



Published in final edited form as:

*Lang Cogn Process.* 2010 January 1; 25(7-9): 905–945. doi:10.1080/01690961003589492.

## Experimental and theoretical advances in prosody: A review

Michael Wagner<sup>1</sup> and Duane G. Watson<sup>2</sup>

<sup>1</sup>Cornell University & McGill University

<sup>2</sup>University of Illinois at Urbana-Champaign

### Abstract

Research on prosody has recently become an important focus in various disciplines, including Linguistics, Psychology, and Computer Science. This article reviews recent research advances on two key issues: prosodic phrasing and prosodic prominence. Both aspects of prosody are influenced by linguistic factors such as syntactic constituent structure, semantic relations, phonological rhythm, pragmatic considerations, and also by processing factors such as the length, complexity or predictability of linguistic material. Our review summarizes recent insights into the production and perception of these two components of prosody and their grammatical underpinnings. While this review only covers a subset of a broader set of research topics on prosody in cognitive science, they are representative of a tendency in the field toward a more interdisciplinary approach.

---

Prosody can be roughly defined as a level of linguistic representation at which the acoustic-phonetic properties of an utterance vary independently of its lexical items. This admittedly vague definition encompasses a variety of phenomena: emphasis, pitch accenting, intonational breaks, rhythm, and intonation. Some aspects of the prosody of an utterance are mere reflexes of processing during speech production, others have been conventionalized and encode grammatical information. In this article, we focus on two aspects of prosody that are central in current research: boundary strength and relative prominence among words. These two components of prosody and the precise way in which various factors influence them have become an important area of research in recent years in various fields, including semantics, syntax, computational linguistics, and psycho- and neuro-linguistics.

A little over ten years before the publication of this paper, Shattuck-Hufnagel and Turk (1996) and Cutler, Dahan, and van Donselaar (1997) each wrote comprehensive reviews of work on prosody in both linguistics and psychology up to that time. Since then, there has been an explosion in the number of studies investigating the role of prosody in cognition and linguistics, as well as improvements in techniques for examining prosody. In this article, we attempt to pick up where those two papers left off and review some of the recent work in prosody. Because it would be impossible to provide a review of the entire field in so small a space, we have tried to cover areas that lie at the interface of theoretical and experimental approaches to prosody, and at the interface of linguistics, psycholinguistics, and computational linguistics.

In so doing, we focus on two aspects of prosody that are central in current research: boundary strength and the relative prominence between words. Within these domains, similar questions have arisen in recent years: What is the relationship between prosody,

discourse, and syntactic structure? What are the acoustic correlates of prosody? What information does prosody convey and are the cognitive processes that underlie it primarily production-centered or comprehension-centered?

## What is Prosody?

Every utterance in human speech comes with certain properties that are referred to as its ‘prosody’. One way to define ‘prosody’ is by its function: ‘Prosody’ is often used to refer to those phonetic and phonological properties of speech that are crucially *not* due to the choice of lexical items, but rather depend on other factors such as how these items relate to each other semantically and/or syntactically, how they are grouped rhythmically, where the speaker places emphasis, what kind of speech act the utterance encodes, whether turn taking in conversation is being negotiated, and they can reflect the attitude and emotional state of the speaker. While these factors can also determine the choice of lexical material, they can affect the signal directly without any mediation by a lexical morpheme with segmental content, and it is this kind of information that is often referred to as the prosody of an utterance (cf. discussions in Ladd 2008, Ferreira 2002).

Another, quite different way to define ‘prosody’ is by its form, which includes its phonetic and phonological substance. A common definition of prosody is that it comprises the ‘suprasegmental’ (Lehiste, 1970) aspects of the speech stream, i.e., properties such as syllable structure, intonation, and reflexes of prosodic structure, which are acoustically reflected in fundamental frequency, duration, and intensity. Both of these definitions, the one that puts more emphasis on the function of prosody and the one emphasizing its form, have their virtues and flaws. An issue with the first definition is that it excludes suprasegmental properties in the lexicon, such as lexical tone, syllable structure, and lexical stress, yet many researchers understand the term ‘prosody’ as including these. An issue with the second definition is that it presupposes an analysis that divides the information in the speech stream cleanly into a segmental and prosodic component, but at the signal level, there is no separation of prosodic and segmental information. Both use the same channel and encode information by the same phonetic correlates, e.g., fundamental frequency, duration, and intensity. Whichever definition one may favor, boundary strength and prominence, the two topics on which the remainder of this review will focus, would count as prosody under either definition.

## Boundaries

An utterance of more than two words in it often has a perceptible sub-grouping (Lehiste, 1973). Prosodic grouping can be produced and perceived even in the absence of identifiable words (cf., Larkey, 1983; de Pijper & Sanderman, 1994). Thus, perceived grouping is not simply due to the semantic relationship or the co-occurrence frequency between words, although of course these factors might add to, or be confounded with, the effects of prosody on perceived grouping in actual speech. Below we discuss the main acoustic correlates of prosodic grouping—duration, fundamental frequency and intensity—and how they signal grouping and boundaries, and we discuss the nature of their relationship to syntax and language processing.

## Phonetic and Phonological Correlates

**Duration**—Lehiste (1973) used ambiguity resolution to study acoustic correlates of prosodic boundaries and to understand the extent to which these correlates reflect syntactic bracketing. Lehiste (1973) identified duration as the most reliable cue in disambiguating syntactic structures based on their bracketing. The main durational cues affecting boundary strength perception are pre-boundary lengthening, pauses, and domain-initial strengthening.

Klatt (1975) showed that segments are lengthened preceding boundaries, even in the absence of pauses. Pre-boundary lengthening has been shown to correlate closely with the strength of the following boundary (Wightman et al., 1992; Price et al., 1991; Shattuck-Hufnagel & Turk, 1996; Byrd & Saltzman 1998). Pre-boundary lengthening correlates with other acoustic cues that reflect that articulatory gestures of segments preceding boundaries are spatially more extreme, i.e., hyperarticulated (Edwards et al., 1991; Fougeron & Keating, 1997; Byrd & Saltzman, 2003), and are spaced further apart (Byrd & Saltzman 2003). Final lengthening affects the syllable left-adjacent to the boundary, and, according to Berkovits (1994), extends to the closest stressed syllable. Turk and White (1999) found lengthening in all material from the boundary to the rime of the syllable carrying main stress. The degree of pre-boundary lengthening of a segment decreases with the distance from the prosodic boundary (Byrd et al., 2006).

Closely related to pre-boundary lengthening are the presence and length of pauses at boundaries. Pre-boundary lengthening and pause duration are closely related and have been argued to contribute to a single percept of pause or juncture, and listeners report hearing pauses even when there are no unfilled pauses in the signal (Martin, 1970). O'Malley et al. (1973) found evidence that different amounts of pause duration can code different degrees of boundary, a finding that was confirmed in Fant and Kruckenberg (1996).

Apart from final lengthening and pausing, a third duration-related phenomenon is *domain-initial strengthening*. Jun (1993), Fougeron and Keating (1997), Lavoie (2001), Cho (2002), and Keating et al. (2003) show that the phonetic realization of segments depends on the strength of a preceding boundary. Evidence from production experiments using electro-palatography suggests that initial strengthening increases cumulatively with the strength of the preceding prosodic boundary. Keating et al. (2003) provides evidence that domain-initial strengthening occurs cross-linguistically in typologically distinct languages with very different prosodic systems. Even languages that are not stress-based such as French, Korean, and Taiwanese, show very similar patterns of domain-initial strengthening to English, although they are quite different when it comes to pitch related cues to phrasing, suggesting that domain initial-strengthening is a general reflex of prosodic organization. In addition to lengthening segments at the beginning of prosodic domains, new segments can also be inserted in order to strengthen the beginning of a domain. Dilley et al. (1996) and Redi and Shattuck-Hufnagel (2001) show evidence that glottal stop insertion is more likely at stronger prosodic domain breaks compared with weaker boundaries.

**Fundamental frequency**—A second important acoustic dimension in cueing prosodic boundaries is fundamental frequency and its perceptual correlate pitch. There are two major sources of information on prosodic phrasing in the pitch curve of an utterance: pitch excursions at prosodic boundaries and the scaling of pitch accents relative to each other.

The first type of pitch cue for boundaries are pitch events that occur at the edges of strong prosodic domains. They are commonly analyzed as boundary tones (following Pierrehumbert, 1980). These boundary tones are aligned relative to the end or beginning of a prosodic domain. Some boundary tones, especially sentence-final ones, are often linked to semantic or pragmatic meaning, and are sometimes treated as intonational morphemes in their own right (Bolinger, 1965; Gussenhoven, 1984; Pierrehumbert & Hirschberg, 1990; Gussenhoven, 2004). It is not clear, however, whether every pitch event at a boundary can be analyzed in this way.<sup>1</sup>

---

<sup>1</sup>Boundary tones also play an important role in negotiating turn-taking. We will not discuss these discourse functions of boundary tones in this review, because our main focus is how boundaries are signaled.

A second type of pitch cue to prosodic phrasing is the relative scaling of pitch accents within an utterance. Pitch accents on individual words are often scaled relative to preceding ones, and the precise scaling pattern depends on the prosodic phrasing (and other factors, e.g. focus, see below). Ladd (1988) and Féry and Truckenbrodt (2004) looked at the following type of coordination structure in English and German respectively, where A, B, and C stand in for sentences:

(1a) A but (B and C)

(1b) (A and B) but C

The pitch accent scaling distinguishes the two types of structures. In structures of type (1a), conjunct C has a lower F0 than B, and B in turn has a lower F0 than A. The pitch level goes down from accent to accent. In structures of type (1b), on the other hand, C and B are at about the same level, but both are set to a lower pitch compared to A.

This contrast in pitch scaling was used to argue for a prosodic representation that reflects the syntactic difference between (2a) and (2b). For example, Ladd (1988) proposes to explain the difference in pitch scaling by a hierarchical metrical representation that allows a recursive nesting of intonational phrases. Within each level of coordinate structure, conjuncts are downstepped relative to the preceding conjunct. In structures of type (1b), conjunct C is downstepped relative to the first conjunct (A and B). This has the effect that the pitch of an accent in C is lower than the maximal pitch in (A and B), but it is not lower than the pitch in a preceding conjunct, in this case B. Further evidence for this kind of scaling is presented in van den Berg et al. (1992), who propose that the pitch level of entire domains containing accents can be downstepped relative to preceding domains, using reference lines.

Related to the relative scaling of pitch accents are resets. They are perceived as discontinuities and are interpreted as a cue for strong boundaries (de Pijper & Sanderman, 1994). Truckenbrodt (2002) argued that phrase-initial F0 reset, i.e., a resetting of the reference pitch line, acts as an additional correlate of intonational phrases in certain dialects of German, and is used in very much the same way as boundary tones to signal phrasing.

There are other cues related to the voice source that correlate with prosodic boundaries. A common phenomenon is voice quality changes at the end of a prosodic domain. For example, creaky voice is a common cue to end a prosodic domain in English and many other languages, as observed already in Lehiste (1973).

**Intensity**—A third source of information for prosodic boundaries apart from fundamental frequency and duration is intensity, although this cue has been less studied as a signal for boundaries. Kim et al. (2004) report that some speakers in the Switchboard corpus show a difference between two boundary types differing in strength (intermediate vs. intonational phrase in ToBI terms), such that the stronger boundary was associated with lower intensity of the material preceding the boundary. This difference, however, was not consistent across speakers.

### Gradient or Categorical?

A common assumption in the linguistic literature is that prosodic boundaries can be categorized according to a very limited inventory of boundary types that are organized in a ‘prosodic hierarchy’. The prosodic hierarchy proposed in Selkirk (1986) includes six categories: the utterance, intonational phrase, phonological phrase, phonological word, foot and syllable. Each utterance contains at least one instance of each category, each category higher up on the hierarchy consists of one or more elements of the next lower category.

Different boundary strengths are interpreted as categorical phonological differences between boundary types.

This assumption of a prosodic hierarchy is shared by the ToBI annotation system of American English (Silverman et al., 1992). It assumes 3 categories, intonation phrase, intermediate phrase, and word—lower prosodic categories are not encoded in the ToBI labeling system since the annotation scheme does not label within-word prosody. The ToBI system has proven useful in making prosodic information available in speech corpora. ToBI was originally developed to transcribe the intonation of American English but has since been adapted to transcribe a wide variety of languages (Jun, 2005).

A problem for the categorical view of boundary strength, however, is that often when boundaries of different strengths can be discerned, the differences are quantitative rather than qualitative. Experiments based on pitch accent scaling and on the various durational cues to prosodic boundary strength reviewed above suggest that many correlates of boundary strength show gradient and cumulative effects. Another source of evidence comes from durational evidence for a relative scaling of the strength of boundaries relative to earlier produced boundaries in production (Wagner, 2005). Some researchers conclude that we need to distinguish between intonational phrases of different strengths above and beyond the categorical distinctions that have been proposed (cf., discussion in Ladd, 2008; Kim et al., 2004).

Since Price et al. (1991), the ToBI system includes a boundary strength annotation, the *break index*, which is based on boundary type differences. Syrdal & McGory (2000), however, found poor inter-labeler agreement in ToBI with respect to the precise boundary type but high agreement with respect to whether or not there is a boundary. And according to de Pijper and Sanderman (1994), both naïve and trained listeners have very similar and very reliable intuitions about *relative* boundary strength, but are not very reliable at categorizing boundaries. A recurring theme in the literature on investigating prosodic boundaries is that researchers decide to annotate whether or not a boundary is present rather than trying to distinguish the precise ToBI type (Pijper & Sanderman, 1994; Watson & Gibson, 2004b).

An alternative annotation system that is compatible with relative notions of boundary strength and prominence is the *Rhythm and Pitch* annotation system developed and tested in Dilley and Brown (2005) and Dilley et al. (2006). This system dissociates the precise nature of the tonal implementation from perceived grouping and prominence relations, and is thus more apt to account for gradient and relative distinctions that are not accompanied by categorical differences. A study comparing inter-labeler agreement of RaP and ToBI annotations (Dilley, et al, 2006) found a higher interlabeler-agreement with respect to boundary type for RaP, a system in which boundary labels are based on perceived degree of disjuncture compared to ToBI where boundary labels are based on perceived disjuncture and the identity of boundary tones.

### Relationship to Syntactic Structure

The relationship between prosodic phrasing and syntactic structure is an area of particularly diverging opinions. Models differ in how closely they assume prosodic phrasing matches up with syntactic constituent structure, and conversely how complicated a mapping function they postulate at the interface between the two representations. Early work in the phonetic and psycholinguistic tradition explored the extent to which the phonetic realization of an utterance directly reflected syntactic structure. It was felt that the surface acoustic form of a sentence might reveal something about the underlying syntactic representation, and this was supported by researchers who found greater segmental lengthening and pause insertion at

points in a sentence that corresponded with phrase structure boundaries (Klatt, 1975; Lehiste, 1973). However, other researchers have found evidence that the relationship is less transparent, and developed models to explain apparent discrepancies between syntax and prosody. We review some of these proposals below.

**Prosody reflects syntax**—According to recent proposals in categorial grammar (Steedman, 1991), the surface prosodic phrasing *is* the syntax. Categorial grammar is a theory of how syntax and meaning composition go hand in hand. It provides a range of operations that can effectively rebracket the phrase structure of an expression in unconventional ways. In English, for example, both (S) (VO) and (SV)(O) are permitted as syntactic bracketings, reflecting the fact that both prosodic phrasings are possible. Categorial grammar thus provides an account of syntax that matches prosodic constituency, and assures surface compositionality even in cases where at first blush the prosodic bracketing seems to contradict the syntactic one.

Compatible with this viewpoint is recent work on bracketing paradoxes (Wagner, 2005; in press), which provides syntactic evidence that at least some apparent cases of mismatches between syntax and prosody actually involve a syntactic structure that in fact matches the prosody. A complex meaning can often be constructed in more than one way, and the choice between structures comes with different prosodies. An apparently mismatching prosodic phrasing may in fact reveal a different syntactic choice about how a complex meaning is constructed. The motivation for the choice between these different structures may ultimately lie in processing factors, e.g., extraposing a relative clause avoids a nested structure, which may be difficult to process.

**Algorithmic approaches**—There are a number of factors that affect prosody that do not appear to be mediated by syntax. Many researchers concluded that the mapping between phrase structure and the acoustics of an utterance is not one to one, and a tradition started that sought to characterize this link, both in the literature on phonological theory (Selkirk, 1984; Selkirk, 1986; Nespor & Vogel, 1986; Truckenbrodt, 1995) and in the processing literature (Cooper & Paccia-Cooper, 1980; Grosjean et al., 1979; Gee & Grosjean, 1983; Ferreira, 1988; Watson & Gibson, 2004b). These approaches were algorithmic, and derived prosodic properties such as pause length based on a syntactic representation and a set of mapping rules.

Developments in prosodic theory in the early 1980s introduced the notion of prosodic and phonological constraints, which were purported to influence pausing and the duration of words independent of syntax, and these principles were incorporated into algorithms (Gee & Grosjean, 1983; Ferreira, 1988; Watson & Gibson, 2004b). For example, pauses are relatively unlikely to occur between phonologically light items like function words and nearby content words that are phonologically heavy (Selkirk, 1984; Nespor & Vogel, 1986). Other work suggested that speakers tend to produce pauses such that the resulting prosodic phrases are roughly the same length (Cooper & Paccia-Cooper, 1980). As a consequence, syntax and prosodic structure can diverge such that pauses might occur at a relatively minor syntactic boundary over a more major boundary. A sentence like (2) with a large pause between “understand” and “the politicians” is well formed with a pause between the verb and the direct object instead of between the subject and the verb.

(2) I don't understand//the politician's policies

Researchers like Gee and Grosjean (1983) (henceforth, GG) incorporated these observations into their algorithm, predicting pauses using both syntactic constraints and prosodic constraints like phonological phrasing and prosodic balancing. This model did quite well in predicting pause lengths, accounting for (in GG's article) 92% of the variance, an

improvement over earlier accounts without prosodic constraints, such as the model in Cooper and Paccia-Cooper (1980) (henceforth, CPC), whose model accounted for only 56% on the same data. However, this model also includes a large number of steps and parameters for building a prosodic representation. GG's algorithm contained a total of 8 steps, and because these steps were highly interrelated, it is difficult to know which aspects of the algorithm were doing the heavy lifting in predicting pause length. As pointed out by GG, the goal of these models was not to provide an explanation of the cognitive mechanisms that underlie speech, but rather to provide a description of where pauses were likely to occur.

Ferreira (1988, 1993) proposed two improvements to models by GG and CPC in her algorithm. First, she introduced the linguistic notion of the prosodic phrase boundary to the algorithmic approach, arguing that psycholinguists should be trying to account for the presence (and absence) of prosodic boundaries rather than pause length. If one assumes that the presence or absence of intonational phrases is binary, aggregating over pause lengths gave the appearance that GG and CPC's models were predicting pause duration. She argued that in reality CPC and GG were predicting the relative likelihood of a boundary occurring at a word boundary. Ferreira (1993) found that the actual extent of pause duration and pre-pausal lengthening was determined by the segmental properties of the pre-boundary word, with segmental duration engaging in a trading relationship with pause length. Words with shorter intrinsic vowel length had a longer pause than words with a longer intrinsic vowel length although the total duration of the word and pause together was roughly the same when controlling for sentence position.

Ferreira's second improvement to the algorithmic approach was the incorporation of semantic constraints into an algorithm. Work by Selkirk (1984) suggests that semantic structure can constrain prosodic structure. She proposed the Sense Unit Condition which roughly states that constituents that do not have a dependency relationship cannot co-occur within the same intonational phrase. For example, (3a) sounds unacceptable because "in the moon" and "is a myth" are not semantically related yet occur within the same intonational phrase. According to Selkirk, if a boundary occurs after "moon" such that the PP and the VP are in separate phrases, the sentence is more acceptable.

**(3a)** The man//in the moon is a myth.

**(3b)** The man//in the moon//is a myth.

However, note that Watson and Gibson (2004a) found that in acceptability surveys, (3a) and (3b) were both unacceptable compared to a sentence in which no boundary occurred. They argue that the poor acceptability of (3a) and (3b) was driven by interrupting the local dependency relationship between the modifier PP and the subject noun.

Ferreira (1988) proposed a model of prosodic phrasing based on X-bar theory (Jackendoff, 1977). Because X-bar theory instantiates different types of dependency relationships into the syntactic representation, she proposed that the likelihood of an intonational phrase boundary could be predicted by the X-bar structure of the sentence, which serves as a proxy for the semantic closeness of dependents. Within her algorithm, boundaries are least likely to occur between semantically related words like a head and its argument while boundaries are more likely to occur between weakly related constituents like a head and an adjunct or between two unrelated adjuncts. In a series of production experiments, Ferreira shows that this model performs significantly better than previous algorithms.

Within the phonological literature, a theory that has gained wide currency is the edge-alignment theory of prosodic phrasing. Based on observations on the phrasing of tone sandhi domains in Taiwanese (Chen, 1987) and related phenomena, Selkirk (1986) proposes that the left and right edge of certain syntactic constituents are aligned with the right and left

edge of certain prosodic constituents. Today, this is often implemented using optimality theory, and output constraints such as ‘Align XP’ (Selkirk, 1995) and ‘Wrap XP’ (Truckenbrodt, 1995; 1999) are used to force certain prosodic phrasings. Differences in the prosodic phrasing between languages are taken to be due to different rankings of the constraints. The edge-alignment theory crucially assumes a set of syntactic categories (e.g., Maximal Projection: XP), and a set of phonological categories (e.g., phonological phrase, intonational phrase), since it is certain *types* of prosodic boundaries that align with certain *types* of syntactic edges.

More recently, there has been a resurgence in trying to understand whether algorithmic approaches can provide a useful account of intonational boundaries. Watson and Gibson (2004b) proposed that much of the success enjoyed by previous theories was due to their incorporation of two factors: 1) predicting a high likelihood of a boundary before a long constituent and 2) predicting a high likelihood of a boundary after a long constituent. In addition, Watson and Gibson showed that an algorithm incorporating these two factors, along with constraints against boundaries occurring when a constituent is not complete and between heads and arguments, did as well as the previous algorithms. Watson and Gibson propose that ultimately, boundary production is related to planning and recovery processes. Boundaries occur before long constituents to give the speaker planning time, and boundaries occur after long constituents to provide speakers with time for recovery. Follow up work by Watson, Breen, and Gibson (2006) suggests that the optionality of a dependent, in addition to its argument status, influences boundary placement. Speakers are reluctant to place boundaries between a head and an obligatory argument. Watson, Breen, and Gibson (2006) argue that the obligatoriness constraint stems from heads and obligatory dependents being more likely to be planned together at the boundary before the head, negating a need for the intervening boundary. Turk (2008) proposes that prosodic phrasing, just like prosodic prominence, reflects local predictability, thus also invoking a processing explanation rather than a grammatical mapping.

The link between planning and prosodic structure has been supported by findings from the literature. Ferreira (1991) found that pauses were longer before syntactically complex object phrases. Wheeldon and Lahiri (1997) also found that initiation times for a sentence increased with the number of phonological words in the subject.

Ferreira (2007) points out that production factors are unlikely to account for all aspects of intonational phrasing, noting that boundaries and pausing may also result from the metrical structure of a sentence. It is also clear that boundaries play a role in the signaling of pragmatic and semantic information as in the case of asides, appositives, and non-restrictive relative clauses (Nespor & Vogel, 1986; Shattuck-Hufnagel & Turk, 1996; Watson & Gibson, 2004b).

Ferreira (2007) has recently challenged the use of the algorithmic approach by itself in understanding boundary placement in production. She points out that in testing these algorithms, different researchers have used different syntactic structures as test sets. Because there is no principled way to select the stimuli to compare these algorithms, it is difficult to evaluate these theories with respect to one another. Both the algorithmic approach and a more traditional approach in which specific properties of matched sentences are manipulated to examine the likelihood of intonational boundaries at specific word boundaries will be important for understanding boundaries in production.

## Boundaries and Parsing

**Resolving Ambiguities**—There is a great deal of work in the literature demonstrating that listeners can take advantage of the close mapping between syntax and prosodic



boundaries to resolve ambiguities in language processing. There is an excellent review by Cutler et al. (1997) that surveys work on listeners' use of prosody in syntactic parsing up to that time.

One of the big questions in the 1990s was understanding whether intonational boundaries can be used to resolve syntactic ambiguities in online processing. This question was asked in the context of a larger debate about the modularity of sentence processing: is syntactic information the only source of information used in the initial stages of processing (e.g. Frazier & Clifton, 1996) or is information from other domains used in these early stages as well (e.g. Tanenhaus et al., 1995)? Studies suggest that non-syntactic information is used very rapidly in processing, though researchers disagree over whether the effects occur immediately or upon re-analysis or re-processing. Research over the past twenty years strongly suggests that prosody is one of the many factors that are rapidly integrated into the linguistic representation (Marslen-Wilson et al., 1992; Grabe, Warren, & Nolan, 1994; Kjelgaard & Speer, 1999; Watson & Gibson, 2005; Snedeker & Trueswell, 2003 to name just a few).

Given that boundaries clearly play a role in sentence processing, researchers have focused on two questions: 1) what sort of information do intonational boundaries provide and 2) Do speakers consistently produce boundaries for the listener? We discuss the second question in the next section, and explore the first question here.

The literature unequivocally demonstrates that boundaries can disambiguate certain types of syntactic structures. Interestingly, certain types of ambiguities are much more easily disambiguated than others. For example, prosody appears to play a stronger role in disambiguating sentences in which the difference between interpretations lies in how the two meanings are grouped. Consider the examples in (4) below.

- (4a) When Roger leaves//the house is dark. (Kjelgaard & Speer, 1999)  
 (4b) When Roger leaves the house//it's dark.

Work by Speer and colleagues (Kjelgaard & Speer, 1999; Speer, Kjelgaard, & Dobroth, 1996) and others (Warren et al., 1995) suggests that boundaries can help to resolve local ambiguities in sentences like (4). Although listeners typically interpret the noun following the verb in the subordinate clause as a direct object (instead of the subject of the main clause), placing a boundary between the noun and the verb such that the noun is grouped with the main clause reduces this bias.

- (5a) Pat//or Jay and Lee convinced the bank president to extend the mortgage. (from Clifton, Frazier, & Carlson, 2006)  
 (5b) Pat or Jay//and Lee convinced the bank president to extend the mortgage.

Similarly, conjunctions like those in (5) have been shown to be disambiguated by prosodic phrasing (Lehiste, 1973; Streeter, 1978; Wagner, 2005; Clifton, Frazier, & Carlson, 2006). An early boundary in (5) groups Jay and Lee together while a later boundary groups Pat and Jay together.

- (6a) Mary maintained//that the CEO lied when the investigation started. (Carlson et al., 2001)  
 (6b) Mary maintained that the CEO lied//when the investigation started.

Finally, boundaries can play a role in signaling the presence of long distance dependencies (Snedeker & Trueswell, 2003; Kraljic & Brennan, 2005; Schafer et al., 2005). The boundary in (6a) biases listeners towards attaching the adverbial phrase to the local verb "lied",

whereas the boundary in (6b) creates a bias towards attachment to the matrix verb “maintained”.

(7) The detective showed the blurry picture of the diamond//to the client.

Boundaries can also facilitate processing unambiguous structures like the one in (7). Listeners rated sentences with a boundary between “diamond” and “to the client” as being easier to understand than sentences without this boundary (Watson & Gibson, 2005). They also found that a boundary between a head and a local dependent in an unambiguous sentence disrupts processing.

When do intonational boundaries not help? Different researchers have made the same general claim: boundaries can resolve ambiguities in which the surface bracketing of a sentence differs across interpretations, but cannot resolve ambiguities in which they do not (Lieberman, 1967; Lehiste, 1973; Shattuck-Hufnagel & Turk, 1996). For example, (8) below is globally ambiguous. “Flying” can be interpreted as either a gerund or an adjective. Lieberman (1967) points out that the prosodic phrasing of this sentence cannot disambiguate it.

(8) Flying airplanes can be dangerous

Lieberman attributes this observation to the similar surface structure but differing deep structure of the two interpretations, arguing that boundaries may only play a role in disambiguation when surface structure differs. The major syntactic breaks in the sentence are in the same location across interpretations even though the sentence meanings differ. This generalization also explains why sentences like (9) have yielded inconsistent results in the literature.

(9a) Mary knows the boy on the bench.

(9b) Mary knows the boy is sleeping.

The structures in (9) contains a local ambiguity at the noun phrase “the boy”. It can either be interpreted as the subject of a sentential complement (9b) or as the direct object of the verb (9a) and listeners have a preference for the latter interpretation. A large number of researchers have investigated whether resolution of this ambiguity might be influenced by prosody, but the results from the literature are mixed. Some studies have found that speakers are more likely to produce a boundary (or at least acoustic correlates of boundaries) after the verb in the sentential complement continuation (Warren, 1985; Nagel et al., 1996). Recent work by Anderson and Carlson (2004) suggests that speakers produce cues in this structure less consistently than they do in early closure/late closure ambiguities like the one in (4). Beach (1991) and Marslen-Wilson et al. (1992) found that listeners interpret a boundary after the verb as indicating the SC continuation. However, Stirling and Wales (1996) and Watt and Murray (1996) found no such difference. To complicate the picture, Gahl and Garnsey (2004) found that lengthening of the verb depends on verb bias. Certain verbs like “believe” occur more frequently with sentential complements than direct objects. Conversely, verbs like “confirmed” occur more frequently with direct objects than sentential complements. Gahl and Garnsey (2004) found that verbs were lengthened and preceded a longer pause when they occurred with the dispreferred interpretation for the verb.

If it is the case that prosodic disambiguation of the structure in (10) is less frequent than other structures, one possible reason is a lack of difference in the surface structure of the two interpretations. Both the sentential complement and direct object continuation enjoy the same type of relationship with the verb. Like the ambiguity in (8), the difference between interpretations lies in their syntactic categories and not in their syntactic relationships. Similarly, classic sentences like (10), which are locally ambiguous and contain a verb

(“raced”) that can be either interpreted as the main verb of the sentence or a verb in a reduced relative clause, cannot be disambiguated prosodically (Fodor, 2002).

**(10a)** The horse raced past the barn fell

**(10b)** The horse raced past the barn and fell

Again, although the surface structure labels differ between these interpretations, the syntactic relations between constituents do not.

There have been two types of explanations for why boundaries can only disambiguate ambiguous sentences that have interpretations that differ in their surface structure. One is that the link between syntactic structure and prosodic breaks is formalized in the grammar as prosodic structure (e.g. Selkirk, 2000; Truckenbrodt, 1999), and prosodic structure provides cues to syntactic structure through knowledge of the grammar. The other is that prosodic boundaries serve as a means by which the listener organizes the incoming linguistic signal for language processing. The latter is discussed below, and of course, the two possibilities are not mutually exclusive.

Psychologists have proposed two types of processing accounts, but both accounts assume that the central role of boundaries in processing is to provide information about how the linguistic signal is organized. One class of theories argues that boundaries group words into processing units (Schafer, 1997; Frazier & Clifton, 1998), while the other argues that boundaries mark points of disjuncture in a sentence (Watson & Gibson, 2005; Pynte & Prieur, 1996; Marcus & Hindle, 1990). At first glance, these theories may appear to make similar predictions, however if we consider sentences like (6) above, the differences become clearer.

Theories based on grouping argue that words that occur within the same intonational phrase are processed together at the same processing stage (e.g. Schafer, 1997). Effects of boundaries on ambiguity resolution are driven by whether ambiguously attached constituents appear in the same prosodic phrase as an attachment site. Thus in (6), a boundary after the matrix verb “maintained” causes the verb in the embedded clause, “lied”, and the adverbial phrase “when the investigation started” to be processed together in the same processing chunk. This temporal proximity in the processing system creates a preference for low attachment. When a boundary occurs after the embedded verb “lied”, the adverbial phrase is not processed in the same unit as either verb. According to grouping theories, a second boundary should not influence attachment preferences. However, this runs counter to findings in the literature that a boundary before an ambiguously attached constituent biases listeners towards high attachment (e.g. Snedeker & Trueswell, 2003; Schafer et al., 2005; Carlson et al., 2001)

Theories based on disjuncture, like Watson and Gibson’s (2004a, 2005) anti-attachment hypothesis, argue that boundaries primarily serve as a signal of non-local attachment. The boundary after “maintained” in (6) signals that the verb is unlikely to receive future attachment, thus facilitating low attachment to the embedded verb. A boundary after the embedded verb “lied” signals that the adverbial phrase does not attach locally. One shortcoming of this theory lies in its explanation of low attachment. A boundary after “maintained” signals that the main verb is unlikely to receive further attachment, but in actuality, it does attach locally to the embedded clause, although it does not receive attachment down stream. This predicts that this boundary should increase sentence difficulty because it provides incorrect information about non-local attachment, while still aiding in resolving the ambiguity. Whether this is actually the case is unclear.

Neither theory provides an adequate explanation of the data, and it is likely that a hybrid of the two approaches will provide a fuller account. That is, the presence of a boundary acts as a signal that two elements are not associated while the absence of a boundary can signal that two elements are grouped together. Listeners may very well use both types of information in parsing syntactic structure.

Another question is understanding whether boundaries provide information about the local syntactic context in which they appear or whether they can provide information about other locations in the sentence. Carlson and colleagues (2001) have argued that listeners use global prosodic structure when processing syntactic information. They found that in sentences like (6), an early boundary after the verb “maintained” reduced the high attachment preference driven by the boundary after the embedded verb “lied.” Similarly a boundary after “lied” reduced the effect of the early boundary on low attachment. In addition, the magnitude of the boundary (i.e. whether the boundary was an intermediate or full intonational phrase) modulated the effect. Snedeker and Casserly (this volume) also find effects of global prosodic structure on ambiguity resolution although they find effects of both the global prosodic structure and absolute boundary size on attachment.

Frazier et al. (2006) argue that this is evidence against approaches that argue that boundaries provide information about the local syntactic context (e.g. Watson & Gibson, 2004a, 2005). However, this line of argument conflates two definitions of locality. On the one hand, locality may describe whether a boundary provides information about its local syntactic context. In another sense, locality may describe whether a listener interprets a boundary only at the position at which it is encountered. These are two independent empirical questions. Watson and Gibson (2004a, 2005) argue that boundaries provide local information in the first sense, whereas Frazier et al. (2006) provide evidence that boundaries are non-local in the second. For example, Carlson et al.’s (2001) evidence is consistent with the early and late boundary providing local information about their syntactic contexts, and the listener integrating over these boundaries across the sentence in making an attachment preference.

**Audience Design**—Recent work in understanding intonational boundaries and parsing has come to focus on the role of boundaries at the interface of language comprehension and language production. Traditionally, psycholinguists have used ambiguity resolution as a tool for understanding the underlying mechanisms of the language comprehension system. A natural question that has followed from this research tradition is understanding whether speakers consistently provide listeners with prosody that disambiguates a sentence.

The evidence thus far has been mixed. Allbritton et al. (1996) found that naïve speakers do not reliably disambiguate sentences prosodically for listeners, while expert speakers did. Findings by Snedeker and Trueswell (2003) are consistent with Allbritton et al.’s findings from naïve speakers. In a referential communication task, speakers instructed their partners to tap objects in a real world display using utterances like (11)

(11) Tap the frog with the flower.

Here, the prepositional phrase “with the flower” can be interpreted as either an instrument of the verb or a modifier of “frog”. Verb attachment is associated with a boundary after “frog”, while NP attachment is associated with a boundary after the verb “Tap”. Snedeker and Trueswell (2003) found that speakers only disambiguated the sentence for listeners if they were aware of the ambiguity. Based on these findings, some have concluded that speakers do not typically disambiguate syntactic structure for listeners during the course of normal conversation.

The above findings conflict with other work in the literature. Schafer and colleagues (2005) found that speakers consistently disambiguate sentences similar to (11) for listeners in a referential communication task. Similarly, Kraljic & Brennan (2005) found that speakers consistently disambiguate structures like (12) below.

**(12)** Put the dog in the basket on the star.

In (12), the listener is being instructed by another subject to move either a dog that is in a basket onto a star, or to move a dog into a basket that is on a star. Kraljic and Brennan (2005) found that speakers consistently disambiguated the sentence by placing an intonational phrase boundary at the right hand constituent boundary of the direct object depending on the intended interpretation, independent of whether they were aware of the ambiguity or not, and independent of whether the context was actually ambiguous or not. They conclude that the production of prosodic structure is constrained by speaker-centered processes rather than listener-centered processes.

A speaker-centered explanation may explain the seemingly inconsistent results of Snedeker and Trueswell (2003). The sentences produced in Snedeker and Trueswell's experiment were shorter than those used in Kraljic and Brennan (2005) or Schafer et al (2005). If speakers' productions of intonational boundaries are largely driven by the length of syntactic constituents (Watson & Gibson, 2004b), then speakers are unlikely to produce intonational boundaries in short sentences, whether they are ambiguous or not. Thus, the effect of speaker awareness found by Snedeker and Trueswell may be due in large part to sentence length. Speakers' awareness of the ambiguity may have encouraged more boundary use than normal in short sentences.

If Kraljic and Brennan's (2005) speaker-centered view is correct, speakers may produce intonational boundaries for reasons related to planning, rather than for the listener. Listeners may be sensitive to the distribution of these boundaries, inferring syntactic and semantic structure from boundary placement. MacDonald (1999) has proposed this type of model to provide a general account of the interface between production and comprehension, arguing that the distribution of syntactic structures constrains listeners' parsing preferences.

Clifton and colleagues (2006) propose a similar model specifically for prosodic structure. Under the Rational Speaker Hypothesis, speakers produce intonational boundaries based on factors such as syntactic structure, constituent length, and pragmatics. Clifton and colleagues argue that listeners are sensitive to the likely underlying causes of boundary placement and weight these boundaries accordingly. Consider (13):

- (13a)** Pat//or Jay and Lee convinced the bank president to extend the mortgage.
- (13b)** Pat or Jay//and Lee convinced the bank president to extend the mortgage.
- (13c)** Patricia Jones//or Jacqueline Frazier and Letitia Connolly convinced the bank president to extend the mortgage.
- (13d)** Patricia Jones or Jacqueline Frazier//and Letitia Connolly convinced the bank president to extend the mortgage.

A boundary in (13) can tell a listener how to group the three referents in the subject. Clifton and colleagues (2006) found that listeners were less likely to use boundaries to group the nouns when the nouns were phonologically longer (as in 13c and 13d) and more likely to use the boundaries for grouping when the referents were shorter. Clifton et al. (2006) argue that in (13c) and (13d), listeners attribute the boundaries to constituent length rather than to disambiguating the grouping of the subject.

**Implicit Prosody**—Recently, there has been a great deal of interest in how the prosodic structure of a language potentially influences language comprehension in reading. With the exception of punctuation, there is no explicit prosody available to readers. Researchers have proposed that listeners construct a prosodic representation on the fly in reading called implicit prosody (Bader, 1998; Fodor, 2002).

Bader (1998) was the first to propose that implicit prosody may be more than an epiphenomenon of reading and may affect parsing decisions. He argued that the degree of difficulty of recovering from a garden path results partly from having to re-analyze both the prosodic structure of a sentence and the syntactic structure of a sentence. Readers experience more difficulty when both syntax and prosodic structure must be re-processed or re-analyzed than when syntax alone must be re-processed or re-analyzed. For example, Bader claims that NP/S ambiguities like (9) result in less severe garden paths than early/late closure ambiguities like (4). According to Nespor and Vogel's (1986) formulation of prosodic structure, the two interpretations in (9) share the same prosodic structure, while the two interpretations in (4) have different prosodic structures. Because re-analysis in the latter interpretation requires an additional change to the prosodic representation, the garden-path is more severe.

Fodor (1998, 2002) has proposed a more direct role for implicit prosody in parsing decisions and has argued that it can provide a broad explanation for an unsolved problem in the field: understanding cross-linguistic differences in attachment preferences. Although there seems to be a general cross-linguistic locality preference for constituents to be attached to more recent material (Gibson, 2000; Frazier & Clifton, 1998), globally ambiguous sentences like the one below result in different preferences in different languages.

(14) Frank knows the secretary of the linguist who quit her job.

The relative clause “who quit her job” can modify either “secretary” or “linguist”. In English, there is a preference toward attaching to the most recent or lowest node in the sentence (low attachment), in this case “linguist”. However in Spanish, there is a preference to attach to the highest node, “secretary”. Although there have been a variety of proposals, there has been no satisfactory explanation as to why speakers of different languages have different attachment preferences (see Fodor, 1998 for a review). Fodor and colleagues have argued that implicit prosody plays a role. The claim is that readers construct a prosodic structure that fits the prosodic constraints of their language, and that this representation, in turn, can influence attachment preferences. For example, shorter relative clauses are associated with low attachment cross-linguistically. Fodor argues that shorter relative clauses are less likely to be preceded by an intonational boundary because listeners disprefer to place a short relative clause in an intonational phrase by itself (see Selkirk, 2000). The absence of a boundary at this location creates a bias towards low attachment. In contrast, a longer relative clause is more likely to be preceded by a boundary in the implicit prosodic representation, and this boundary creates a bias towards high attachment. Fodor (1998, 2002) argues that languages in which one sees high attachment preferences are those in which prosodic constraints require prosodic breaks at the beginning of constituents (English is not one of these languages). Boundaries are constructed before relative clauses because these breaks are required by the grammar (in Fodor's theory), and these boundaries induce high attachment. Fodor and colleagues (e.g. Lovric, 2003; Hirose, 1999) are testing these hypotheses across a wide class of languages.

Although implicit prosody is providing interesting potential explanations for a variety of phenomena, this type of approach faces a large set of challenges. Because implicit prosody cannot be directly measured, a manipulation of implicit prosody in reading necessarily requires a manipulation of another linguistic factor that is assumed to affect prosody in

spoken speech. Thus, it is difficult to know whether implicit prosody or another linguistic factor is the underlying source of any effect that is found. A cross-linguistic approach like the one offered by Fodor in which one studies links between varying distributions of prosodic structure and syntax across languages, rather than manipulating linguistic variables within a language, may offer the best approach.

## Prosodic Prominence

### Phonetic and Phonological Correlates

Just as in the case of boundary strength and phrasing, the phonetic cues related to prominence include duration, fundamental frequency, and intensity, as shown by Fry (1955, 1958) and Lieberman (1960). Which cues are the most important is still controversial, and also varies vastly between languages. The following discussion is mostly based on research on English.

Duration has been shown in many studies to correlate with prominence in English both to signal word stress, and, at the phrasal level, to signal phrasal prominence. Given that duration also plays a crucial role in signaling phrasing, it seems that the same channel has to do double duty (more than that, really, since duration is also crucial in conveying lexical contrasts between words). However, the precise durational changes related to prominence were argued in Beckman and Edwards (1992) to differ from the durational changes affected by boundary strength. While the durational correlate of prominence (e.g., under focus) is often achieved by increasing phase, that is, decreasing the overlap between gestures, the durational lengthening at the end of prosodic constituents tends to be achieved by a slowing down of the gestures, i.e., a decrease in stiffness.

Fundamental frequency can encode prominence in a number of ways. The location of primary prominence in a word is reflected in the temporal alignment of pitch accents. The degree of prominence on a word can be increased by a higher pitch excursion (Eady et al. 1986, Rietveld & Gussenhoven, 1985). Furthermore, an increase in pitch range was shown in Ladd and Morton (1997) to be interpreted as encoding a distinction as to whether or not an accented constituent was especially emphasized. Post-focal material is often realized with a reduced pitch-range (Xu & Xu, 2005). Relative pitch scaling is thus a main cue in English for encoding relative prominence.

Kochanski et al. (2005), however, question whether pitch is a reliable cue even if the mean values in experimental studies often show differences. Kochanski et al. (2005) point out the variability of pitch cues across utterances. They demonstrate that pitch is not a good predictor of prominence in an English corpus, and find evidence that loudness is the best acoustic correlate. Beckman (1986) found that duration and intensity together form a reliable correlate of prominence, a finding that was confirmed in a perception study based on synthetic stimuli in Turk and Sawusch (1997).

Finally, spectral slope has also been linked to cueing prominence (Sluijter & van Heuven, 1996; Heldner, 2001). Sluijter and van Heuven found that syllables show a flatter spectral tilt in words carrying contrastive stress.

### Relative Prominence Relations and Categorical Prominence Levels

A controversial issue in the domain of prominence is whether linguistically significant degrees of prominence are gradient and relative or whether they correspond to phonological categories—a debate similar to the one on boundary strength. The categorical view holds that phonology provides a small number of discrete levels of prominence, whose use is determined by grammatical constraints. This controversy can be traced at least to the

discussion of the prominence assignment algorithm presented in Chomsky and Halle (1968), in which each syllable was assigned an integer value reflecting the degree of prominence as determined by its place in the syntactic structure. This algorithm was criticized for providing an unrealistic number of degrees of prominence (Vanderslice and Ladefoged 1972), even though it was explicit in not wanting to provide a phonetic theory of prominence in the first place. The crux of the gradient stress representation in Chomsky & Halle (1968), and also more recent gradient representations such as the ‘metrical grid’ (Liberman 1975, Liberman & Prince 1977), was to allow a fine-grained encoding of *relative* prominence relations between syllables and words, because the idea was that relative prominence is what is influenced by syntactic and other factors. The phonetic realization was assumed to be *compatible* with these relations, but would not necessarily reflect them in all their detail.

A narrower inventory of phonetic prominence distinctions seems plausible at least at the word-stress level, where there is evidence for a small set of distinctions that are phonologically motivated, such as reduced, unreduced, and stressed vowels, although this does not mean that there are no gradient distinctions within these categories. Similar suggestions of a limited set of categorical options have been made for prominence at the sentence level, e.g., whether or not a word is pitch-accented is often taken to be a categorical distinction. Related to this is the question of whether there are categorically different types of pitch accents, each associated with distinct semantic/pragmatic import (e.g. Pierrehumbert & Hirschberg, 1990).

Even if it remains controversial how prominence should be represented, it is uncontroversial that relative prominence plays an important role in grammar, e.g. in the information structure of English. The wrong relative prominence relation between words often leads to infelicity:

- (15)      **A.** Would you like some coffee?  
             **B.** I would love some *coffee*.

If the prominence of ‘coffee’ in (15B) is higher than that of ‘love’, this sentence sounds odd, for reasons to be discussed below. Relative prominence is very salient at the end of phrases and utterances, which is why much of the impressionistic data reported in the literature on sentence-level prominence discusses where the last and main prominence falls in a sentence. Shifting the prominence away from the end of a sentence in cases when it is not motivated or failing to do so when it is can lead to strong pragmatic deviance. Xu and Xu (2005) argue that such shifts in prominence are achieved by a relative pitch range suppression rather by a categorical deaccentuation. They show evidence that pitch suppression after (but not before) focused constituents is observed in English, but within the subordinated domain there are still remnants of the pitch movements usually associated with word stress.

Thus the extent to which prominence is mediated by the placement and omission of accents or by gradient adjustments of relative prominence such as pitch range modulation remains controversial, but the fact that prominence encodes important information about the structure of an utterance and its context is not.

### Relationship to Syntactic Structure

Sentence-level prominence can be affected by a number of syntactic factors. In fact, Chomsky and Halle (1968) argued that the relative prominence of words in a sentence is entirely determined by a recursive algorithm that translates phrase structures into a phonological transcription which includes relative prominence. Bolinger (1972) challenged this syntactic approach and argued that accents directly reflect the intentions of the speaker and the information flow of the sentence. Schmerling (1976) also questioned the feasibility



of a purely syntactic account of sentence stress assignment, and proposed a number of semantic constraints that affect sentence stress instead. One factor she proposes is that ‘predicates receive lower stress than their arguments’, which accounts for why in German and English a direct object is always more prominent, independent of whether or not the word order is OV or VO. This semantic generalization is integrated in the theory of sentence stress proposed in Gussenhoven (1984). Gussenhoven presents a system that captures the circumstances under which words are phrased into a single domain carrying a single accent. The decisive factor that makes one constituent integrate into a single accent domain with a second constituent is semantic: If a constituent is the argument of a second argument, they form a single accent domain. The only syntactic condition in this theory is that the two constituents be adjacent to each other. Effects of argument structure on prominence are also discussed in Selkirk (1984).

It is not clear though that a semantic approach to sentence prominence is sufficient. Whether or not a predicate is prosodically subordinated and forms a single accent domain with an adjacent argument depends on its syntactic relation to the argument. Truckenbrodt and Darcy (2008) report that complement clauses in German differ from complement NPs in that the selected predicate remains accented when taking a clausal complement. This observation can be related to the fact that complement clauses in German obligatorily ‘extrapose’, that is, they are treated very differently from other syntactic complements, e.g. they follow the verb instead of preceding it, in contrast to nominal arguments in German. Two influential syntactic approaches that place a higher emphasis on syntax in the negotiation of prominence are Cinque (1993) and Truckenbrodt (1995).

### Prominence and Information and Discourse Structure

It is widely accepted that pitch accents play a role in signaling the information and discourse status of the words and constituents in which they appear, although how to characterize this information is under considerable debate. Researchers have proposed that a pitch accent is placed on constituents that are new (e.g. Halliday, 1967; Prince, 1981; Chafe, 1987), important (Bolinger, 1972) or non-given (Schwarzschild 1999), focused (Jackendoff 1972, Rooth 1992, Selkirk, 1995; Büring, 2007), less accessible (Arnold 2008), or unpredictable (Gregory, 2001; Aylett & Turk, 2004). Below, these differing perspectives are reviewed.

**Givenness**—Work in the literature suggests a link between the givenness of a referent and whether it is produced with an accent. New information is argued to be accented while given information is argued to be de-accented, and work from the psycholinguistic literature supports this characterization. This pattern is also reflected in preferences in comprehension. Both Terken and Nootboom (1987) and Dahan, Tanenhaus, and Chambers (2002), found that listeners’ comprehension of instructions was better when new information was accented and given information was de-accented. Listeners also rate sentences with accented new information and de-accented given information as being more acceptable (Birch & Clifton, 1995).

However, it is also clear that previous mention is not sufficient for a word to be interpreted as given. In a corpus study, Hirschberg (1993) found that previously mentioned information is often accented. De-accented words were very likely to be given, but given words were only de-accented approximately half of the time. Terken and Hirschberg (1994) found that whether speakers produced a word with an accent depended on whether the referent was new and whether it had changed syntactic position: given information produced in a different syntactic role was accented. This may be related to evidence in Williams (1997) and Wagner (2006) that shows that in many cases marking a constituent as given is only possible when there is an antecedent that includes a plausible alternative to its sister

constituent. Dahan et al. (2002) found that listeners preferred accents on given information if that information was not salient in the discourse.

To illustrate the issues for a theory in which previously mentioned material is de-accented and new material is accented, consider the following examples (based on examples in Schwarzschild, 1999):

(16a) Patty's cousin likes classical music, but PATTY likes rock 'n' roll.

(16b) Who is Patty's cousin arguing with? She's arguing with PATTY.

(16c) Cathy likes to Waltz, but her partner HATES dancing.

In (16a), the second instance of "Patty" is accented to contrast her with her cousin even though "Patty" was previously mentioned. Example (16b) demonstrates that accenting previously mentioned material may be felicitous even without contrast. In (16c), new material like "partner" and "dancing" can forgo accenting even though they are mentioned for the first time.

Researchers have tried two different approaches to solve this problem. One has been to re-define given and new so that it captures the data (see Venditti & Hirschberg, 2003 for a review). Halliday (1967) argues that given information should be defined as information that is recoverable from the discourse while new information is not inferable from the discourse and violates the expectations of the conversational participants. Schwarzschild (1999) argues that the givenness of information depends on whether it is entailed by the discourse. Under these accounts, "Patty" is accented in (16a) and (16b) because it is not predictable from the discourse. "Partner" and "dancing" can be de-accented because it is inferable from the discourse.

**Focus Structure**—Closely related to theories of 'givenness' are theories of 'focus structure'. Jackendoff (1972), e.g., suggested the following: "As Working definitions, we will use the 'focus of a sentence' to denote the information in the sentence that is assumed by the speaker not to be shared by him and the hearer, and 'presupposition of a sentence' to denote the information in the sentence that is assumed by the speaker to be shared by him and the hearer. Intuitively, it makes sense to speak of a discourse as 'natural' if successive sentences share presuppositions, that is, if the two speakers implicitly agree on what information they have in common."

One particularly influential theory of focus is that of Rooth (1992), who proposed to capture the notion of focus using the alternatives that a focused constituent evokes. In his theory, every constituent in a sentence is assigned a meaning and a set of alternatives. If there is no focused information inside of a constituent, the set of alternatives is the unitary set of that constituent. If there is a constituent marked as focused, then the alternative set includes all constituents that can be built by replacing that constituent with contextually relevant alternatives of the same semantic type. Rooth's theory has the nice property that it accounts for different kinds of focus phenomena, such as question-answer congruence, contrastive stress, and givenness-marking with a single formalism. Variations of this type of approach were presented in Schwarzschild (1999), Büring (2003), and Wagner (2006).

It remains controversial, however, whether it is indeed the case that a unique theory can account for the full range of 'focus' phenomena. A theory that makes a distinction between focus and anaphoric distressing is proposed in Reinhart (2006), and this distinction is also commonly made in recent Optimality Theory literature on the topic (e.g., Samek-Lodovici, 2002).

**Focus Projection**—Acoustic prominence of a word is often assumed to reflect the information status of the word itself. As has been pointed out in the literature (e.g., in Chomsky, 1970, see also Ladd, 2008 for a discussion), this assumption is problematic because of examples like the one below:

(17a) What does Patty do in her free time?

(17b) Patty likes to ROCK ‘n ROLL.

Here “rock ‘n roll” is accented, but the new information, or focused material, is actually the entire VP “likes to rock ‘n roll”. An entire constituent may require highlighting in the discourse, yet frequently, an accent might only occur on only one of the words in the constituent. There appear to be constraints on which word in a focused constituent can carry an accent. For example, (18b) is a well formed response to the question in (18a), while (18c) is less acceptable.

(18a) What does Mary like?

(18b) Mary likes the picture of the VASE.

(18c) Mary likes the picture near the VASE.

Selkirk (1984) and Gussenhoven (1983) argue that to prosodically focus a constituent, an accent must occur on an (internal) syntactic argument of the constituent (in this case, “vase”), but there need not be an accent on the head (in this case, “picture”). But adjuncts differ from arguments in that if the head remains unaccented, focus cannot project from a syntactic adjunct of a head. Thus, in (18a), “vase” can project focus to the entire object noun phrase while in (18b) it cannot.

Experimental work by Birch and Clifton (1995, 2002) confirm these claims. Participants rated sentences in which arguments carry an accent to focus a constituent as being more acceptable than sentences in which adjuncts carry the accent. Similar findings are reported by Welby (2003). However, Breen et al. (2009) found consistent differences between narrow and broad focus in the production of accented words, and point out that the difficulty between perceptually differentiating broad and narrow focus in the case of VP vs. O focus may be that the two conditions have similar prominence relations between verb and object.

**Prominence and Accessibility**—Another strategy to better understand ‘givenness’ and ‘focus’ effects has been to use accounts of discourse focus and accessibility to describe givenness constraints on accenting (Dahan et al., 2002; Venditti & Hirschberg, 2003; Watson, Arnold, & Tanenhaus, 2005). These accounts have traditionally been used to explain speakers’ word choice preferences, particularly pronoun usage (see Arnold, 2008 for a review). The assumption is that as a discourse proceeds, information varies in its overall level of activation. Information that is highly activated tends to be highly accessible and is referred to using pronouns or shortened expressions while less accessible material is referred to with a full referring expression (Brennan, 1995; Gundel, Hedberg, & Zacharski, 1993; Grosz & Sidner, 1986; Grosz, Joshi, & Weinstein, 1995). Accessibility is influenced by syntax, topicality, recency of mention, and other factors. It is possible that accessibility plays a role in accenting as well as in the choice between full noun phrases and pronouns: accessible information is de-accented while non-accessible information is accented. This approach can capture given/new effects since given information is likely to be accessible while new information is not. It also captures the apparent exceptions: non-previously mentioned information that can be inferred or derived from the discourse context is likely to be accessible. In addition, previously mentioned information is not always accessible and would thus require an accent.

**Prominence and Predictability**—There is a close relationship between frequency and probability and quantitative measures of prominence (Bard et al., 2000; Jurafsky et al., 2001; Bell et al., 2009; Aylett, 2000; Aylett and Turk, 2004; Gregory et al., 2002). Watson, Arnold, and Tanenhaus (2008) found that words that are predictable from task-based constraints are more likely to be shortened, and Gahl and Garnsey (2004) found that verbs occurring with dispreferred arguments tend to be longer in duration. The acoustic length of a syllable correlates with its likelihood in context. The precise modeling of this probabilistic relationship is still an open question.

A model to explain these effects is proposed in Aylett (2000) and Aylett and Turk (2004), whose Smooth Signal Hypothesis holds that the degree of prominence of a syllable depends on its degree of redundancy, and various processes in speech conspire to evenly spread redundancy throughout an utterance. The effects that can be accounted for by the Smooth Signal Hypothesis involve quantitative differences in the strength of certain phonetic cues but also categorical choices, such as whether or not a syllable carries a pitch accent.

More recent research by Jaeger (2006) and Levy and Jaeger (2007) show that choices between different lexical and syntactic options (e.g., the choice between inserting a relative pronoun or not) can be explained by this idea--by inserting a relative pronoun, the redundancy of the signal is increased, and this insertion/omission of the additional morpheme was shown to depend on the redundancy of the surrounding material. They argue that lengthening may be part of a general strategy speakers use to optimally communicate with listeners, the ‘Uniform Information Density Hypothesis’ (Levy & Jaeger 2007, Jaeger, 2006).

There is also a clear link between predictability and pitch accenting. Highly predictable words are less likely to be accented than non-predictable words (Bell et al., 2002; Gregory, 2001). Gregory (2001) found that both low frequency and low transitional probabilities predict pitch accenting in corpora of spoken speech. What underlies this relationship? One possibility is that predictability effects simply reflect the statistics of discourse structure. Arnold (1998) found that changes in discourse accessibility are relatively infrequent in conversation. Thus, accenting unpredictable information may simply reflect the effects of accessibility discussed above. Predictability might also have a more direct effect on pitch accenting. Words that are infrequent might be more difficult to produce and require hyperarticulation (see Gahl & Garnsey, 2004 for a discussion). Watson, Arnold, & Tanenhaus (2008) found that word lengthening due to unpredictability also correlated with disfluency, suggesting that production difficulty may underlie both. In addition, Fowler and Housum (1987) found that repeated words tend to be reduced, and more recent work by Bard and Aylett (1999) found that this reduction is not always the result of de-accenting, suggesting that it may be the result of something other than marking information for the listener. Bell et al. (2009) argue that lengthening may relate to difficulties with lexical access. Understanding whether pitch accents are primarily speaker or listener centered faces the same challenge this question has faced in other areas of the literature (see Arnold, 2008 for discussion of the debate in the word choice literature). Because production- and listener-centered preferences are intimately linked, untangling the two poses a challenge for the field.

**Integrating Focus Structure, Accessibility, Predictability**—There is relatively little overlap between the tradition of work on focus and givenness on the one hand and the work on accessibility and predictability on the other. Often, similar terms used in these traditions can mean very different things, which can be confusing to the novice who is just beginning to explore the literature. For example, the use of the term “focus” in the focus structure tradition differs from its usage in the discourse accessibility tradition discussed

above. In the latter tradition, focus refers to given, highly accessible information. In the focus structure tradition, it refers to new, non-accessible information.

It is very clear that future work in pitch accenting will require cross-talk between these traditions. Researchers in the pragmatic and discourse accessibility tradition will need to integrate findings from focus structure into their work if they wish to successfully predict and understand how pragmatics affects the distribution of accenting. Some of the researchers in this tradition are interested in the computational problem of predicting accent distribution, and integrating focus structure constraints into current models is inevitable. It is also the case that researchers in the focus structure condition who abstract away from the discourse and cognitive constraints that govern importance and pitch accenting do so at their own peril. Evidence for performance constraints on accenting like referent accessibility (e.g. Venditti & Hirschberg, 2003) and repetition effects on pitch accenting (Bard & Aylett, 1999) suggest that a full understanding of information structure and pitch accenting may not be possible if one assumes that the link between accenting and information structure lies only in the grammar. Recent work on coherence relations (Kehler et al., 2008) can be seen as one attempt at integrating the more formal theory of focus structure with a more elaborate theory of how discourse works.

**Pitch Accents and On-line Processing**—A great deal of controversy surrounds the representations and usages underlying pitch accents, including whether there are different types of accents (e.g., H\* vs. LH\*), whether these differences are categorical or continuous, and what acoustic properties of pitch accents drive their perception.

One source of this controversy has been the lack of an implicit measure of listeners' perception of an accent. Much of the early work on accent perception relied on the intuitions of researchers or naïve subjects. In either case, judgments are potentially influenced by top down perception, and in many instances, the differences on which the judgments turn may depend on potentially complex discourse representations, which may be too subtle to intuit with any consistency.

Until recently, researchers have not had the tools to implicitly measure a listener's perception of an accent as it is heard in real time (see Watson, Gunlogson, & Tanenhaus, 2006 for a review). There has been an explosion of recent work using eye-tracking in a visual world paradigm to answer questions about the comprehension of pitch accents. In this paradigm, participants are presented with a visual array, either on a computer screen or a real life display. The participant is typically given auditory instructions while their fixations are monitored by an eye-tracking system, allowing the researcher to manipulate some aspect of the linguistic signal. Because fixations are driven by a wide array of factors that may have little to do with the language, the participant's task is designed so that verbal instructions play a critical role in completion of the task. The visual world paradigm has been used by the psycholinguistics community to investigate a wide range of questions about both language production and comprehension, including syntactic ambiguity resolution (e.g. Tanenhaus et al., 1995), verb processing (e.g. Altmann & Kamide, 1999), semantic and discourse processing (e.g. Sedivy et al., 1999), conversation (Brown-Schmidt et al., 2005), lexical access (Allopenna, Magnuson, & Tanenhaus, 1998), categorical perception (McMurray, Tanenhaus, & Aslin, 2002), and many others.

The visual world paradigm is ideally suited for investigating questions related to prosody. Fixations are highly sensitive to the acoustic phonetic signal, making it an ideal tool for investigating prosody. Dahan et al. (2002) used this paradigm to investigate how quickly listeners process pitch accent information in on-line comprehension. On a computer display containing eight pictures of objects, participants were instructed to move two objects to

different locations in the screen. Critically, two of the objects were a phonetic cohort, sharing their initial segments (e.g. “candy” and “candle”).

Typically, when listeners encounter temporarily ambiguous cohorts competitors, participants fixate on each cohort picture equally until disambiguating information is encountered. In order to see whether a pitch accent can resolve this ambiguity, and to understand whether pitch accents are rapidly processed in comprehension, Dahan et al. manipulated both the given/new status of one of the cohorts and whether it was produced with a pitch accent.

(19a) Put the candle/candy below the triangle.

(19b) Now put the CANDLE/candle above the square.

When participants heard a pitch accent on the target word in (19b), they fixated on the new referent more often than the given referent. When the critical cohort was de-accented, participants looked to the given referent more than the new referent. These differences appeared approximately 300ms after the onset of the target cohort, suggesting that not only are pitch accents rapidly detected, but they can be integrated into the discourse representation in the first moments of processing.

The success of Dahan et al.’s (2002) study has led researchers to adopt this technique to investigate the semantic and acoustic properties of pitch accents. Using the visual world paradigm, Ito and Speer (2008) and Weber, Braun, and Crocker (2006) found that listeners fixated potential contrast objects when they heard an L+H\* on an adjective. Watson, Tanenhaus, and Gunlogson (2008) have found that the interpretation of L+H\* and H\* may overlap instead of having complimentary distributions. Chen, den Os and de Ruyter (2007) have used the paradigm to investigate the semantic properties of different accents. Isaacs and Watson (this volume) manipulated the acoustic properties of pitch accents in a paradigm similar to that used by Dahan et al. (2002) and found that F0 slope contributed more to the perception of accenting and de-accenting than overall F0 and duration.

## Conclusions

In this review we have attempted to summarize some of the advances in our understanding of pitch accents and prosodic boundaries over the past decade. Several themes emerge. There are ongoing debates surrounding how to characterize the acoustic-phonetic properties of both pitch accent and boundaries. Controversy also surrounds how they are linked to discourse, syntactic, and semantic structure. However, there is general agreement that both boundaries and accents are sensitive to a variety of factors. For example, prominence can result from a word being focused in the discourse, important, and unpredictable. A challenge for the next decade will be building theories in which these factors can be understood in a unified framework.

Much of the interdisciplinary work in prosody has focused on pitch accents and intonational phrasing, so it is in these areas that we have focused this review. However, there are wide swaths of research in prosody that we have not mentioned.

There remain questions about the role the utterance level contour or tune plays in signaling pragmatic and semantic information (e.g. Pierrehumbert & Hirschberg, 1990; Gunlogson, 2003) as well as questions about the representations that underlie rhythmic adjustments of prominence (e.g., Selkirk, 1984; Gussenhoven 1991; Dilley, 2005) and phrasing (e.g., Nespore & Vogel, 1986; Ghini, 1993; Post, 1999; Prieto, 2005). More generally, the basic representations that underlie prosodic structure are still controversial. Although much of the reviewed work above assumes that the acoustic properties of prosody map onto an abstract phonological representation, this view is not shared by all researchers in the field (e.g. see

Ladd, 2008 for discussion), some of whom argue that there is a direct mapping between acoustic features and linguistic meaning (e.g., Xu & Xu, 2005). There is debate surrounding the representations that underlie prosody more generally as well as questions about the assumptions that underlie coding systems like ToBI. There are also questions about how prosodic structure fits into models of language production.

Note that the basic questions we review here about intonational phrasing and pitch accents: questions about acoustic correlates, function, and underlying representations, are also fundamental questions in these other areas of prosody. An interdisciplinary approach, which has been so fruitful in furthering our understanding of pitch accenting and intonational breaks, may also be fruitful in these other areas.

## Acknowledgments

We would like to thank Mara Breen, Ted Gibson, and Laura Dilley for comments on earlier drafts of this paper. The first author was supported by FQRSC grant NP-132516, SSHRC Canada Research Chair 212482, and NSF grant 0642660, and the second author by NIH grant R01DC008774 from the National Institute on Deafness and Other Communication Disorders during the writing of this paper.

## References

- Allbritton DW, McKoon G, Ratcliff R. Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1996; 22(3):714–735.
- Alloppenna PD, Magnuson JS, Tanenhaus MK. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*. 1998; 38(4):419–439.
- Altmann GT, Kamide Y. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*. 1999; 73(3):247–264. [PubMed: 10585516]
- Anderson, C.; Carlson, K. Prosodic phrasing in DO/SC and closure sentences. Paper presented at The Seventeenth Annual CUNY Conference on Human Sentence Processing.; College Park, MD. 2004.
- Arnold, J. Unpublished PhD. Stanford University; 1998. Reference form and discourse patterns.
- Arnold JE. Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*. 2008; 23(4):495.
- Aylett, M. Stochastic suprasegmentals: Relationships between redundancy, prosodic structure and care of articulation in spontaneous speech. Sixth International Conference on Spoken Language Processing; 2000.
- 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(1):31–56. [PubMed: 15298329]
- Bader, M. Prosodic influences on reading syntactically ambiguous sentences. In: Fodor, J.; Ferreira, F., editors. *Reanalysis in sentence processing*. Dordrecht: Kluwer; 1998. p. 1-46.
- Bard, EG.; Aylett, MP. The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech. *Proceedings of the XIVth International Congress of Phonetic Sciences*.; 1999. p. 1753-1756.
- Bard EG, Anderson AH, Sotillo C, Aylett M, Doherty-Sneddon G, Newlands A. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*. 2000; 42:1–22.
- Beach CM. The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language*. 1991; 30:644–663.
- Beckman, ME.; Edwards, J. Intonational categories and the articulatory control of duration. In: Tohkura, Y.; Vatikiotis-Bateson, E.; Sagisaka, Y., editors. *Speech perception, production and linguistic structure*. Tokyo: Ohmsha; 1992. p. 359-375.
- Beckman M, Pierrehumbert J. Intonational structure in English and Japanese. *Phonology Yearbook*. 1986; 3:255–310.

- Bell, A.; Gregory, M.; Brenier, J.; Jurafsky, D.; Ikeno, A.; Girand, C. Which predictability measure affect content and durations?. Proceedings of ISCA Tutorial and Research Workshop on Pronunciation Modeling and Lexicon Adaptation for Spoken Language; 2002.
- 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(1):92–111.
- Berkovits R. Durational effects in final lengthening, gapping, and contrastive stress. *Language and Speech*. 1994; 37:237–250. [PubMed: 7861912]
- Birch S, Clifton C Jr. Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech*. 1995; 38(4):365–392. [PubMed: 8816086]
- Birch S, Clifton C. Effects of varying focus and accenting of adjuncts on the comprehension of utterances. *Journal of Memory and Language*. 2002; 47:571–588.
- Bolinger D. Accent is predictable (if you're a mind reader). *Language*. 1972; 48:633–644.
- Bolinger, D.; Abe, I.; Kanekiyo, T. *Forms of English: Accent, morpheme, order*. Cambridge, Ma: Harvard University Press; 1965.
- Breen M, Fedorenko E, Wagner M, Gibson E. Acoustic correlates of information structure. Submitted to *Language and Cognitive Processes*. 2009
- Brennan SE. Centering attention in discourse. *Language and Cognitive Processes*. 1995; 102:137–167.
- Brown-Schmidt, S.; Campana, E.; Tanenhaus, MK. Real-time reference resolution by naïve participants during a task-based unscripted conversation. In: Trueswell, JC.; Tanenhaus, MK., editors. *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge, Mass: MIT Press; 2005.
- Büring, D. Focus projection and default prominence. In: Molnár, V.; Winkler, S., editors. *The architecture of focus*. Berlin: Mouton De Gruyter; 2003.
- Büring, D. Intonation, semantics, and information structure. In: Ramchand, G.; Reiss, C., editors. *The Oxford handbook of linguistic interfaces*. Oxford: Oxford University Press; 2007. p. 445-473.
- Byrd D, Krivokapic J, Lee S. How far, how long: On the temporal scope of prosodic boundary effects. *The Journal of the Acoustical Society of America*. 2006; 120:1589. [PubMed: 17004481]
- Byrd D, Saltzman E. Intra-gestural dynamics of multiple phrasal boundaries. *Journal of Phonetics*. 1998; 26:173–199.
- Byrd D, Saltzman E. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*. 2003; 31(2):149–180.
- Carlson K, Clifton J, Charles &, Frazier L. Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*. 2001; 45(1):58–81.
- Chafe, W. Cognitive constraints on information flow. In: Tomlin, R., editor. *Coherence and grounding in discourse*. Amsterdam: John Benjamins; 1987. p. 21-51.
- Chen A, den Os E, de Ruyter JP. Pitch accent type matters for online processing of information status: Evidence from natural and synthetic speech. *The Linguistic Review*. 2007; 24:317–344.
- Chen MY. The syntax of Xiamen tone sandhi. *Phonology Yearbook*. 1987; 4:109–149.
- Cho, T. *The effects of prosody on articulation in English*. New York: Routledge; 2002.
- Chomsky, N. Deep Structure, Surface Structure, and Semantic Interpretation. In: Steinberg, DD.; Jakobovits, LA., editors. *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*. Cambridge University Press; 1971.
- Chomsky, N.; Halle, M. *The sound pattern of English*. New York: Harper & Row; 1968.
- Cinque, Guglielmo. A null theory of phrase and compound stress. *Linguistic Inquiry*. 1993; 24:239–298.
- Clifton CJ, Frazier L, Carlson K. Tracking the what and why of speakers' choices: Prosodic boundaries and the length of constituents. *Psychonomic Bulletin & Review*. 2006; 13(5):854–861. [PubMed: 17328385]
- Clifton CJ, Karlson K, Frazier L. Informative prosodic boundaries. *Language and Speech*. 2002; 45:87–114. [PubMed: 12613557]
- Cooper, WE.; Paccia-Cooper, Jeanne. *Syntax and speech*. Cambridge, Mass: Harvard University Press; 1980.



- Cutler A, Dahan D, van Donselaar W. Prosody in the comprehension of spoken language: A literature review. *Language and Speech*. 1997; 40(2):141–201. [PubMed: 9509577]
- Dahan D, Tanenhaus MK, Chambers CG. Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*. 2002; 47(2):292–314.
- de Pijper JR, Sanderman AA. On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *The Journal of the Acoustical Society of America*. 1994; 96:2037.
- Dilley, L. Unpublished PhD. MIT; 2005. The phonetics and phonology of tonal systems.
- Dilley, L.; Breen, M.; Bolivar, M.; Kraemer, J.; Gibson, E. A comparison of inter-coder reliability for two systems of prosodic transcriptions: RaP (Rhythm and Pitch) and ToBI (Tones and Break Indices). *Proceedings of the International Conference on Spoken Language Processing*; Pittsburgh, PA. 2006.
- Dilley L, Brown M. The RaP (rhythm and pitch) labeling system, version 1.0. Ms. 2005
- Dilley L, Shattuck-Hufnagel S, Ostendorf M. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*. 1996; 24:423–444.
- Eady SJ, Cooper WE, Klouda GV, Mueller PR, Lotts DW. Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech*. 1986; 29:233–251. [PubMed: 3695759]
- Edwards J, Beckman ME, Fletcher J. The articulatory kinematics of final lengthening. *The Journal of the Acoustical Society of America*. 1991; 89:369. [PubMed: 2002175]
- Fant G, Kruckenberg A. On the quantal nature of speech timing. *Spoken Language*, 1996. ICSLP 96.Proceedings. 1996
- Féry C, Truckenbrodt H. Tonal scaling and the sisterhood principle. 2004
- Ferreira F. Prosody and performance in language production. *Language and Cognitive Processes*. 2007; 22:1151–1177.
- Ferreira, F. Unpublished doctoral dissertation. Amherst, University of Massachusetts; Amherst: 1988. Planning and timing in sentence production: The syntax-to-phonology conversion.
- Ferreira F. Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*. 1991; 30(2):210–233.
- Ferreira F. Creation of prosody during sentence production. *Psychological Review*. 1993; 100:233–253. [PubMed: 8483983]
- Ferreira, F. *Encyclopedia of Cognitive Science*. London, U.K: Macmillan Reference Ltd; 2002. Prosody.
- Fodor JD. Learning to parse? *Journal of Psycholinguistic Research*. 1998; 27:285–319.
- Fodor, JD. *Speech Prosody*. Aix-en-Provence; France: 2002. Psycholinguistics cannot escape prosody; p. 83-90.
- Fougeron C, Keating P. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*. 1997; 101:3728–3740. [PubMed: 9193060]
- Fowler C, Housum J. Talkers signaling of new and old words produced in various communicative contexts. *Language and Speech*. 1987; 28:47–56.
- Frazier, L.; Clifton, C. Sentence reanalysis and visibility. In: Fodor, JD.; Ferreira, F., editors. *Reanalysis in sentence processing*. Dordrecht: Kluwer; 1998. p. 143-176.
- Frazier L, Carlson K, Clifton C Jr. Prosodic phrasing is central to language comprehension. *Trends in Cognitive Science*. 2006; 10:244–249.
- Frazier, L.; Clifton, C. *Construal*. Cambridge, Mass: MIT Press; 1996.
- Fry DB. Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*. 1955; 27:765–768.
- Fry DB. Experiments in the perception of stress. *Language and Speech*. 1958; 1(2):126–152.
- Gahl S, Garnsey SM. Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*. 2004; 80(4):748–775.
- Gee J, Grosjean F. Performance structures: A psycholinguistic appraisal. *Cognitive Psychology*. 1983; 15:411–458.

- Ghini M. Phi-formation in Italian: A new proposal. *Toronto Working Papers in Linguistics*. 1993; 12:41–78.
- Gibson, E. The dependency locality theory: A distance-based theory of linguistic complexity. In: Miyashita, Y.; Marantz, A.; O'Neil, W., editors. *Image, language, brain*. Cambridge, MA: MIT Press; 2000. p. 95-126.
- Grabe E, Warren P, Nolan F. Resolving category ambiguities - evidence from stress shift. *Speech Communication*. 1994; 15:101–114.
- Gregory, M. Linguistic informativeness and speech production: An investigation of contextual and discourse-pragmatic effects on phonological variation. University of Colorado; Boulder: 2001.
- Gregory, ML.; Healy, AF.; Jurafsky, D. Unpublished Manuscript. University of Colorado; Boulder: 2002. Common ground in production: Effects of mutual knowledge on word duration.
- Grosjean F, Collins M. Breathing, pausing, and reading. *Phonetica*. 1979; 36:98–114. [PubMed: 482374]
- Grosz BJ, Joshi AK, Weinstein S. Centering: A framework for modeling the Local Discourse. *Computational Linguistics*. 1995; 21(2):203–225.
- Grosz BJ, Sidner C. Attention, intentions, and the structure of discourse. *Computational Linguistics*. 1986; 12:175–204.
- Gundel JK, Hedberg N, Zacharski R. Cognitive status and the form of referring expressions. *Language*. 1993; 69:274–307.
- Gunlogson, C. *True to form: Rising and falling declaratives as questions in English*. New York: Routledge; 2003.
- Gussenhoven, C. A semantic analysis of the nuclear tones of English. Bloomington, Ind: Indiana University Linguistics Club; 1983.
- Gussenhoven, C. On the grammar and semantics of sentence accents. Cinnaminson, N.J., U.S.A: Foris Publications; 1984.
- Gussenhoven C. The English rhythm rule as an accent assignment rule. *Phonology*. 1991; 8:1–35.
- Gussenhoven, C. *The phonology of tone and intonation*. Cambridge: CUP; 2004.
- Halliday MAK. Notes on transitivity and theme in English: Part 2. *Journal of Linguistics*. 1967; 3:199–244.
- Heldner M. Spectral emphasis as an additional source of information in accent detection. *Prosody 2001: ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding*. 2001:57–60.
- Hirose, Y. Unpublished doctoral dissertation. CUNY; 1999. Resolving reanalysis ambiguity in Japanese relative clauses.
- Hirschberg J. Pitch accent in context predicting intonational prominence from text. *Artificial Intelligence*. 1993; 63(1–2):305–340.
- Ito K, Speer SR. Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*. 2008; 85(2):541–573. [PubMed: 19190719]
- Isaacs, A.; Watson, DG. *Language and Cognitive Processes*. Accent detection is a slippery slope: Direction and rate of F0 change drives listeners' comprehension. this volume
- Jackendoff, R. *Semantic Interpretation in Generative Grammar*. MIT Press; 1972.
- Jackendoff, RS. *X'-syntax: A study of phrase structure*. Cambridge, MA: MIT Press; 1977.
- Jaeger, TF. Unpublished doctoral dissertation. Stanford University; 2006. Redundancy and syntactic reduction in spontaneous speech.
- Jun, SA. Unpublished doctoral dissertation. The Ohio State University; 1993. The phonetics and phonology of Korean prosody.
- Jun, SA. *Prosodic Typology: The Phonology and Intonation and Phrasing*. Oxford: OUP; 2005.
- Jurafsky, D.; Bell, A.; Gregory, M.; Raymond, WD. Probabilistic relations between words: Evidence from reduction in lexical production. In: Bybee, J.; Hopper, Paul, editors. *Frequency in the emergence of linguistic structure*. Amsterdam: John Benjamins; 2001. p. 229-254.
- Keating, P.; Cho, T.; Fougeron, C.; Hsu, C. Domain-initial strengthening in four languages. 2003. p. 143-161.

- Kehler A, Kertz L, Rohde H, Elman JL. Coherence and coreference revisited. *Journal of Semantics*. 2008; 25.1:1–44.
- Kim, H.; Chavarría, S.; Yoon, TJ.; Cole, J.; Hasegawa-Johnson, M. Acoustic differentiation of ip and IP boundary levels: Comparison of L- and L-L% in the switchboard corpus. *Proceedings of Speech Prosody*; 2004.
- Kjelgaard MM, Speer SR. Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*. 1999; 40(2):153–194.
- Klatt DH. Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*. 1975; 3:129–140.
- Kochanski G, Grabe E, Coleman J, Rosner B. Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*. 2005; 118:1038. [PubMed: 16158659]
- Kraljic T, Brennan SE. Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology*. 2005; 50:194–231. [PubMed: 15680144]
- Ladd DR. *Declination and ‘Reset’ and the hierarchical organization of utterances*. *JASA*. 1988; 84
- Ladd, DR. *Intonational phonology*. 2. Cambridge, England; New York, NY, USA: Cambridge University Press; 2008.
- Ladd D, Morton R. The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics*. 1997; 25:313–342.
- Larkey LS. Reiterant speech: An acoustic and perceptual validation. *The Journal of the Acoustical Society of America*. 1983; 73:1337. [PubMed: 6853845]
- Lavoie, L. *Consonant strength: Phonological patterns and phonetic manifestations*. New York: Garland; 2001.
- Lehiste, I. *Suprasegmentals*. Cambridge, MA: MIT Press; 1970.
- Lehiste I. Phonetic disambiguation of syntactic ambiguity. *Glossa*. 1973; 7:107–122.
- Levy R, Jaeger TF. Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems (NIPS)*. 2007; 19:849–856.
- Lieberman M, Prince A. On Stress and Linguistic Rhythm. *Linguistic Inquiry*. 1977; 8(2):249–336.
- Lieberman P. Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*. 1960; 32(4):451–454.
- Lieberman, P. *Intonation, perception, and language*. Cambridge: M.I.T. Press; 1967.
- Lovic, N. *Implicit prosody in silent reading: Relative clause attachment in Croatian*. Unpublished CUNY; New York, NY: 2003.
- MacDonald, M. Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In: MacWhinney, B., editor. *The Emergence of Language*. Mahwah, NJ: Lawrence Erlbaum; 1999. p. 177-196.
- Marcus, M.; Hindle, D. Description theory and intonation boundaries. In: Altmann, G., editor. *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. Cambridge, MA: MIT Press; 1990. p. 483-512.
- Marslen-Wilson W, Tyler L, Warren P, Grenier P, Lee C. Prosodic effects in minimal attachment. *The Quarterly Journal of Experimental Psychology Section A*. 1992; 45(1):73–87.
- Martin JG. On judging pauses in spontaneous speech. *Journal of Verbal Learning and Verbal Behavior*. 1970; 9:75–78.
- McMurray B, Tanenhaus MK, Aslin RN. Gradient effects of within-category phonetic variation on lexical access. *Cognition*. 2002; 86(2):B33–42. [PubMed: 12435537]
- Nagel HN, Shapiro LP, Tuller B, Nawy R. Prosodic influences on the resolution of temporary ambiguity during on-line sentence processing. *Journal of Psycholinguistic Research*. 1996; 25(2): 319–344. [PubMed: 8667301]
- Nespor, M.; Vogel, I. *Prosodic phonology*. Dordrecht: Foris; 1986.
- O’Malley M, Kloker D, Dara-Abrams B. Recovering parentheses from spoken algebraic expressions. *IEEE Transactions on Audio and Electroacoustics*. 1973; 21(3):217–220.
- Pierrehumbert, J. Unpublished doctoral dissertation. MIT; 1980. *The phonology and phonetics of english intonation*.

- Pierrehumbert, J.; Hirschberg, J. The meaning of intonational contours in the interpretation of discourse. In: Cohen, PR.; Morgan, J.; Pollack, ME., editors. *Intentions in communication*. Cambridge, Ma: MIT Press; 1990. p. 271-311.
- Post B. Restructured phonological phrases in French: Evidence from clash resolution. *Linguistics*. 1999; 37(1):41–63.
- Price PJ, Ostendorf S, Shattuck-Hufnagel S, Fong C. The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*. 1991; 9:2956–2970. [PubMed: 1787237]
- Prieto, P. Syntactic and eurhythmic constraints on phrasing decisions in Catalan. In: Horne, M.; van Oostendorp, M., editors. *Studia Linguistica*. Vol. 59. 2005. p. 194-222. special issue on 'Boundaries in Intonational Phonology
- Prince, E. Toward a taxonomy of given-new information. In: Cole, Peter, editor. *Radical pragmatics*. New York: Academic Press; 1981. p. 223-255.
- Pynte J, Prieur B. Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes*. 1996; 11(1):165.
- Redi L, Shattuck-Hufnagel S. Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*. 2001; 29(4):407–429.
- Reinhart, T. *Interface strategies: Focus: The PF-interface*. Cambridge, Ma: MIT Press; 2006.
- Rietveld ACM, Gussenhoven C. On the relation between pitch excursion size and prominence. *Journal of Phonetics*. 1985; 13:299–308.
- Rooth M. A theory of focus interpretation. *Natural Language Semantics*. 1992; 1:75–116.
- Samek-Lodovici V. Prosody-syntax in the expression of focus. *Natural Language & Linguistic Theory*. 2002; 23:687–755.
- Schafer, AJ.; Speer, SR.; Warren, P. Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In: Tanenhaus, M.; Trueswell, J., editors. *Approaches to studying world situated language use: Psycholinguistic, linguistic and computational perspectives on bridging the product and action tradition*. Cambridge: MIT Press; 2005.
- Schafer, A. Doctoral dissertation. University of Massachusetts Amherst; 1997. *Prosodic parsing: The role of prosody in sentence comprehension*.
- Schmerling, SF. *Aspects of English sentence stress*. Austin: University of Texas; 1976.
- Schwarzschild R. Givenness, AVOIDF and other constraints on the placement of accent. *Natural Language Semantics*. 1999; 7:141–177.
- Sedivy JC, Tanenhaus MK, Chambers CG, Carlson GN. Achieving incremental semantic interpretation through contextual representation. *Cognition*. 1999; 71(2):109–147. [PubMed: 10444906]
- Selkirk, EO. *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press; 1984.
- Selkirk EO. On derived domains in sentence phonology. *Phonology Yearbook*. 1986; 3:371–405.
- Selkirk, EO. Sentence prosody: Intonation, stress and phrasing. In: Goldsmith, JA., editor. *The Handbook of Phonological Theory*. Cambridge, Mass., USA: Blackwell; 1995.
- Selkirk, EO. The interaction of constraints on prosodic phrasing. In: Horne, M., editor. *Prosody: Theory and experiment*. Dordrecht: Kluwer; 2000. p. 231-262.
- Shattuck-Hufnagel S, Turk AE. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*. 1996; 25(2):193–247. [PubMed: 8667297]
- Silverman, KEA.; Beckman, M.; Pitrelli, JF.; Ostendorf, M.; Wightman, C.; Price, P.; Pierrehumbert, J.; Hirschberg, J. TOBI: A standard for labeling English prosody. *Proceedings of the 1992 International Conference on Spoken Language Processing; Banff, Canada*: 1992. p. 867-870.
- Sluijter AMC, van Heuven VJ. Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America*. 1996; 100:2471. [PubMed: 8865652]
- Snedeker J, Casserly E. Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. *Language and Cognitive Processes*. this volume.
- 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.

- Speer SR, Kjelgaard MM, Dobroth KM. The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*. 1996; 25(2):249–271. [PubMed: 8667298]
- Steedman M. Structure and intonation. *Language*. 1991; 67(2):260–296.
- Steedman, M. *Surface Structure and Interpretation*. Cambridge: MIT Press; 1996.
- Stirling L, Wales R. Does prosody support or direct sentence processing? *Language and Cognitive Processes*. 1996; 11:193–212.
- Streeter LA. Acoustic determinants of phrase boundary perception. *The Journal of the Acoustical Society of America*. 1978; 64(6):1582–1592. [PubMed: 739094]
- Syrdal AK, McGory J. Inter-transcriber Reliability of ToBI Prosodic Labeling. *Proc ICSLP*. 2000; 3:235–238.
- Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JC. Integration of visual and linguistic information in spoken language comprehension. *Science*. 1995; 268(5217):1632–1634. [PubMed: 7777863]
- Terken J, Hirschberg J. Deaccentuation of words representing :”given” information: Effects of persistence of grammatical function and surface position. *Language and Speech*. 1994; 37(2): 125–145.
- Terken J, Nooteboom S. Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes*. 1987; 2(3):145–163.
- Truckenbrodt, H. *Phonological phrases: Their relation to syntax, focus, and prominence*. MIT; 1995.
- Truckenbrodt H. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry*. 1999; 30(2):219–255.
- Truckenbrodt H. Upstep and embedded register levels. *Phonology*. 2002; 19(01):77–120.
- Truckenbrodt, H.; Darcy, I. Object clauses and phrasal stress. To appear in *Object clauses, movement, and phrasal stress*. With Isabelle Darcy. In: Shir, Nomi; Rochman, Lisa, editors. *The sound patterns of syntax*. Oxford University Press; 2008.
- Turk AE, Sawusch JR. The processing of duration and intensity cues to prominence. *Journal of the Acoustical Society of America*. 1996; 99(6):3782–3790.
- Turk AE, White L. Structural influences on accentual lengthening in English. *Journal of Phonetics*. 1999; 27(2):171–206.
- Turk, AE. Paper presented. *Labphon: University of Wellington*; 2008. Prosodic constituency signals relative predictability.
- van den Berg, R.; Gussenhoven, Carlos; Rietveld, Toni. Downstep in dutch: Implications for a model. In: Docherty, Gerlad; Ladd, Robert, editors. *Papers in laboratory phonology, vol. II: Gesture, segment, prosody*. Cambridge: Cambridge University Press; 1992. p. 335-358.
- Vanderslice R, Ladefoged P. Binary suprasegmental features and transformational word-accentuation rules. *Language*. 1972; 48(4):819–838.
- Venditti, J.; Hirschberg, J. Intonation and discourse processing. *Proceedings of the International Congress of Phonetic Sciences; Barcelona, Spain*. 2003.
- Wagner, M. Unpublished doctoral dissertation. MIT; 2005. Prosody and recursion.
- Wagner, M. Givenness and Locality. In: Gibson, M.; Howell, J., editors. *Proceedings of SALT XVI*. Ithaca, NY: CLC Publications; 2006. p. 295-312.
- Wagner M. Prosody and recursion in coordinate structures and beyond. *Natural Language & Linguistic Theory*. in press.
- Warren, P. *The temporal organization and perception of speech*. Unpublished University of Cambridge; Cambridge, U.K: 1985.
- Warren P, Grabe E, Nolan F. Prosody, phonology and parsing in closure ambiguities. *Language and Cognitive Processes*. 1995; 10(5):457–486.
- Watson, DG.; Arnold, JE.; Tanenhaus, MK. Not just given and new: The effects of discourse and task based constraints on acoustic prominence. *The 2005 CUNY Human Sentence Processing Conference*;; Tucson, AZ. 2005.

- Watson DG, Arnold JE, Tanenhaus MK. Tic tac toe: Effects of predictability and importance on acoustic prominence in language production. *Cognition*. 2008; 106(3):1548–1557. [PubMed: 17697675]
- Watson DG, Breen M, Gibson E. The role of syntactic obligatoriness in the production of intonational boundaries. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2006; 32(5):1045–1056.
- Watson DG, Gibson EA. Making sense of the sense unit condition. *Linguistic Inquiry*. 2004a; 35:508–517.
- Watson DG, Gibson EA. The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*. 2004b; 19(6):713–755.
- Watson D, Gibson EA. Intonational phrasing and constituency in language production and comprehension. *Studia Linguistica*. 2005; 59(2–3):279–300.
- Watson, DG.; Gunlogson, C.; Tanenhaus, MK. Online methods for the investigation of prosody. In: Saube, A., editor. *Methods in empirical prosody research*. Leipzig: Mouton de Gruyter; 2006. p. 259-282.
- Watson DG, Tanenhaus MK, Gunlogson C. Interpreting pitch accents in on-line comprehension: H\* vs. L+H\*. *Cognitive Science*. 2008; 32:1232–1244. [PubMed: 21585451]
- Watt SM, Murray WS. Prosodic form and parsing commitments. *Journal of Psycholinguistic Research*. 1996; 25(2):291–318. [PubMed: 8667300]
- Weber A, Braun B, Crocker MW. Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*. 2006; 49(3):367–392. [PubMed: 17225671]
- Welby P. Effects of pitch accent position, type, and status on focus projection. *Language and Speech*. 2003; 46(1):53–81. [PubMed: 14529111]
- Wheeldon L, Lahiri A. Prosodic units in speech production. *Journal of Memory and Language*. 1997; 37:356–381.
- Wightman CW, Shattuck-Hufnagel S, Ostendorf M, Price PJ. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*. 1992; 92:1707–1717. [PubMed: 1564206]
- Williams E. Blocking and anaphora. *Linguistic Inquiry*. 1997; 28:577–628.
- Xu Y, Xu CX. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*. 2005; 33(2):159–197.