Does prosodic constituency signal relative predictability? A Smooth Signal Redundancy hypothesis

Citation for published version:
Turk, A 2010, 'Does prosodic constituency signal relative predictability? A Smooth Signal Redundancy hypothesis', *Laboratory Phonology*, vol. 1, no. 2, pp. 227-262. https://doi.org/10.1515/LABPHON.2010.012

Digital Object Identifier (DOI):
10.1515/LABPHON.2010.012

Published In:
Laboratory Phonology

Does prosodic constituency signal relative predictability? A Smooth Signal Redundancy hypothesis

ALICE TURK

University of Edinburgh

Abstract

This paper explores issues relating to signaling word boundaries from the perspective of Aylett's Smooth Signal Redundancy proposal (Aylett 2000, Aylett and Turk 2004) that language has evolved to spread redundancy, i.e. recognition likelihood, evenly throughout utterances. In Aylett's proposal, information that enables listeners to identify sequences of elements in an utterance (signal redundancy) comes from two sources: a) language redundancy, recognition likelihood based on lexical, syntactic, semantic, pragmatic and other factors, and b) acoustic redundancy, recognition likelihood based on acoustic salience. Smooth signal redundancy is achieved by a complementary relationship between language redundancy and acoustic redundancy that is implemented via prosodic structure.

While Aylett and Turk (2004) present the case for prosodic prominence as a lever for modulating the acoustic salience of syllables, the current paper proposes that prosodic constituency also fulfils this function for words. The current paper proposes that the signal redundancy, or recognition likelihood, of words can be manipulated by signaling their boundaries, and that the occurrence and strength of these boundary markers correlates inversely with language redundancy. Prosodic constituency implements the complementary relationship between language redundancy and word boundary salience.

Smooth Signal Redundancy provides an integrated explanation for a set of properties relating to prosodic constituent structure.

1. Introduction

It is often observed that spoken language lacks reliable cues to word boundaries in all contexts. However, speakers have other ways of signaling lexical words in speech, namely through the contrastive elements of which they are formed (phonemes/gestures/distinctive features, tones etc). For example, even without an overt phonetic correlate of the word boundary after *Tom* in the utterance *Tom went home,* speakers signal the word sequence *Tom went* by signaling the phonemes /t/,

JLP 1 (2010), 227–262
DOI 10.1515/LABPHON.2010.012
1836-6346/10/01–227

/ɔ/, /m/, /w/, /ɛ/, /n/, /t/. From this perspective, then, the question is slightly different: Given that words can be communicated via their distinctive features, why and when do speakers choose to use overt phonetic correlates of word boundaries?

This paper explores issues relating to signaling word boundaries from the perspective of Aylett's Smooth Signal Redundancy proposal (Aylett 2000; Aylett and Turk 2004; Aylett and Turk 2006). Following Shannon (1948) and Pierce (1961), Aylett proposes that language has evolved to spread redundancy evenly throughout utterances, where redundancy refers to the multiple clues to the identity of linguistic elements and equates with recognition likelihood. This even spreading of redundancy is argued to ensure robust, efficient communication in a potentially noisy environment. In the Smooth Signal Redundancy proposal, clues to the identity of sequences of elements in an utterance (signal redundancy) come from two main sources: a) language redundancy, that is, clues to identity based on e.g. lexical, syntactic, semantic and pragmatic factors, and b) acoustic redundancy, i.e. clues to identity based on acoustic salience. In Aylett (2000) and Aylett and Turk (2004), smooth signal redundancy is achieved by an inverse, complementary relationship between language redundancy and acoustic redundancy implemented via prosodic prominence structure. That is, speakers produce high acoustic redundancy (saliency) for unpredictable sections of speech, and less saliency when predictability is high. For example, in the sentence *Who's the author?*, as schematized in Figure 1, the language redundancy of the first two syllables *who's the* is relatively high. *Who*, *is*, and *the* are high frequency words; *who* is relatively often followed by *is*, and *who's* is relatively often followed by *the. Who* may have lower language redundancy than *is* or *the* since it begins the sentence and is less frequent. In contrast, the language redundancy of the first syllable in *author* is lower than both *who's* and *the*: it is relatively less frequent and could be one of any number of open class words that could have followed *the.* In addition, as a word-initial syl-

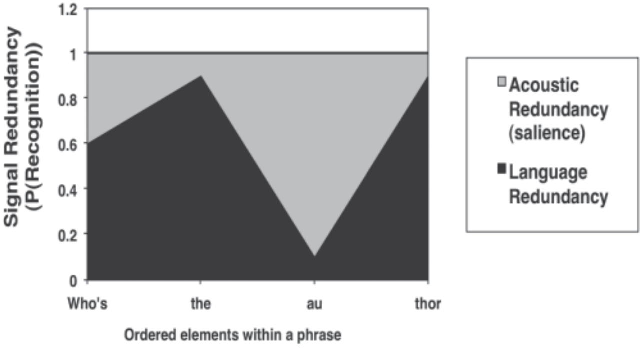

Figure 1. *Schematic diagram of the complementary relationship between language redundancy and acoustic redundancy. Language redundancy and acoustic redundancy combine to give signal redundancy.*

lable, *au(th)-* is relatively unpredictable compared to the second syllable *–(th)or;* the syllable *au(th)-* therefore has lower language redundancy than *–(th)or*. Prosodic prominence structure implements the redundancy profile by assigning a full vowel to *who's*[1] and by putting primary phrasal stress (nuclear pitch accent) on the syllable with lowest language redundancy. The acoustic correlates of prominence, i.e. full vowel quality, longer duration, F0 movement and less steep spectral balance, together with the segmental phonetic attributes, provide the acoustic redundancy. Because acoustic redundancy varies inversely with language redundancy, the information conveyed by the utterance (signal redundancy) is evenly distributed, and maximizes its likelihood of recognition even in the presence of randomly occurring noise. Figure 1 schematizes the inverse, complementary relationship between language redundancy and acoustic redundancy that yields smooth signal redundancy.

To understand why an even distribution of recognition likelihood maximizes correct recognition, let us consider the following example. Let us compare strings of elements whose overall probability of recognition, p(recognition), add up to 1: string AB, whose elements each have p(recognition) of 0.5, and thus has a smooth redundancy profile, and string CD whose elements have p(recognition) of 0.1 and 0.9 respectively, and thus does not have a smooth redundancy profile. The probability of correct recognition of both elements in the string in the correct order is a product of the p(recognition) of each element in the string (assuming that the p(recognition) of each element is independent of the other). The probability of correct recognition of string AB is $(0.5 * 0.5) = 0.25$, whereas the probability of correct recognition of string CD is less than half that of AB: $(0.1 * 0.9) = 0.09$.

Figure 2 shows how the Smooth Signal Redundancy view fits with more traditional views of the relationship of prosodic structure with the rest of grammar.

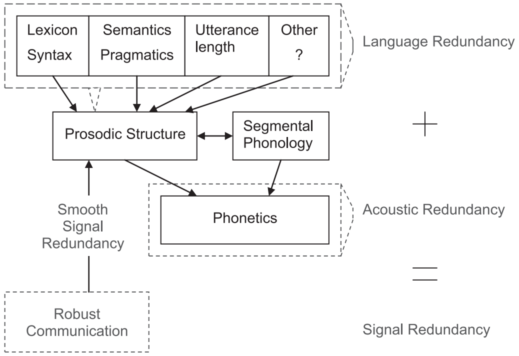

Figure 2. *Schematic diagram of the relationship of Smooth Signal Redundancy with a more traditional view of the relationship of prosodic structure with the rest of grammar. Based on a figure in Aylett (2000), which in turn was based on a figure in Shattuck-Hufnagel & Turk (1996).*

The inverse relationship between language redundancy and acoustic redundancy is well documented in the literature (cf. Lieberman 1963; Fowler and Housum 1987; Lindblom 1990; Jurafsky et al. 2001; Son and Pols 2003; Pluymaekers et al. 2005a, 2005b; Bell et al. 2009 among others, but see Kuperman et al. 2008). For example, Jurafsky et al. (2001) found that function words are more likely to reduce when either predictable in context or highly frequent in the language. The Smooth Signal Redundancy hypothesis contributes to this discussion by providing an explanation for this complementary relationship: The smooth distribution of information that it yields gives a communicative advantage.

Additionally, it proposes that prosodic (prominence) structure provides a lever with which to increase or decrease the acoustic redundancy provided by segmental structure. Results from Aylett (2000) and from Aylett and Turk (2004) are consistent with the view that prosodic prominence structure implements the complementary relationship between language and acoustic redundancy for syllables: In a corpus study of spontaneous speech, the durational exponents of phrase-medial word and phrasal stress correlate inversely with measures of language redundancy (word frequency, syllabic trigram probability and givenness).

The Smooth Signal Redundancy hypothesis as presented in Aylett (2000) and Aylett and Turk (2004) therefore consists of the following claims. First, there should be an inverse relationship between language redundancy (recognition likelihood based on the text) and acoustic redundancy (acoustic salience). Second, prosodic prominence structure shapes how this inverse relationship is realized, with the result that there is an even distribution of signal redundancy (overall recognition likelihood) across the syllables of an utterance or discourse.

The principles and mechanisms of Smooth Signal Redundancy are likely to have evolved for the benefit of both speaker and listener(s), because they allow the speaker to conserve effort while maximizing the likelihood that the listener(s) will recognize the utterance (cf. Lindblom 1990; Kuperman et al. 2008). While this originating impetus seems clear, the selectional forces could favor any of several mechanisms for balancing the needs of the speaker and hearer. One possibility is that the synchronic, online computation of redundancy, as well as the online planning of the phonetic characteristics that depend on it, proceed with the listeners' needs in mind as the speaking process unfolds. However, a growing body of evidence suggests that speech is often produced without much attention to listener requirements (see Bard et al. 2000; Schafer et al. 2004; and Bard and Aylett 2005 for experimental evidence). Crucially, the Smooth Signal Redundancy proposal does not require that the speaker necessarily take the listeners into account during the online speaking process. The speaker's language redundancy computation can be made on the basis of his or her own language experience. While not necessarily optimal for the listener, this type of language redundancy computation may represent a reasonable approximation to the language redundancy of the listener. Information about the listeners' knowledge or experience can be incorporated in the computation, but doesn't have to be.

When the speaker uses language redundancy information, his or her goal is to produce signal redundancy that is evenly distributed throughout the utterance(s). This smooth spread of signal redundancy represents the optimal *distribution* of signal redundancy for the listener, because it maximizes recognition likelihood (see also Levy and Jaeger 2007) for a given energy expenditure on behalf of the speaker. The extent to which the resulting signal meets the listener's needs will depend on the appropriateness of the language redundancy computation, the care and time with which the speaker chooses to implement it, and characteristics of the listener (e.g. attentional focus, hearing ability, distance from the speaker).

While Aylett and Turk (2004) present the case for prosodic prominence as a lever for modulating the acoustic salience of *syllables*, the current paper advances the hypothesis that prosodic constituency also fulfils this function for *words*. I claim that the acoustic redundancy, or relative salience, of lexical words can be manipulated by signaling their boundaries. Along with Bell et al. (2009), I ask whether the occurrence and strength of these boundary markers correlates inversely with language redundancy. Prosodic constituency is proposed to implement the complementary relationship between language redundancy and word boundary salience.

Additionally, I propose that the word boundary salience modulations that prosodic constituency governs are implemented within the constraints of global settings of effort and rate. For example, a generally slower global rate of speech and/or more careful or effortful speech can result in more and/or stronger boundaries (Caspers 1994; Strangert 2003). These global settings are proposed to add (or take away) a fixed amount of acoustic redundancy to all boundaries, and therefore raise (or lower) signal redundancy by a fixed amount. Nevertheless, the complementary relationship between language and acoustic redundancy is maintained.

The goal of this paper is to motivate Smooth Signal Redundancy and its implementation within global effort and rate settings as possible explanations for the use of prosodic constituency and its phonetic correlates. It is speculative. Background sections of this paper review facts about word boundary correlates and their hierarchical organization, and discuss current challenges in predicting how and when speakers choose to signal word boundaries. I then motivate the boundary-correlates-as-salience assumption, and discuss evidence for the relationship between predictability and boundary signaling, both crucial for establishing the relevance of the Smooth Signal Redundancy proposal for prosodic constituency. Sections 4.3 and 4.4 show how the Smooth Signal Redundancy framework can account for aspects of prosodic boundary occurrence that are unexplained by theories of syntax/prosodic mapping. Section 5 discusses the role of clarity and rate in modulating the strength of boundary markers. Section 6 suggests that together with variations in effort and rate, Smooth Signal Redundancy provides several sources of optionality in constituent boundary placement. Final sections discuss possible alternatives to the Smooth Signal Redundancy explanation for prosodic boundary occurrence, possible redundancy computation mechanisms, issues relating to the need for a separate prosodic component of grammar, and the possibility that Smooth Signal Redundancy operates on levels other than that of the syllable and word.

2. Background: Facts about word boundary correlates and their hierarchical organization

2.1. *A repertoire of boundary-signaling and grouping techniques*

Although speakers often produce strings of words without any noticeable cues to the boundaries between many of them, it is nevertheless clear that speakers have a large repertoire of boundary-signaling skills. These include

- constituent-initial and final voice quality modifications (e.g. Pierrehumbert and Talkin 1992; Ogden 2004; Dilley et al. 1996; Redi and Shattuck-Hufnagel 2001; Tanaka 2004)
- supralaryngeal articulatory modifications, (e.g. phrase-initial strengthening, syllable-final lenition, Fougeron and Keating 1997; Keating et al. 2003; Lavoie 2001),
- temporal modifications (e.g. initial and final lengthening, polysyllabic shortening, Lehiste 1972; Lindblom 1968; Nooteboom 1972; Wightman et al. 1992; Beckman and Edwards 1990; Turk and Shattuck-Hufnagel 2000; White 2002),
- the use of word- or phrasal-prominence near the beginnings or ends of constituents (e.g. Shattuck-Hufnagel et al. 1994; Astésano et al. 2007),
- pausing (Gee and Grosjean 1983),
- intonational phenomena, e.g. phrase-final lowering, phrase-initial reset (cf. Beckman and Pierrehumbert 1986; Ladd 2008 among others).

All of these phenomena target the edges of constituents, and can thus be termed ‘Domain limit phenomena’ (Selkirk 1980). Other types of phonetic attributes appear to signal the membership of elements within constituents. These include vowel harmony and anticipatory co-articulation (pharyngealisation in Arabic: Ghazeli 1977; Bukshaisha 1985; nasalization in English: Krakow 1999), and have been termed ‘Domain Span phenomena’ (Selkirk 1980), although many don’t necessarily span throughout an entire domain. Still other boundary-related correlates seem to function like sticky tape in the way that they adjoin adjacent constituents. These so-called Domain Juncture phenomena (Selkirk 1980) include segmental and tone sandhi, where one or more properties of the edge of one constituent are influenced by one or more properties of an immediately adjacent constituent. Other examples can be found in Nespor and Vogel (1986) and elsewhere.

2.2. *Boundary-related phenomena as correlates of a hierarchy of constituents*

Several lines of evidence suggest that these boundary-related phenomena are correlates of a hierarchy of prosodic constituents.

Some segmental phenomena apply with reference to relatively small constituents, e.g. syllable-final s → h in some varieties of Spanish. Other phenomena are influenced by potentially larger constituents (e.g. intonational boundary tone occurrence), see Nespor and Vogel (1986) for examples. In addition, the magnitude of some phonetic parameters varies in a gradient way, as predicted by a hierarchical representation. For example, the magnitude (but not the extent, Cambier-Langeveld 1997, 2000), of final lengthening can be related to a hierarchy of prosodic constituents, with more final lengthening at the boundaries of higher level constituents (Wightman et al. 1992 for English). The magnitude of tongue-palate contact for coronal consonants in constituent-initial position patterns in a similar way, in many languages of the world (Keating et al. 2003), as does pause duration (Gee and Grosjean 1983; Ferreira 1993). Evidence for the hierarchical structuring of boundary-related correlates also comes from the likelihood of occurrence of these correlates. For example, Astésano et al. (2007) showed that the likelihood of French initial accent occurrence varied with prosodic boundary strength.

Prosodic constituency is word-based in the sense that higher level constituents group words into larger constituents, and lower level constituents (e.g. syllables) sub-divide words. Crucially, constituents of the prosodic hierarchy never group word-fragments from different words together in single constituents. By virtue of the fact that higher constituent boundaries delimit *words*, the prosodic hierarchy represents a coarse-grained hierarchy of relative word boundary strength. On this view, the boundaries of higher level constituents (e.g. Intermediate Intonational Phrases, Full Intonational Phrases), are strong word boundaries, and the boundaries of lower level constituents (e.g. Minor phrase and Prosodic word boundaries) are weaker.

2.3. *Non-isomorphism of prosodic and syntactic hierarchies*

Many of the word boundaries signaled by overt phonetic means have been related to morpho-syntax. The close correspondence between syntax and word-boundary correlates is illustrated by the well-documented fact that word-boundary correlates often provide cues to listeners that allow them to recover the intended syntactic structures of strings that are potentially ambiguous. For example, Lehiste (1973) and Price et al. (1991) cite examples such as *when you learn, gradually you worry more vs. when you learn gradually, you worry more*, whose left vs. right syntactic attachment of e.g. *gradually* is typically disambiguated by prosody (one of several types of examples cited in Shattuck-Hufnagel and Turk 1996).

However, it is equally clear that syntactic and prosodic hierarchies are different. The non-isomorphism of prosodic and morpho-syntactic trees is shown by the non-occurrence of prosodic boundaries where syntax would predict them (e.g. boundaries between subject pronouns and verbs are rare, Gee and Grosjean 1983), by the occurrence of boundaries where syntax would predict none (e.g. *Sesame Street is brought to you by, the Children's Television Workshop*, cited in Shattuck-Hufnagel and Turk 1996), and by restrictions on recursivity in prosodic structure as compared to syntax (cf. the childrens' rhyme discussed in Chomsky and Halle 1968; Ladd 1996 *this is the dog that chased the cat that killed the rat that ate the malt . . . that lay in the house that Jack built.* which shows indefinite syntactic embedding, but a flatter, less-recursive prosodic structure).

2.4. *Challenges for the prosody-syntax mismatch problem and proposed solutions*

To deal with some of these discrepancies between syntactic and prosodic constituent structure, some researchers have proposed prosody-syntax relationships which involve aligning single edges of prosodic constituents with single edges of syntactic constituents (e.g. Selkirk 1986, 1993; Hale and Selkirk 1987; Truckenbrodt 1999). Others have argued that a better syntax-prosody mapping can be achieved by adopting alternative theories of syntax; cf. combinatorial categorial grammar (Steedman 2000) that include discourse semantic constraints on constituent formation.

But all theories of syntax-prosody mapping are challenged by the following two facts: First, prosodic boundary realisation/grouping is optional in that its occurrence cannot be fully predicted by text. Remijsen and Ladd (2008) present an example of the optional application of a tone sandhi rule in their materials: whereas 3 of 3 studied Dinka speakers changed a falling tone (HL) on a noun to level (H) before an adjective carrying an initial low tone (L), only one of the three produced this type of sandhi at the juncture between the same noun and an adverb. Other examples of optionality are cited in Shattuck-Hufnagel and Turk (1996); Jun (1993), and Fougeron (1998).

Second, the occurrence and realization of boundaries is influenced by length. Longer utterances contain more, higher level, prosodic boundaries (Watson and Gibson 2004; Watson et al. 2006, and Astésano et al. 2007). This evidence is discussed in more detail in Section 4.3.

The Smooth Signal Redundancy hypothesis offers a possible integrated explanation for the influence of length on prosodic boundary signaling, as well as other properties of boundary realisation, such as the general hierarchical nature of boundary correlates, the symmetrical nature of the prosodic hierarchy, and (some of) the optionality of prosodic boundary occurrence.

In the following section, I motivate boundary signaling as a way of increasing the acoustic salience of lexical words.

3. Smooth Signal Redundancy and word boundaries

Although the Smooth Signal Redundancy theory was originally developed to account for the similar effects of language redundancy and prominence on acoustic measures of syllable salience, what is proposed here is that this framework can explain when and how speakers choose to signal word boundaries.

A key assumption of the current proposal is that boundary signaling, like prominence, is a technique for signaling relative salience. But, whereas prominence signals the relative salience of *syllables*, boundary signaling signals the additional relative salience of *words*. On this view, word boundary correlates, like prominence correlates, relate inversely to language redundancy (see also Bell et al 2009): if a sequence is unpredictable (e.g. a two-word sequence spanning two, non-topically related sentences) then speakers will be likely to signal the boundary between the words. On the other hand, boundary signals may not be required between e.g. *he* and a following verb. Verbs often follow subject pronouns, and recognition of the sequence is likely without any additional boundary-signaling acoustic redundancy. On this view, the acoustic redundancy given by the phonetic segments of the words and the relative prominence of their component syllables would be sufficient for word recognition.

Language redundancy would therefore make it possible for speakers to communicate without overt phonetic boundary correlates for many words. A great deal of information about word boundary location is provided by the acoustic correlates of the contrastive elements that make up words in the mental lexicon, and a great deal of information about word sequences is also contained in the lexicon (i.e. selectional restrictions on preceding and/or following context). The word boundaries that speakers are most likely to signal, given sufficient effort and time are those of word sequences that are the least predictable from e.g. lexical, syntactic, semantic and pragmatic factors.

One way of manipulating salience in speech is by modulating the relative prominence of syllables and segments. Another way to increase the salience of a word is to signal its boundaries, making it easier for listeners to pick a word out from its neighbours, and therefore to recognize it. In this section, we present evidence supporting the view that signaling word boundaries is one of a set of techniques used to make words more salient.

3.1. *Word boundaries are signaled under contrastive stress and focus*

Turk and Sawusch (1997), Turk and White (1999), Cambier-Langeveld and Turk (1999) showed that speakers signal the locations of word boundaries when signaling contrastive prominence in both Dutch and English, and Turk and Shattuck-Hufnagel (2000) showed that speakers are more likely to use word-boundary signaling correlates in contrastively stressed contexts. In these experiments, target segments in different contrastive stress conditions were placed in two-word phrases differing in word boundary location. Target segments included consonants, e.g. /f/ in *beef arm vs. bee farm*; unstressed syllables, e.g. [ən] in *bacon force* vs. *bake enforce*, and lexically stressed syllables, e.g. *–fore* in *therefore square* vs. *there foresquare*. Target segments were elicited in sentences with or without contrastive stress either on the word containing the target segment. For example, Turk and White (1999) used sentences such as *Say "BACON force", don't say "REGAL* *force"; Say "bacon FORCE", don't say "bacon HOARD"; SAY "bacon force", don't SHOUT "bacon force"* (capitals indicate contrastive phrasal stress).

Results showed reliable effects of word boundary position on measurements of the target medial consonant or syllable. When adjacent to a phrasally-stressed vowel or syllable, target segments were longer if they belonged to a phrasally stressed word. For example, [ən] was longer in *BACON force*, than in *BAKE enforce*[2]. The duration of [ən] therefore signals the location of the word boundary by signaling that it belongs to a phrasally-stressed word. On the assumption that the function of contrastive phrasal stress is to make words more salient, these results suggest that signaling word boundaries is one of the techniques speakers use to accomplish this goal, and therefore to increase acoustic redundancy.

3.1.1. *Durational mechanisms for signaling word boundaries in prominent words* Lengthening mechanisms are often difficult to determine in studies of mono- and di-syllabic words (as presented above). For example, longer durations of both syllables of *BACON* as compared to *bacon* can be interpreted as lengthening on the entire phrasally stressed word, or as phrasal stress-related lengthening on the primary lexically stressed syllable in addition to word-final lengthening in phrasally stressed words. Cases where all syllables are lengthened under contrastive phrasal stress (e.g. Turk and White's 3-syllable words like *catapult*), are consistent with the "entire word" view. Findings from four-syllable words (Dimitrova et al. 2006; Turk and Dimitrova 2007), however, suggest an edge marking interpretation, accompanied by occasional findings of "spill over" of lengthening effects from lexically stressed syllables onto adjacent syllables (Turk and White 1999). Contrastively phrasally-stressed words like *presidency* show longer first and fourth syllables as compared to their non-phrasally stressed counterparts. And, as shown in Figure 3, contrastively phrasally-stressed words like *demo'cratic* on average show longer 3rd syllables (primary lexically stressed syllables) as well as longer first syllable onsets and final syllables.

The patterns of lengthening on the first and last syllables shown here are similar to the distributions of phrase-initial and phrase-final lengthening, where constituent-onset-lengthening appears to be localized primarily on onsets, and constituent-final lengthening occurs primarily on rimes.

Although these findings for English support the view that signaling contrastive-phrasal stress involves using duration to signal the boundaries of phrasally-stressed words, findings from Swedish and Finnish long words suggest that speakers of these languages use a different mechanism. In Swedish and Finnish, the durational marking of contrastive stress includes the lexically stressed syllable and one or two following unstressed syllables, but the final syllable in long words can be unaffected (Heldner and Strangert 2001; Suomi 2007).

To what extent, then, is edge-marking under focus found cross-linguistically? Although Swedish speakers don't appear to use word-final lengthening to make words more salient, they do insert pauses when words are contrastively focused

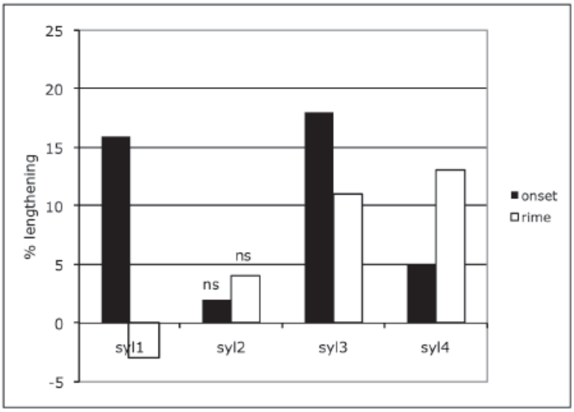

Figure 3. *Percentage lengthening on onsets and rimes in four syllable words with primary stress on the 3rd syllable, and secondary stress on the first syllable, e.g. condescending (2010 mnemonic).* ns *indicates the lack of statistically significant lengthening as diagnosed by a t-test comparison of contrastively phrasally stressed vs. non-phrasally stressed words. From Turk & Dimitrova 2007.*

(Strangert 2003). And a growing body of evidence from other languages suggests that boundary signaling under focus is common. Chen's (2006) study of the durational correlates of focus in Mandarin showed that focus domains are marked by longer durations at their edges than their non-focused counterparts. And focus-marking in Japanese and Korean also appears to involve optional boundary marking: the initial boundaries of focused constituents can block lexical pitch accent downstep in the same way that intonational phrase-boundaries do (for Japanese: Pierrehumbert and Beckman 1988; Venditti, Maekawa and Beckman 2008; but see Kubozono 2006; for Korean: Jun 1993; Jun et al. 2006; Kenstowicz and Sohn 1997, Kim 1997). Japanese speakers can also insert a prosodic boundary tone at the end of the focused constituent to set it of from a following constituent; this type of boundary insertion is often found when the focused word is utterance-initial (Venditti et al. 2008). Other findings suggest that the Japanese analogy to post-focal deaccenting that occurs in languages like English is to weaken the prosodic boundary between a focused constituent and a following non-focused constituent (Pierrehumbert and Beckman 1988; Venditti et al. 2008; Kori 1997, cited in Venditti et al. 2008).

Other languages are also claimed to have focus-related boundary marking. For example, Inkelas (1990) suggests that a phrasal prosodic boundary is inserted on the right edge of focused verbs in Chichewa, as evidenced by several markers of phrasal prosodic constituents (penultimate vowel lengthening, H tone retraction and high tone spread). Hayes and Lahiri (1991) propose that a high phonological phrase boundary tone is inserted at the right edge of focused words in Bengali.

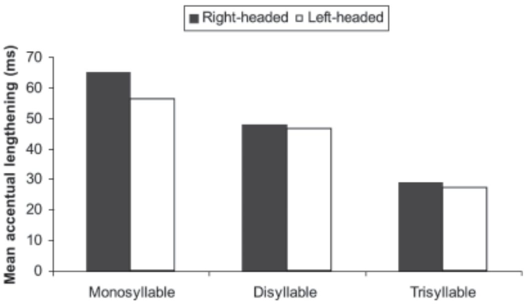

Figure 4. *The magnitude of phrasal-stress-related (accentual) lengthening of the primary stressed syllable in monosyllables, disyllables and trisyllables. All right-headed keywords are included except* port*; left-headed keywords included are* fish*,* mace*,* sense *and* ten*. From White & Turk (in press).*

3.1.2. *Making words salient by signaling how many syllables they contain* The evidence reviewed above shows that words can be made more salient by signaling their edges. Signaling the number of syllables contained in a word is another technique to make words more salient. White (2002) and White and Turk (2010) tested the durational effects of phrasal stress on triplets of left- and right-headed words varying in syllable number, e.g. *mace*, *mason*, *masonry*, and *mend*, *commend*, *recommend,* while keeping the total number of syllables in a carrier phrase constant. Figure 4 from White and Turk (in press) shows that the magnitude of phrasal-stress-related lengthening on the primary lexically stressed syllables in these words varies as a function of the number of syllables in the word.

These results support the view that the duration of the stressed syllable can be used to signal word structure. The general inverse relationship between syllable duration and the number of syllables in a word has been termed *polysyllabic shortening*[3] and has been observed in other studies and languages, including Swedish (Barnwell 1971; Lehiste 1972; Klatt 1973; Port 1981; Nakatani et al. 1981; Cooper et al. 1985, for English; Nooteboom 1972 for Dutch; Lindblom et al. 1981 for Swedish). White and Turk's English study found no effect of number of syllables in the word on the duration of the lexically main stressed syllable when the words were non-phrasally stressed. This finding, as well as those in Turk and Shattuck-Hufnagel (2000), suggest that the use of polysyllabic shortening for signaling word boundary location is most likely to surface in phrasally prominent contexts.

Generally, these findings support the idea that the phonetic correlates of boundaries and prominences can all be viewed under a single saliency umbrella. According to the Smooth Signal Redundancy hypothesis, this saliency makes words, syllables and segments more recognizable precisely when required by language redundancy. The edge-marking technique for marking word boundaries under prominence appears to target the same segments as those targeted by phrase-boundary-related lengthening (phrase-initial and phrase-final lengthening). The second, polysyllabic shortening technique targets stressed syllable nuclei, operates preferentially in prominent contexts (White 2002; White and Turk in press; Turk and Shattuck-Hufnagel 2000), and appears more likely to operate at lower levels in the hierarchy (White 2002; Huggins 1975).

3.2. *Phonetic correlates of word boundaries facilitate online word recognition*

Several experiments suggest that listeners make use of boundary signaling cues during online word recognition (e.g. Salverda et al. 2003; Christophe et al. 2004; Shatzman and McQueen 2006a, 2006b; Cho et al. 2007). For example, Salverda et al. (2003) and Shatzman and McQueen (2006b) showed the relevance of durational cues that signal the distinction between mono- and di-syllabic words, e.g. *ham* vs. *hamster*. In these eye tracking experiments, listeners' distraction by e.g. *ham* when listening to e.g. *hamster* was measured by their looking times to a picture of *ham* vs. a picture of *hamster* vs. an unrelated picture. Listeners were more likely to be distracted by *ham* when *hamster* contained a longer *ham-* initial syllable than when it contained a shorter *ham-*. These results suggest that appropriate boundary signals (i.e. that signal that *ham-* is initial in a di-syllabic word) help listeners to decide between potential lexical competitors, e.g. *ham vs. hamster* in this case). These findings support the view that word-boundary signals provide acoustic redundancy in the sense that they make words easier to recognize.

4. Smooth Signal Redundancy and word-boundary-related phenomena

The Smooth Signal Redundancy hypothesis proposed here claims that speakers use prosodic constituent structure to implement the complementary relationship between language redundancy and the acoustic redundancy of words. In other words, speakers vary prosodic boundary likelihood and strength to compensate for variations in language redundancy. This mechanism for implementing language redundancy complements the prosodic prominence mechanism presented in Aylett and Turk (2004).

In this section, I discuss two key predictions of the Smooth Signal Redundancy hypothesis as it applies to prosodic constituency, namely that there should be a hierarchy of boundary strengths, as well as an inverse correlation between language and acoustic redundancy. I further argue that the Smooth Signal Redundancy hypothesis as applied to prosodic constituency provides a potential explanation for a number of seemingly unrelated phenomena. These include effects of length and symmetry, aspects of strict layering, and aspects of prosodic optionality.

4.1. *A hierarchy of prosodic boundary strengths*

The Smooth Signal Redundancy proposal predicts that a range of language redundancy values should correlate inversely with a range of acoustic redundancy values. On the assumption that lexical word boundary markers are acoustic redundancy parameters in the sense that they make words more salient, the prediction that there should be a range of boundary strengths appears to hold, at least in a general way. Although there is a great deal of observed variability in measures of final lengthening and initial strengthening, many studies have found that several distinct degrees of final lengthening and initial strengthening can be observed. For example, Wightman et al. (1992) distinguished 4 levels in English on the basis of final lengthening (but listeners are reported to be able to distinguish 6), Kainada (2009) distinguished 4 levels in Greek, and Cambier-Langeveld (1997, 2000) distinguished 3 in Dutch. Fougeron (1998) reviews results from several studies and reports 2–5 levels distinct levels on the basis of various initial phenomena in various languages (the most was reported for Korean (5), documented by Cho and Keating 2001).

On the assumption that there is a potentially continuous range of language redundancy values for words in utterances, the evidence for a relatively limited number of prosodic levels in studies like these may suggest that speakers may perform only a coarse-grained evaluation of language redundancy when they plan their utterances. Alternatively, if their redundancy computation is fine grained, they would need to map ranges of different redundancy values onto a relatively small number of boundary strengths.

4.2. *A complementary relationship between language and acoustic redundancy*

The Smooth Signal Redundancy hypothesis predicts an inverse relationship between language redundancy and measures of acoustic salience (acoustic redundancy). The proposal put forth here is that speakers have two means of implementing this inverse relationship: 1) via prosodic prominence structure, as proposed in Aylett (2000), Aylett and Turk (2004), and 2) via prosodic constituent structure. In this section, I present evidence supporting the relationship between language and acoustic redundancy, and motivate the use of prosodic constituent structure as a means of implementing it. This evidence is a key prerequisite for establishing the plausibility of the Smooth Signal Redundancy hypothesis. It should be noted though, that the Smooth Signal Redundancy hypothesis is by no means the only possible account for these findings; other accounts are discussed in Section 7.

Aylett and Turk's study showed that word frequency and the predictability of syllables based on previous information (the predictability of a syllable based on two previous syllables, and the number of times a referent containing that syllable had been mentioned before) correlated with syllable duration. When they investigated a phrase-medial subset of their data (words followed by a GlaToBI break index of 1 (Mayo et al. 1997), they found that the variance predicted by language redundancy and prominence structure was largely shared. When they included materials followed by the full range of boundary strengths, they found that the proportion of shared variance was smaller, and moreover found a rather large unique contribution of prosodic variables (including boundary strength) to the prediction of the durational variance of syllables. This finding seems to run counter to the view presented here: that is, that prosodic boundary strength implements the relationship between language redundancy and acoustic salience as instantiated by duration. However, Aylett and Turk did not investigate the relationship between syllable duration and predictability based on *following* material, nor did they investigate measures of word *sequence* probability.

To what extent, then, is there a correlation between predictability and acoustic salience when prominence information is controlled and when word sequence predictability is considered? Available studies show significant negative correlations between predictability and acoustic salience, as predicted by the Smooth Signal Redundancy hypothesis presented here. Bell et al.'s (2009) study of word durations in the Switchboard corpus of conversational telephone speech found significant unique contributions to their regression models of frequency and the conditional probability of a word given the word that *follows* it, when, among other things, intonational accent, whether the word began or ended an intonational phrase, average word length, and rate of speech were controlled. Jurafsky et al.'s (2001) study of the predictability of function words in telephone conversations showed inverse correlations between function word duration and several predictability factors: frequency, conditional probability given the previous word, conditional probability given the following word, and conditional probability given the preceding and following words together. Their findings were gradient even after controlling for vowel reduction and rate of speech (among other things). Pluymaekers et al.'s (2005b) study of stem and *-lijk* durations in Dutch adverbs showed similar findings for a corpus of spontaneous dialogue. They controlled for a variety of factors, including pitch accent on the stem, speech rate, and the number of syllables in pause-pause chunks. Their results showed an inverse correlation between mutual information (log(Frequency of a word sequence XY)/(log(Frequency (word X)) × log(Frequency (word Y))) and stem and suffix duration, as well as between the number of realized segments in the words. Effects of following context affected all words in their test; effects of previous context affected the duration and number of segments in some words. An effect of repetition on stem duration (shorter stems for repeated words) was also observed.

Many of the words preceding the strongest boundaries were either excluded from these studies or controlled. Bell et al. (2009) excluded pre-pausal words, prefilled pause words, words at turn beginnings and ends, and controlled for position with respect to an intonation phrase boundary; Jurafsky et al. (2001) excluded pause adjacent and turn-initial and final function words, and Pluymaekers et al. (2005b) excluded words adjacent to pauses or disfluencies. However, Jurafsky et al. (2001) and Pluymaekers et al. (2005b) are likely to have included some forms adjacent to non-prepausal intonational phrases, and all three studies will have included words adjacent to lower level boundaries. Although it is possible that prominence degree may have affected the results to some extent, all three studies included some degree of prominence control. Jurafsky et al.'s (2001) study of function words controlled for the full vs. reduced vowel distinction and is likely to have included mostly non-phrasally stressed forms. Both Bell et al. (2009) and Pluymaekers et al. (2005b) controlled for the presence vs. absence of intonational accent (phrasal stress). The findings from all three are therefore at least consistent with the hypothesis presented here, namely, that language redundancy correlates inversely with acoustic redundancy (duration), and raise the possibility that prosodic boundaries may have implemented the observed duration variation.

Gahl and Garnsey (2004) showed that syntactic predictability can also affect segment and pause durations. They studied two types of transitive verbs: those with a high likelihood of being followed by a direct object (e.g. *confirmed*), and those with a high likelihood of being followed by a sentential complement (e.g. *believed*). They found that the durations of the verbs, noun phrases, and pauses varied systematically according to the syntax of their carrier utterances. In particular, verbs with a high likelihood of being followed by a direct object were ca. 10 ms longer in (relatively unpredictable) sentential complement contexts than in (relatively predictable) direct object contexts. That is, some verbs were longer in e.g. *The CIA director confirmed the rumor should have been stopped* (sentential complement syntax) than in *The CIA director confirmed the rumor once it had spread widely* (direct object syntax). Similarly, they found that the nouns following verbs with sentential complement biases were longer in direct object contexts than in sentential complement contexts. That is, e.g. *interviewer* was on average 25 ms longer in *The job applicant believed the interviewer when she discussed things with her* than in *The job applicant believed the interviewer had been dishonest with her.* A possible explanation for the longer durations of direct objects in these latter cases may be that the material following them is optional (less predictable) than the material following the subject of the sentential complement (obligatory, and therefore more predictable). Participants may therefore have inserted a minor prosodic boundary after the noun phrase in the direct object case.

4.2.1. *Syntax and language redundancy both relate to prosodic boundary occurrence* The results of studies reviewed above show the predicted complementary relationship between language and acoustic redundancy, even when prosodic prominence is controlled. However, although results are consistent with the view that prosodic boundaries have implemented the observed acoustic redundancy (durational) effects, none of these have explicitly tested whether the observed variance in acoustic redundancy was implemented via prosodic constituency. An empirical link between syntax, language redundancy, and the occurrence of prosodic boundaries can be found in Watson et al. (2006). The predictability (on a scale of 1–7) of an adjunct or argument dependent was assessed by asking participants to judge the completeness of a sentence preamble without the dependent, shown on the screen in brackets, e.g. *The reporter investigated [the crash]*. Completeness in this task implies lower predictability of a dependent. As expected, subjects were more likely to report that the preamble was complete when a (less predictable) adjunct dependent followed (as for intransitive verbs). They showed in a controlled study that the likelihood of intonational boundary insertion after the preamble was greater when the presence of a word's dependent was optional (less predictable) than when it was judged to be obligatory (more predictable), as for transitive verbs. For example, an intonational boundary was more likely after *arrived* in *The reporter arrived after the crash* . . . , (less predictable occurrence of the dependent *after the crash*, since *the reporter arrived* is a complete description of the event) than after e.g. *investigated* in *The reporter investigated the crash* . . . (more predictable occurrence of the dependent *the crash* since *The reporter investigated* was judged to be a less complete description of the event).

This study shows that syntactic relationships that are encoded prosodically can be interpreted in terms of predictability relationships between sequences of words and/or syntactic categories. Although Watson et al. (2006) was a study of a particular type of boundary *occurrence* (ToBI 3 or 4 vs. no boundary), rather than boundary *strength*, the observed relationship between predictability and the likelihood of intonational boundary occurrence is consistent with the view that prosodic boundaries implement the relationship between language redundancy (predictability) and acoustic redundancy. The Smooth Signal Redundancy hypothesis would further predict a correlation between boundary *strength* and language redundancy (e.g. greater final lengthening, initial lengthening, initial strengthening, F0 reset, etc.), i.e. that lower rates of predictability should correlate with *stronger* intonational phrase boundaries, and that higher rates of predictability should correlate with the acoustic correlates of lower boundaries (e.g. initial/final lengthening etc.).

In summary, the studies reviewed here provide good evidence for one of the key predictions of the Smooth Signal Redundancy hypothesis: an inverse correlation between language (predictability) and acoustic redundancy (salience). They also show that many different types of redundancy (frequency, repetition, word sequence likelihood, syntactic) can influence the acoustic characteristics of words. Although the studies did not explicitly test the relationship between prosodic boundary strength and acoustic redundancy, results are nevertheless consistent with the possibility that prosodic boundary strength has implemented this relationship, particularly in cases where prosodic prominence was controlled, and in Watson et al.'s (2006) case, where the likelihood of a higher prosodic boundary correlated inversely with sequence predictability.

4.3. *Smooth Signal Redundancy can explain: Effects of length and symmetry*

4.3.1. *Effects of length* Another piece of evidence in favor of the view that the function of prosodic boundaries is to compensate for language redundancy comes from effects of length on prosodic boundary occurrence (e.g. Astésano et al. 2007; Watson and Gibson 2004; Watson et al. 2006). For example, Astésano et al. varied the number of syllables on a target adjective in sentences like *les bonimenteurs et les baratineurs fabulateurs* 'fibbing smooth-talkers and cheaters' and *Les* *bonimenteurs et les baratineurs fameux* ‘famous smooth-talkers and cheaters’ and found that the likelihood of initial accent increased quasi-linearly from ca. 10% to 35% as the number of syllables on the adjective increased from 2 to 4 when the adjective modified a single noun, and from ca. 35% to 65% when the adjective modified a conjoined noun phrase. The following examples illustrate how effects of length such as these can be explained from the perspective of the Smooth Signal Redundancy hypothesis.

All things being equal, without considering different possible syllable structures, if an utterance consists of two syllables, there are two ways it can be parsed into words, as shown below (parentheses indicate word boundaries). If an utterance consists of three syllables, there are four possible parsings; if it consists of four syllables, there are eight, and so on. The language redundancy of shorter utterances is therefore higher in this respect than the language redundancy of longer utterances:

Possible parsings of a two-syllable utterance, where material enclosed in parentheses represents a single word:

(σ)	(σ)
(σ	σ)

Possible parsings of a three-syllable utterance

(σ)	(σ)	(σ)
(σ)	(σ	σ)
(σ	σ)	(σ)
(σ	σ	σ)

Possible parsings of a four-syllable utterance

(σ)	(σ)	(σ)	(σ)
(σ	σ)	(σ)	(σ)
(σ)	(σ)	(σ	σ)
(σ	σ)	(σ	σ)
(σ)	(σ	σ)	(σ)
(σ)	(σ	σ	σ)
(σ	σ	σ)	(σ)
(σ	σ	σ	σ)

If we place a salient word boundary somewhere within our four-syllable sequence (indicated by |), we increase the signal redundancy by reducing the number of possible parsings to four, for example:

Boundary after Syllable 1:

(σ)|	(σ)	(σ)	(σ)
(σ)|	(σ)	(σ	σ)
(σ)|	(σ	σ)	(σ)
(σ)|	(σ	σ	σ)

Boundary after Syllable 2:
(σ) (σ)| (σ) (σ)
(σ) (σ)| (σ σ)
(σ σ)| (σ) (σ)
(σ σ)| (σ σ)

In this way, the increased occurrence of phonetic correlates of word boundaries in longer utterances (the use of acoustic redundancy) can be seen as a technique for partially compensating for their lower language redundancy.

Evidence from an artificial language segmentation experiment supports this view. Frank et al. (2007) created an artificial language by concatenating synthesized syllables (e.g. *ba, bi, da, du, ti, tu*, etc.) into words of different lengths (2–4 syllables) and by randomly concatenating these words into sentences of different lengths (1–24 words per sentence), without pauses, F0 perturbations, or coarticulatory attenuation between words. Subjects were assigned to different sentence length conditions, and were asked to listen to 15 minutes of speech before answering questions about the likelihood that a given sequence was a word in their language. Results showed a significant, gradient decrease in word identification accuracy as sentence length grew, supporting the view that signal redundancy was lower for longer utterances. On the view presented here, because the acoustic redundancy of the synthetic words in these utterances was not modified as utterance length increased, it failed to compensate for the lower language redundancy in longer utterances. The resulting lower signal redundancy compromised word recognition. The Smooth Signal Redundancy hypothesis predicts that in natural speech, word boundaries would be signaled more strongly precisely in these cases, to compensate for low language redundancy. This prediction seems to be borne out by the data, mentioned above.

4.3.2. *Effects of symmetry* In the Astésano et al. example, speakers' increased likelihood of inserting phrasal boundaries before longer words (as diagnosed by initial accent use), can be seen as a technique for making the number of possible parsings in these longer utterances more similar to the number of possible parsings in the shorter utterances (produced in the same experimental block). Boundary insertion therefore equalizes or smoothes the overall signal redundancy of utterances of different lengths.

Smoothing the distribution of signal redundancy can also be seen as the rationale for effects of symmetry on the occurrence of boundaries, observed in Gee and Grosjean (1983), and in Goldhor (1976), cited in Klatt (1976). Gee and Grosjean asked participants to pause between words when reading aloud and found that the hierarchies he constructed from the different pause durations (which he called *performance structures*) often appeared to be symmetrical. To give an extreme example, it was rarely the case that a single-word NP subject (e.g. *he*) occurred on its own in a phrasal-level constituent when multi-word phrases followed, possibly because this type of parsing would result in a highly skewed signal redundancy profile. Producing a word or phrase boundary after the first syllable in long utterance would increase the signal redundancy of that particular syllable (only one one-word parsing is possible), while leaving the number of possible parsings in the following constituent high, and therefore its signal redundancy, and recognition likelihood, low. In contrast, the Smooth Signal Redundancy hypothesis predicts that a more even, symmetrical distribution of boundaries within an utterance would smooth out the signal redundancy, and hence raise the likelihood of recognition.

Frazier, Carlson, and Clifton (2006) suggest that listeners evaluate relationships between boundary strengths within utterances when judging the syntax of syntactically ambiguous sentences. That is, they found that the high attachment reading of the ambiguous sentence *John said Susan telephoned # after the party* was only obtained when the boundary after *telephoned* was the only one in the utterance, or stronger than all the others. The mere presence of a boundary was not sufficient to yield the high attachment interpretation. These results suggest that listeners evaluate boundary strength in relation to other boundaries in the utterances they hear, and is thus consistent with the hypothesis that listeners use their expectations about smooth signal redundancy in interpreting boundary strength.

4.4. *Strict layering: Combined result of Smooth Signal Redundancy and the coarse granularity of acoustic redundancy?*

It may be possible to explain some of the discrepancy between allowable types of syntactic and prosodic trees, where syntactic trees permit indefinite levels of embedding or recursivity, but embedding in prosodic hierarchical structures is limited (see discussions of Strict Layering, e.g. Selkirk 1996). If prosodic structure matched syntactic structure, all possible levels of syntactic embeddings would have to be distinct in terms of prosodic correlates. This doesn't appear to be the case. As discussed in Section 4.1, the number of distinct prosodic levels seems to be limited to ca. 5–6. Chomsky and Halle (1968) attribute this constraint on the number of prosodic levels to performance factors; the Smooth Signal Redundancy proposal suggests that the relevant performance factors may either be limitations in calculating fine-grained differences in redundancy, in producing fine-grained differences in acoustic boundary correlates, or both.

The Smooth Signal Redundancy hypothesis may offer an explanation for another characteristic of prosodic trees typically lumped under Strict Layering: the fact that prosodic constituents tend to be exhaustively parsed into constituents one level down in the hierarchy. Exhaustive parsing throughout an utterance yields symmetric trees. On the other hand, syntactic parsings of utterances can be very asymmetric, with many embeddings in one part of the utterance and few embeddings in the other. For example, in *this is the dog that chased the cat that killed the rat that ate the malt . . . that lay in the house that Jack built,* the syntactic structure is infinitely right-branching, predicting weaker boundaries between words as the end of the utterance gets nearer (cf. Figure 5).

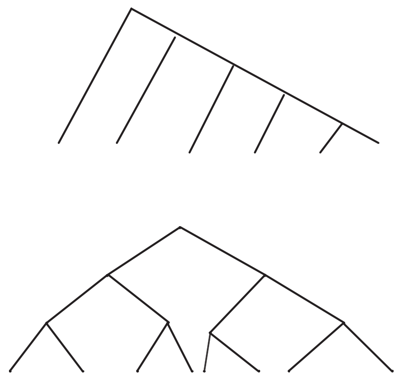

Figure 5. *Schematic diagrams of 1) a syntactic structure with embedding towards the end of the utterance (top) and of 2) a strictly layered prosodic structure (bottom).*

The Smooth Signal Redundancy principle of equalizing the distribution of signal redundancy throughout speech predicts a more symmetric structure.

In summary, the Smooth Signal Redundancy hypothesis leads us to expect phonetic markers of constituency predicted by a symmetrical hierarchical structure with relatively few layers (as predicted by Strict Layering).

5. Clarity and slow rates increase the likelihood and magnitude of boundary signals

On the Smooth Signal Redundancy view, speakers manipulate acoustic redundancy to compensate for relatively low levels of language redundancy. The overall frequency with which they use boundary-signaling correlates and/or the magnitude of these correlates is predicted to depend on the overall effort they use to produce their speech and on the time they have to produce it. Experimental evidence in support of this claim with respect to time comes from Beckman and Edwards (1990), Caspers (1994), and Sugahara and Turk (2009). Caspers (1994) found that fewer intonational phrase boundaries were produced under fast rates of speech as compared to slow rates, as diagnosed by F0 movements. The boundaries that were left out at slow rates were those predicted to be optional by Nespor and Vogel (1986), examples theirs: *Lions, $[_I$as you know$]_I$, are dangerous*. (obligatory intonational phrase boundary), vs. *Our next door neighbour truly believes $[_I$that black cats bring bad luck$]_I$* (optional intonational phrase boundary).

Beckman and Edwards (1990) and Sugahara and Turk (2009) found that slow rates of speech tend to magnify the expression of boundary cues. Beckman and Edwards (1990) found that temporal markers of word boundaries were stronger at slower rates, and Sugahara and Turk (2009) found that a slow rate of speech can even encourage speakers to signal the boundary between a stem and a level II suffix. For example, under slow (but not normal or slowest) rates of speech VC durations distinguish e.g. [ek] in *baking* from [ek] in *bacon*, [ʌk] in *tucks* vs. [ʌk] in *tux,* with longer VC durations before Level II suffixed stems[4]. These durational differences could improve recognition of the two grammatical units (stem and suffix), but crucially only if the other boundaries in the utterance were magnified in similar proportion, in order to maintain smooth signal redundancy across the utterance. Magnification of all boundaries to a similar degree would lead to overall improved clarity.

Taken together, these findings suggest that rate can lead to variation in boundary occurrence and strength. Although rate change doesn't necessarily imply a change in speech clarity (see e.g. Son 1993), boundary strength can be magnified at slow rates of speech, and boundaries may be weaker (or absent) at faster rates of speech. In order to maintain smooth signal redundancy, the magnification of a boundary in one part of an utterance would co-occur with magnification of all boundaries to maintain their relative salience across an utterance (Frazier et al. 2006).

Variation in effort seems to work in a similar way, that is, with more, stronger boundaries with greater care, clarity, and effort. Strangert (2003) found a gradient effect of longer pauses after focused words when these were produced with greater effort.

Further research will be needed to determine whether the granularity of prosodic strength distinctions is also modified with rate and/or clarity. That is, do speakers signal finer-grained redundancy distinctions (more prosodic levels) in slow or clear speech? Or do they only magnify the distinctions that would normally be there at normal rates and/or effort levels? In either case, subtle language redundancy distinctions are more likely to be observable in clear speech and at slow rates. For example, all other things being equal, one might predict that word boundary locations for words that can be embedded in longer words (e.g. the right boundary of *ham*, where *ham* can occur in *hamster*) might be more salient than the boundaries of words that can't be parts of other words[5], because the predictability of word boundary location should be lower. But this type of language redundancy difference is likely to be subtle, and may require the magnification of careful and/or slow speech to be observable.

6. Sources of prosodic optionality

6.1. *Variance in effort and rate*

As discussed above, prosodic boundaries are generally weaker at faster rates and with less effort. These differences due to rate and effort provide one potential source of differences in apparent prosodic boundary realization.

6.2. *Differences in language redundancy*

Another potential source of prosodic optionality comes from differences in language redundancy due to individual differences in language and world experience.

For example, Southern Sudanese often produce no overt boundary markers in the sequence *Comprehensive peace agreement for Southern Sudan*. However, people less familiar with the political situation in that part of the world would be more likely to signal some of the boundaries.

It is also possible that individuals may have different ways of computing redundancy. That is, redundancy of different types of units (syllables, words, etc.) can be computed over different sizes and types of domains (likelihood of e.g. a syllable based on preceding or following 1 vs. 2 vs. 3, etc., or likelihood of e.g. a word based on a preceding syntactic unit or semantic category). It is possible therefore that different individuals apply different strategies or algorithms to the task of estimating redundancy. Cases of syntactic ambiguity illustrate this point. For example, in the Astésano et al. utterances, e.g. *les bonimenteurs et les baratineurs fameux 'famous smooth-talkers and cheaters'*, a strategy of computing the language redundancy of the adjective *fameux 'famous'* on the basis of the preceding words will give the same likelihood of a boundary before the adjective when it modifies a single noun (e.g. *baratineurs 'cheaters'*) as compared to when it modifies a conjoined noun phrase. It is only when the syntactic status of the preceding noun is considered that the language redundancy of *fameux* becomes different in the two different syntactic structures. Cases where speakers avoided producing a phrase boundary before *fameux* when it modified the conjoined noun phrase may have been due to computing language redundancy without taking syntax into account.

Yet another potential source of individual differences/optionality may come from differences in the granularity of the redundancy distinctions that are computed and/or signaled. For example, some speakers may compute rougher estimates of redundancy than others (and therefore may signal fewer degrees of boundary strengths). Individual differences in the number of distinct boundary strengths are common, even in tightly controlled experimental settings (e.g. Fougeron 1998). The results could be explained by differences in redundancy granularity computation, or in terms of prosodic boundary implementation (possibly due to rate or style differences), or both.

And finally, there may be differences in the extent to which listeners' language redundancy is taken into account. If, for example, the speaker knows that the listener is unfamiliar with a particular topic, s/he might choose to highlight some of the relevant words or phrases either by making them more prominent, by using stronger boundaries, or both.

7. Other accounts of predictability effects and boundary occurrence

The Smooth Signal Redundancy Hypothesis is by no means the only account of predictability effects on acoustic salience, or of boundary occurrence in speech. Bybee and Hopper (2001) suggest that articulatory practice will shorten durations; frequency effects and effects of previous mention on duration may be accounted for to some extent by this mechanism.

Bell et al. (2009) argue that longer durations may be used to adjust for potential asynchronies between planning and implementation processes. On their view, planning involves lexical access, syntactic organization, and phonological encoding, whereas phonetic implementation involves recruiting and activating task-oriented sets of muscles to produce the airflows and constrictions required for speech. These two activities may well take different amounts of time; longer articulations arise when the articulatory processes must wait in order for planning to catch up. A corollary of this view might be that shorter articulations occur when planning processes are quick. This asynchrony explanation provides a straightforward account of the lengthenings and pausings (filled an non-) associated with hesitations: Speech is slowed down or paused to give the speaker time to find the right word or plan the appropriate phrase. In addition, it provides an interesting potential account for shorter durations that occur in words containing more syllables: Articulations may be shorter in longer words to catch up with the planning process.

Bell et al. (2009) presented this mechanism as a way for accounting for the durational effects observed in their study, and did not propose it as a way of accounting for prosodic boundary occurrence or strength. Could their account provide an explanation for prosodic boundary occurrence or strength? Along these lines, Watson and Gibson (2004) and Watson et al. (2006) suggest that boundary occurrence may relate to recovery from resource expenditure due to the production of preceding material, and/or to the time it takes to plan upcoming material. On this view, longer boundaries are expected with longer and or more complex preceding material, since more resources will have been used to produce them, and more time for recovery is required. At the same time, longer boundaries are expected with longer and/or more complex following material, since more time will be required to plan.

Several facts relating to prosodic boundaries are difficult to explain on planning/recovery views. First, final lengthening is localized: Although smaller effects can be observed on other syllables, final lengthening in stress-accent languages like English, Estonian, German and Dutch and Finnish is mainly restricted to the rime of the phrase-final syllable and (to a lesser extent) to the rime of the phrase-final primary stressed syllable (Cambier-Langeveld 1997; Turk and Shattuck-Hufnagel 2007; Berkovits 1994; Kohler 1983; Krull 1997; Nakai et al. 2009; Wightman et al. 1992). Initial lengthening is mainly restricted to phrase-initial onset consonants (Cho 2001; but see Fougeron 2001), and polysyllabic shortening effects appear to be localized mainly on stressed syllable nuclei (White 2002; White and Turk 2010; Turk and Shattuck-Hufnagel 2000). There is no principled reason to expect that planning should be restricted to these temporal targets, or to the pause between them. Second, although pause duration depends on the length and complexity of following material (Krivokapic 2007; Ferreira 1993, 2007), final lengthening does not appear to be (Ferreira 1993). Ferreira (2007) therefore argues that prosodic boundary realization should be seen as a process qualitatively different from planning. Third, boundary occurrence seems to increase (rather than decrease) at slow speech rates. Slow speech should presumably give the planning process more time to catch up with articulation; with extra time, one might expect fewer boundaries. By similar logic, slow speech might be expected to obviate the need for recovery-related boundaries, but instead more, stronger boundaries are observed. And most importantly, prosodic boundaries are implemented not only with temporal correlates (final lengthening, pausing, initial lengthening), but also with tonal, prominence, and other correlates that have no principled explanation under the recovery or planning/implementation synchrony views. What all of this suggests, then, is that speakers may be able to exploit the temporal correlates of boundaries for recovery and planning purposes, but that they are unlikely *causes* for boundary occurrence.

Yet another possible explanation for boundary occurrence is Aylett's (2000) account: Boundaries are signalled in order to give speakers and listeners the opportunity to check to make sure the message has been transmitted properly. While it is possible that speakers and/or listeners would use strong phrase or utterance boundaries as opportunities for checking, weaker, phrase-medial boundaries are less well-suited to this task, and therefore less well-motivated on this view. The checking account is therefore difficult to reconcile with the hierarchical nature of boundary occurrence.

What I propose here is that boundary signalling has evolved to implement smooth signal redundancy by varying the acoustic salience of words in a way that complements language redundancy. This inverse relationship between acoustic salience and language redundancy aids recognition by yielding an optimal distribution of signal redundancy for the listener. This view predicts that many different types of boundary-signaling techniques should be available to speakers, as long as all of them increase word salience. Indeed there seem to be a great variety of available techniques both within and across languages. These include temporal mechanisms (e.g. final and initial lengthening, pause) but are not restricted to them. The temporal correlates of boundaries may be prevalent cross-linguistically for several reasons, including the fact that they give speakers the time to recover and plan, and for speakers and listeners to check that information has arrived safely. That is, speakers may use the opportunity given by a phrase boundary to plan and/or check for safe information arrival, and may therefore pause longer at this boundary in order to do so. At the same time, listeners may use the same opportunity to process the information they have just heard (Krivokapic 2006), and/or to confirm that the information has arrived safely. Although speakers and listeners no doubt take advantage of the temporal properties of boundaries for multiple purposes, it is doubtful that these temporal mechanisms are the driving force behind the use of prosodic boundaries.

7.1. *Possible mechanisms for implementing the relationship between language redundancy and boundary occurrence*

The view presented here is that it is language redundancy, rather than factors such as syntax or discourse semantics, that predict boundary occurrence for given rates and styles of speech. If this theory is correct, a mechanism must exist that enables speakers to compute the language redundancy of a given word sequence. Pierrehumbert 2002 suggests that language redundancy (predictability) may correlate with ease of retrieval of words from memory. On this view, frequent words would be easier to retrieve than less frequent words because of higher resting activation levels. And referents mentioned before in discourse might prime the activation of repeated entities (Pluymaekers et al. 2005b; citing Pickering and Garrod 2004). Some aspects of word sequence predictability could be accounted for by the ease of retrieval mechanism, given that selection restrictions on e.g. arguments are thought to be stored along with a word's lexical entry. Other aspects might be accounted for by memory traces of word sequences.

What is more difficult to account for by the ease of retrieval mechanism is the computation of the lower language redundancy in longer utterances. Some of the length effects observed in the literature may be confounded with other, more traditional measures of predictability, and language redundancy might therefore be computable solely on the basis of ease of retrieval. For example, in Astésano et al.'s 2007 experiment, some of the longer words may have been less frequent than some of the shorter words. It will be important to find out to what extent length effects can be de-confounded from frequency effects in these experiments. However, other studies show length effects that are independent of word sequence predictability. For example, in Watson and Gibson's (2004) experiment, length effects on intonational boundary likelihood were observed even for identical word sequences, e.g. boundaries between *judge* and *who*, and between *ignored* and *fired* in *The judge who the reporter ignored fired the secretary* vs. *The judge who the reporter who attacked the senator ignored fired the secretary.* These effects are more difficult to account for on an ease of retrieval view, and would seem to require a mechanism for integrating more global length-related language redundancy, with more local language redundancy information. Similar effects of length on final lengthening magnitude, for identical word sequences, are found in Kainada (2009).

8. Do we need a separate prosodic component of grammar if the occurrence and magnitude of boundary correlates is predicted by language redundancy, effort and rate?

Why not propose a direct link between language and acoustic redundancy, without a prosodic structure intermediary? Three lines of evidence support a separate prosodic component of grammar: 1) language-specific correlates of stress and boundaries, 2) language specific conventions about stress placement, and 3) categorical differences in the realization of different prosodic levels. Although language-specific correlates of stress and boundaries could be implemented by having direct, but yet language-specific, implementation of language redundancy, the other types of evidence are more difficult to reconcile with a direct language-redundancy-to-phonetics implementation view.

8.1. *Language-specific stress and boundary correlates*

If correlates of language redundancy were universal, there would clearly be no need for a prosodic component of grammar. Although there are some proposed universal correlates of prosodic structure (e.g. final lengthening and pausing, as correlates of boundaries Vaissière 1983), there is a large body of evidence that suggests cross-linguistic differences in the way prosodic boundaries and prominences are implemented. At the very least this evidence requires language- and language variety-specific implementation of language redundancy. American English flapping and British English glottal stop use is an example. Whereas Americans can flap or tap the /t/ in *city*, British English speakers either produce a glottal stop or produce the /t/ with less aspiration (Gussenhoven 1986). In this case, the same syllable structure and/or position with respect to lexical stress is produced with different phonetic correlates in the different varieties. Word-level stress also shows language-specific variation. For example, Tunisian Arabic uses F0, spectral balance and F1 to signal lexical stress in non-phrasally stressed contexts. In contrast, English uses duration, spectral balance and vowel centralization (Bouchhioua 2008)[6].

Even types of boundary correlates that are proposed universals (e.g. final lengthening, (Vaissière 1983) show language specific differences in the correlates with which they co-occur. For example, although Finnish and Japanese both show final lengthening in spite of the fact that both are quantity languages, Finnish final lengthening often co-occurs with phrase-final breathiness, whereas Japanese final lengthening often co-occurs with glottalization (see Nakai et al. 2005). Here again, either language-specific implementation of phrase structure or language-specific implementation of redundancy would be required.

Additionally, there are language-specific differences in how these so-called universal cues are implemented. For example, Nakai et al. (2009) found that sentence-final lengthening in Finnish interacts with quantity: on Finnish half-long vowels (final V in CVCV sequences) lengthening was restricted compared to lengthening on short (final V in CVVCV) and long (final V in CV(V)CVV) vowels. In contrast, no differences in proportional final lengthening behaviour were observed for Japanese short vs. long vowels.

8.2. *Categorical differences between hierarchical levels*

Prosodic hierarchies traditionally have labeled levels (e.g. prosodic word, phonological phrase, intonational phrase, etc.) that are proposed to be categorically distinct from each other, in the sense that different levels can be associated with qualitatively different phonetic phenomena. Nespor and Vogel's (1986) book provides many proposed examples: American English flapping is proposed to occur within the Utterance, Greek nasal assimilation is proposed to occur within the Phonological Phrase. On the other hand, a direct language-redundancy-phonetics mapping would suggest a hierarchy of phonetic behaviour that is qualitatively similar at all levels in the hierarchy, with gradient reflection of boundary strength, either in terms of likelihood of occurrence, or in terms of magnitude of expression of a particular feature. For example, final lengthening supports this view: Final lengthening is known to increase in magnitude at the boundaries of higher level constituents (Wightman et al. 1992).

Although phonetic research may show that some (or even many) of the phenomena proposed to associate with particular levels in the hierarchy are in fact gradiently expressed at many levels, some phenomena may associate categorically with particular prosodic levels. Greek stop voicing may be an example: Kainada's (2009) experimental phonetic study of Greek has shown that stop voicing associates with a low level in the hierarchy but is blocked at higher levels (see also Arvaniti and Joseph 2000, 2004 for the discussion of an additional phenomenon associated with stop voicing that may also be categorical). Another example of the categorical realization of particular prosodic levels is that of intonational boundary tones. These appear to associate only with higher levels in the hierarchy, and do not occur phrase-medially. And yet another set of examples comes from Korean. In this language, particular types of segmental sandhi phenomena are associated with particular levels in the prosodic hierarchy (Jun 1996, 1998). Such phenomena suggest a hierarchy of categorically distinct levels, like the one proposed by prosodic phonologists (e.g. Nespor and Vogel 1986), that is not just a hierarchy of boundary strengths, but additionally a collection of categorically distinct domains. Language redundancy measures do not provide an obvious way of accounting for categorical phenomena associated with particular domains.

On the view presented here, we would expect most effects of language redundancy to be implemented via prosodic structure. However, some independent effects of redundancy may remain, including articulatory practice effects, and some planning effects (e.g. hesitations). The idea is that the temporal correlates of prosodic structure will give speakers some time to plan upcoming words and phrases, but in cases where word finding or planning is difficult, speakers may require extra time. They may then either produce an overt hesitation or prolong pauses beyond what might be expected from boundary strength alone (cf. Ferreira 2007).

9. Summary and discussion

When and why do speakers signal word boundaries? Following the Smooth Signal Redundancy hypothesis of Aylett (2000) and Aylett and Turk (2004, 2006), I suggest that they are most likely to do so when word sequences are least predictable, given sufficient effort and time. I further claim that they do this in order to smooth out the overall information in the signal so that recognition is equally likely for each of the words in the string. The hierarchy of prosodic constituents is proposed to implement a coarse-grained range of language-redundancy measures using a set of language-variety-specific phonetic word boundary markers.

If plausible, this proposal provides a unified explanation for a diverse set of well-documented phenomena: a hierarchy of phonetic word boundary strengths, aspects of syntax-prosody mapping, effects of length and symmetry on prosodic boundary realization, aspects of strict layering, as well as aspects of prosodic optionality. The proposal put forward here supplements Aylett's proposal for syllables (Aylett 2000; Aylett and Turk 2004), by suggesting that Smooth Signal Redundancy also applies to words, and is controlled by prosodic constituent structure, in addition to the prominence structure that controls the acoustic redundancy of syllables.

It is possible that the principles of Smooth Signal Redundancy may also apply at the sub-syllabic level, where syllable structure is used to implement language redundancy. The sub-syllabic control of segmental, gestural or featural language redundancy may provide an account for differences in the articulatory and acoustic realization of constituent-initial consonants as compared to medial and final consonants. Syllable-initial consonants tend to be longer, and have stronger and tighter constrictions than syllable-final or ambisyllabic consonants (Lavoie 2001; Jong 1998). Findings of this type fit with predictions of Smooth Signal Redundancy if we assume that syllable-initial consonants are less predictable/redundant than final or ambisyllabic consonants, given that initial consonants appear to carry the most acoustic redundancy. Also consistent with this view are findings that the salience of sub-syllabic units can be manipulated under contrastive focus, at least for some speakers (Heuven 1994).

The plausibility of the current proposal needs to be tested via correlations of language redundancy measures with measures of boundary strength. One challenge of such tests will be to determine appropriate ways of calculating redundancy. It is unclear whether speakers' computed redundancy measures reflect the redundancy of categories of units (e.g. all nouns, all fricatives, etc.), or of tokens. In addition, there are many possible sizes and types of domains over which to compute redundancy, e.g. redundancy given the previous/following 1 vs. 2 vs. 3 elements, redundancy given the preceding syntactic category, etc. And finally, if redundancy applies to different levels (syllables, words, and possibly sub-syllabic units), the interaction of redundancy computation for these nested levels must be determined (see Pluymaekers et al. 2005b). For example, it is well-known that some aspects of prominence placement relate to boundary placement (e.g. English early accent and nuclear stress rule, Chomsky and Halle 1968; Shattuck-Hufnagel et al. 1994), and it may be important first to predict boundary strength before predicting the relative prominence of stressed syllables. In addition, it is possible that the language redundancy measures required for prominence prediction may be different from those required to predict boundary strength (see Section 4.2). However, on the assumption that the Smooth Signal Hypothesis is correct, it should be possible to determine the appropriate ways of computing redundancy by measuring acoustic salience (using measures of the acoustic correlates of prominence and boundary strength) and seeing which language redundancy measures best predict it.

Acknowledgements

Thanks to Evia Kainada, Satsuki Nakai, Inga McKendry and Bert Remijsen for examples, to Matthew Aylett and Stefanie Shattuck-Hufnagel for inspiration, and to Peter Ackema, Laurie Bauer, Sasha Calhoun, Bert Remijsen, Stefanie Shattuck-Hufnagel and the members of the University of Edinburgh Phonetics and Phonology Workshop for helpful comments and discussion. I am especially grateful to Mary Beckman and Mirjam Ernestus for their comments on a previous version of this paper, and to Paul Warren for editorial help. Any flaws in logic or other mistakes are mine. Some of the research reported here was supported by funding from the AHRC, grant number B-RE-AN9288-APN13614 for which I am grateful. The presentation to Laboratory Phonology 11 was supported by funds from the British Academy.

Correspondence e-mail address: turk@ling.ed.ac.uk

Notes

1. Ladefoged (1982) argues that full vowels are more prominent than reduced vowels and are assigned to the first level of a prominence hierarchy.
2. Word-initial syllables in words like *ENFORCE* are longer than their counterparts in non-phrasally-stressed words, but the size of the effect for these initial syllables is smaller than the effect for final syllables (see Cambier-Langeveld and Turk 1999).
3. It is not altogether clear which, or how many, mechanisms are involved in signaling the number of syllables in a word. Five proposed mechanisms are discussed in Turk and Shattuck-Hufnagel (2000), and White and Turk (2010); polysyllabic shortening is therefore used here as a cover term.
4. But because the effects were not observed at slowest rates suggests that there may be a ceiling on cumulative lengthening effects in a particular context (that is, duration due to a very slow rate combined with duration due to morphological context may have been constrained in this phrase-medial context).
5. Thanks to Mirjam Ernestus for this idea.
6. Sandhi phenomena are also relevant in showing the language-particular nature of word-boundary-related phenomena. For example, although sequences of two low tones are found in both Dinka and Mixtec, neither of these languages show the L → rising / ____ L sandhi found in Mandarin Chinese at the juncture of closely affiliated words. And although both Dinka and Mixtec have HLH sequences that participate in sandhi processes (Remijsen and Ladd 2008; McKendry in prep.), the nature of the sequences and their sandhi behaviour is different in the two languages. In Dinka, word-final contour HL sequences change to H before a word-initial H in sequences of closely grouped words. In Mixtec, each tone associates to a different syllable. In this language, when a constituent boundary occurs in an HLH sequence, the Low tone is upstepped to Mid.

References

Arvaniti, Amalia & Brian D. Joseph. 2000. Variation in voiced stop prenasalisation in Greek. *Glossologia* 11–12. 131–166.

Arvaniti, Amalia & Brian D. Joseph. 2004. Early Modern Greek /b d g/: evidence from Rembetika and Folk songs, *Journal of Modern Greek Studies* 22. 73–94.

Astésano, Corine, Ellen Gurman Bard & Alice Turk. 2007. Structural influences on initial accent placement in French. *Language and Speech* 50(3). 423–446.

Aylett, Matthew P. 2000. *Stochastic suprasegmentals: Relationships between redundancy, prosodic structure and care of articulation in spontaneous speech*. PhD dissertation, University of Edinburgh. http://www.cogsci.ed.ac.uk/~matthewa/thesis_sum.html

Aylett, Matthew P. & Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1). 31–56.

Aylett, Matthew P. & Alice Turk. 2006. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America* 119(5). 3048–3059.

Bard, Ellen Gurman & Matthew P. Aylett. 2005. Referential form, duration, and modelling the listener in spoken dialogue. In J. Trueswell and M. Tanenhaus (eds.), *Approaches to Studying World-situated Language Use: Bridging the Language-as-Product and Language as Action Traditions*, 173–191. Cambridge, MA: MIT Press.

Bard, Ellen Gurman, Anne H. Anderson, Catherine Sotillo, Matthew Aylett, Gwyneth Doherty-Sneddon & Alison Newlands. 2000. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42 (1). 1–22.

Barnwell, Thomas P. 1971. An algorithm for segment durations in a reading machine context. *Massachusetts Institute of Technology Research Laboratory of Electronics Technical Report*. 479.

Beckman, Mary E. & Jan Edwards. 1990. Lengthenings and shortenings and the nature of prosodic constituency. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: between the grammar and physics of speech*, 152–178. Cambridge: Cambridge University Press.

Beckman, Mary E. & Janet Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* (3). 255–309.

Bell, Alan, Jason Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60. 92–111.

Berkovits, Rochele. 1994. Durational effects in final lengthening, gapping, and contrastive stress. *Language and Speech*, 37(3). 237–250.

Bouchhioua, Nadia. 2008. *The acoustic correlates of stress and accent in Tunisian Arabic: A comparative study with English*. PhD dissertation, Université de 7 Novembre, Carthage, Tunisia.

Bukshaisha, Fouzia. 1985. An Experimental Phonetic Study of Some Aspects of Qatari Arabic. PhD dissertation, University of Edinburgh.

Bybee, Joan & Paul Hopper. 2001. Introduction to frequency and emergence of linguistic structure. In Joan Bybee & Paul Hopper (eds.) *Frequency and the emergence of linguistic structure*, 1–24. Amsterdam: John Benjamins.

Cambier-Langeveld, Tina. 1997. The domain of final lengthening in the production of Dutch. In Helen de Hoop and Jane Coerts (eds.) *Linguistics in the Netherlands*, 13–24. Amsterdam: John Benjamins

Cambier-Langeveld, Tina. 2000. Temporal marking of accents and boundaries. PhD dissertation, University of Amsterdam, LOT series 32.

Cambier-Langeveld, Tina & Alice Turk. 1999. A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics* 27. 171–206.

Caspers, Johanneke. 1994. Pitch movements under time pressure. PhD dissertation, University of Leiden. The Hague: Holland Academic Graphics.

Chen, Yiya. 2006. Durational adjustment under corrective focus in Standard Chinese. *Journal of Phonetics* 34(2). 176–201.

Cho, Taehong. 2001. *Effects of prosody on articulation in English*. PhD dissertation, University of California, Los Angeles. Published in UCLA Dissertations in Linguistics 22.

Cho, Taehong & Patricia Keating. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* 29. 155–190.

Cho, Taehong, James McQueen & Ethan Cox. 2007. Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, 35. 210–243.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Christophe, Anne, Sharon Peperkamp, Christophe Pallier, Eliza Block & Jacques Mehler. 2004. Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language* 51(4). 523–547.

Cooper, William, Stephen J. Eady & Pamela R. Mueller. 1985. Acoustical aspects of contrastive stress in question-answer contexts, *Journal of the Acoustical Society of America*, 77(6). 2142–2156.

Dilley, Laura, Stefanie Shattuck-Hufnagel & Mari Ostendorf. 1996. Glottalization of vowel-initial syllables as a function of prosodic structure. *Journal of Phonetics*, 24. 423–444.

Dimitrova, Snezhina, Alice Turk & Mariko Sugahara. 2006. English phrasal accent can lengthen multiple sites. Poster presented at the *Tenth International Conference on Laboratory Phonology*, Paris.

Ferreira, Fernanda. 1993. Creation of prosody during sentence production. *Psychological Review* 100 (2). 233–253.

Ferreira, Fernanda. 2007. Prosody and performance in language production. *Language and Cognitive Processes*, 22(8). 1151–1177.

Fougeron, Cécile. 1998. Variations articulatoires en début de constituents prosodiques de différents niveaux en français. PhD dissertation, Université Paris III-Sorbonne Nouvelle.

Fougeron Cécile. 2001. Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics,* 29(2). 109–135.

Fougeron, Cécile & Patricia Keating. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101(6). 3728–3740.

Fowler, Carol & Jonathan Housum. 1987. Talkers' signaling of "New" and "Old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26. 489–504.

Frank, Michael, Sharon Goldwater, Vikash Mansinghka, Tom Griffiths & Joshua Tenenbaum. 2007. Modeling human performance in statistical word segmentation. In Danielle S. McNamara and J. Greg Trafton (eds.), *Proceedings of the 29th Cognitive Science Society,* 281–286. Austin, TX: Cognitive Science Society.

Frazier, Lyn, Katy Carlson & Charles Clifton. 2006. Prosodic phrasing is central to language comprehension. *TRENDS in Cognitive Sciences*, 10(6), 244–249.

Gahl, Susanne & Garnsey, Susan. 2004. Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language* 80. 748–775.

Gee, James Paul & François Grosjean. 1983. Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology 15,* 411–458.

Ghazeli, Salem. 1977. *Back consonants and backing coarticulation in Arabic*. PhD Dissertation, University of Texas at Austin.

Goldhor, Richard. 1976. *Sentential determinants of duration in speech.* MIT ms.

Gussenhoven, Carlos. 1986. English plosive allophones and ambisyllabicity. *Gramma* 10:22. 119–141.

Hale, Kenneth & Elizabeth Selkirk. 1987. Government and tonal phrasing in Papago. *Phonology Yearbook 4.* 151–183.

Heldner, Mattias & Eva Strangert. 2001. Temporal effects of focus in Swedish. *Journal of Phonetics* 29. 329–361.

Heuven, Vincent van. 1994. What is the smallest prosodic domain? In Patricia Keating (ed.) *Papers in Laboratory Phonology III*, 76–98. Cambridge: Cambridge University Press.

Inkelas, Sharon. 1990. *Prosodic Constituency in the Lexicon*. Outstanding Dissertations in Linguistics Series. New York: Garland Publishing Co.

Jong, Ken de. 1998. Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics, 26*. 283–310.

Jun, Jongho, Jungsun Kim, Hayoung Lee, & Sun-Ah Jun. 2006. The prosodic structure and pitch accent of northern kyungsang Korean, *Journal of East Asian Linguistics* 15(4). 289–317.

Jun, Sun-Ah. 1993. *The phonetics and phonology of Korean prosody.* PhD dissertation, The Ohio State University.

Jun, Sun-Ah. 1996. *The Phonetics and Phonology of Korean Prosody: intonational phonology and prosodic structure*. New York: Garland Publishing Inc.

Jun, Sun-Ah. 1998. The Accentual Phrase in the Korean prosodic hierarchy, *Phonology*. 15(2). 189–226

Jurafsky, Dan, Alan Bell, Michelle Gregory & William Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In Joan Bybee and Paul Hopper (eds) *Frequency and the Emergence of Linguistic Structure*, 229–254. Amsterdam: John Benjamins.

Hayes, Bruce & Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language & Linguistic Theory* 9. 47–96.

Huggins, A. William. 1975. On isochrony and syntax. In Gunnar Fant & Mark Tatham (eds.) *Auditory Analysis and Perception of Speech*, 455–464. New York: Academic Press.

Kainada, Evia. 2009. *Effects of prosodic boundary strength on segmental and suprasegmental processes in Modern Greek*. PhD dissertation, University of Edinburgh.

Keating, Patricia, Taehong Cho, Cécile Fougeron & Chai-Shune Hsu. 2003. Domain-initial strengthening in four languages. In John Local, Richard Ogden, & Ros Temple (eds.) *Phonetic interpretation: Papers in laboratory phonology VI*, 143–161. Cambridge: Cambridge University Press.

Keller, Frank, Mirella Lapata & Olga Ourioupina. 2002. Using the web to overcome data sparseness. Paper presented at the *Conference on Empirical Methods in Natural Language Processing*, Philadelphia.

Kenstowicz, Michael & Hyang-Sook Sohn. 1997. Phrasing and focus in Northern Kyungsang Korean. In Pier Marco Bertinetto, Livio Gaeta, Georgi Jetchev & David Michaels (eds.) *Certamen Phonologicum III*, 137–156. Turin: Rosenberg and Sellier.

Kim, No-Ju. 1997. *Tone, segments, and their interaction in North Kyungsang Korean: A correspondence theoretic account*. PhD dissertation,The Ohio State University.

Klatt, Dennis. 1973. Interaction between two factors that influence vowel duration, *Journal of the Acoustical Society of America*, 54. 1102–1104.

Klatt, Dennis. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America 59.* 1208–1220.

Kohler, Klaus. 1983. Prosodic boundary signals in German. *Phonetica*, 40. 89–134.

Kori, Shiro. 1997. Nihongo no intoneeshon: kata to kinoo [Japanese intonation: form and function]. In Tetsuya Kunihiro (ed.) *Akusento, Intoneeshon, rizumu to poozu* [Accent, intonation, rhythm, and pause], *vol. 2 of Nihongo Onsei*, 169–202. Tokyo: Sanseido.

Krakow, Rena. 1999. Physiological organization of syllables: a review. *Journal of Phonetics* 27(1). 23–54.

Krivokapic, Jelena. 2007. Prosodic planning: effects of phrasal length and complexity on pause duration. *Journal of Phonetics* 35, 162–179.

Krull, Diana. 1997. Prepausal lengthening in Estonian: Evidence from conversational speech. In Ilse Lehiste & Jaan Ross (eds.), *Estonian Prosody: Papers from a symposium*, 136–148. Tallinn: Institute of Estonian Language.

Kubozono, Haruo. 2006. Focus and intonation in Japanese: Does focus trigger pitch reset? In Shinichiro Ishihara, Michaela Schmitz and Anne Schwarz (eds.), *Working Papers of the SFB 632: Interdisciplinary Studies on Information Structure 9*, 1–27. Potsdam, Germany: Universitätsverlag Potsdam.

Kuperman, Victor, Mirjam Ernestus & Harald Baayen. 2008. Frequency distributions of uniphones, diphones, and triphones in spontaneous speech. *Journal of the Acoustical Society of America* 124(6). 3897–3908.

Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge University Press.

Ladd, D. Robert. 2008. *Intonational Phonology. 2nd Edition*. Cambridge University Press.

Ladefoged, Peter. 1982. *A Course in Phonetics, 2nd edition*. New York: Harcourt, Brace, Jovanovich.

Lavoie, Lisa. 2001. *Consonant Strength: Phonological Patterns and Phonetic Manifestations*. London: Routledge.

Lehiste, Ilse. 1972. The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 51. 2018–2024.

Lehiste, Ilse. 1973. Phonetic disambiguation of syntactic ambiguity, *Glossa,* 7, 107–121.

Levy, Roger & T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. In Bernhard Schlökopf, John Platt and Thomas Hoffman (eds.), *Advances in Neural Information Processing Systems (NIPS) 19*, 849–856. Cambridge, MA: MIT Press.

Lieberman, Philip. 1963. Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech* 6, 172–187.

Lindblom, Björn. 1968. Temporal organisation of syllable production. *Quarterly Progress and Status Report* 2–3/1968, Speech Transmission Laboratory, Stockholm.

Lindblom, Björn. 1990. Explaining phonetic variation: a sketch of the H & H theory. In William Hardcastle & Alain Marchal (eds.) *Speech Production and Speech Modeling*, 403–439. Dordrecht: Kluwer Academic.

Lindblom, Björn, Bertil Lyberg & Karin Holmgren. 1981. *Durational Patterns of Swedish Phonology: Do They Reflect Short-Term Memory Processes?* Indiana University Linguistics Club, Bloomington, Indiana.

Mayo, Catherine, Matthew Aylett & D. Robert Ladd. 1997. Prosodic transcription of Glasgow English: an evaluation study of GlaToBI, in Antonis Botinis, Georgios Kouroupetroglou, & George Carayannis (eds) *Intonation: Theory, Models and Applications*, 231–234. Athens, Greece: ISCA Archive.

McKendry, Inga (in prep.) Tone, prominence and prosodic structure in Santo Domingo Nuxaá Mixtec (provisional title). Edinburgh: University of Edinburgh PhD dissertation.

Nakai, Satsuki, Sari Kunnari, Alice Turk, Kari Suomi, K. & Riikka Ylitalo. 2005. Final lengthening and quantity in Finnish and Japanese. Poster presented at the *Between Stress and Tone Conference*, Leiden.

Nakai, Satsuki, Sari Kunnari, Alice Turk, Kari Suomi & Riikka Ylitalo. 2009. Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics* 37. 29–45.

Nakatani, Lloyd, Kathleen O'Connor & Carletta Aston. 1981. Prosodic aspects of American English speech rhythm. *Phonetica*, 38. 84–106.

Nespor, Marina & Irene Vogel. 1986. *Prosodic Phonology*. Dordrecht: Foris Publications.

Nooteboom, Sieb. 1972. *Production and Perception of Vowel Duration: A Study of Durational Properties in Dutch.* PhD dissertation, University of Utrecht.

Ogden, Richard. 2004. Non-modal voice quality and turn-taking in Finnish. In Elizabeth Couper-Kuhlen & Cecilia Ford (eds.) *Sound Patterns in Interaction*, 29–62. Amsterdam: John Benjamins.

Pickering, Martin & Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27, 169–226.

Pierce, John R. 1961. *Symbols, signals and noise: The nature and process of communication*. New York: Harper.

Pierrehumbert, Janet. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory phonology 7*, 101–139. Berlin/New York: Mouton.

Pierrehumbert, Janet & Mary Beckman. 1988. *Japanese Tone Structure*, Linguistic Inquiry Monograph 15. Cambridge, MA: MIT Press.

Pierrehumbert, Janet & David Talkin. 1992. Lenition of /h/ and glottal stop. In Gerry J. Docherty and D. Robert Ladd (eds), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody,* 90–115. Cambridge: Cambridge University Press.

Pluymaekers, Mark, Mirjam Ernestus & R. Harald Baayen. 2005a. Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America* 118. 2561–2569.
Pluymaekers, Mark, Mirjam Ernestus & R. Harald Baayen. 2005b. Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62. 146–159.
Port, Robert. 1981. Linguistic timing factors in combination. *Journal of the Acoustical Society of America* 69. 262–274.
Price, Patti, Mari Ostendorf, Stefanie Shattuck-Hufnagel & Cynthia Fong. 1991. The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America* 90. 2956–2970.
Redi, Laura & Stefanie Shattuck-Hufnagel. 2001. Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, 29 (4). 407–429.
Remijsen, Bert & D. Robert Ladd. 2008. The tone system of the Luanyjang dialect of Dinka. *Journal of African Languages and Linguistics* 29(2). 173–213.
Salverda, Anne Pier, Delphine Dahan & James M. McQueen. 2003. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 90. 51–89.
Schafer, Amy J., Shari Speer & Paul Warren. 2004. Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In John C. Trueswell and Michael K. Tanenhaus (eds.) *Approaches to Studying World-Situated Language Use*, 209–225. Cambridge, MA: MIT Press.
Selkirk, Elisabeth O. 1980. Prosodic domains in phonology: Sanskrit revisited. In Mark Aronoff & Richard Oehrle (eds.), *Language Sound Structure*, 107–136. Cambridge, MA: MIT Press.
Selkirk, Elisabeth O. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3. 371–405.
Selkirk, Elisabeth O. 1993. Modularity in constraints on prosodic structure. ms., presented at the *ESCA Workshop on Prosody*, Lurid.
Selkirk, Elisabeth O. 1996. The prosodic structure of function words. In James L. Morgan and Katherine Demuth (eds), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition,* 187–214. Mahnwah, NJ: Lawrence Erlbaum Associates.
Shannon, Claude E. 1948. *A mathematical theory of communication*. Bell System Technical Journal, 27. 379–423 and 623–656.
Shattuck-Hufnagel, Stefanie, Mari Ostendorf & Ken Ross. 1994. Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics* 22. 357–388.
Shattuck-Hufnagel, Stefanie & Alice Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25(2). 193–247.
Shatzman, Karen & James M. McQueen. 2006a. Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics* 68(1). 1–16.
Shatzman, Karen & James M. McQueen. 2006b. Prosodic Knowledge Affects the Recognition of Newly Acquired Words. *Psychological Science* 17 (5). 372–377.
Son, Rob van. 1993. *Spectro-temporal features of vowel segments*. PhD. Dissertation, University of Amsterdam. Studies of language and language use 3. Amsterdam: IFOTT.
Son, Rob van & Louis Pols. 2003. How efficient is speech?, Institute of Phonetic Sciences, University of Amsterdam, *Proceedings of the Institute of Phonetic Sciences* 25. 171–184.
Steedman, Mark. 2000. Information structure and the syntax-phonology interface. *Linguistic Inquiry* 31(4). 649–689.
Stevens, Kenneth N. 2002. Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America* 111 (4). 1872–1891.
Strangert, Eva. 2003. Emphasis by pausing. In M. J. Solé, D. Recasens & J. Romero (eds.) *Proceedings of the 15th International Congress of Phonetic Sciences*, 2477–2480. Rundle Mall: Causal Publications.
Sugahara, Mariko & Alice Turk. 2009. Durational correlates of sub-lexical constituent structure in English. *Phonology* 26. 477–524.
Suomi, Kari. 2007. On the tonal and temporal domains of accent in Finnish. *Journal of Phonetics* 35. 40–55.

Tanaka, Hiroko. 2004. Prosody for marking transition-relevance places in Japanese conversation: The case of turns unmarked by utterance-final objects. In Elizabeth Couper-Kuhlen & Cecilia Ford (eds.), *Sound Patterns in Interaction: Cross-linguistic Studies From Conversation*, 63–96. Amsterdam: John Benjamins.

Truckenbrodt, Hubert. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30(2). 219–255.

Turk, Alice & Snezhina Dimitrova. 2007. English phrasal stress targets multiple, optional, lengthening sites. In Jürgen Trouvain and William Barry (eds.) *Proceedings of the 17th International Congress of Phonetic Sciences*, 1177–1180. Saarbrücken.

Turk, Alice & James Sawusch. 1997. The domain of accentual lengthening in American English. *Journal of Phonetics* 25, 25–41.

Turk, Alice & Stefanie Shattuck-Hufnagel. 2000. Word-boundary-related duration patterns in English. *Journal of Phonetics* 28, 397–440.

Turk, Alice & Stefanie Shattuck-Hufnagel. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics* 35, 445–472.

Turk, Alice & Laurence White. 1999. Structural effects on accentual lengthening in English. *Journal of Phonetics* 27, 171–206.

Vaissière, Jaqueline. 1983. Language independent prosodic features. In Anne Cutler & D. Robert D. Ladd (eds.). *Prosody: Models and Measurements*, 53–65. Berlin: Springer Verlag.

Venditti, Jennifer, Kikuo Maekawa & Mary Beckman. 2008. Prominence marking in the Japanese intonation system. In Shigeru Miyagawa & Mamoru Saito (eds.). *The Oxford Handbook of Japanese Linguistics*, 456–512. Oxford: Oxford University Press.

Watson, Duane, Mara Breen & Edward Gibson. 2006. The role of syntactic obligatoriness in the production of intonational boundaries. *Journal of Experimental Psychology: Learning, Memory and Cognition* 32. 1045–1056.

Watson, Duane & Edward Gibson. 2004. The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes* 19(6). 713–755.

White, Laurence. 2002. *English Speech Timing: a Domain and Locus Approach.* PhD dissertation, University of Edinburgh. (http://www.cstr.ed.ac.uk/projects/eustace/dissertation.html).

White, Laurence & Alice Turk. English words on the Procrustean bed: polysyllabic shortening revisited. *Journal of Phonetics* 38. 459–471.

Wightman, Colin, Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91(3). 1707–1717.