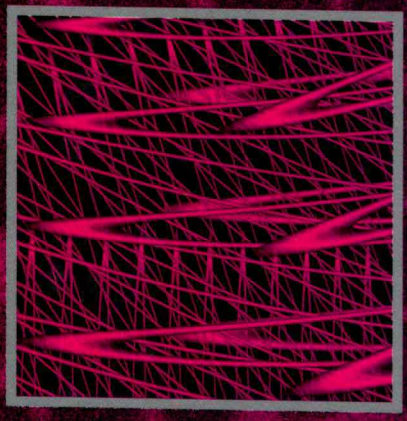
Experimental and Theoretical Advances in Prosody



Edited by Duane G. Watson, Michael Wagner, Edward Gibson

Psychology Press
Taylor & Francis Group

A special issue of the Journal
Language and Cognitive Processes

A Special Issue of
Language and Cognitive Processes

Experimental and Theoretical Advances in Prosody

Edited by

Duane G. Watson
University of Illinois, USA

Michael Wagner
McGill University, Canada

and

Edward Gibson
Massachusetts Institute of Technology, USA

Psychology Press
Taylor & Francis Group
HOVE AND NEW YORK

Published in 2010 by Psychology Press
27 Church Road, Hove, East Sussex BN3 2FA

www.psypress.com

Simultaneously published in the USA and Canada
by Psychology Press
270 Madison Avenue, New York NY 10016

Psychology Press is part of the Taylor & Francis Group, an Informa business

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN 978-1-84872-740-3
ISSN 0169-0965

Cover design by Jim Wilkie
Typeset in India by Datapage (India) Private Limited, Chennai
Printed in the UK by Henry Ling Ltd., Dorchester
Bound in the UK by TJ International, Padstow, Cornwall

Contents*

*This book is also a special issue of the journal *Language and Cognitive Processes*, and forms issues 7, 8, & 9 of Volume 25 (2010). The page numbers are taken from the journal and so begin with p. 905

LANGUAGE AND COGNITIVE PROCESSES
2010, 25 (7/8/9), 905–945

Psychology Press
Taylor & Francis Group

Experimental and theoretical advances in prosody: A review

Michael Wagner[1,2] and Duane G. Watson[3]

[1]*Cornell University, Ithaca, NY, USA,* [2]*Department of Linguistics, McGill University, Montreal, QC, Canada,* [3]*Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, USA*

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 summarises 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, these topics are representative of a tendency in the field towards a more interdisciplinary approach.

Keywords: Prosody; Intonation; Comprehension; Production.

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

Correspondence should be addressed to Duane G. Watson, Department of Psychology, University of Illinois at Urbana-Champaign, 603 E Daniel St., Champaign, IL 61820, USA. E-mail: dgwatson@illinois.edu

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.

http://www.psypress.com/lcp DOI: 10.1080/01690961003589492

processing during speech production, others have been conventionalised 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 10 years before the publication of this paper, Cutler, Dahan, and van Donselaar (1997) and Shattuck-Hufnagel and Turk (1996) 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 centred or comprehension centred?

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 Ferreira, 2002; Ladd, 2008).

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 emphasising 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 favour, boundary strength and prominence, the two topics on which the remainder of this review will focus, would count as prosody under either.

BOUNDARIES

An utterance of more than two words in it often has a perceptible subgrouping (Lehiste, 1973). Prosodic grouping can be produced and perceived even in the absence of identifiable words (cf. de Pijper & Sanderman, 1994; Larkey, 1983). 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 (Byrd & Saltzman, 1998; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Shattuck-Hufnagel & Turk, 1996; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). Pre-boundary lengthening correlates with other acoustic cues that reflect that articulatory gestures of segments preceding boundaries are spatially more extreme, i.e., hyperarticulated (Byrd & Saltzman, 2003; Edwards, Beckman, & Fletcher, 1991; Fougeron & Keating, 1997), 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, Krivokapic, & Lee, 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, Kloker, and Dara-Abrams (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*. Cho (2002), Fougeron and Keating (1997), Jun (1993), Keating, Cho, Fougeron, and Hsu (2003) and Lavoie (2001) show that the phonetic realisation 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 organisation. 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, Shattuck-Hufnagel, and Ostendorf (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 is pitch events that occur at the edges of strong prosodic domains. They are commonly analysed 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 (Gussenhoven, 1984, 2004; Pierrehumbert & Hirschberg, 1990). It is not clear, however, whether every pitch event at a boundary can be analysed in this way.[1]

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). Féry and Truckenbrodt (2005) and Ladd (1988) 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 (1a) and (1b). 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

[1] 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.

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, Gussenhoven, and Rietveld (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, Yoon, Cole, and Hasegawa-Johnson (2006) 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 categorised according to a very limited inventory of boundary types that are organised 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 three categories, intonation phrase, intermediate phrase, and word—lower prosodic categories are not encoded in the ToBI labelling 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 Kim et al., 2006; Ladd, 2008).

Since Price et al. (1991), the ToBI system includes a boundary strength annotation, the *break index*, which is based on boundary type differences. Syrdal and McGory (2000), however, found poor inter-labeller 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 categorising 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 (de 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, Breen, Bolivar, Kraemer, and Gibson (2006) and Dilley and Brown (2005). 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-labeller agreement of RaP and ToBI annotations (Dilley et al., 2006) found a higher inter-labeller 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 realisation 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. Categorical grammar is a theory of how syntax and meaning composition go hand in hand. It provides a range of operations that can effectively re-bracket 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, 2010), 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 characterise this link, both in the literature

on phonological theory (Nespor & Vogel, 1986; Selkirk, 1984, 1986; Truckenbrodt, 1995) and in the processing literature (Cooper & Paccia-Cooper, 1980; Ferreira, 1988; Gee & Grosjean, 1983; Grosjean & Collins, 1979; 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 (Ferreira, 1988; Gee & Grosjean, 1983; 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 (Nespor & Vogel, 1986; Selkirk, 1984). 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 eight 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 (2006a) 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 et al. (2006a) 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 signalling of pragmatic and semantic information as in the case of asides, appositives, and nonrestrictive 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, Spivey-Knowlton, Eberhard, & Sedivy, 1995)? Studies suggest that nonsyntactic 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 20 years strongly suggests that prosody is one of the many factors that are rapidly integrated into the linguistic representation (Grabe, Warren, & Nolan, 1994; Kjelgaard & Speer, 1999; Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992; Snedeker & Trueswell, 2003; Watson & Gibson, 2005 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, Grabe, & Nolan, 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. (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 (Clifton et al., 2006; Lehiste, 1973; Streeter, 1978; Wagner, 2005). 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, Clifton, Charles, & Frazier, 2001)
(6b) Mary maintained that the CEO lied // when the investigation started.

Finally, boundaries can play a role in signalling the presence of long distance dependencies (Kraljic & Brennan, 2005; Schafer, Speer, & Warren, 2005; Snedeker & Trueswell, 2003). 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.