. Discourse and information structure. *Journal of Logic, Language and Information* 2003, 12:249-259.
69. Rooth M. A theory of focus interpretation. *Natural Language Semantics* 1992, 1:75-116.
70. Katz J, Selkirk E. Contrastive focus vs. discourse-new: evidence from phonetic prominence in English. *Language* 2011, 87:771-816.
71. Breen M, Fedorenko E, Wagner M, Gibson E. Acoustic correlates of information structure. *Language and Cognitive Processes* 2010, 25:1044-1098.
72. Calhoun S. How does informativeness affect prosodic prominence? *Language and Cognitive Processes* 2010, 25:1099-1140.
73. Calhoun S. The theme/rheme distinction: Accent type or relative prominence? *Journal of Phonetics* 2012, 40:329-349.
74. Bock JK, Mazzella, JR. Intonational marking of given and new information: Some consequences for comprehension. *Memory & Cognition* 1983, 11:64-76.
75. Nooteboom SG, Kruyt JG. Accents, focus distribution, and the perceived distribution of given and new information: An experiment. *The Journal of the Acoustical Society of America* 1987, 82:1512-1524.
76. Birch S, Clifton Jr C. Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech* 1995, 38:365-391.
77. Birch S, Clifton Jr C. Effects of varying focus and accenting of adjuncts on the comprehension of utterances. *Journal of Memory and Language* 2002, 47:571-588.
78. Dahan D, Tanenhaus MK, Chambers CG. Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language* 2002, 47:292-314.
79. Jackendoff R. *Semantic interpretation in generative grammar*. Cambridge, MA: The MIT press; 1972.
80. Ito K, Speer SR. Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language* 2008, 58:541-573.
81. Weber A, Braun B, Crocker MW. Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech* 2006, 49:367-392.

82. Watson DG, Tanenhaus MK, Gunlogson, CA. Interpreting pitch accents in online comprehension: H\* vs. L H\*. *Cognitive Science* 2008, 32:1232-1244.
83. Selkirk EO. On derived domains in sentence phonology. *Phonology* 1986, 3:371-405.
84. Gussenhoven C. *A semantic analysis of the nuclear tones of English*. Indiana University Linguistics Club; 1983.
85. Beckman ME. The parsing of prosody. *Language and Cognitive Processes* 1996, 11: 17-68.
86. Welby P. Effects of pitch accent position, type, and status on focus projection. *Language and Speech* 2003, 46:53-81.