UNIVERSITY OF CALIFORNIA

Los Angeles

On the Question of Accent Domains in English

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in Linguistics

by

Molly Susan Shilman

2008

The dissertation of Molly Susan Shilman is approved.

Daniel Büring

Jody Kreiman

Kie Ross Zuraw

Sun-Ah Jun, Committee Chair

University of California, Los Angeles

2008

I dedicate this dissertation to my family: my mother Mary, my father Lee, and my brother Michael. Thank you for always pushing me to be better than I am.

Dear Sun-Ah,

Thanks for everything!

-Molly

Table of Contents

List of Figures

List of Tables

Acknowledgments

Without the help and support of many people, this dissertation would never have come to be. I owe all of them a great deal and have used this space to try to thank them all adequately.

First, of course, I thank my advisor, Sun-Ah Jun for her advice and guidance. Sun-Ah's acoustic phonetics course changed the way I looked at linguistics, and her course on intonation changed it again. Indeed, this dissertation grew out of a term paper written for a seminar co-taught by Sun-Ah and by Daniel Büring. Sun-Ah's knowledge of intonation systems and the mapping between their phonology and phonetics—both broad and deep—has proven an invaluable resource to me as I worked my way through this project. Her unvarnished honesty has also been a great help, since I have no sense of subtlety to speak of.

The feedback and support I received from the other members of my committee, Daniel Büring, Kie Zuraw, and Jody Kreiman, have also been critical to the success of this project. Daniel has offered his advice and enthusiasm from the beginning of it, and has continued to remind me of the progress I have made when all I see is the work that remains. Kie's questions and suggestions have always tremendously improved any part of this work that she has commented on; my thanks to her for her lucid, insightful, and kind observations.

Thanks also to everyone who completed my experiments, including the friends who helped me test items for readability, the pilot subjects, and the actual subjects. Special thanks to everyone who participated multiple times—it was probably very dull, and I appreciate your willingness to suffer a little. On a related note, J'aime Roemer and

Byron Ahn, who helped out by transcribing speaker data, provided something I never could have: other eyes (and ears) on my data.

I also owe a great deal to Sandra Chung, whose Syntax 1 course at the University of California at Santa Cruz was so absolutely compelling and fascinating that I couldn't help becoming a Linguistics major, and to Armin Mester and Junko Ito, who made me into a phonologist. My gratitude also to Geoffrey Pullum, who lead by example in teaching the ruthless dissection of linguistic formalism, and to Angela Elsey, one of the first people to call me serious and mean it as a compliment.

My supervisors and boss at Consumer Financial Service Corporation, Ivan Rascanin, Stacy Leier-Valentine, and Loy Shefflot taught me a great deal—Loy most of all. Working for and with you improved my communication skills, my computer skills, and my confidence and helped me let go of any remaining emotional connection to my own prose—and a lot more, besides. Everyone should be so lucky.

During the course of my graduate studies, I have also had the good fortune to benefit from the support, linguistic and otherwise, of my fellow graduate students. To the senior students who offered mentoring and encouragement, including Christina Esposito, Jeff Heinz, Shabnam Shademan, Tim Arbisi-Kelm, Kuniko Nielsen, and Rebecca Scarborough: each of you has taught me a lot, and has done so painlessly and with great patience. To the other students in my cohort, Lawrence Cheung, Sameer Khan, Christina Kim, Tuan Le, Leonard Chacha Mwita, Reiko Okabe, and Lauren Varner: I have a great deal of respect and admiration for each of you and will miss you all. Sameer: you've been a fantastic friend, often more supportive than I deserve. If you weren't mostly in charge of making me have fun, it wouldn't get done. Much appreciation also to all of the other students who have made me feel at home here, including Jennifer Michaels, Kristine Yu, Vincent Homer, Tomoko Ishizuka, Ingvar Lofstedt, Nicole Gfroerer, and J'aime Roemer.

To Kathryn Roberts, Mandana Hashemzadeh, Lisa Harrington, and Melanie Levin: your company has been a pleasure and your assistance with the practical things that stump me utterly is appreciated more than I can say.

To the friends who have been with me the longest, Hab, Autumn, Huan-Hua, and Anna: you have all left your mark on me, and I hope that you continue to do so. In the years that I have known each of you, you have changed me in every way: body, mind, heart, and soul. As far as I can tell, it has always been for the better (in this, as in my dissertation, all remaining faults are entirely mine). As I've worked on this project, each of you has, in your own way, reminded me that there is life outside of academia and, at the same time, that there is real wonder in curiosity and the life of the mind.

Naturally, I owe a great and obvious debt to my family. To my mother Mary and father Lee: apart from the fact that I would not exist without you, I am also grateful for the all the ways you have supported me in the course of my life and my education and for all the time you spent trying to convince me to stop studying, go outside, and play. Mom: I would never have thought that trying to draw an egg could tell me so much. Dad: I promise to spend more time having fun now than I have been during the researching and writing of my dissertation. To my brother Michael: you have always pushed me, and your effective and persistent nagging has been instrumental in propelling me through the later stages of preparing my dissertation. It was, under the circumstances, the best thing anyone could have done for me, and it means a lot to me that you did it.

Finally, as a product of the public education system of the state of California, my entire education through high school was entirely funded by the taxpayers of California—in particular, by all of the Californians who paid property taxes during the years 1983–1996. I was able to go to college because of funding provided through the University of California at Santa Cruz. My graduate education, including the production of this dissertation, was funded by fellowships and teaching assistant positions from the University of California at Los Angeles. For my post-secondary education, I will always be grateful both to my home state's taxpayers and to everyone who has contributed to the University of California system. Thank you all for giving me something that I could never have obtained on my own: the experience—from kindergarten through the research and writing of my dissertation—has been truly priceless.

VITA

June 1, 1979	Born, Oakland, California
2001	B.A., Linguistics, *with departmental and college honors, PBK* University of California, Santa Cruz Santa Cruz, California
2001–2003	Project Manager and Editor Consumer Financial Service Corporation Emeryville, California
2004–2006	Teaching Assistant, Associate Department of Linguistics University of California, Los Angeles
2006–2007	President's Fellowship University of California, Los Angeles
2007 (Spring)	Teaching Fellow Department of Linguistics University of California, Los Angeles
2007–2008	Dissertation Year Fellowship University of California, Los Angeles
2008 (Spring)	Teaching Fellow Department of Linguistics University of California, Los Angeles

PUBLICATIONS AND PRESENTATIONS

Jun, Sun-Ah and Molly Shilman. 2008. Default Phrasing in English RC Attachment Data. Proceedings of Speech Prosody 2008

Shilman, Molly. 2007. Levels of the Prosodic Hierarchy in English. *Talk presented at the International Congress of Phonetic Sciences XVI,* Saarbrucken, Germany.

__. 2007. Levels of the Prosodic Hierarchy in English. Proceedings of ICPhS XVI.

__. 2006. Does English have accent domains? *Poster presented at the 152nd meeting of the Acoustical Society of America,* Honolulu, Hawaii.

ABSTRACT OF THE DISSERTATION

On the Question of Accent Domains in English

by

Molly Susan Shilman

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 2008

Professor Sun-Ah Jun, Chair

Influential generative theories of prosody rely on phrasing to generate all sentence-level prominence: prominence can come from the edges or heads of phrases, but not from any other source. For English, one result of this approach is that pitch accent is treated as the head of a level of prosodic phrase just above the word. However, the evidence for this level of phrasing is inconsistent, and there is no evidence of a linkage between this level of phrasing and pitch accent placement. The rationale for linking the two together is generally that there are pitch accents (heads), so there must be phrases for them to head.

This dissertation investigates the question of whether there is support for including a level of prosodic phrase in the analysis of English prosody that is headed by pitch accent—that is, whether English has Accent Domains. To this end, two production experiments were conducted, both in American English and using read speech. Pitch

accent position and syntactic structure were varied to create sentences that Accent Domain theories predict to have three different levels of prosodic unit: word, Accent Domain, and intermediate phrase. The segmental material around the boundaries of these units is kept identical, allowing for measurement of fine phonetic detail around the predicted boundary locations. The recorded items were measured for known phonetic markers of prosodic boundaries, including final lengthening.

The findings of these experiments are inconsistent with the proposals that the experiments were designed to test: final lengthening of the appropriate degree is found, but the distribution of the boundary marking is not consistent with these phrases being headed by pitch accent. Therefore, the use of this level of phrasing to generate pitch accent in English prosody is rejected and an alternative proposal—one that generates pitch accent independent of phrasing—is outlined. This proposal is formalized in the generative phonological framework of Optimality Theory ; the crucial addition to the grammar that allows it to generate pitch accents without Accent Domains is the X > Y constraint type, which directly imposes conditions on the relative prominence of words in the output. The general properties of these constraints when incorporated into larger grammars for English and their possible uses in grammars for other languages are also discussed.

Chapter 0: Introduction

As speakers and listeners of language, we have an intuitive sense that spoken utterances contain groupings of words. In some cases, these groupings are so clear to us that they make their way into the punctuation conventions of the language, as in the case of non-restrictive relative clauses in written English which are, by convention, set off with commas from the nouns they modify. In the case of non-restrictive relative clauses, the breaks that we hear line up well with phrase boundaries in the syntax but, as has been pointed out by many linguists, this is not always the case. Perhaps the example most often reproduced to illustrate this point is from Chomsky and Halle (1968: 372), shown below with its original bracketing, "where the bracketed expressions are the three noun phrases in the predicate":

(1) This is [the cat that caught [the rat that stole [the cheese]]].

In the neutral spoken production of this sentence, "the major breaks are after *cat* and *rat,*" which Chomsky and Halle describe as having "the three-part structure *this is the cat—that caught the rat—that stole the cheese.*" In this example, as in many cases, the phrasing of the utterance as it is spoken is not the same as its syntactic phrasing.

The study of how spoken utterances are phrased is one of the major components of the study of prosody, but these groupings—prosodic phrases—are not the only components of prosody. The study of prosody is also concerned with relative prominence. This view of prosody—that it deals in both phrasing and relative prominence—is nicely supplemented by the "working definition" of prosody put forth by Shattuck-Hufnagel & Turk (1996):

1. acoustic patterns of F0, duration, amplitude, spectral tilt, and segmental reduction, and their articulatory correlates, that can be best accounted for by reference to higher-level structures, and

2. the higher-level structures that best account for these patterns. p. 4

In what follows, we will be concerned with prosody in all of these senses, with the relationships between prosodic units, and with the relationship of prosody to other parts of the grammar. In particular, we will examine a proposal about the higher-level structure of prosody in English to determine whether there are any acoustic patterns that are best explained by this proposal, or whether it would require a more purely abstract phonological unit.

The proposal that I investigate here concerns the relationship between prosodic units and relatively prominent elements above the word level (henceforth "prominences"): prominences are often hypothesized to be the heads of prosodic phrases, as in Beckman's (1996) definition of prosody as the "hierarchically organized structure of phonologically defined constituents and *heads*" (emphasis mine). To define prominences this way is to assert that the existence of prominences is dependent on the existence of prosodic units for them to head. This hypothesis is so common for relative prominence within the word, where the prosodic units are feet and words and the prominences are secondary and primary stresses, that it is rarely questioned. Above the word level, the hypothesis works quite well in languages identified as edge marking languages by Jun (2005), which have a prosodic constituent slightly larger than a content word that serves as the domain of assignment for pitch accent, the lowest level of prominence above the word. Head marking languages, such as English, are thought to lack such a constituent but have clear pitch accent marking (thus "head marking"); this constitutes something of a challenge for the hypothesis that prominences are always heads, as there does not seem to be any prosodic unit for the pitch accents of English and other head marking languages to head. This naturally leads to the question of whether any evidence can be found for a prosodic unit headed by the pitch accent in English. The chapters that follow present an investigation of this question. Chapter 2 is devoted to a review of the relevant literature that provides a context for two experiments, presented in Chapters 3 and 4, and for an interpretation of the experimental results and resulting phonological analysis presented in Chapter 5.

Chapter 2: Background

This chapter provides background on proposals about the phonology of prosody, both general and English-specific (i.e. a review of proposals about part 2 of Shattuck-Hufnagel & Turk's definition, as on p. 2). The goal of this review is to clarify the nature of the question addressed here and the reasons for collecting the particular types of data presented as part of the original work that follows. To understand exactly how the question can be answered empirically, a short summary of work done on the phonetic realizations of prosodic entities (i.e. a review of work pursuing part 1 of the definition on p. 2) is then presented.

2.1 Prosodic structure

There have been any number of proposals about what the formal linguistic treatment of prosody should be; there are several schools of thought on the matter that are relevant here, and these are reviewed below. As mentioned above, there seem to be two types of prosodic entities—phrases and prominences (heads). All of the theories described below address both phrasing and relative prominence; because they differ considerably more in their phrasing proposals than in their treatment of heads, the section on phrasing (2.1.1) is sub-divided by theoretical approach, while the section on prosodic heads (2.1.2) is sub-divided by the level of structure (putatively) headed.

2.1.1 Prosodic phrasing

In The Sound Pattern of English, Chomsky and Halle use a non-hierarchical representation of prosody, indicating the presence of prosodic breaks using boundary symbols; these symbols are initially placed according to the syntax and are later moved by readjustment rules that are responsible for the differences between syntactic and prosodic phrasings. Such a system has been shown to over generate (Selkirk 1980) and to require more outright stipulation than the prosodic phonology approach that came after it (Hayes 1989). For these reasons, the boundary symbol approach has been superseded by an approach pioneered in work by Selkirk. Starting with this work, the prevailing approach to encoding prosody at the phrasal level has been to use a prosodic tree, which is constructed with reference to the surface structure generated by the syntax. All of the approaches described employ a tree representation for prosody, although they differ in the constituents and types of structures they permit, as well as the way in which they derive these tree structures.

2.1.1.1 Prosodic Phonology

Prosodic Phonology—the research program initiated in the early work of Selkirk—constructs the prosodic tree from the syntactic s-structure; in the most rigid proposals of this type, the syntactic structure completely determines the prosodic structure. The prosodic tree, like the syntactic tree, is hierarchically organized. However, prosodic trees differ from syntactic trees in a number of ways; in much work in prosodic phonology, prosodic trees are:

1. Composed exclusively of a (small) finite set of prosodic categories
2. Strictly hierarchical, with a fixed order among the prosodic categories used (see Table 1)
3. Strictly layered (exhaustively parsed at every level and non-recursive)
4. N-ary (rather than binary) branching

As a result of these conditions, the prosodic tree, unlike the syntactic tree, is of a fixed, finite depth. Such structures are prefigured in Chomsky and Halle (1968), in which it is observed that an example like their (124)—reproduced as example (1) on p. 1 of this paper—"with its multiply embedded sentences," could be converted to a structure that more closely matches its prosodic structure, "where each embedded sentences is sister-adjoined in turn to the sentences dominating it." The resulting structure would be "a conjunction of elementary sentences (that is, sentences without embeddings)," which would allow for a rule placing "intonation breaks...preced[ing] every occurrence of the category S (sentence) in the surface structure" (372). This course of action is not pursued in Chomsky and Halle (1968)—and differs in important ways from the course pursued in Prosodic Phonology. Perhaps most notably, Chomsky and Halle seem to have envisioned the transformed structure they mention as a late-stage syntactic representation rather than a purely phonological structure. Accordingly, they mention the use of syntactic category nodes rather than developing an inventory of phonological categories.

In proposals that fall under the umbrella of Prosodic Phonology, syntactic structure plays a key role—often the key role—in determining prosodic structure, but the tree so constructed is conceived of as a purely phonological entity, using phrasal categories that are fully separate from the categories employed in the syntax. Although proposals about the exact inventory of constituents making up the prosodic tree vary in their details, the following levels are often included in models of prosodic phonology:

Table 1—Categories of the prosodic hierarchy and their approximate descriptions

Category	*Description*
Utterance	The largest prosodic grouping, its right edge is marked by a clear pause in the phonetic realization.
Intonation phrase	Its boundaries often, but not always, line up with syntactic clause boundaries; its right edge is marked by a tonal sequence.
Phonological phrase	It canonically includes a lexical syntactic head and the material within its maximal projection on the non-branching side.
Clitic group	The smallest grouping above the word, it is made up of a content word and the surrounding function words that are part of the same syntactic phrase(s).
Prosodic word	This is related to the grammatical word, although it may include more material; for Hayes (1989), this is the lowest level in the prosodic hierarchy.
Foot	A grouping of syllables within the word; for Selkirk (1986), this is the lowest level in the prosodic hierarchy.
Syllable	A grouping of segments within the foot.

Considerably more rigorous definitions of these units are proposed in specific works (e.g. Selkirk 1986, Nespor & Vogel 1986, Hayes 1989). These definitions are algorithmic, specifying a particular process for forming a prosodic tree on the basis of the syntactic s-structure; the more specific definitions differ depending on the source. This is because the research program of Prosodic Phonology is not simply to identify prosodic constituents for their own sake, but to come up with a principled relation between syntax and prosody that will derive the domains relevant to phonological rules. In many cases, the evidence advanced for the size and content of a particular phrase is its use as a bounding domain for phonological rules or the ability of its edges to trigger a phonological rule. Some rules used to diagnose prosodic phrasing are segmental and some concern the assignment of head status to one of the daughter nodes within the domain; there is some overlap between these two groups. For example, one argument advanced by Selkirk (1986) for the existence of levels of structure above the word is the stress / vowel length pattern in Chi Mwi:ni, which she claims can only be accounted for in a principled way by looking at a level of prosodic structure above the word. This argument involves both segmental alternation and assignment of prosodic head location: Selkirk takes a segmental alternation in vowel length to be an indication of the location of a prosodic head and then argues for a prosodic constituent on the basis that its head will be positioned as needed to drive the phrase-level vowel-length alternation.

2.1.1.2 Intonational Phonology and ToBI

Like Prosodic Phonology, Intonational Phonology assumes that the prosodic tree is composed of a small fixed set of prosodic phrases, hierarchical, strictly layered, and n-ary branching. Unlike Prosodic Phonology, Intonational Phonology does not attempt to derive the prosody from syntax; instead, the focus is on defining the phonologically important aspects of the prosody of a particular language as instantiated in actual utterances (as discussed in Jun 1998). Thus, the prosodic units of Intonational Phonology are defined by intonation itself rather than by syntax. Intonational Phonology assumes an Autosegmental-Metrical view of intonation; the approach is autosegmental in its view of tones, which Intonational Phonology regards as autosegments linked up to the segmental string according to principles relating these tones to the prosodic structure of the sentence. The metrical component of the Intonational Phonology approach comes from its representation of prominence (headedness), which is based on the type of metrical grid discussed in section 2.1.2.1. An analysis developed within the Intonational Phonology framework generally includes an inventory of phrase types, an inventory of tones used to mark the edges of the different phrase types, and (for languages with pitch accents) an inventory of pitch accent types.

Intonational Phonology, unlike Prosodic Phonology, does not assume that there is a universal inventory of prosodic phrase types that occur in every language. However, the primitive types used in these models are the same: all use a sparse inventory of tonal targets (often just High and Low), which can appear alone or in combination and are associated with the edges or the heads of prosodic constituents, and an inventory of similar prosodic constituents. There is always at least one prosodic constituent above the word; most often, in Jun's (2005) sample, there are two such units. With the exception of the lexical tone languages included in the sample (Mandarin and Cantonese), all of the languages discussed have two levels of prosodic structure marked by intonation, although the exact nature of these units and the tone inventories used to mark them differs by language. Some trends also emerge in relationships between the lexical prominence type of a language (stress, lexical pitch accent, tone, or none) and the post-lexical prominence marking type of the language (whether a language marks the head or the edge of a prominent word with a pitch event or other supra-segmental phonetic realization of prominence): stress languages tend to be head-marking, lexical pitch accent languages tend to mark both the head and the edge, and languages with no lexical prominence marking tend to be edge-marking (Jun 2005: 446).

English is typical in this sense: it is a language with lexical stress that uses post-lexical pitch accent marking rather than edge marking at the word level. Like its close relative German—and unlike some other lexical stress and post-lexical pitch accent languages, such as Spanish and Greek—English employs a wide variety of different tonal types in its pitch accent marking and does not have pitch accent on every content word (*ibid,* 447). Because the research proposed below focuses on English, we next briefly review the Intonational Phonology analysis of English, following the analysis in Beckman & Pierrehumbert (1986).

In this analysis, English has two levels of prosodic structure above the word—the intermediate phrase and the intonation phrase. The intonation phrase is the highest level of structure proposed and is roughly the same size as the intonation phrase of Prosodic Phonology. The intonation phrase in Beckman & Pierrehumbert's analysis, however, is not defined by the syntactic structure from which it might be derived, but rather by the presence of a boundary tone that marks its right (and, in some cases, left) edge. The realization of the boundary tone marking the edge of the intonation phrase is localized to the syllable at the edge of the intonation phrase, as is the lengthening that obligatorily marks the right edge of an intonation phrase. The end of an intonation phrase may also be marked by a non-hesitation pause.

The intermediate phrase is above the prosodic word but below the intonation phrase, and the phrase tone that marks its right edge is subject to a rule of tone spreading—the phrase tone aligns with the right edge of the intermediate phrase and spreads leftward to the end of the last pitch accented word, so that the shift from the last pitch accent target to the phrase tone target is "relatively abrupt"; after this transition, there is a plateau in the f0 contour, as the phrase tone target is "maintained over the remainder of the phrase" (p. 288). The intermediate phrase is also identified as the domain of catathesis (downstep)—a phonological rule that lowers and compresses pitch range after a bitonal pitch accent[1]—in English. While the effect of downstep is most visible on high tones, the phenomenon is said to affect the entire pitch range for all material within the same phrase that is to the right of the bitonal pitch accent. In a phrase with multiple bitonal pitch accents, the downstep rule will apply multiple times, sometimes producing a pitch contour for the phrase that resembles a staircase. The intermediate phrase is also the smallest unit that may have its own pitch range—a whole intermediate phrase can have a lowered, compressed pitch range (as in Beckman & Pierrehumbert's discussion of various kinds of tags) and the pitch range is reset at the beginning of a new intermediate phrase. The last accent in a phrase is often not subject to catathesis even when preceded by a bitonal pitch accent and tends to be especially prominent when compared to other pitch accents (see below); the relevant phrasal domain for these generalizations also appears to be the intermediate phrase. The prosodic unit below the intermediate phrase is the prosodic word, although Beckman & Pierrehumbert do raise the possibility that there is an accentual phrase—a unit between the word and the intermediate phrase—in English, just as there is in Japanese (for more on this, see section 2.1.2.2).

In addition to the inventory of prosodic domains discussed, Beckman & Pierrehumbert propose an inventory of tonal autosegments that associate to words with sentence level stress; this intonational marking of sentence level stress is called pitch

[1] In Beckman & Pierrehumbert's analysis, the rule of catathesis can be stated in this way, as the inventory of bitonal pitch accents is larger than that employed in current MAE_ToBI (Beckman & Ayres-Elam 1997). Specifically, the Beckman & Pierrehumbert (1986) analysis includes the bitonal pitch accent H*+L, where the role of the L (which is abstract in the sense that it is not realized as lowered f0 on the surface) is to trigger downstep. In more current analyses of English, which lack bitonal pitch accents of the form H+L, catathesis is still regarded as a phonological process but is optional and may occur at any point in a phrase provided that there is an earlier H tone within the same phrase.

accenting. The simplest pitch accents are single high and low tones, and these tend to be realized on the primary stressed syllable of the prominent word. The other pitch accents are bitonal, with one of the tones marked as aligning with the primary stressed syllable of the prominent word and the other either leading or trailing this tone and realized on the syllable immediately preceding or following the main stress of the pitch accented word. One explicitly made link between phrasing and prominence in the Intonational Phonology model of English connects the relative prominence of some pitch accents to the level of the intermediate phrase. In particular, the last pitch accent in an intermediate phrase is often the most prominent pitch accent in that phrase; this pitch accent is called the nuclear pitch accent and is often referred to as the head of the intermediate phrase. (For more on the relationship between prominence and phrasing, see section 2.1.2.2.)

The phrase and tone inventories proposed in Intonational Phonology have clear phonetic markings, often realized not only in the pitch (f0) contour but also by cues in other phonetic dimensions, such as the relative durations of segmental material. The Intonational Phonology approach is associated with a very direct view of the mapping between the phonological representation of intonation and its phonetic realization, as in Pierrehumbert (1980), in which it is claimed that "there is no level of systematic phonetic representation for intonation"—in other words, "there is no well-defined level of representation in between the underlying representation as it is before any rules apply and the F0 contour which is output" (p. 28). In such a system, we expect the elements of the phonological system that captures the intonation of a language to have clear phonetic effects, making the expectation of being able to identify phonological entities through phonetic investigation quite reasonable in this framework. In addition, it has been shown that these domains are also the domains of phonological rules, as in Jun (1998) for Korean post-obstruent tensing, Baltazani (2002) for Greek hiatus resolution across word boundaries, and Hayes & Lahiri (1991) for Bengali r-assimilation and voicing assimilation, so that the domains of application for phonological rules may also constitute supporting evidence of intonationally defined prosodic domains in some cases.

The tones and prosodic units proposed by Intonational Phonology have been employed in the development of language-particular transcription systems for Tones and Break Indices, often referred to as ToBI transcription systems. These transcription systems are used to transcribe the phonologically relevant prosodic features of utterances; the systems for some languages also include transcription conventions for more purely phonetic information. ToBI systems are language-specific transcription systems for intonation; the single largest difference between ToBI systems and the other approaches reviewed in this section is that ToBI is a transcription system rather than a generative grammar.

The ToBI system for English is called Mainstream American English ToBI (MAE_ToBI). This transcription system uses only high (H) and low (L) values for tone targets, as proposed in Pierrehumbert (1980). There are three types of tonal markings (T): pitch accents (T*) aligned with the stressed syllable of a pitch accented word, phrase accents (T-) to mark the right edges of intermediate phrases, and boundary tones (T%) to mark the right edges of intonation phrases. The realizations of these tone types is as laid out in Beckman & Pierrehumbert (1986), although the current MAE_ToBI pitch accent inventory is smaller than that proposed in Pierrehumbert (1980) and Beckman &

Pierrehumbert (1986). Because transcriptions are generally annotations on a text string accompanied by a pitch track and a waveform (and, in some cases, a spectrogram), ToBI systems use numerical break indices to mark right phrase edges rather than requiring that a full prosodic tree be drawn for each utterance. The phrase types that are employed in MAE_ToBI are listed in Table 2. Each phrase type is shown with its Break Index (BI), a numerical marker indicating the strength of the break at the edges of the phrase, and a description of what the phrase type is used to transcribe in English. Note that these descriptions are similar but not identical to those used for these same phrase types in ToBI systems for other languages.

Table 2—Phrase types in the MAE_ToBI system

Category	*BI*	*Description*
Word	1	Every orthographic word is usually considered its own word in ToBI transcription. When the juncture between two words is judged to be more like the boundary between two segments within a word, a break index of 0 is used instead.
Intermediate phrase	3	The intermediate phrase (ip) is the domain of downstep and of nuclear pitch accent assignment. Its end is marked by a phrasal tone (L- or H-), which extends from the end of the last pitch accented word to the end of the ip. The end of the ip is also marked by a moderate degree of final lengthening.
Intonation phrase	4	The intonation phrase is the largest constituent in the MAE_ToBI hierarchy. Its beginning may be marked by an initial high tone (%H), which is realized locally on the segmental material at its very left edge. Its end is marked by a final high or low tone (H% or L%), also very locally realized, and by quite a lot of final lengthening.

As indicated in Table 2, MAE_ToBI does not include transcription of syllable or foot boundaries—that is, MAE_ToBI labels structure above the word level only.

As the astute reader will have noticed, the break index column in Table 2 includes the numerical indices 1, 3, and 4, but has no index 2; the 2 index in English is used to mark a mismatch between the perceived break strength and the tonal boundary marking (e.g. the break perceived by the transcriber is no greater than a word break, but the tonal contour in this area indicates that there must a phrase tone T- at that juncture; or the word preceding the break is lengthened but the tonal contour indicates the absence of a phrase tone T-). In more recent usage, some transcribers have replaced the 2 break index of the English ToBI system, instead labeling the break with the degree of perceived break strength (1, 3, or 4) and adding an "m" to indicate that the tonal marking does not match.[2] The 2 break index is not used in these transcriptions.

In some other ToBI systems (e.g. Greek, Korean, Serbo-Croatian), a different marking is used for mismatches between perceived break strength and tonal marking (often "m"), leaving the break index 2 available for other uses. One common use for the 2 break index is to mark the degree of juncture corresponding to the boundary after an accentual phrase. An accentual phrase is larger than a word—often, like the Prosodic Phonology Clitic Group, it is a content word and surrounding function words—whose edges are marked with tone autosegments and with the phonetic boundary markers described in section 2.2.1. Some languages, like Korean (Jun 1993, 2005), have accentual phrases but not pitch accents. Other languages, like Japanese (Venditti 1995) and French (Jun & Fougeron 2002), have both accentual phrases and pitch accents, in which case the pitch accent may be analyzed as the head of the accentual phrase (see section 2.1.2). English does not have accentual phrases in this sense, although the accent domain proposals discussed in 2.1.1.3 are not alone in positing that there is some domain for post-lexical accent assignment in English.

[2] Complete, revised MAE_ToBI labeling guidelines incorporating this change are not currently available. Some discussion of this proposal is linked from the English ToBI homepage (http://anita.simmons.edu/~tobi/ as of October 2008) in the section on the 2004 Workshop on ToBI for Spontaneous Speech.

Because we will be particularly interested in the link between prosodic phrasing and phonetic realization—and because it will be necessary to employ transcriptions of the prosody of actual utterances—the MAE_ToBI phrasal categories and other conventions have been employed in the work described below.

2.1.1.3 Accent domain proposals

In the Prosodic Phonology proposals reviewed in section 2.1.1.1, the only way to generate prominence is to mark some constituent of a prosodic domain as its head. Assigning prominence on the basis of phrasing is also a part of the proposal of Beckman & Edwards (1994), who show that different levels of prominence—interpreted by the authors as the heads of different levels of phrasing—are marked by different phonetic cues. However, as Shattuck-Hufnagel & Turk (1996) observe in their discussion of Beckman & Edwards' (1994) results, "the proposal leaves several levels in the prosodic constituent hierarchy without well-defined heads, and suggests no specific constituent for which Prenuclear Pitch Accents could serve as heads, at least in American English" (p. 224). Beckman & Pierrehumbert (1986), who consider but ultimately reject the use of a domain assigning the pre-nuclear pitch accent as its head in American English, point out that the existence of such a phrase is also crucial for "some...versions of metrical theory", where the only way to identify an element as prominent that is consistent with "the way pitch accent placement works" is to mark that element as the "designated terminal element...of some prosodic domain" (p. 268). In some recent proposals for deriving the prosody of an utterance from other independently necessary (syntactic and semantic) structure, one of the most important functions of lower levels of phrasing is to correctly place pre-nuclear pitch accents; several such proposals are reviewed below.

All of these proposals use a node labeling system that is closer to that of Prosodic Phonology than to Intonational Phonology—as expected given their use of syntactic and semantic structure to define prosodic constituency. Perhaps the most notable of these proposals is the work of Truckenbrodt (1999), who formalizes the syntax / semantics-prosody mapping using Optimality Theory (OT, Prince & Smolensky 1993). Just as the Prosodic Phonology grammars include algorithms for building a prosodic tree based on the information in the syntactic tree, Truckenbrodt (1999) makes use of OT constraints that favor particular relations between syntactic and prosodic phrases. The constraint that favors folding two syntactic entities into a single prosodic entity is called WRAP-XP and requires that a syntactic maximal projection XP be contained within a single phonological phrase (although note that the grammar as a whole need not actually require that this hold for every optimal mapping).

Important for current purposes is that Truckenbrodt utilizes what he calls the Lexical Category Condition (based largely on previous work by Selkirk and Nespor & Vogel), reproduced below:

(2) **Lexical Category Condition**
Constraints relating syntactic and prosodic categories apply to lexical syntactic elements and their projections, but not to functional elements and their projections, or to empty syntactic elements *and their projections*."
p. 226, emphasis in original

Because of the Lexical Category Condition, Truckenbrodt's WRAP-XP applies only to lexical projections, not to functional projections; there is, for example, no impetus to wrap together the subject and verb phrase of a regular transitive, active sentence, which are in the specifier and complement positions of the functional head T° (assuming that the sentence is headed by tense, and its root node is TP[3]). WRAP-XP is in competition with constraints that require alignment of the edges of each syntactic maximal projection (XP) with the edges of prosodic phrases (in the cases that Truckenbrodt discusses in detail, it is the right edges that align) and with a constraint against recursive phrasing.

Because Truckenbrodt's (1999) case studies are mostly concerned with the positioning of tone rather than stress, his other alignment constraints deal with alignment between tones and phrase edges, and the only constraint clearly concerned with headedness is one that requires right alignment of each focused constituent with a prosodic boundary (p. 248). However, in Truckenbrodt's other work (e.g. 2006, which discusses the head-marking languages English and German), he employs a constraint called STRESS-XP, which "requires that each XP receive phrase level stress, i.e., that it contain the head of a prosodic phrase," (Büring & Gutiérrez-Bravo 2001: 9). The position of each prosodic head is determined by other constraints that mandate particular relationships between the heads and edges of prosodic phrases (e.g. alignment or non-finality constraints). In other words, the positioning of prosodic heads in general is handled in the same way that the positioning of focused elements is handled in Truckenbrodt (1999). Note that even without the reference to prosodic phrasing in the alignment constraints, the notion of prominence used here is inherently phrasal because STRESS-XP itself makes reference to phrase level stress,

[3] I have used TP as the root node of a sentence in the syntax not to make any deep syntactic claim, but to be relatively neutral while avoiding confusion with the abbreviation IP, which is used for the Intonation Phrase.

characterized by Truckenbrodt (2006) as "stress on the level of the phonological phrase (Nespor and Vogel 1986, 1989)".

A similar system is employed by Büring & Gutiérrez-Bravo (2001), who use a constraint XP = pP, which requires that a lexical XP be aligned with a phonological phrase (the level of phrase between word and Intonation phrase, p. 11) to do the work that WRAP-XP does for Truckenbrodt. Although there are substantial differences between Büring & Gutiérrez-Bravo's grammar—which includes constraints concerned with focus and focus-related movement, for example—and Truckenbrodt's grammar, the basics of Büring & Gutiérrez-Bravo's grammar is closely related to Truckenbrodt's. Therefore, both make the same assumption about prominences above the word level: that they are the heads of smaller phrases. Both proposals make use of the Lexical Category Condition, as well, thus sharing the prediction that material dominated by the maximal projection of a functional head (and not by the maximal projection of a lexical head) should not belong to the same phonological phrase. Büring & Gutiérrez-Bravo distinguish two levels of phrasing above the word, which they call the phonological phrase and the intonation phrase. They also use the term accent domain to refer to the phonological phrase and, as suggested by this alternate name, analyze the pre-nuclear pitch accent as the head of the phonological phrase. The head of the intonation phrase is, in their terms, nuclear pitch accent.

As discussed above, these systems share with their Prosodic Phonology predecessors the property that there is no means of assigning head (prominent) status without reference to phrasing. Because the constraints mandating the presence of a head and regulating its position refer explicitly to phrasal stress and to phrase edges, there is no way for the grammar to prefer prominence in unphrased material. The lowest level of prominence above the word is, therefore, generated as the head of the smallest level of phrase above the word and is positioned with respect to the edges of this phrase.

2.1.1.4 Summary

We have seen several different types of proposals about prosodic phrasing. All of these approaches are related, and each contributes in an important way to our understanding of prosody. The Prosodic Phonology approach provides the fundamentals: the prosodic tree itself and the Strict Layer Hypothesis. Intonational Phonology, with its focus on the range of spoken utterances that actually occur, contributes both the insight that the syntax does not absolutely determine the prosody and the knowledge that prosodic units often have phonetic correlates. Furthermore, it is from research conducted in the tradition of Intonational Phonology that we know that no phonetic evidence has been found for the edges of a domain with the size and function of the accent domains predicted by the accent domain proposals mentioned above. These accent domains, then, can be justified only by their use in determining the locations of pitch accents, which are otherwise heads without phrases—the very configuration that most generative proposals predict should be impossible. The proposals of Truckenbrodt (1999) and Büring & Gutiérrez-Bravo (2001) contribute not only a clear picture of how important accent domains can be—providing, as they do, a way to predict the position of pitch accents—but also a way to predict the position of accent domain boundaries that is attentive to syntax but can also incorporate other factors.

2.1.2 Prominences (or prosodic heads)

As mentioned above, prominent elements above the word level are often hypothesized to be the heads of prosodic phrases. This hypothesis is entrenched enough that it might also be stated that the head, or strong daughter node, is the constituent with the greatest (perceived) prominence within the domain. The terms most often used to describe prominence of various degrees are *stress* and *accent,* with the use of both terms varying considerably depending on the author (see Shattuck-Hufnagel & Turk 1996 for a review of this confusing situation). In this work, prominence at the word level and below is called stress; prominence above the word level is called accent.[4] Just as there are different levels of stress (e.g. primary vs. secondary), there are different levels of accent (pre-nuclear vs. nuclear). The sections below summarize previous work on the assignment of prosodic heads. Because the largest division here seems to be by domain size (within word versus larger than word), this section is sub-divided by domain size rather than by theoretical approach (although differences in approach will be discussed as needed).

2.1.2.1 Prosodic heads within the word

For the lower domains of the hierarchy—syllable, foot, and prosodic word—the assignment of prominence has to do with word-level stresses. Two formal representations of prominence assignment are particularly useful: a tree representation with nodes marked strong and weak and the bracketed grid (both reviewed in Hayes 1995). The tree representation of stress assignment is illustrated in Figure 1 using the English word

[4] The original uses of these words have been preserved in quotations. An attempt has been made to include notes clarifying usage in quoted material when it varies significantly from the convention used here.

Apalachicola, which has the stress pattern Àpalàchicóla.[5] A bracketed grid representation of the same word is shown in Figure 2.

Figure 1—Tree representation of stress (* marks strong daughter, weak daughters unmarked)

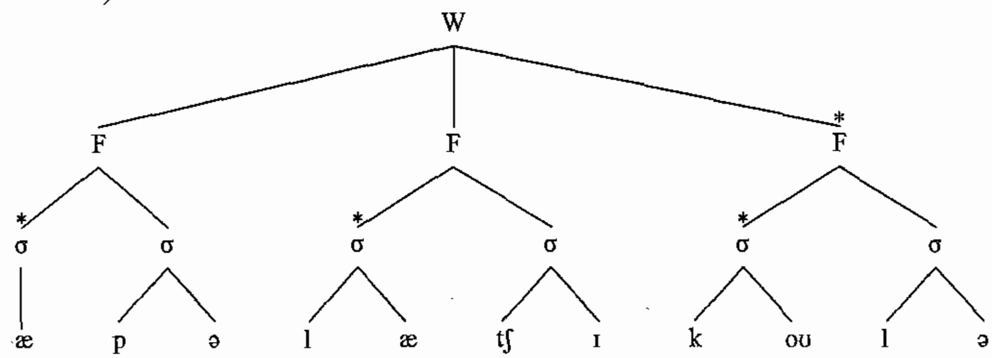


Figure 2—Bracketed grid representation of Apalachicola

```
Word      (                                            x           )
Foot      (x          )  (     x          )  (         x           )
Syllable  (x)   (     x)  (     x)  (     x)  (        x)  (       x)
           æ    p     ə   l     æ    tʃ   ɪ   k        oʊ    l     ə
```

Note: Syllable bracketing is often not included in bracketed grids, which are more often built over orthographic representations.

The mapping between these two representations is fairly intuitive: all of the material dominated by a node at some level, *n,* in the tree representation is enclosed in brackets at layer *n* of the bracketed grid representation. Above the syllable layer, every strong node marker in the tree is a grid mark; this mark is placed within the strong constituent, and is positioned within the strong constituent directly over a sub-constituent that has a grid mark at the next layer down. It is in order to allow the grid marks in the foot layer to be placed in this way that each syllable receives a grid mark over its nucleus.[6]

[5] In this orthography-based representation, the grave accent (`) indicates secondary stress; the acute accent (´) indicates primary stress.

[6] Assigning a strong node marker in the tree and a grid mark in the bracketed grid to the syllable nucleus is consistent with Nespor & Vogel's (1986) assertion that the syllable, as a unit of the prosodic hierarchy, should have a single strong daughter node.

In both of these representations, headedness is, in some sense, cumulative—the strongest syllable in the word is the strong syllable of the strong foot of the prosodic word. This syllable bears main word stress not because it receives a single designation as the main stress syllable, but because it is strong in every layer of the grid (or at every level of the tree). In languages with documented secondary stresses, these stresses are the strong syllables of weak feet within the word. In fact, in a well-formed representation, a constituent that is strong in layer $n + 1$ must also be strong in layer n. This is best formalized using the bracketed grid representation; the formal statement of this condition is called the Continuous Column Constraint (the statement of the condition in (3) is from Hayes 1995: 34, although the idea originates with Prince 1983):

(3) **Continuous Column Constraint**
A grid containing a column with a mark on layer $n + 1$ and no mark on layer n is ill-formed. Phonological rules are blocked when they would create such a configuration.

Having established that prominence at layer $n + 1$ can only be assigned to certain constituents of layer n rules out some patterns of prominence assignment but is not enough to determine which syllable of a foot or which foot of a word will be prominent. In general, the heads of these units are assigned by rule, and these rules make crucial reference to the edges of the domains. For example, the most prominent syllable in a foot is most often at its left or right edge, and the most prominent foot of a word is, likewise, usually positioned either initially or finally within the word.[7] In Figure 1 (and Figure 2), prominence within the foot is assigned to the left-most syllable of the foot and

[7] The most common type of exception to this is antepenultimate main word stress, often described as involving final extrametricality (see Hayes 1995). Nespor & Vogel (1986), who claim that syllables and feet are subject to the same exhaustive parsing as higher domains, would be forced to analyze these cases as ternary feet at the right word edge.

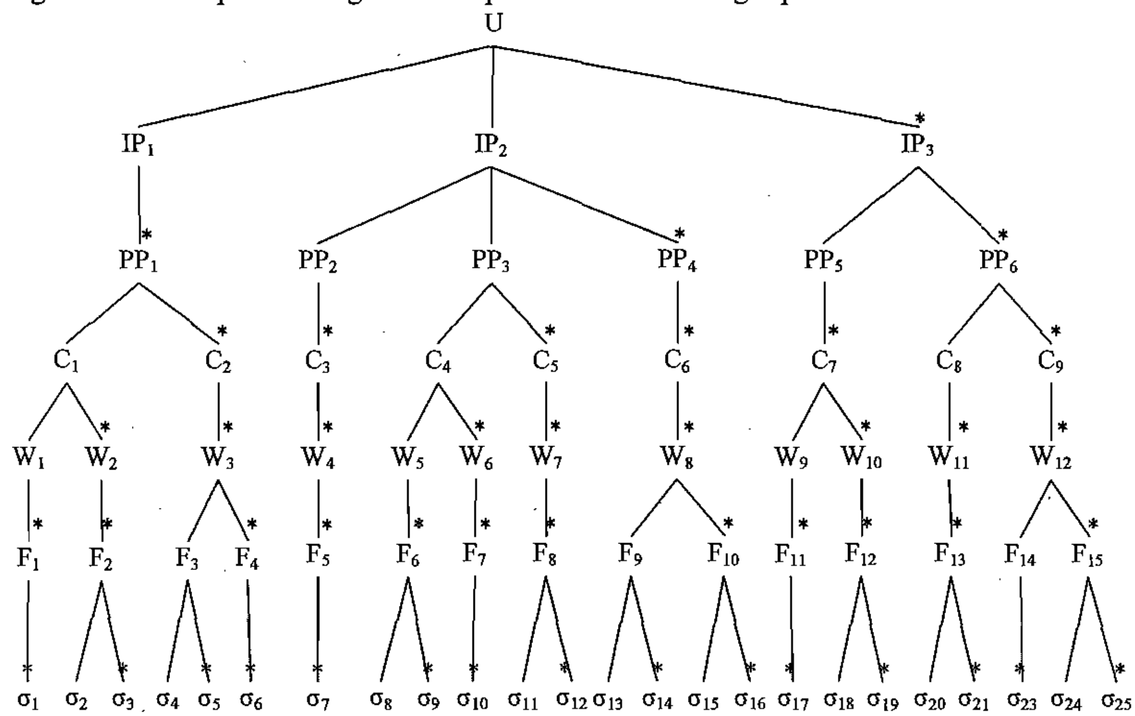
prominence within the word is assigned to the right-most foot in the word. The result, as shown, is penultimate stress.

Of the sources on sentence-level prosody referenced here, only Nespor & Vogel (1986) consider the prosodic hierarchy to extend all the way down to the level of the syllable. I have discussed word-level prominence and the formal apparatus most often used to describe it not only in deference to Nespor & Vogel's (1986) inclusion of these units in the prosodic hierarchy, but also because there is reason to believe that prominence at the phrase level and prominence at the word level share at least some properties. Perhaps most notable is that sentence level prominences seem to be subject to the Continuous Column Constraint—that is to say, a sentence-level prominence must fall on a word-level prominence. When a sentence-level prominence is placed on a syllable that does not receive main word stress in the neutral reading of the word, this unusual placement of sentence-level prominence is always accompanied by a shift of the main word-level prominence onto the same syllable. This alone requires that word-level prominence and phrase-level prominence be considered strongly related, if not formally identical.

2.1.2.2 Prosodic heads at the phrasal level

At the higher levels of phrasing—the clitic group, phonological phrase, and intonation phrase—the prominences assigned are pitch accents and particularly prominent pitch accents. The most prominent pitch accent in a phrase with multiple pitch accents is generally called the nuclear pitch accent (NPA). In many languages, prominence can be assigned in exactly the same way at higher levels that it is at lower levels—by a rule that assigns head status to a daughter that sits in a particular relation to one edge of the phrase. Such a system is illustrated in Figure 3 (schematic only) using the tree representation illustrated in Figure 1 and all seven of the categories in Table 1.

Figure 3—Example of assignment of prominence in a larger prosodic tree



U = Utterance, IP = Intonation phrase, PP = Phonological phrase, C = Clitic group, W = Word, F = Foot, σ = Syllable

Examination of Figure 3 tells us that prominence at every level has been assigned to the right-most daughter of each node. This makes σ_{25} the strongest element of the tree. Just as word-level prominence (stress) was built up cumulatively, so is prominence in the larger tree. In other words, σ_{25} is the strongest syllable in the entire Utterance: it is the strong syllable of the strong foot of the strong word of the strong Clitic Group of the strong Phonological Phrase of the strong Intonation Phrase of the Utterance—σ_{25} is strong at every level of the tree.

In some systems (e.g. that of Selkirk 1986), it is not required that every constituent in the tree have a head; in others, a prosodic tree can only be well-formed if every node in the tree has a single strong daughter. One such system is that of Nespor & Vogel (1986), which includes the requirement that "[t]he relative prominence relation defined for sister nodes is such that one node is assigned the value strong…and all the other nodes are assigned the value weak…" among the "principles that establish the geometry" (p. 7) of prosodic trees. Accordingly, Nespor & Vogel (1986) discuss the rule(s) assigning head status to a daughter at each level of the tree. Most of these rules make crucial reference to phrase edges.[8]

The accent domain proposals of Truckenbrodt (1999) and Büring & Gutiérrez-Bravo (2001) are similar to Nespor & Vogel's (1986) system in that they assign a single head to every phrase, using this mechanism to build up higher level prominences.[9] In these grammars, the tight relation between heads and phrases does not arise from a well-formedness condition on prosodic trees, but rather seems to be built in as an assumption

---

[8] Clitic group heads are discussed in much the same way as Word heads, even to the point of being identified as stress; the rules governing the location of Clitic Group stress are language-specific, but, given their similarity to word stress rules, almost certainly must make reference to phrase edges. Phonological phrase heads are assigned according to the principle

> In languages whose syntactic trees are right branching, the rightmost node of φ [the Phonological phrase] is labeled *s* [strong]; in languages whose syntactic trees are left branching, the leftmost node of φ is labeled *s*. All sister nodes of *s* are labeled *w* [weak].
> p. 168

While Nespor & Vogel's (1986) rule of Intonational Phrase Relative Prominence declares that the strong daughter of the Intonation phrase is assigned "on the basis of its semantic prominence" (p. 191) rather than with reference to a phrase edge, the rule they use to assign prominence within the phonological Utterance states that "[t]he rightmost node dominated by *U* is strong; all other nodes are weak" (p. 223). In other words, the only level at which the location of phrase edges is not crucial in determining the location of the phrasal head in this system is the Intonation phrase, where Nespor & Vogel see the effects of focus and other semantic properties of the utterance coming into play; even so, there will be a necessary relation of some sort between the head and the phrase in that there can only be one head per Intonation phrase, so that every pair of Intonation phrase heads (e.g. every pair of focused items within an utterance) must be separated by an Intonation phrase boundary according to this proposal.

[9] For Truckenbrodt (1999) and Büring & Gutiérrez-Bravo (2001), however, the semantic properties that Nespor & Vogel (1986) make reference to can be incorporated directly into the grammar that constructs the prosodic tree.

that prominence arises only from headedness, which is an inherently phrase-related notion. In this sense, the accent domain proposals operate in much the same way as foot-based stress systems: there is no stress outside of a foot or a phrase because, as discussed in section 2.1.1.3, there is no force in the grammar that would encourage the presence of a head without its domain.[10]

For Intonational Phonology systems, whether there is a single prominence within a phrase depends on the type of phrase. As mentioned above, Intonational Phonology assumes a metrical grid–type representation of prominence, complete with Continuous Column Constraint (see p. 23). In edge-marking languages—those with Accentual Phrases (AP)—it is often possible to have an AP without an accent, which can be seen clearly in Korean (Jun 1993), which has APs but no pitch accents at all. An AP with no pitch accent is also possible even in a language that does have pitch accents (e.g. Tokyo Japanese, Venditti 1995), but it is not possible to have multiple pitch accents within a single AP. Thus the AP can be considered the domain headed by pitch accent, with headless phrases sometimes permitted (which should be possible in the OT framework). The next level of phrasing—the intermediate phrase—contains multiple APs and thus multiple pitch accents, and the level above this is the intonation phrase; in some languages, only one phrase level above the AP exists. The head of a phrase with multiple APs is generally the leftmost AP, which is more prominent than other APs in its phrase.

In head-marking languages—those without APs—like English, the pitch accent is not the head of any domain; it seems to be an independent entity. There can still be

[10] Alternatively, the absence of a head without a phrase might be built into Gen.

multiple pitch accents in an intermediate phrase, and pitch accents still mark some words within an intermediate phrase as more prominent than others. The intermediate phrase still has its own head (the nuclear pitch accent, NPA), most often positioned at the right edge of the intermediate phrase. To the best of my knowledge, the Intonation Phrase is not believed to have its own head in this system (i.e. it is not claimed that the NPA at the end of an Intonation phrase containing multiple intermediate phrases is more prominent than earlier NPAs).[11] This state of affairs is summarized below, in Table 3.

Table 3—Types of prominence and the phrases they head (head marking languages only)

Prominence	*Head of...*
Secondary stress	Foot
Primary stress	Word
Pitch accent	(nothing)
NPA	ip

What is of particular interest about the fact that the pitch accent is not the head of any phrase in head-marking (non-AP) languages in Intonational Phonology analyses is that it is the Intonational Phonology systems that rely most heavily on phonetic marking of the prosodic entities they employ. In other words, while the Nespor & Vogel (1986) model of Prosodic Phonology and the models of Truckenbrodt (1999) and Büring & Gutiérrez-Bravo (2001) have a one-to-one relationship for heads and phrases, the Intonational Phonology systems for head-marking languages do not: there is no domain for the pitch accent to head, and the absence of such a phrase is based on the lack of phonetic evidence for its existence (beyond the presence of the pitch accents themselves).

[11] It is possible that the Intonation Phrase does have such a head; however, this topic will not be addressed here.

2.1.2.3 Summary

In the preceding sections, we have reviewed the representations that relate to the assignment of prominences, most often analyzed as the heads of prosodic domains. The relationship between word and above-word level prominences has been discussed, which led to a clearer view of the problem mentioned in section 0—namely, that there is a sort of gap in the prosodic phrase inventory of head-marking languages, which thus have a type of prominence (pitch accent) without a phrase.

2.1.3 Summary: Prosodic structure

We have briefly reviewed several types of proposals about prosodic structures, all of which share the same basic types of prosodic entities: domains and prominences. While both types of generative proposals (Prosodic Phonology in 2.1.1.1 and the accent domain proposals in 2.1.1.3, respectively) have a phrase for every head and, in some cases, a head for every phrase, the more data-driven Intonational Phonology systems do not preserve this relationship for head-marking languages, allowing the level of prominence labeled as pitch accent to occur on its own, rather than as the head of any domain. This is clearly contrary to the predictions of both generative approaches and, as such, suggests that further experimental research on the phrasal structure of head marking languages might be productive.

In order to conduct such research, it is necessary to understand both the predictions of the generative theories being tested and the types of phonetic evidence needed to test them. From the preceding sections, we can already see that the general prediction to be tested is that there is some domain, larger than a word and smaller than an intermediate phrase, of which pitch accent is the head, and that this domain appears in head-marking languages, like English; more specific assumptions are laid out in section 2.3. However, no background has yet been provided to the reader on the kinds of phonetic cues that have been found to mark prosodic structures (especially those used in Intonational Phonology systems). On the assumption that cues used to mark known constituents are likely to be used in the marking of other phrases, it is important to understand which acoustic properties have been found to mark prosodic entities. Accordingly, background on this subject is provided below.

2.2 Phonetic cues to prosodic entities

Like smaller phonological entities—segments and features—prosodic entities are phonetically realized. In many cases, there are clear and well-documented phonetic cues to the presence of one or both edges of a phrase and to the presence of a prosodic head. A summary of these cues is presented in sections 2.2.1 and 2.2.2, respectively.

2.2.1 Cues to prosodic boundaries

Unlike syntactic boundaries, which are generally believed to be without direct phonetic marking,[12] the presence of a prosodic boundary is marked by the presence of one or more phonetic cues. The higher the phrase is in the hierarchy, the more pronounced the phonetic cues that mark its edges are. Although there are cues that seem to be used to mark prosodic boundaries in many languages, the type and strength of the cues used can vary by language; the sections below include information on multiple

[12] Some syntactic entities, such as relative clauses, are often set off in the prosody and thus have phonetically marked boundaries. However, the marking in these cases is indirect—mediated by the prosody. In the majority of cases, syntactic phrases are not said to have phonetic edge marking. That is, we do not expect to assign or argue for syntactic structures on the basis of phonetic boundary cues, etc.

languages but focus most on English. There is marking associated with both the left and right edges of phrases, as well as indications of cohesion within a phrase whose absence across boundaries is also a cue to boundary presence and strength. Each cue receives its own short section below.

2.2.1.1 Final lengthening

Final lengthening is the name given to the phenomenon first described by Oller (1973), wherein the phonetic material preceding a boundary has a greater duration than it would have in the middle of a prosodic phrase. Wightman et al (1992) used a break index scale of 0–6 and found the degree of lengthening to distinguish four levels of phrasing in English: prosodic word (0–1), "a grouping of words within a larger unit" (2),[13] intermediate phrase (3), and Intonation Phrase or larger (4–6). In previous work on English phrasing and durational cues, Beckman and Edwards (1990) also found regular word-final lengthening (although the existence of word-final lengthening is challenged by Turk and Shattuck-Hufnagel 2000).

2.2.1.2 Initial strengthening

The realization of initial stop consonants has been found to be more extreme as the strength of the preceding boundary increases (Fougeron & Keating 1997, Fougeron 2001, Keating et al 2003 among others), a phenomenon called initial strengthening. For non-continuants, this means that the degree and duration of closure is greater for consonants at the beginning of larger domains than for consonants at the beginning of smaller domains. For voiceless stops, voice onset time (VOT) is longer at the beginning

[13] Wightman et. al. 1992:1710

of larger phrases in some languages. This type of result is reported for Korean in Keating et al (2003), in which the VOT of /t/ is described across languages as "not especially sensitive to prosodic position" across languages, as it varies "with prosodic position in Korean but not in French or Taiwanese," (p. 15).[14] Fougeron (2001: 110) reports that glottal opening "as approximated by VOT" shows strengthening effects in English based on the data of Pierrehumbert & Talkin (1992). While most of Pierrehumbert & Talkin's data concerns the sounds /h/ and /ʔ/, recordings made for this study allowed for the measurement of the VOT of English /t/ in four prosodic positions, and it is reported that VOT for English /t/ tends to be longer phrase-initially than phrase-medially. Assuming that lengthening the VOT of English /t/ will enhance the contrast between /t/ and /d/, this is consistent with the findings of Cho & Jun (2000) on Korean,[15] which suggest that VOT in Korean is used to enhance the paradigmatic contrast between the aspirated, lenis, and tense categories rather than the syntagmatic contrast between initial consonant and following vowel.

2.2.1.3 <u>Vowel-to-vowel coarticulation</u>

Cho (2004), looking at tongue body position in the production of American English /i/ and /ɑ/, found that the degree of vowel-to-vowel coarticulation diminishes as boundary strength between the two vowels increases. This lessening of coarticulation is specifically realized on vowels in prosodically strong locations (in accented syllables and

---

[14] Keating et al also show some initial strengthening data on English /n/ but do not present data on English /t/.

[15] Cho & Jun found that the VOT of Korean aspirated and lenis stops was longer when the stop appeared at the beginning of progressively larger prosodic domains, but that the VOT of the tense stop, which has the shortest VOT of the three stop categories in general, did not show this effect.

at domain boundaries), which show less influence from neighboring vowels; the effect of boundary strength is clearer for domain-initial position, where it is seen for both /i/ and /ɑ/ (p. 154).[16] This is consistent with the findings of Fougeron & Keating (1997), looking at extent of tongue contact with the roof of the mouth in American English /o/, who also found that the realization of the first vowel in a phrase is progressively more extreme (less contact) as the strength of the preceding phrase boundary increases.

2.2.1.4 Pitch

Pitch provides multiple types of cues to phrasing. These include pitch excursions at phrase edges, which are viewed in ToBI systems as the phonetic realizations of phonological edge tones. The type and complexity of the edge tones, as well as other properties of their realization, can mark differences in the sizes of the phrases so marked. Pitch range and reset are also used to mark domains. A particular size of domain may form the domain of downstep (the gradual lowering of high tone targets relative to previous high targets over the span of the phrase), as the intermediate phrase does in English. At the beginning of the next phrase of this size, the speaker's pitch range is reset. The particular prominence of the nuclear pitch accent, coming at the right edge of the intermediate phrase, might also be considered a cue to its end. There are also cases in which a prosodic phrase may be produced with its own lowered, reduced, or flattened pitch range, as is often the case with parentheticals, non-restrictive relative clauses, and tags.

[16] In domain-final position, an effect of boundary strength was observed for /□/ but not for /i/ (Cho 2005: 156).

2.2.2 Cues to prosodic heads

As discussed in section 2.1, there are two types of prosodic entities: phrases and heads. In the preceding section, we reviewed some of the phonetic cues to prosodic boundaries; this section gives a brief summary of the cues to prosodic head status, focusing, as before, on English. Before reviewing which phonetic cues are employed in marking prominence, it is worth noting that there are probably four levels of prosodic head in English (secondary word stress, primary word stress, pre-nuclear pitch accent, and nuclear pitch accent, as in Table 3, p. 28), but most of the studies mentioned here contrast only two levels: primary word stress and nuclear pitch accent. The exceptions are Beckman & Edwards (1994), who also include reduced (i.e. completely unstressed) vowels and de Jong (1995), who includes pre-nuclear pitch accent.[17] Because most of the studies cited here make only this two-way distinction in prominence, this section addresses what phonetic dimensions are generally used to mark prominence and, in some cases, how these cues are used differently to mark prominence at the word level and at the level of nuclear pitch accent. Unfortunately, the acoustic studies (those whose measures can be employed in the current study) make only the two-way distinction between unaccented and nuclear pitch accented, leaving us without a good understanding of the full range of acoustic cues that mark pre-nuclear pitch accent.

There are four cues to prosodic head status that are commonly listed in the literature: duration, pitch excursion, loudness, and an increase in the vowel formant space. There is considerable evidence (e.g. Beckman & Edwards 1994, Okobi 2006) that

[17] It seems likely that the choice to examine only the two more extreme cases is based on the difficulty of reliably eliciting pre-nuclear accented, which can be seen in de Jong's results, which show fewer pre-nuclear accented tokens than nuclear accented or unaccented tokens.

the heads of different prosodic constituents in the same language can be marked by different cues in the acoustic signal. In other words, not every level of prominence is marked by all four cues, although it is possible for a single cue to be used to mark prominence at more than one level. In addition, the cues may be given different weights—in general, at a particular level of the hierarchy, or in different languages. For example, vowel space expansion is often said to be the least important of the four cues listed above in general (e.g. de Jong 1995, Sluijter et al. 1997), and it is said by Sluijter & van Heuven (1996) to be of greater importance in English than in Dutch. It is worth noting that the importance of this cue is uncertain even in English—Sluijter & van Heuven find it to be a correlate of prominence in English and de Jong (1995) finds hyperarticulation under accent that he suggests may result in more extreme formant values; however, Okobi (2006) finds little evidence of such expansion under either pitch accent or lexical stress, with only F1 of the vowel [ɑ] showing any difference (p. 59).

Duration and pitch excursion, which are also used to mark phrase boundaries, are generally acknowledged to be more reliable cues to prominence than vowel space expansion. Duration is found to be a correlate of prominence in many studies (e.g. Beckman & Edwards 1994, de Jong 1995, Okobi 2006); the correlation between the two is positive, so that greater length indicates greater prominence. This seems at first to closely resemble the use of duration in marking phrase boundaries, where a greater degree of lengthening indicates greater boundary strength. However, results from de Jong (1995) show that lengthening under accent is not the same as lengthening to mark a boundary. In particular, de Jong finds that prominence is realized as localized hyperarticulation, with more extreme positions of the jaw and tongue, implying that the lengthening found under accent may be a result of the greater time taken to move the articulators to a more extreme position. This hyperarticulation is likely responsible for any difference in the vowel formant space in accented words and does not seem to be a property of phrase-final lengthening English. In acoustic studies of lengthening under accent such as those conducted on English by Turk and colleagues (Turk & Sawusch 1997, Turk & White 1999, Cambier-Langeveld & Turk 1999, Turk & Shattuck-Hufnagel 2000), the degree and span of lengthening under accent is examined; it is found to extend not only beyond the stressed syllable of the accented word, but even beyond the word itself. Given such spreading of accent marking over a domain larger than the syllable to which the phonological models discussed above would assign the formal marking of prominence, the potential of duration-related head marking and boundary marking (at least) to interact with each other is clear.

Earlier sources refer to pitch as a correlate of lexical stress (in stress languages) and this conclusion was considered solid enough to appear in textbooks (e.g. the results of Fry 1955, 1958 as described in Hayes 1995—see also for sources claiming correlation between pitch excursion and lexical stress). However, as noted by Beckman & Edwards (1994) and others, pitch is actually used to cue prominence only at the post-lexical level in stress languages; the earlier conclusions are based on experiments that confounded post-lexical prominence with lexical prominence. More recent work (see, e.g., Sluijter & van Heuven 1996, Heldner 2001, and Okobi 2006) shows that pitch excursion in stress languages marks pitch accent but not lexical stress. For some languages (e.g. Spanish and

Greek, as discussed in Jun 2005), the pitch accent inventory is small enough that it might be possible to get a sense of whether a word has received pitch accent using a raw measure of f0 or a simple transformation of same. In English, however, the pitch accent inventory allows for either a low or a high tone to be associated with the primary stressed syllable of a pitch accented word, so that we can only say that pitch excursion—not a particular shape in the pitch contour or a particular percentage increase in the pitch—marks post-lexical prominence in English.

Loudness can also be used to mark phrase-level prominence, just as it is one of the phonetic correlates of stress (Fry 1955, 1958 as characterized in Hayes 1995). Using overall intensity of the stressed vowel as the acoustic correlate of loudness gives the impression that loudness is not a very consistent cue to prominence (see, e.g., Heldner 2001), and the conventional wisdom on this point has been that loudness has "the least effect on stress perception, despite its intuitive status as the most natural correlate of stress" (Hayes 1995: 6). However, more recent work on loudness-related cues to prominence has found the intensity of noise in higher frequencies (often frequencies above 500 or 1000 Hz or above some transformation of the pitch) to be an effective acoustic cue to prominence that is related to loudness. A number of studies that have examined the efficacy of this type of acoustic cue (e.g. Sluijter and van Heuven 1996, Heldner 2001, Okobi 2006) have found that measures of the intensity of higher frequency energy does correlate with prominence on at least one level. Heldner (2001) finds this cue to be approximately as good as duration in cueing prominence. However, the work of Heldner (2001) and Sluijter & van Heuven (1996) was conducted on Dutch and Swedish, not on English, where results have been more mixed: Sluijter et al (1997) find that intensity in higher frequencies is a good cue to prominence, and Okobi (2006) finds that a measure of spectral tilt (H1* - A3*) to be a cue to English lexical stress. However, Iseli et al. (2003) found a correlation between another measure of spectral tilt and f0, and a relationship between f0 and pitch accented status, but no actual relation between tilt and pitch accented status. In other words, the question of whether loudness is a correlate of prominence, and at what level, seems to depend heavily on the exact nature of the experiment—perhaps what is recorded and almost certainly the fine details of what is measured, which varies considerably by study.

2.3 Assumptions made for the project

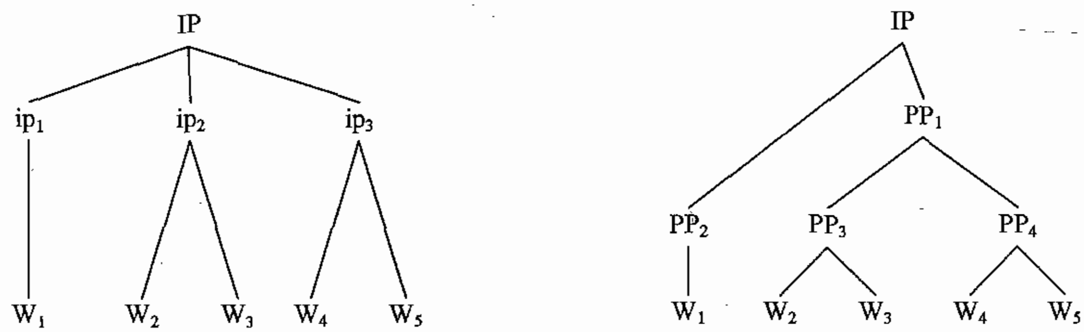
In what follows, I have assumed the prosodic categories employed in the Intonational Phonology model of English adopted in MAE_ToBI rather than the Prosodic Phonology inventory; I have also assumed, following Intonational Phonology, that the prosodic structure of utterances is best studied by examining actual output (that is, I have assumed an intonationally defined prosody rather than a syntactically defined prosody). In keeping with the mainstream view of prosodic structures—and thus the majority of work done on phonetic cues to prosodic entities—I have taken prosodic trees to be strictly layered (see section 2.3.1, below, for more on this). I have followed Prosodic Phonology and both of the accent domain proposals discussed above (Truckenbrodt 1999 and Büring & Gutiérrez-Bravo 2001) in maintaining that there are crucial relationships between syntax and prosodic phrasing, and between pitch accent placement and prosodic phrasing. Also following the accent domain proposals, I have started from the premise that only material dominated by the maximal projection of a single lexical head can form a single accent domain; this would crucially make the boundary between the subject of a sentence and the verb phrase special, in that the subject-verb boundary in the syntax should always correspond to at least an accent domain boundary. This last assumption was crucial in constructing experimental items; however, it is not transparently consistent with the previously mentioned assumption that prosody is defined on the surface. It is important to remember that the syntactic structures used in the construction of experimental items are assumed to permit—not to guarantee—the desired prosodic structures. As discussed below, the tokens produced by speakers were screened by accent pattern as actually produced, so that the ultimate criterion for whether an utterance constituted an example of the desired type is consistent with an intonationally defined (or surface-defined) prosody rather than a syntax-defined prosody.

2.3.1 A brief excursus on recursive phrasing

One property of Truckenbrodt's (1999) proposal that we have not yet discussed is its use of recursive phonological phrasing, which is necessary for one of his case studies (Kimatuumbi). This use of recursive structure mirrors in some ways earlier work in prosodic phonology (e.g. Selkirk 1980), in which the structure of prosodic trees was closer to that of syntactic trees, being both recursive and more prone to binary branching. Recursive prosodic structure is also employed in Jacobs (1991 / 1992) and argued for at the level of the IP in Ladd (1986). The differences between strictly layered and recursive prosodic structures are illustrated below, in Figure 4.

Figure 4—Strictly layered versus recursive prosodic architectures (schematic)[18]

a. Strictly layered structure (ToBI labeled) b. Recursive structure



For present purposes, the important difference between these two types of systems is this: in a strictly layered system, to be at the end of a higher phrase is always to be at the end of every type of phrase that can come below this; it is also always to be at the end of each such phrase type once and only once. Thus, a phonetic cue to being at the end of an Intonation Phrase is actually a cue to being at the end of a word, intermediate phrase, and Intonation Phrase simultaneously, while a cue to being at the end of an intermediate phrase is a cue to being at the end of a word and an intermediate phrase, but not necessarily at the end of an Intonation Phrase.

In recursive systems like that of Truckenbrodt, because it is possible to have recursive phrasing at a level intermediate between word and Intonation Phrase, it is not clear that there should be a single set of phonetic cues that mark material at the edge of this level of phrasing. In other words, W_1 and W_3 are in similar positions with respect to right phrasal edges in Figure 4a but not in Figure 4b. The use of strictly layered prosodic trees is pervasive enough that most (perhaps all) research into the phonetic correlates of prosodic structure assumes that the prosodic structures in question are non-recursive, thus

[18] The labels in the recursive structure are from Prosodic Phonology: IP = Intonation phrase and PP = Phonological phrase.

there are known phonetic cues for the presence of boundaries in strictly layered prosodic systems (see section 2.2.1), but our knowledge of such cues for recursive prosodic trees is considerably less developed. This makes direct phonetic evaluation[19] of the full system proposed by Truckenbrodt difficult—and such an evaluation is beyond the scope of the proposed work.

However, Truckenbrodt's theory makes interesting predictions about which syntactic elements should form a single prosodic phrase containing only one pitch accent, which will be of great value in designing the experiments described in chapters 3 and 4. These predictions are also made by the proposal of Büring & Gutiérrez-Bravo (2001), who do not employ (or discuss) recursive formation of phonological phrases. This project assumes non-recursive formation of accent domains (phonological phrases in the terms of Truckenbrodt and Büring & Gutiérrez-Bravo), which is possible but not necessary for the former and entirely consistent with the latter.

2.3.2 Predictions of accent domain proposals

Using accent domains means, by definition, that there is a level of structure between the word and the intermediate phrase formed in the prosody. Based on what we know about languages that clearly use such domains and what we know about other prosodic domains in English, we would expect to see some marking of not only the heads of these domains (pitch accents), but also their edges. The phonetic properties used to mark such edges are discussed in section 2.2.1; the most important of these cues in the experiments discussed below will be final lengthening (section 2.2.1.1, p. 31).

[19] Assuming that such a direct evaluation would involve taking phonetic measurements to determine the locations of phrase boundaries.

Based on the nature of accent domains (that is, the fact that they are prosodic domains headed by pitch accent) and on the way that they are said to relate to the syntax in the accent domain proposals reviewed in section 2.1.1.3, there are two places that we expect to see accent domain junctures:

- Between any two pitch accented words; and,
- Between two syntactic constituents that are not dominated by the maximal projection of a single lexical head.

The two experiments presented in Chapters 3 and 4 exploit both of these factors: pitch accent location is varied across experiment and syntactic structure is varied within experiment. The data collected (described in considerably more detail in Chapters 3 and 4) are designed to trigger the presence of word, accent domain, and intermediate phrase boundaries if the accent domain proposals are correct. The boundary locations are surrounded by identical segmental strings to allow measurement of fine phonetic detail surrounding them.

Both experiments make use of the fact that the juncture between the subject and the main verb of a sentence is a juncture between two syntactic constituents that are not dominated by the maximal projection of a single lexical head, and should therefore always trigger an accent domain boundary, regardless of whether the words around the boundary are pitch accented or not. The juncture between the subject and main verb of a sentence is compared to the juncture between an adjective and the noun it modifies, a construction in which the two words are (in many syntactic analyses) dominated by the maximal projection of a single lexical head, and should therefore be capable of belonging to a single accent domain when only one of the words is pitch accented.[20] A third type of juncture, between the end of a restrictive relative clause modifying a subject and the main verb of the sentence, was also included in the data collected. This construction was selected because it tends to induce production of an intermediate phrase boundary and therefore provided a point of comparison for the noun-verb juncture—a way to make sure that the noun-verb boundary was not being produced with an intermediate phrase break between the two.

If the accent domain proposals are correct, we would expect the noun-verb boundary and the adjective-noun boundary to be different when only one of the words is pitch accented, with the noun-verb boundary showing evidence of a prosodic break but the adjective-noun boundary showing no evidence of a break (see Experiment 1, Chapter 3). When both words are pitch accented, however, we would expect evidence of an accent domain break at both the noun-verb and adjective-noun boundaries (see Chapter 4). Furthermore, upon comparing the two experiments, we would expect the boundary between adjective and noun with only one of the surrounding words pitch accented to look unlike the other three boundaries, because this should be the only case in which there is no accent domain boundary present in the span being examined.

Looking at the pitch accent status of the two words surrounding the boundary (immediately before and immediately after), we can make specific predictions about how strong we expect the boundaries between these two words to be. Using the abbreviations

[20] Syntactic proposals that do not analyze adjective-noun sequences in this way should still preserve the observation that an adjective and the noun it modifies are more closely linked—in both structural and informational terms—than the sentence of a subject and its main verb. It is this difference in degree of connection that is most critical here.

Adj-N (adjective-noun boundary), NP-VP (noun-main verb boundary), and RC-VP (relative clause end-main verb boundary), the accent domain proposals predict the following in terms of degree of juncture:

- One word pitch accented, one word not pitch accented:
  Adj-N < NP-VP < RC-VP

- Both words pitch accented:
  Adj-N, NP-VP < RC-VP

The next two chapters present two experiments conducted to test the predictions about boundary strength shown above. Both are production experiments that rely on audio recordings of native speakers reading aloud. While experiments of this type certainly have shortcomings (for example, those described in Snedecker & Trueswell 2003), the items that speakers had to be induced to utter in order to test for phonetic marking of the edges of a prosodic unit between word and intermediate phrase are probably unobtainable without using a script (to see an example of what these items look like, see section 3.1.3, where the Experiment 1 script items are described).

As noted above, both pitch accent location (across experiment) and syntactic structure (within experiment) were varied in the script items; both of these factors presented their own challenges. As discussed in section 3.1.3, varying the syntactic structure of the region of interest required the use of items that were quite different outside the region of interest. Varying the pitch accent pattern proved to be even more difficult: the common method of moving pitch accents around in a sentence using contrastive or corrective focus was entirely unsuccessful in pilots of Experiment 1[21] because the presence of focus caused so much lengthening on the pitch accented item that any more subtle differences in lengthening that might have been due to the presence of a boundary smaller than an intermediate phrase were completely masked. Thus, a more subtle manipulation of pitch accent placement was attempted in Experiment 1; no manipulation of this sort was used in Experiment 2.

The goal of these manipulations was, of course, to produce script items that would trigger the production of three different sizes of break according to a grammar with the properties described above—word, accent domain, and intermediate phrase—while keeping the segmental material around the boundaries identical. These predictions about boundary strength are translated into specific predictions about phonetic properties laid out in the introduction to each experiment.

[21] In fact, this means of varying pitch accent placement was also attempted separately in another experiment intended as a follow-up to Experiments 1 and 2, which was to have kept the syntactic structure constant and varied only the pitch accent placement (thus allowing for the use of syntactic structures that could not be incorporated into a design like that of the experiments reported here). However, this experiment was eventually abandoned due to the same problems encountered with the use of focus in the pilot of Experiment 1.

Chapter 3: Experiment 1

Two experiments were conducted to test the predictions of a grammar for English prosody that has the properties laid out in section 2.3; the first of these is presented in this chapter and the second in the following chapter. This experiment is designed to establish whether there is evidence of a boundary that is intermediate in size between a word boundary and an intermediate phrase boundary. The design of the experiment assumes that the location of accent domain (AD) boundaries is related to syntactic structure; for this reason, finding evidence of a boundary of the appropriate size is indicative not only of its existence, but of its relation to the syntactic structures used. This experiment tests two specific hypotheses:

1. The degree of final lengthening will distinguish the AD from the word and from intermediate phrase, with the least lengthening for the word and the most for the intermediate phrase.
2. The degree of initial strengthening will distinguish the AD from the word and the ip, with the least strengthening for the word and the most for the intermediate phrase.

The following sections report in detail how Experiment 1 was designed and carried out, as well as describing its results. An interpretation of the results of Experiment 1, taken in combination with those of Experiment 2, is presented in Chapter 5.

3.1 Methods

3.1.1 Speaker population

All speakers recorded for this experiment were native speakers of Mainstream American English; native speaker status was determined by the experimenter after a brief conversation about each speaker's language background. Speakers ranged in age from early twenties to early thirties. Most had studied linguistics but none had studied prosody. Nine speakers were originally recorded (six female, three male) but three (one female, two male) were eliminated because large numbers of their tokens were unusable. A token was considered unusable if it was disfluent during the target portion of the utterance, markedly different from the sentence in the script, or produced with a prosodic pattern dramatically different from the one desired.

3.1.2 Recording procedure

Script items were presented to speakers in printed booklets, with one item per page. Each booklet began with two filler items, followed by a sequence of alternating target and filler items, and ended with a filler; each booklet gave the items in a different pseudo-random order. Speakers were digitally recorded in the sound-attenuated booth at the UCLA Phonetics Laboratory. The recordings were saved in WAV format, sampled at 22 kHz and quantized at 16 bits. Once recorded, the sound files were segmented and annotated using Praat (Boersma & Weenink 2005).

3.1.3 Data collected

As discussed above, it was necessary to construct sets of sentences that would contain varying degrees of prosodic break in otherwise homophonous strings. In each set, there was a sequence of two words (W_1 and W_2 in what follows) that were held constant across three different syntactic constructions. These constructions, shown in Table 4, were chosen to induce different degrees of prosodic break: the Adj-N construction, in which both the adjective and the noun are within the maximal projection of the noun, was hypothesized to have no prosodic break between W_1 and W_2. The NP-VP construction was hypothesized to have an accent domain boundary between W_1 and W_2, as the subject NP of a sentence and the VP are not contained within the maximal projection of a lexical head. The RC-VP construction was intended to cause speakers to produce an intermediate phrase (ip) break between W_1 and W_2; this was based on observation rather than on the prediction of any particular theory of the syntax-prosody mapping.

Table 4—Target sentence types, Experiment 1[22]

Name	*Schematic*	*Break*	*Example*
1) Adj-N	$\ldots W_1 \quad W_2]_{NP} \ldots$	Word	For almost all swindles, the natural targets are the greediest of the novice investors.
2) NP-VP	$\ldots W_1]_{NP} W_2 \ldots$	AD	In almost every case, a natural targets the greediest new investors for his stock scams.
3) RC-VP	$\ldots W_1]_{RC} \quad W_2 \ldots$	ip	A con artist who was a natural targeted the greediest investors for his stock scam.

In all sets, W_1 is a Noun-Adjective homophone with three syllables, antepenultimate stress, and a final voiced continuant segment (*fugitive, natural, radical*); W_1 is also always capable of being an agent in its noun reading. The length and stress pattern of W_1 were selected to keep the lengthening associated with stress as far from the right word edge—the area of greatest interest—as possible. The voicing of the final segment helped to ensure a good segmentation point between W_1 and W_2. The use of only those Noun-Adjectives homophones that were capable of being agents in their nominal uses was required by the words that worked well in W_2 position. W_2 is a Noun-Verb homophone with two syllables, initial stress, and an initial voiceless stop (*pilot, target, partner*). It was necessary to use only words with initial stress in this position to ensure a fully voiced vowel in the first syllable for formant measurement. The initial voiceless stops allowed for the measurement of known acoustic correlates of initial strengthening; in

[22] W_1, W_2 = word 1, 2 in the string of words that remained identical across the sentences in a set. RC = restrictive relative clause.

combination with the restrictions on the end of W_1, this restriction on segmental form also helped create a good segmentation point between the two words. The end of W_2 differed some across conditions because it was necessary to inflect W_2 when it appeared as a verb. The number of syllables in the sentence and the position of the target in the sentence were controlled both within and across sets: each sentence contained 24 syllables and the target sequence began with the eighth syllable in the sentence.

Constructing the target sentences in this way provided sentences in which phonologically identical words (W_1 and W_2) were separated in one case by no boundary, in one case by an AP boundary, and in one case by an ip boundary. Homophones were used to keep both the segmental material and the word boundaries around the (expected) prosodic breaks constant, thus avoiding the potential confounds of inherently varying segment lengths and word-final lengthening (as found by Beckman & Edwards 1990).

Because the experiment was intended to diagnose the presence of accent domains—not just a level of phrase with final lengthening between word and ip—it was necessary to control the placement of accents on W_1 and W_2. In the NP-VP and RC-VP constructions, the there is already some tendency to put no accent on W_2 (the verb in these sentences); in the Adj-N construction, the default is for both W_1 and W_2 (adjective and noun) to have separate accents. If the relevant domain is truly an AD (with at most one accent per AD), the default accent pattern of the Adj-N construction would not result in both W_1 and W_2 phrased in a single accent domain. In order to prevent a pitch accent from appearing on W_2, each target sentence was preceded by a context sentence that contained W_2 (or a morphologically related word) but not W_1, so that W_2 would be given

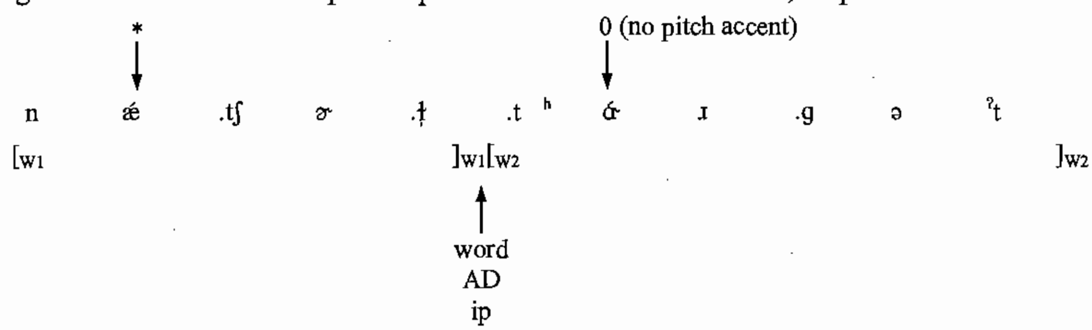
and thus deaccented.[23] This resulted in script items made up of pairs of sentences, as shown in Table 5 (W_1 and W_2 underlined in both context and target):

Table 5—Examples of context-target sentence pairs, Experiment 1

Name	*Example*
1) Adj-N	Picking the right targets can make all the difference in whether a stock scam is successful. For almost all swindles, the natural targets are the greediest of the novice investors.
2) NP-VP	Who a con artist targets can make all the difference in whether a swindle is successful. In almost every case, a natural targets the greediest new investors for his stock scams.
3) RC-VP	Targeting the right person can make all the difference in whether a swindle is successful. A con artist who was a natural targeted the greediest investors for his stock scam.

The expected locations of pitch accents and prosodic domain boundaries in these utterances are marked in Figure 5, below.

Figure 5—Locations of expected pitch accents and boundaries, Experiment 1

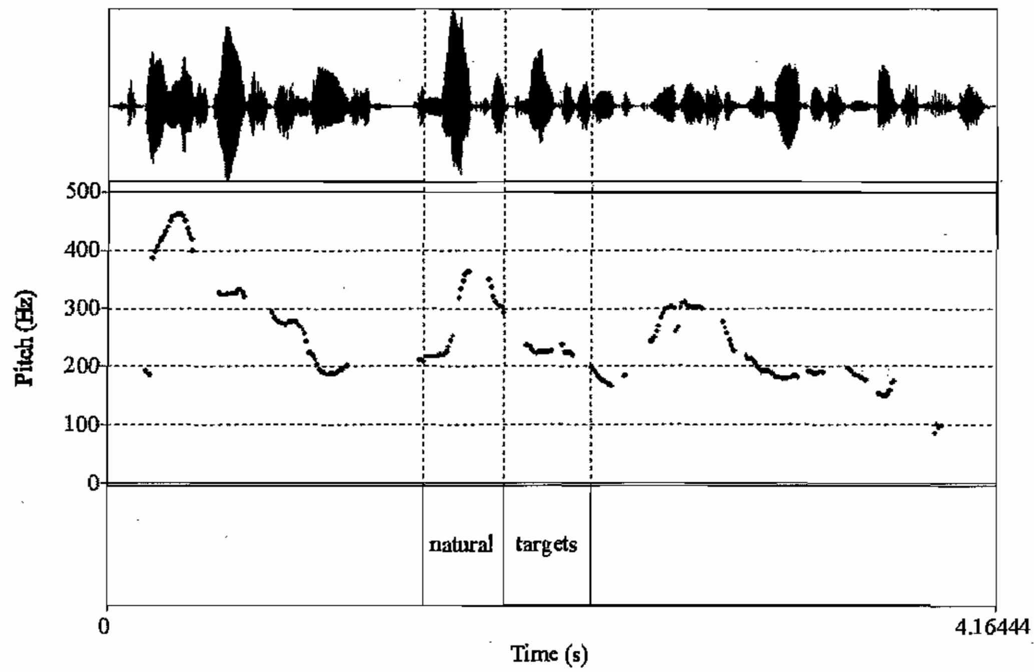


An example of an utterance representing each construction is shown in the figures below. These examples—utterances of the target sentences in Table 5 (target sequence annotated in Figure 5 marked in the examples)—are all from the same female speaker. This speaker seems to have used L+H* accents on *natural* in all three cases, with the L tone on the preceding determiner continuing into the [n] of *natural* and followed by a sharp rise to the H* tone during the stressed syllable. In the Adj-N example, Figure 6, the pitch after the peak on *natural* seems to fall fairly evenly to the leading L of another

[23] Many thanks to Daniel Büring and Sun-Ah Jun for this excellent suggestion.

L+H* on *greediest;* this, along with downstep on the H of *greediest,* suggest that *natural* and *greediest* are part of a single intermediate phrase.[24]

Figure 6—Example utterance of an Adj-N target sentence

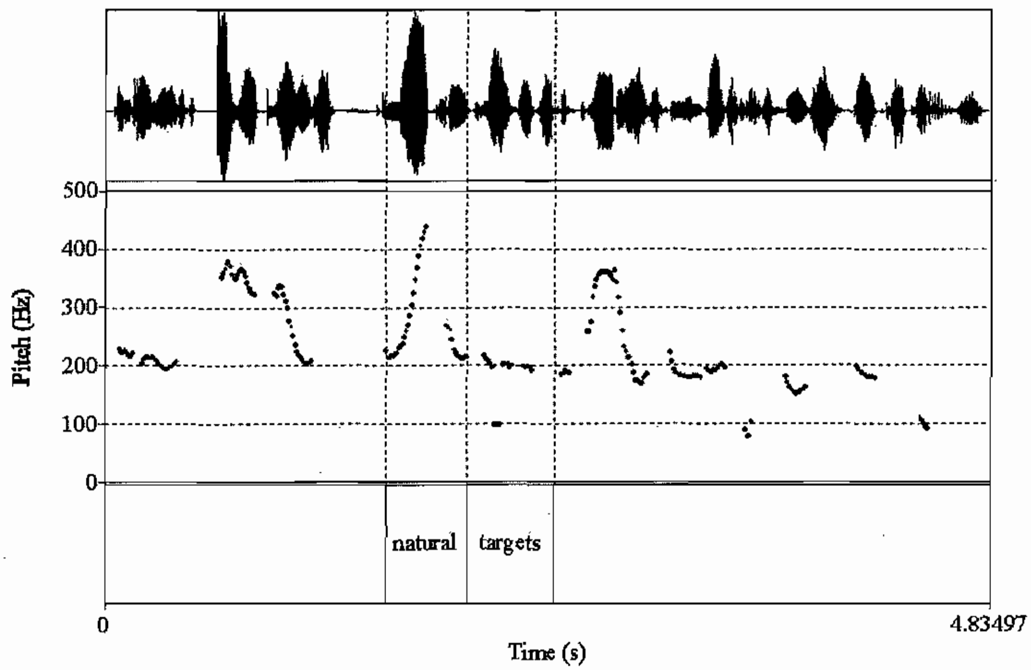


The deaccenting on the verbs in the NP-VP and RC-VP conditions was effected by the use of a morphologically related word in the context sentence (the underlined word in the first sentence of the examples in Table 5). The forms of *target* used in the NP-VP context and target sentences are identical (*targets* in both) in the example shown here, although this was not true for the other script items. In the RC-VP example, the forms used had different inflectional endings (*targeting* in the context sentence and *targeted* in the target sentence), which was more typical. In Figure 7, *natural* again has a L+H*

[24] An alternative transcription for this portion of the sentence would employ the same pattern of pitch accents on *natural* and *greediest,* but with an ip break with a low (L-) phrase tone marking its right edge. While this transcription explains the rapidity of the fall from the H* of *natural,* it cannot explain the mid (not low) pitch of *targets*. In either case, the sequence *natural targets* is within the same intermediate phrase.

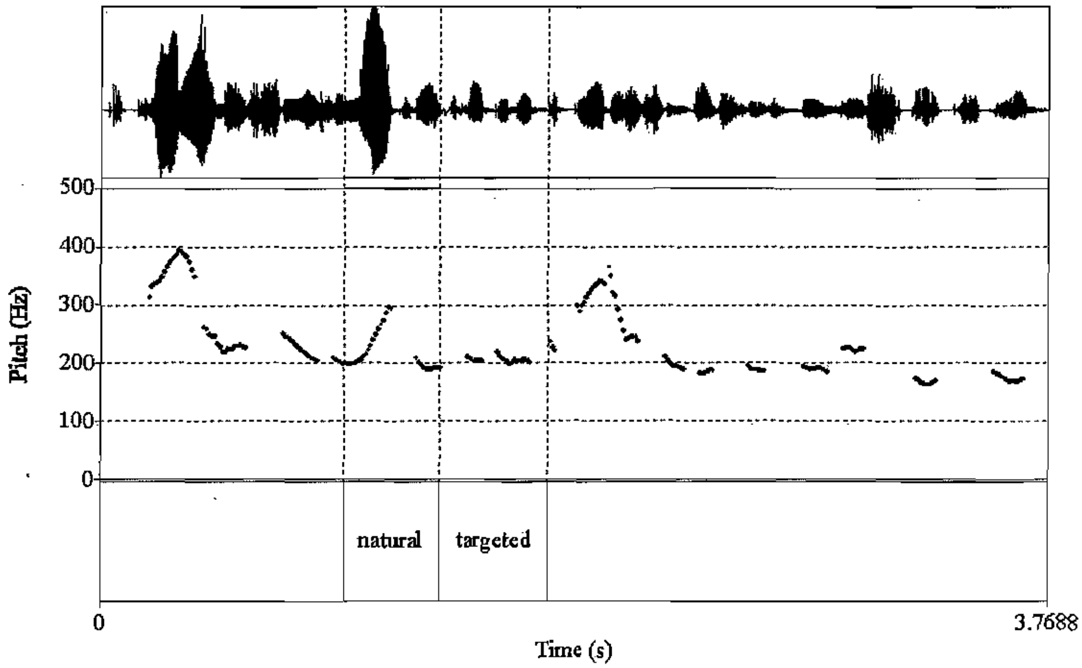
accent followed by a clear low plateau that seems to result from the leftward spreading of a L- marking an intermediate phrase boundary at the end of *targets*, which is then followed by another L+H* on *greediest.*

Figure 7—Example utterance of a NP-VP target sentence



The differences in amplitude between the first vowel of *natural* and the first vowel of *target,* both of which are low vowels, can be seen in the waveforms of all three examples. These differences are secondary to the pitch in the Adj-N and NP-VP examples, but are more important in looking at the RC-VP example in Figure 8, where the lowered and reduced pitch range of the relative clause ending in *natural* make the pitch peak on the pitch-accented *natural* less dramatic than in the other examples. Here, the sharp drop from the high of the L+H* on *natural* is due to a low phrase tone (L-) at the end of *natural.* The pitch on *targeted* remains relatively low (although it is discernibly higher than at the end of *natural),* probably in anticipation of the leading low of the L+H* that appears on *greediest.*

Figure 8—Example utterance of a RC-VP target sentence



3.1.4 Measurement and annotation procedures

The prosody of the target portions of the recorded utterances was transcribed. The recordings were screened according to these transcriptions, and acoustic measures were taken from tokens that passed the screening criteria. The sections below describe in more detail the transcription, screening, and measurement procedures.

3.1.4.1 Prosodic coding and screening

The two-word target sequence of each recorded utterance was annotated for the presence or absence of pitch accent and for the break strength and tone type between the two target words according to MAE_ToBI labels by two trained transcribers. The author (hereafter T1 in this capacity) coded only the target sequence. The second ToBI-trained labeler (T2) coded a larger section of each utterance; the portion of each script item coded by T2 is shown in Table 6.

Table 6—Portion of each script item coded by T2 (second transcriber)

Fugitive	
Adj-N	fugitive pilot flew his plane
NP-VP	fugitive piloted a plane
RC-VP	fugitive piloted a plane
Natural	
Adj-N	natural target is the greediest of the novice investors
NP-VP	natural targets the greediest new investors
RC-VP	natural targeted the greediest investors
Radical	
Adj-N	radical partner supported the responsible purchase
NP-VP	radical partnered with the lead conservative
RC-VP	radical partnered with the lead conservative

Tokens that did not meet the following criteria according to either transcriber were removed from the data set: W_1 pitch accented, W_2 not pitch accented in the Adj-N and NP-VP conditions,[25] and break strength of 1 for the Adj-N and NP-VP conditions and of 3 for the RC-VP condition.[26] Because this produces an uneven number of tokens for each script item, quantitative measures from all the repetitions of a single item by a single speaker have been averaged to produce a single value of each measure for that item and that speaker. These are the figures used in descriptive and predictive statistics unless otherwise noted.

3.1.4.2 Measurement procedures

From utterances of the sort schematized in Figure 5 (examples in Figure 6–Figure 8), the measurements shown schematically in Figure 9 were taken; following Figure 9, Figure 10 shows an actual token as segmented for measurement and the definitions for each labeled measurement are given in Table 7.

[25] The original intent was to enforce this screening criterion for the RC-VP tokens, as well. However, speakers were very resistant to deaccenting W_2 when it was initial in an intermediate phrase (see section 3.2.1). Because it seemed unfeasible to do otherwise, the criterion was dropped for this condition.

[26] In some cases, a speaker did not produce any usable tokens of a given script item. All data from these speakers has been removed from the set.

Figure 9—Measurements taken from tokens in Experiment 1, schematic

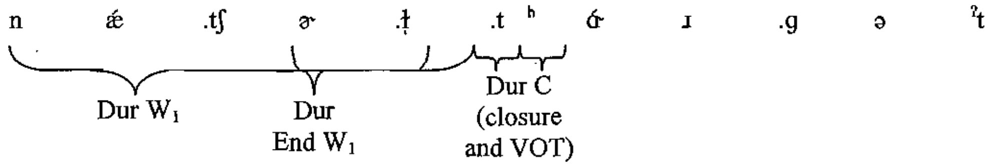


Figure 10—A token of *natural targets* in the Adj-N condition, segmented for measurement

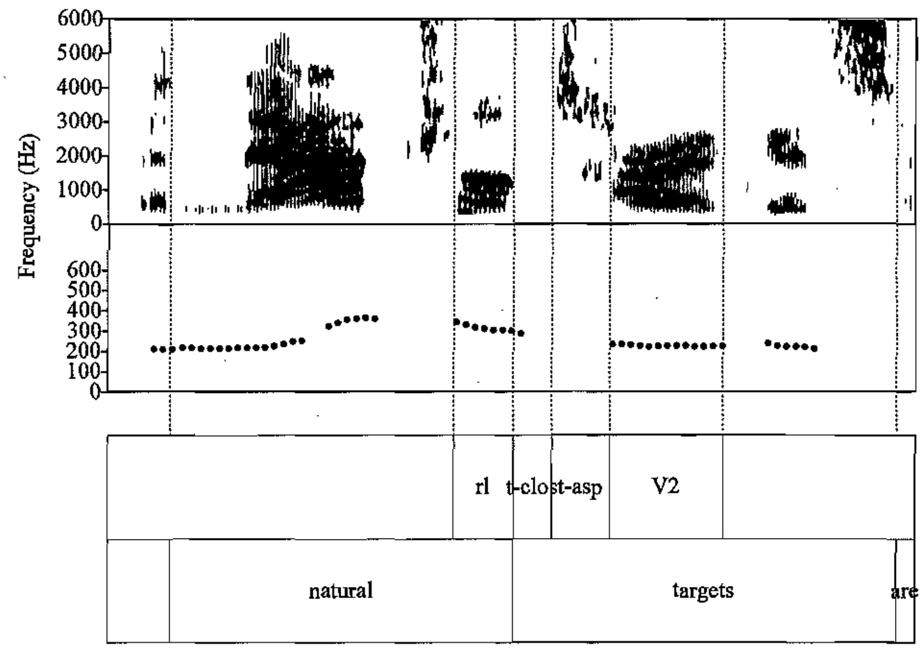


Table 7—Definition of measurements

Measurement	*Ex. in Figure 10*	*Description*
Dur W_1	natural	Duration of the entire W_1
Dur End W_1	rl	Duration of the part of W_1 starting with the last reliable segmentation point and ending at the end of W_1 (same point as the end point used for Dur W_1)
Dur C (clos)	t-clos	Duration of the closed portion of the initial consonant in W_2. For some tokens of *target*, there was frication rather than closure; in this case, the duration of the frication was measured
Dur C (VOT)	t-asp	Time between the release of closure for the initial consonant in W_2 and the start of the following vowel (based on start of voicing and start of the vowel's F2)

3.2 Results

The results of the experiment are divided into transcription-related results (labeler agreement only, for this experiment) and the acoustic measures taken.

3.2.1 Transcription (agreement)

Overall labeler agreement for the two-word target portion is 82%. Agreement is notably lower (57%) for the break index between W_1 and W_2 in the NP-VP group. This is to be expected if these tokens tend to fall somewhere in between prototypical MAE_ToBI breaks 1 and 3. In keeping with previous comparisons of the coding styles of T1 and T2, T2 tended to hear more and larger breaks than T1 when there was disagreement.

3.2.2 Acoustic measures

The acoustic measures reported below are divided into two sections: initial strengthening measures (the duration measures for the first consonant of W_2) and final lengthening measures (the duration measures from W_1).

3.2.2.1 Initial strengthening measures

Closure and aspiration duration measures from tbe first consonant of W_2 (C_2) were taken separately; the combination of the two was also examined, and it is this measure (total duration of the first consonant of W_2) that is shown in Figure 11, below. Note that what holds true for the total duration of C_2 also holds true of its two component measures.

Figure 11—Duration of C_2 in ms separated by word and overall

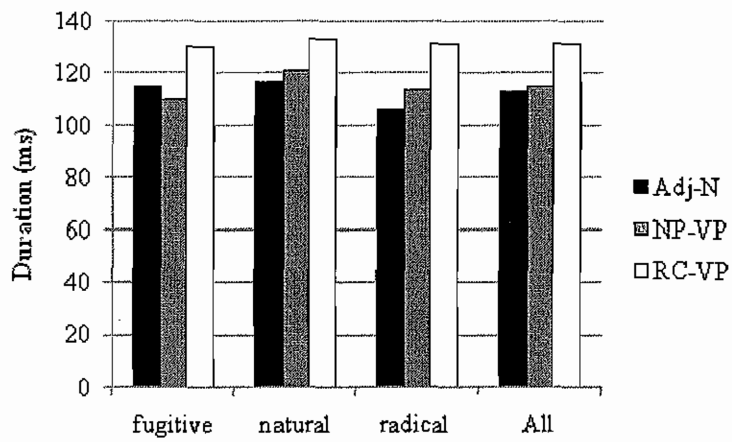


Figure 11 suggests that all three sets of items behaved similarly, with a difference in C_2 duration between the RC-VP group and the other two, but not between the Adj-N and NP-VP groups. This is supported by the results of a two-way RM-ANOVA with factors construction (three levels: Adj-N, NP-VP, RC-VP) and lexical item (three levels: *fugitive, natural, radical*), which shows an effect of construction but no effect of lexical item and no interaction between the two factors (see Table 8).

Table 8—Results of RM-ANOVA on duration of C_2 (* indicates significance, $\alpha = 0.05$)

Source	Result	
construction (const)	$F(2.0, 10.0) = 16.9, p < 0.01$	*
lexical item	$F(1.3, 6.9) = 0.8, p = 0.4$	
const * lexical item	$F(4.0, 20.0) = 0.7, p = 0.6$	

Paired t-tests with Bonferroni correction were also carried out, one comparing the Adj-N and NP-VP groups and one comparing the NP-VP and RC-VP groups (results in Table 9). These show a two-way distinction in initial strengthening, between the RC-VP group and the other two (a difference of about 17 ms from the NP-VP group overall), but no distinction between the Adj-N and NP-VP groups (about 2 ms difference overall).

Table 9—Results of paired t-tests on duration of C_2

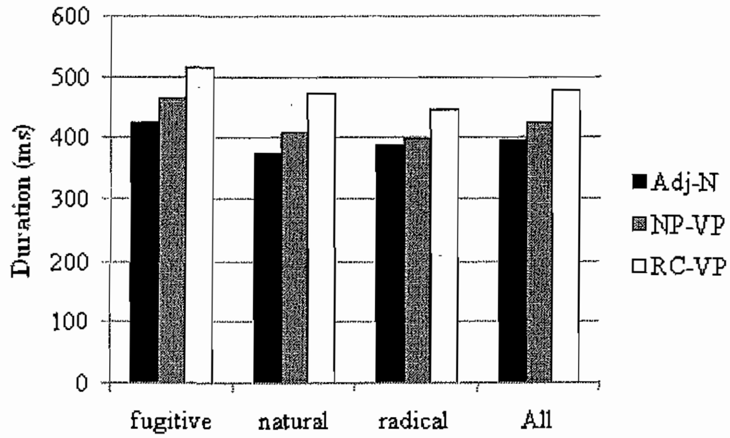
Comparison	Result	
Adj-N vs. NP-VP	$t(5) = 0.6, p = 0.6$	
NP-VP vs. RC-VP	$t(5) = 6.1, p < 0.01$	*

Because the RC-VP tokens were not screened for pitch accent status of W_2, the difference between the NP-VP and RC-VP tokens cannot be clearly attributed either to pure domain-initial strengthening (as in Pierrehumbert & Talkin 1992, Keating et al 2003) or to the sort of increase of VOT under pitch accent found by Cole et al (2003). The essential finding from this measure is the absence of any evidence of initial strengthening to cue the beginning of an AD.

3.2.2.2 <u>Final lengthening measures</u>

The use of ADs in the phonological representation predicts that final lengthening will distinguish three levels of boundary—word, accent domain, and intermediate phrase—and that the strength of the break between W_1 and W_2 will be crucially influenced by syntactic construction. Therefore, the prediction is that the duration measurements taken from W_1 will show the pattern Adj-N < NP-VP < RC-VP. As shown in Figure 12, this prediction is correct for the measurements of the full duration of W_1. This is true for each individual word and for all of the tokens combined.

Figure 12—Duration of W_1 in ms separated by word and overall



The same pattern is shown by the duration of the end of W_1, both in milliseconds in Figure 13 and as a proportion of the entire word in Figure 14.

Figure 13—Duration of EndW_1 in ms separated by word and overall

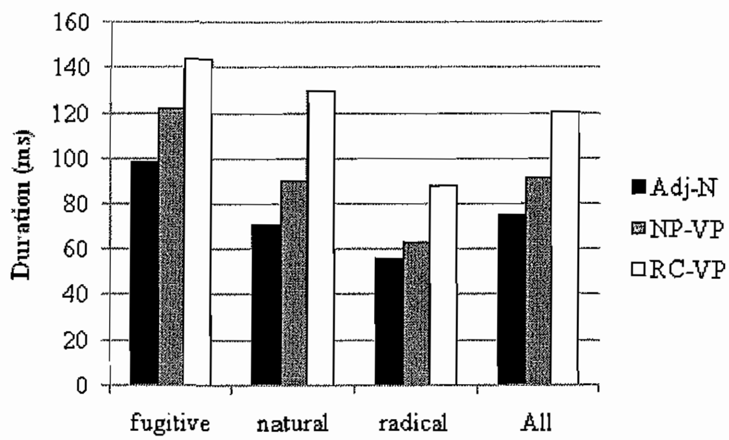
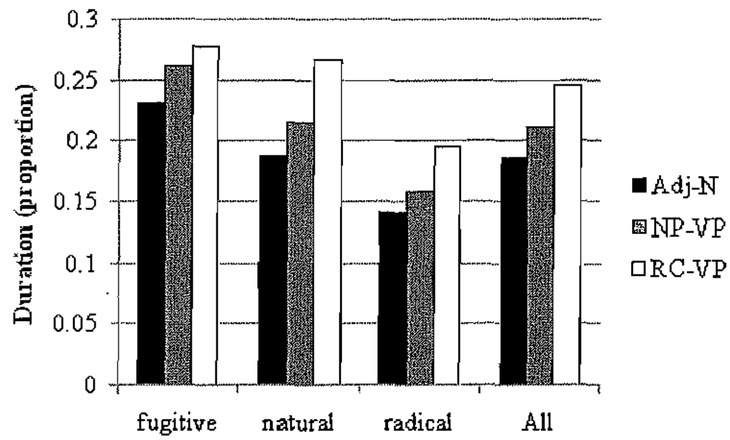


Figure 14—Duration of EndW_1 as a proportion of W_1 separated by word and overall



A two-factor repeated measures ANOVA with Huynh-Feldt correction was run on these measurements, with factors construction (three levels: Adj-N, NP-VP, RC-VP) and lexical item (three levels: fugitive, natural, radical). These tests show main effects of both syntactic construction and lexical item for both sets of measurements, as shown in Table 10 (entire W_1) and Table 11 (end of W_1 in ms). For the measurements of the full W_1, there is also a significant interaction between the two factors.

Table 10—Results of RM-ANOVA on duration of W_1 (* indicates significance, $\alpha = 0.05$)

Source	Result	
construction (const)	$F(1.8, 8.9) = 53.8, p < 0.01$	*
lexical item	$F(1.4, 6.9) = 21.4, p < 0.01$	*
const * lexical item	$F(3.5, 17.4) = 3.8, p < 0.05$	*

Table 11—Results of RM-ANOVA on duration of the end of W_1 in ms

Source	Result	
construction (const)	$F(1.2, 5.8) = 25.7, p < 0.01$	*
lexical item	$F(2.0, 10.0) = 29.3, p < 0.01$	*
const * lexical item	$F(4.0, 20.0) = 2.5, p = 0.07$	

The effect of lexical item is to be expected, reflecting inherent differences in the lengths of the words *fugitive, natural,* and *radical.* The effect of syntactic construction is the effect that the experiment is designed to detect—this result indicates that the syntactic construction in which the sequence of W_1 and W_2 is placed has some effect on the duration of W_1. The interaction between these two factors is not necessarily an expected result; this likely reflects the different patterns shown by the whole word measurements of *radical* (little difference between the Adj-N and NP-VP groups) in contrast with *fugitive* and *natural* (clearer difference between these groups) as shown in Figure 12.

In order to see whether the effect of syntactic construction revealed by the RM-ANOVA indicated the expected three-way distinction, paired t-tests with Bonferroni correction were carried out for the planned comparison—two for each set of measurements, comparing the Adj-N items to the NP-VP items and comparing NP-VP items to the RC-VP items. As shown in Table 12 and Table 13, all four t-tests showed significant differences between the groups compared.

Table 12—Results of paired t-tests on duration of W_1

Comparison	Result	
Adj-N vs. NP-VP	$t(5) = 5.7$, $p < 0.01$	*
NP-VP vs. RC-VP	$t(5) = 6.3$, $p < 0.01$	*

Table 13—Results of paired t-tests on duration of the end of W_1 in ms

Comparison	Result	
Adj-N vs. NP-VP	$t(5) = 6.2$, $p < 0.01$	*
NP-VP vs. RC-VP	$t(5) = 4.3$, $p < 0.01$	*

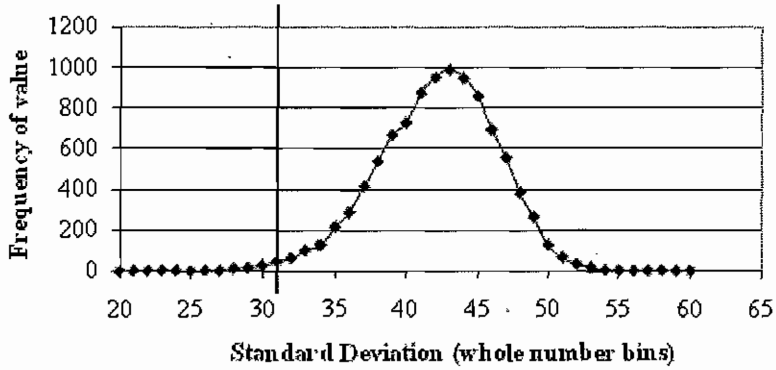
Both sets of duration measurements from W_1—the full length of W_1 and the length of the end of W_1—pattern as expected under an AD analysis: Final lengthening makes a three-way distinction between the Adj-N, NP-VP, and RC-VP conditions and thus, according to the assumptions laid out above, between word-final, AD-final, and ip-final material. Furthermore, as W_1 gets longer, progressively more of its duration comes from its end. This distribution of lengthening is consistent with final lengthening, as discussed in Wightman et al (1992) and Shattuck-Hufnagel & Turk (2000).

Because the NP-VP group had the lowest labeler agreement and a mixture of break labels (1, 3-, and 3m), the possibility that this group might be made up of some tokens with duration patterns like the Adj-N group and some with duration patterns like the RC-VP group must be considered. The Monte Carlo method was used to evaluate the plausibility of this explanation. In particular, the standard deviation[27] of the measure $EndW_1$ for the actual NP-VP group was compared to the standard deviations of the simulated groups. The simulated groups were the same size as the actual group of NP-VP tokens (46 tokens) and were generated by drawing values from the Adj-N and RC-VP groups. The composition of the simulated groups matched the actual group in the number of tokens of each word contributed by each speaker; apart from this matching, the selection of values in the simulated groups was random.

Figure 15 shows the results of the simulation (the curve) and, for comparison, the actual NP-VP group (the vertical line). The probability that the standard deviation of the actual NP-VP group comes from the distribution of standard deviations generated in the simulation is equal to the number of times the simulated groups had standard deviations less than or equal to the actual standard deviation divided by the total number of groups. In this case, the probability of a mixed group having a standard deviation as low as the actual NP-VP group is 0.01—in other words, the NP-VP group is more tightly grouped than a mixed group is likely to be.

[27] The standard deviation was used as a measure of how dispersed the set of data is. In principle, some other property of the group (e.g. standard error) could have been used instead.

Figure 15—Comparison of simulated and actual standard deviations of the NP-VP tokens



3.3 Summary

Two-word homophonous strings embedded in three different syntactic constructions were compared. Two of the syntactic constructions selected—an adjective and the noun it modifies (Adj-N) and the subject of a sentence and the following main verb (NP-VP)—are predicted by some generative grammars to show different prosodic phrasings, with the former capable of forming a single accent domain and the latter incapable. A third construction—the end of a restrictive relative clause modifying the subject of a sentence and the following main verb (RC-VP)—was included for comparison. The tokens were screened for pitch accent pattern and break strength between the two target words.

Duration measures were taken from the homophonous string, with particular attention to the juncture between the two words. Measures of the initial consonant in the second word showed only a two-way distinction, with the beginning of the new phrase after the relative clause showing an initial strengthening effect relative to the other two constructions, but no difference between the Adj-N and NP-VP tokens. Thus, the second hypothesis—that a three-way distinction in phrasing would be reflected in initial strengthening—is not supported. Measures of the final rhyme of the first word in the sequence showed a three-way distinction in length, with the Adj-N group shortest (MAE_ToBI 1), the RC-VP group longest (MAE_ToBI 3), and the NP-VP group in between the other two (a mix of 1, 3-, and 3m). This three-way distinction in length, with one group in between the levels of word and MAE_ToBI intermediate phrase, means that the first hypothesis is supported. Thus, the final lengthening results are consistent with the grammars used in constructing the script items for the experiment.

Chapter 4: Experiment 2

The results of Experiment 1 are consistent with theories that employ accent domains, but they are not conclusive. Specifically, because the accent pattern in Experiment 1 was kept constant, its results cannot demonstrate that the apparent boundary observed must stand in a particular relation to nearby pitch accents. Experiment 2 was conducted to allow a more confident interpretation of the results from Experiment 1 and is very similar in structure to Experiment 1.

Because the pitch accent pattern used in Experiment 2 is different—both words surrounding the boundaries are pitch accented—the accent domain proposals predict that the Adj-N and NP-VP conditions should have similar outcomes in Experiment 2 (in contrast to Experiment 1, where they were different). Specifically, Experiment 2 tests the hypothesis that final lengthening will distinguish only two degrees of lengthening in the items recorded: accent domain (AD) and intermediate phrase (ip). Initial strengthening was not measured in Experiment 2, as it did not distinguish three domain sizes in Experiment 1, so only one specific hypothesis is being tested here (cf. the two hypotheses tested in Experiment 1, as stated in the introduction to Experiment 1 in Chapter 3, beginning on p. 46).

4.1 Methods

Because Experiment 2 is closely linked to (and based on) Experiment 1, many details of its methods are the same as those of Experiment 1 (see section 3.1). An attempt was made to include as many of the same speakers as possible, but only four of the six were able to participate in the second experiment. Two speakers (from the same general speaker population) were added for a total of six participants.

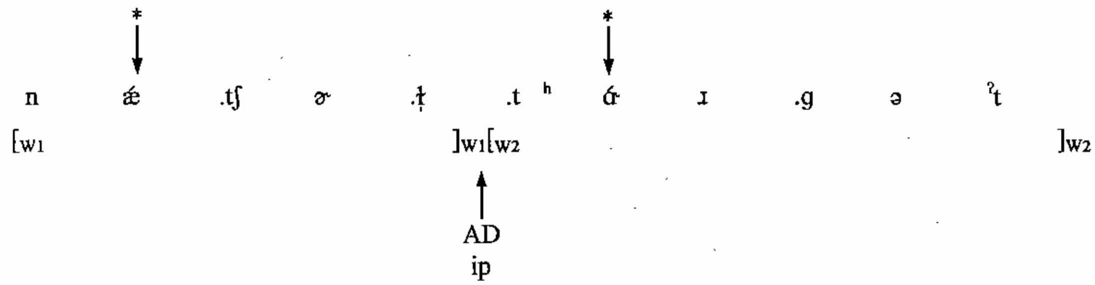
The crucial difference between the two experiments is the pitch accent pattern on the target portion of the script items: In Experiment 2, both W_1 and W_2 were pitch accented.[28] While a context sentence was needed to prompt speakers to deaccent W_2 in Experiment 1, out of the blue reading resulted in the accent pattern desired in Experiment 2. Because removal of the context sentences approximately halved the length of each script item, the length of the filler items was reduced to match, and twice the number of tokens of each item was recorded (six per item).[29]

If the difference in the length of W_1 in the Adj-N and NP-VP conditions is due to an AD edge, this difference should not appear in the same items read with pitch accent on both W_1 and W_2. Just as Figure 5 (p. 50) is a schematic of the crucial prosodic properties of the script items of Experiment 1, Figure 16 is a schematic of the crucial prosodic properties of the script items in Experiment 2.

[28] The determiner of the phrase containing the target noun was changed from definite to indefinite as a way to encourage the pitch accenting of the noun (as presumably new information). For this reason, the subject of the Adj-N condition *natural target* sentence was changed from plural to singular, which was necessary to keep the number of syllables in that sentence the same as the others.

[29] This left more leeway for screening out tokens, which had become a concern after Experiment 1. However, the target pattern for Experiment 2 proved much easier to obtain; many more tokens for Experiment 2 met the criteria for inclusion in the measured data set.

Figure 16—Locations of expected pitch accents and boundaries, Experiment 2 (cf. Figure 5)



For comparison with the examples in Chapter 3, example pitch tracks of utterances from Experiment 2 are shown in Figure 17–Figure 19.

Figure 17—Example utterance of an Adj-N target sentence

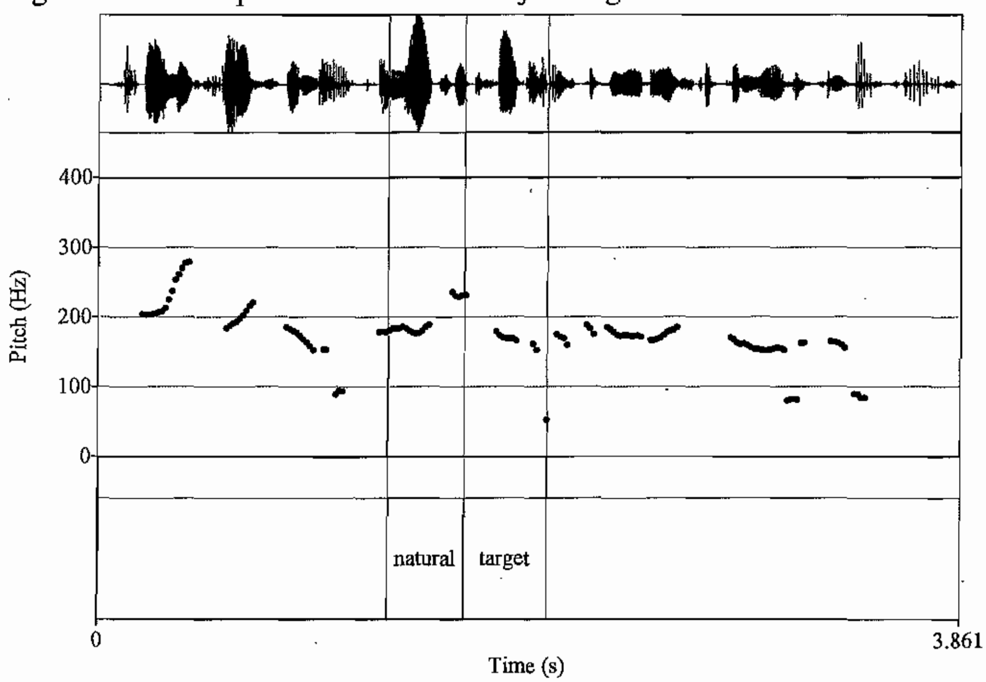


Figure 18—Example utterance of a NP-VP target sentence

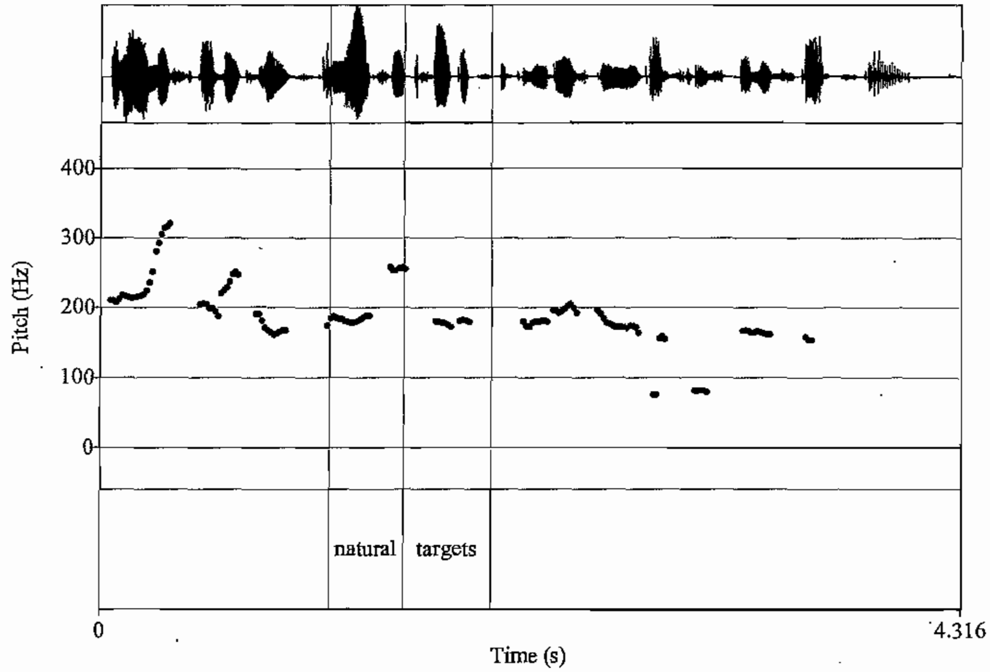
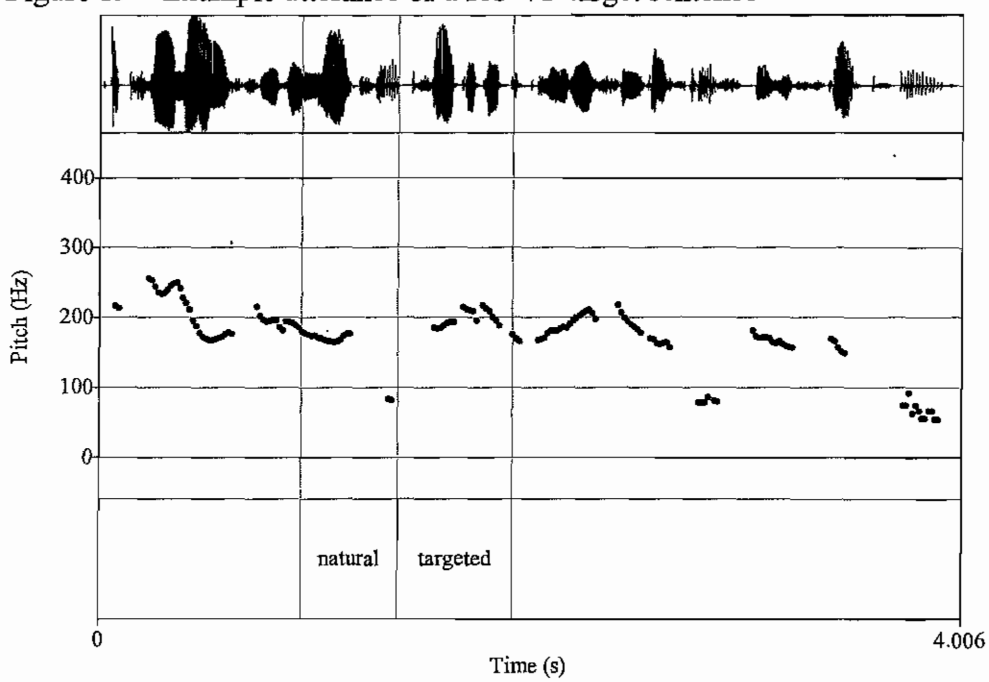


Figure 19—Example utterance of a RC-VP target sentence



4.2 Results

The results of the experiment are divided into transcription-related results (including labeler agreement and some additional characterization of the pitch accent patterns transcribed) and the acoustic measures taken.

4.2.1 Transcription (agreement)

Overall labeler agreement for the two-word target portion is 83%. Agreement was high for break strength (95%), but unexpectedly low (53%) for the pitch accent status of W_1 in the Adj-N condition. To better understand the pitch accent status of W_1 in these tokens, the judgments of a third ToBI-trained labeler (T3) were solicited. In general, T3's pitch accent judgments agreed more often with T1's (82%) than with T2's (74%). The difference is more pronounced in the case of tokens on which T1 and T2 disagreed; in these cases, T3 agreed with T1 68% of the time and with T2 only 32% of the time. All the labelers agree that W_2 is often more prominent in the Adj-N condition; the source of the disagreement is that T1 and T3 more often considered both W_1 and W_2 to be pitch accented, while T2 often perceived only W_2 as accented. Because T3's coding supported the judgment that there are valid transcriptions for the disputed utterances in which W_1 is pitch accented, the screening threshold from Experiment 1 (at least one transcriber gives the expected transcription to the utterance) was also used for Experiment 2.

4.2.2 Acoustic measures

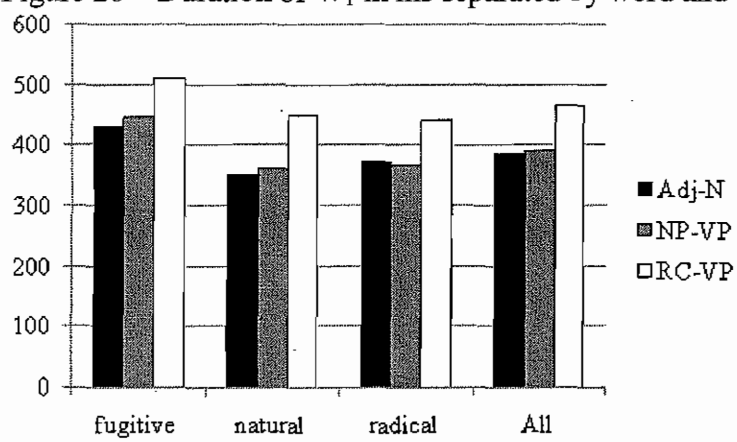
Only the final lengthening-related measures from Experiment 1 (see section 3.2.2.2, p. 58) were taken from the tokens recorded for Experiment 2. The measures from

Experiment 2 are presented in section 4.2.2.1; they are compared with the results of Experiment 1 at the beginning of chapter 5.

4.2.2.1 Final lengthening measures from Experiment 2

The use of ADs in the phonological representation predicts that final lengthening will distinguish only two levels of boundary—accent domain (this time in both the Adj-N and NP-VP groups) and intermediate phrase—and that the strength of the break between W_1 and W_2 will be crucially influenced by syntactic construction. Therefore, the prediction is that the duration measurements taken from W_1 will show the pattern Adj-N, NP-VP < RC-VP. As shown in Figure 12, this prediction is correct for the measurements of the full duration of W_1. This is true for each individual word and for all of the tokens combined.

Figure 20—Duration of W_1 in ms separated by word and overall



The same pattern is shown by the duration of the end of W_1, both in milliseconds in Figure 13 and as a proportion of the entire word in Figure 14.

Figure 21—Duration of $EndW_1$ in ms separated by word and overall

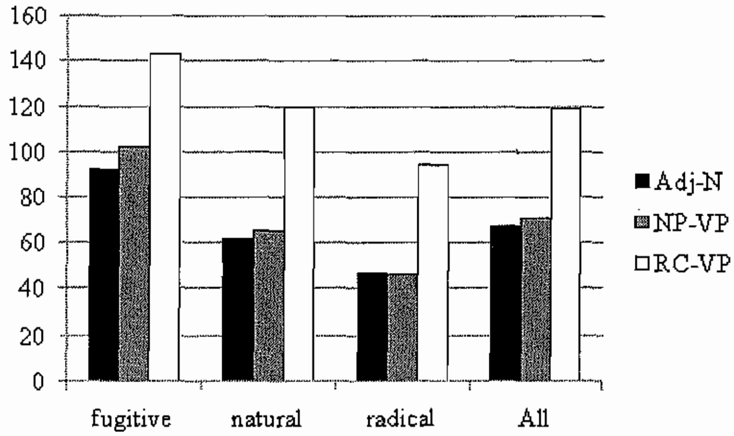
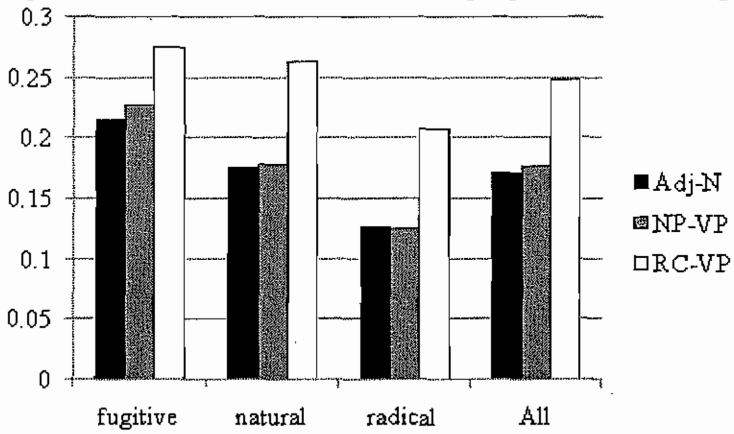


Figure 22—Duration of $EndW_1$ as a proportion of W_1 separated by word and overall



The same statistical tests were applied to these measures as were used in Experiment 1 (see section 3.2.2.2, p. 58). RM-ANOVAs show main effects of both syntactic construction and lexical item for both sets of measurements, as shown in Table 10 (entire W_1) and Table 11 (end of W_1 in ms). Neither test shows significant interaction between the two factors.

Table 14—Results of RM-ANOVA on duration of W_1 (* indicates significance, $\alpha = 0.05$)

Source	Result	
construction (const)	$F(1.1, 5.3) = 36.2$, $p < 0.01$	*
lexical item	$F(2.0, 10.0) = 41.4$, $p < 0.01$	*
const * lexical item	$F(4.0, 20.0) = 1.7$, $p < 0.2$	

Table 15—Results of RM-ANOVA on duration of the end of W_1 in ms

Source	Result	
construction (const)	$F(1.2, 6.2) = 20.3, p < 0.01$	*
lexical item	$F(1.3, 7.9) = 19.5, p < 0.01$	*
const * lexical item	$F(4.0, 20.0) = 1.2, p = 0.4$	

As in Experiment 1, the effect of lexical item is to be expected, reflecting inherent differences in the lengths of the words *fugitive, natural,* and *radical.* The effect of syntactic construction indicates that the syntactic construction in which the sequence of W_1 and W_2 is placed has some effect on the duration of W_1.

As before, paired t-tests with Bonferroni correction were used to better understand the source of the effect of syntactic construction. As before, two t-tests were run for each set of measurements, comparing the Adj-N items to the NP-VP items and comparing NP-VP items to the RC-VP items. As shown in Table 12 and Table 13, only the t-tests comparing the NP-VP and RC-VP items showed significant differences.

Table 16—Results of paired t-tests on duration of W_1

Comparison	Result	
Adj-N vs. NP-VP	$t(5) = 1.4, p = 0.2$	
NP-VP vs. RC-VP	$t(5) = 5.1, p < 0.01$	*

Table 17—Results of paired t-tests on duration of the end of W_1 in ms

Comparison	Result	
Adj-N vs. NP-VP	$t(5) = 1.1, p = 0.3$	
NP-VP vs. RC-VP	$t(5) = 4.2, p < 0.01$	*

Both the full duration of W_1 and the duration of the end of W_1 pattern as expected under an AD analysis: Final lengthening makes a two-way distinction between the Adj-N and NP-VP conditions on one hand and RC-VP condition on the other.

4.3 Summary

In Experiment 2, the target sentences from Experiment 1 were recorded with a different accent pattern on the two-word sequence from which acoustic measures were taken. Because both of the words in this sequence were pitch accented in Experiment 2, accent domain proposals predict that there will be a an accent domain boundary between these two words and, therefore, that the duration difference between the Adj-N and NP-VP conditions found in Experiment 1 will disappear. This prediction is borne out—there is a difference between the NP-VP and RC-VP conditions, but not between the Adj-N and NP-VP conditions. The hypothesis that Experiment 2 was designed to test is supported. However, while the overall pattern of results from Experiment 2 is consistent with AD proposals, just as the results from Experiment 1 are, a comparison of the two sets of results is not consistent with AD proposals, as discussed in section 5.1.1.

Chapter 5: Discussion and Analysis

The previous two chapters have reported two experiments that were designed to allow for the detection of fine phonetic cues to the presence of prosodic boundaries larger than a word but smaller than an intermediate phrase. While a tentative interpretation for each experiment has been presented along with its results, it is necessary to view the evidence of both together, as well, and to draw a more definite conclusion using this more complete picture. This is the starting point for this chapter—a comparison of the results of the two experiments—which is presented in section 5.1. The subsequent sections are devoted to exploring what this interpretation might mean for phonological grammars of English prosody: section 5.2 sketches out what a grammar consistent with the experimental results might look like, and section 5.3 discusses some of the more notable properties of a proposed constraint type. Finally, section 5.4 is a brief summary of the experimental and theoretical work reported in this and the preceding chapters.

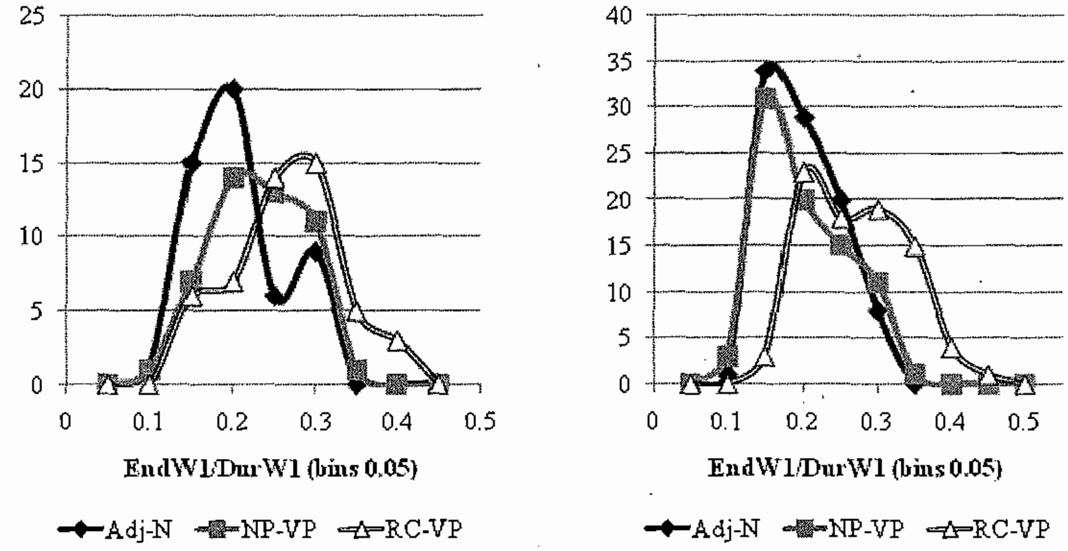
5.1 The question of Accent Domains in English

5.1.1 Comparing Experiment 1 and Experiment 2

In order to answer the question of whether we can support the use of accent domains in American English prosody, we must take into account the results of both experiments, comparing the two. The difference between the pattern of final lengthening results found in Experiments 1 and 2 can be seen clearly by comparing Figure 23a and Figure 23b, which show the duration of the end of W_1 as a proportion of the duration of W_1 in histogram format.

Figure 23—Histograms of $EndW_1$ / duration of W_1, Experiment 1 vs. Experiment 2

a. Experiment 1 b. Experiment 2

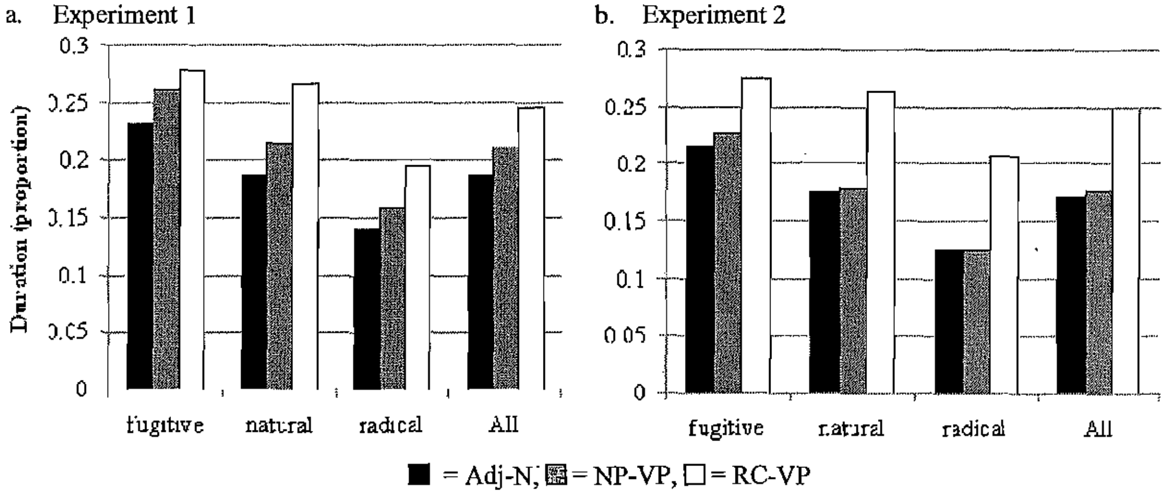


Note: The graphs above have different y-axes; more tokens were collected in Experiment 2.

In short, there are three length categories in Experiment 1, with considerable overlap between the two end-point categories (Adj-N and RC-VP), and the middle category (NP-VP) centered in the region of this overlap. In contrast, only two length categories were produced in Experiment 2, with the Adj-N and NP-VP tokens overlapping almost completely in their distributions and the RC-VP tokens longer as a group, although still showing overlap with the Adj-N and NP-VP tokens.

While the overall pattern of each experiment is consistent with the existence of the hypothesized AD, further comparison casts serious doubt on this interpretation. The crucial comparison is visible in the histogram representations of Figure 23, but even clearer in representations that highlight the central tendency of the groups, as the bar graphs of Figure 24 (reproductions of Figure 14, p. 60 and Figure 22, p. 71, placed side by side for convenience) do.

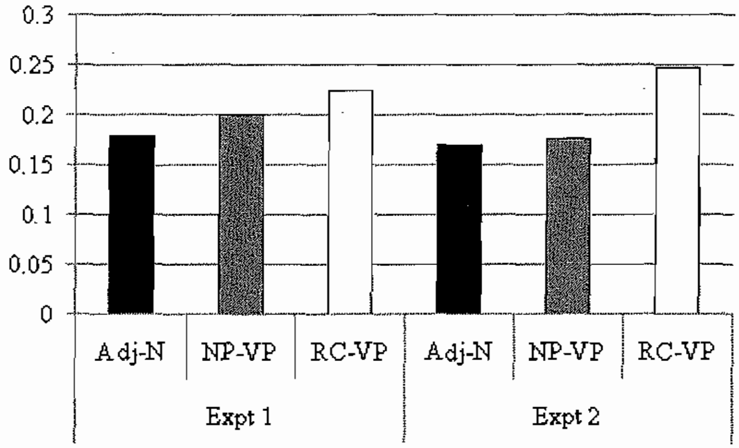
Figure 24—Bar graphs of $EndW_1$ / duration of W_1, Experiment 1 vs. Experiment 2



Recall from Chapters 3 and 4, where the originals of these figures appeared, that a three-way distinction was produced in Experiment 1 (Adj-N < NP-VP and NP-VP < RC-VP both statistically significant, as shown in Table 13, p. 61) but a two-way distinction was produced in Experiment 2 (Adj-N = NP-VP, NP-VP < RC-VP, as shown in Table 17, p. 72). Note also that the speaking rate across the two experiments was similar.

The comparison at issue is the durations of the Adj-N and NP-VP groups across the two experiments. If the NP-VP condition in Experiment 1, the Adj-N condition in Experiment 2, and the NP-VP condition in Experiment 2 all triggered AD boundaries and the Adj-N condition in Experiment 1 did not, we would expect the first three to be similar in duration and the last to be noticeably shorter. However, in the actual results, the two shorter categories in Experiment 2 are closer in length to the Experiment 1 Adj-N tokens than to the Experiment 1 NP-VP tokens. (In fact, they are shorter than the Experiment 1 Adj-N tokens.) As Figure 25 shows, this pattern holds true of the sub-group of four speakers who participated in both experiments; it cannot be attributed to the influence of speakers who participated in only one of the experiments.

Figure 25—$EndW_1$ / duration of W_1, Experiment 1 vs. Experiment 2, four overlapping speakers



Given these outcomes, the same logic that allows us to conclude that the three degrees of lengthening in Experiment 1 indicate the presence of three different levels of prosodic phrasing obliges us to conclude that the two levels of phrasing in Experiment 2 are the smallest and the largest from Experiment 1 (word and ip), not the middle and largest (AD and ip).

In the Experiment 2 data, we are therefore seeing word-size domains and ips without any apparent accent domains. However, there are clearly still pitch accents present in the Experiment 2 data, which suggests that pitch accents can be (and are) generated without these domains. Therefore, the domain between word and ip, as marked by final lengthening in the Experiment 1 tokens, cannot properly be considered an accent domain in the relevant sense: it is not a domain that generates pitch accent as its head. If we wish to continue using accent domains, we must do so without evidence of consistent edge marking—that is, the final lengthening seen in Experiment 1 cannot be considered a characteristic or reliable marking of the edge of an accent domain.

This leaves us with a situation in which the only reliable cue to the presence and distribution of accent domains would be the presence of pitch accents. Although this may seem logical at first glance, it is important to remember that the latter cues the presence of the former only under the assumption that all non-edge prominence is head assignment.[30] To maintain the claim that all prominence is head assignment in the absence of evidence for accent domain edges is to claim not only that the construction of different levels of prosodic phrase accompanied by the assignment of different levels of prosodic head is build into universal grammar as the only way to generate non-edge prominence (because the presence of these levels of phrasing—and thus the one-to-one mapping between phrases and heads—cannot be learned from available evidence), but that all the particulars of how accent domains are constructed and distributed is also a part of universal grammar (as this—the locations of accent domain edges and their relations to other structure—cannot be learned from the available evidence either).[31]

While the requirement that universal grammar must inform speakers of English of nearly all of the properties of accent domains is not absolute proof of the absence of accent domains, it is somewhat troubling. Furthermore, the results of Experiments 1 and 2 require us to believe considerably more than that a great deal of work is being done here by universal grammar. Given the results of Experiments 1 and 2, to continue using accent domains would require us to believe that learners are exposed to a degree of lengthening

[30] In this form, the claim is essentially circular: There must be accent domains because all prominence is generated via head assignment, and the existence of accent domains is demonstrated by the presence of pitch accents for the same reason.

[31] As a part of this claim, we would expect any language without overt accentual phrase marking to construct its accent domains in the same way, or in some way that is predietable from the same independent factor.

between the word and intermediate phrase levels (i.e. of the right degree to mark accent domains), and that this degree of lengthening appears in at least some of the places where they expect accent domain boundaries, as in the Experiment 1 data. However, because we know from Experiment 2 that this edge marking is not a necessary marker of accent domain edges, learners must correctly attribute this lengthening to some other source,[32] while also constructing prosodic structures that include accent domains for which there are no cues beyond the presence of the pitch accent itself, as they would for data of the type seen in Experiment 2. This state of affairs is troubling, suggesting that we might be better off pursuing a strategy that treats the generation of prominence as something other than pure head assignment—that is, in which prominence is not entirely dependent on phrasing, although the two may be related to each other in specific ways by the grammar.

5.1.2 Proposed prosodic phrasings for experimental items as produced

The analysis of the phrasing patterns that is favored by the above discussion is illustrated in this section in the interest of clarity. The phrasing that was expected in Experiment 1 is shown in Figure 26, using the *fugitive pilot* items as examples.[33] The expected phrasings for the same items in Experiment 2 are shown in Figure 27.

---

[32] They may also attribute the lengthening to a combination of the accent domain boundary and some other influence, but cannot be using the accent domain boundary alone, as this would predict lengthening at every accent domain boundary, which Experiment 2 has shown us is not present.

[33] The *fugitive pilot* items have been used in these figures because they are shorter than the other examples, which allows them to fit better into the figures. The structures shown can also be considered to apply to the other items.

Figure 26—Expected prosodic phrasing for the *fugitive pilot* items in Experiment 1

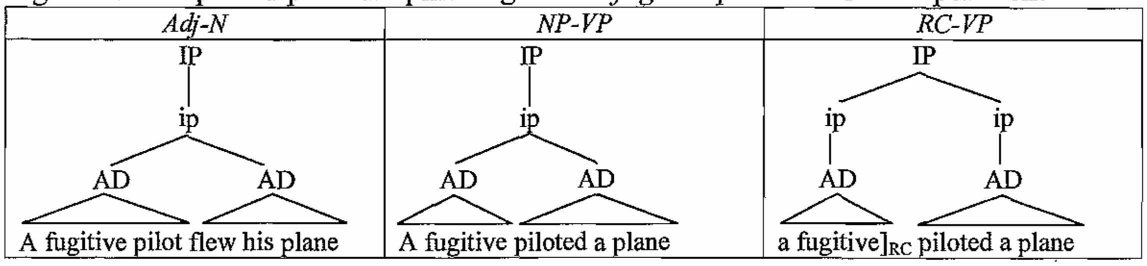
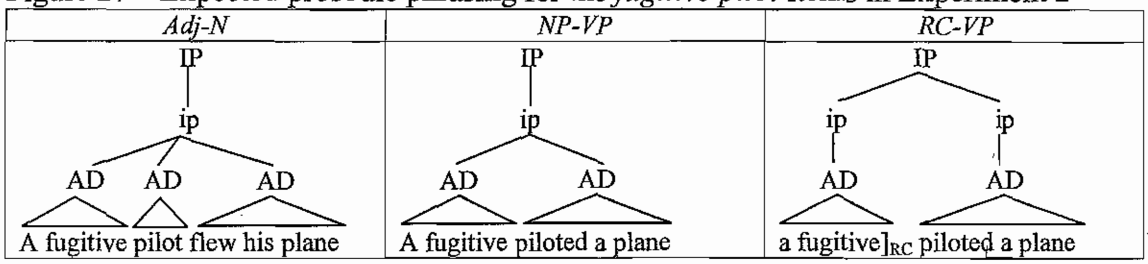
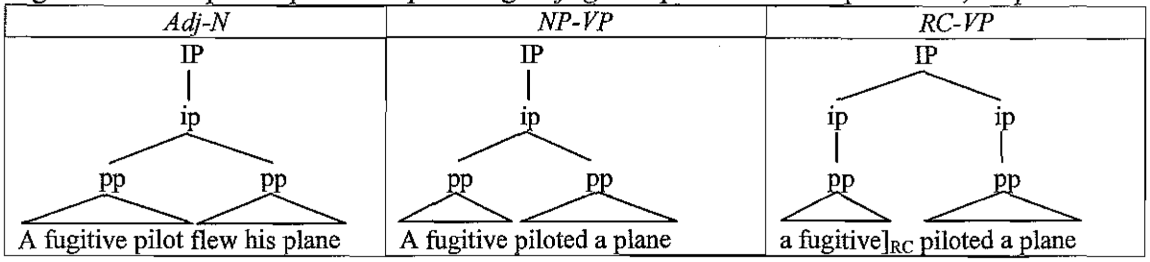


Figure 27—Expected prosodic phrasing for the *fugitive pilot* items in Experiment 2



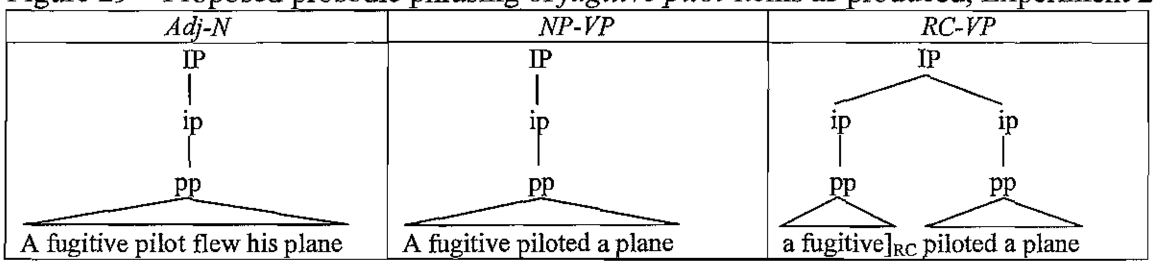
The phrasings suggested by the actual results of the experiments are similar in most respects, but not exactly as expected. In particular, the target span (*fugitive pilot,* in the examples used in Figure 26 and Figure 27) shows fewer phrasal divisions in Experiment 2 than expected. Thus, the actual structures are similar to those in Figure 28 for Experiment 1 and Figure 29 for Experiment 2. Note that the main difference between Figure 26 and Figure 28 is in the node labels: the label AD has been replaced with pp (for small prosodic phrase) in Figure 28. Other than this, the two structures are identical. Because a pp (formerly AD) division was produced in the NP-VP items in Experiment 1, I have assumed that a similar division was produced between the subject and verb of the Adj-N sentences in Experiment 1. Note that pp (small prosodic phrase) will be used instead of AD in much of what follows. The capitalized PP (prosodic phrase) will continue to be used as the generic—that is, to mean a prosodic phrase of any level or size.

Figure 28—Proposed prosodic phrasing of *fugitive pilot* items as produced, Experiment 1



The differences between Figure 27 and Figure 29 are more than notational: the entire sentence is shown as one pp and one ip in both the Adj-N and NP-VP conditions. For the NP-VP condition, this is because the measurements from the target span indicate no additional lengthening at the end of the subject. For the Adj-N condition, the same logic is followed as in the Experiment 1 structures: the end of the subject in the two conditions is treated consistently within the experiment.

Figure 29—Proposed prosodic phrasing of *fugitive pilot* items as produced, Experiment 2



5.2 Generation without accent domains

Having done away with accent domains, we are now left with the question of how to accomplish the work that this level of structure performs in theories that use it. To begin with, abandoning the accent domain means that our formal notion of prominence[34] can no longer be limited to a kind of general head marking, as we now have a head (the pitch accent) without a corresponding level of structure (the accent domain). As a result,

[34] Unless otherwise noted, *prominence* is used in the discussion to refer to head prominence, rather than to edge prominence or to prominence more generally.

we risk losing the ability to formally treat all prominence as simply more (or less) of the same thing: heads in the most traditional sense are defined as more prominent that the rest of the material in their phrase, a definition that cannot be trivially modified to avoid reference to phrasing.

The goal of the following discussion is to show what types of constraints might be used in a grammar that does not view prominence strictly as head assignment. Because there is little point in developing a grammar without understanding the patterns in the data set to be derived, the first order of business will be to show what the most commonly transcribed patterns were for the core sentences from all of the items in Experiment 2 (out of the blue reading), in section 5.2.1; this will be accompanied by a description of the key differences between the patterns produced in the first and second experiments. Once we have a better idea of the patterns to be derived, section 5.2.2 lays out a proposal for deriving these common patterns. This discussion will begin by establishing the view of prominence to be taken, as well as other theoretical preliminaries that must be addressed before specific constraints are introduced. Following this, constraimts on phrasing are introduced, followed by constraints relating prominence and phrasing and then by constraints requiring prominence. Following this is a discussion of remaining issues for the grammar proposed.

5.2.1 Test cases: transcriptions from experiments 1 and 2

The generative grammars used to construct the items in these experiments (see chapter 2 in general on this, and section 2.3 in particular) do not fit the results well: they all use something like accent domains to position prominence at every level.[35] Therefore, we must ask what type of generative grammar would fit the results better. Because these grammars deal in phonological representations, not phonetic ones, we need to know where pitch accents and breaks of various sizes were, at least for the regular SVO core sentences within the longer script items. The transcriptions of the most common patterns produced for each item are drawn mainly from T2's transcriptions of the longer portions of the recorded items. However, as discussed in section 4.2.1, T2 differed systematically from both T1 and T3 in rarely (rather than frequently) transcribing pitch accent on the initial adjective in the Adj-N sentences. As mentioned in sections 3.2.1 and 4.2.1, agreement was otherwise quite high. To explore the question of whether these differences in transcription were systematic (for example, whether T1 would almost always perceive pitch accent on adjectives and T2 almost never perceive same), T1 carefully reviewed T2's transcriptions. In general, T1 agreed with T2's transcriptions with the exception of the prominence status of W_1 in the Adj-N items. The common patterns listed in Table 18, therefore, show the T1 (and T3) majority pattern from Experiment 2 for the item with regards to W_1 in the Adj-N condition (and general consensus about the status of W_1 in the NP-VP and RC-VP items), general consensus on W_2 (which all transcribers agreed was pitch accented most of the time), and T2's transcriptions (verified by T1) for the remainder of the sentence. Note that the proposed grammar can derive the deaccenting of the initial adjective in the Adj-N sentences by re-ranking the proposed constraints.

[35] The MAE_ToBI system does not suffer from this problem, but is not actually a generative model.

Table 18—Most common pitch accent patterns for all items, Experiment 2

Item	Condition	Pattern	Alternative pattern
fugitive pilot	Adj-N	x x x x A fugitive pilot flew his plane…	
	NP-VP	x x x x A fugitive piloted a plane…	
	RC-VP	x x x x x a fugitive]$_{RC}$ piloted a plane…	
natural target	Adj-N	x x x x x A natural target is the greediest of the novice investors.	
	NP-VP	x x x x x A natural targets the greediest new investors…	x x x x A natural targets the greediest new investors…
	RC-VP	x x x x x a natural]$_{RC}$ targeted the greediest investors…	
radical partner	Adj-N	x x x x A radical partner supported the responsible purchase…	
	NP-VP	x x x x A radical partnered with the lead conservative…	
	RC-VP	x x x x x x a radical]$_{RC}$ partnered with the lead conservative…	

The most important difference between the Experiment 1 and Experiment 2 pitch accent patterns is that W_2 was usually not pitch accented in Experiment 1. There is also some variation in which words outside the target sequence are pitch accented—the general tendency is for adjectives that are infrequently pitch accented in Experiment 2 to be more often pitch accented in Experiment 1. In addition, there are more productions of the Adj-N *natural target* item in which *investors* is pitch accented; *novice* is pitch accented somewhat less often in Experiment 1.

The general patterns in the break index transcriptions are not included in Table 18 because they can be described quite simply. The most common break strength after every word is 1, with two exceptions: after W_1 in the RC-VP condition the most common break index is 3, and at the end of the Adj-N *natural target* item (see Table 18) the most common break index is 4. The main difference between Experiment 1 and Experiment 2 with regards to break strength is the difference at the end of W_1 in the NP-VP items, which may be interpreted as the occurrence of a break greater than a word and less than an intermediate phrase.[36]

The following section (5.2.2) lays out a grammar that can account for most of these common patterns without using accent domains. The common patterns that cannot be selected by the grammar presented—and the reasons that they cannot be derived by the system used—will also be discussed.

5.2.2 Developing a grammar

Accent domains in accounts of English prosody are used to position the level(s) of prominence above the word and below the intonation phrase. A grammar that does not use phrasing to position this level of prominence—or to generate or license it—must accomplish these goals another way. In a rule-based system, rules might add x marks to the prosodic grid under certain conditions, not all of which must refer to phrase boundaries. In an OT grammar, what we need are constraints that will favor the presence of increased prominence on some words. Before discussing how these constraints should be formulated, I will first discuss how prominence might be represented and treated by the grammar.

5.2.2.1 Representation of prominence

Although both Prosodic Phonology and Intonational Phonology (see Chapter 1 for brief overviews) employ the metrical grid (see section 2.1.2), which I will also use here,

[36] Note that, as always, other interpretations of this result are also possible. However, as discussed in Chapter 3, there are good reasons to believe that this is phrase-final lengthening.

they differ in whether they view prominence as unitary or binary. Specifically, prosodic phonology accounts have traditionally taken the unitary view of prominence—that it is a single property that can be assigned in differing degrees, resulting in differing levels of both word stress and sentence-level stress. Intonational Phonology accounts, on the other hand, have taken a binary view of prominence, with the word-level and above word–level systems related (e.g. by the Continuous Column Constraint, p. 23) but distinct (see Beckman & Edwards 1994 for a quick review). The latter view is supported, at least in the case of English, by phonetic evidence (Beckman & Edwards 1994) that word-level stress and sentence-level prominence are realized differently. In what follows, I have followed the phonetic evidence in regarding these as related but separate systems. What this means in terms of the representations and constraints used is that

- The only marks shown on the metrical grid will be for prominence above the word level and constraint violations will be assessed according to what is shown on the grids
- It will be assumed that the constraints that place accent cannot move word stress (either because they cannot see it or because they are outranked by the constraints that place word stress); this holds for the data considered here, although there are circumstances in which it does not hold, as when contrastive focus is placed on an unstressed syllable
- All words will be considered to have equal above-word prominence unless assigned grid marks above the word level—that is, two words whose highest columns on the metrical grid are different heights at the word level but not above word level will not be considered to have different levels of prominence above the word; therefore, the grid information about word stress that has been omitted from the representations used here is not vital to the grammar developed below

I have also assumed that the grid structure, the phrasing, and relationships between the two are simultaneously visible to the evaluating grammar, along with a fair amount of other information.

Separating below-word and above-word prominence from each other is not the only choice that will affect what constraints can refer to. At the level of above-word prominence, which we are concerned with here, we must again make the choice between viewing different levels of prominence as different degrees of the same formal type and viewing them as formally distinct—that is, giving the grammar the power to explicitly refer to different prominence types (e.g. pitch accent, nuclear pitch accent) in phonological constraints. The most extreme version of the first position would be a metrical grid without a bounded number of layers and without any labeling of layers. The most extreme form of the latter position is to abandon the grid entirely and use features like [+pitch accent] and [+nuclear pitch accent]. Staying within the metrical grid formalism (that is, not going so far as to use features), the first view can be represented as in Figure 30a, while the second can be represented as in Figure 30b.

Figure 30—Labeled and unlabeled grid representations[37]

a. Unlabeled grid	b. Labeled grid[38]
x x n ǽ .tʃ ɚ .ɫ̩ .t ʰ ɚ́ ɪ .g ə ˀt	NPA x PA x n ǽ .tʃ ɚ .ɫ̩ .t ʰ ɚ́ ɪ .g ə ˀt

The system in Figure 30b allows constraints to refer explicitly to something that has prominence at the level of a nuclear pitch accent, even if this prominence is not the last in its phrase and not the only prominence of this level in its phrase. While this may offer increased convenience, it also includes language-specific structures as part of

[37] The Nuclear Pitch Accent (NPA) level is included in this diagram for illustrative purposes; in fact, the pitch accent on W_1 in Experiment 1 was very rarely the last in an intermediate phrase.

[38] The representation in Figure 30b uses the MAE_ToBI conventions for degrees of prominence, although the use of such a system does not presuppose or require use of these particular categories; they have been shown here because they are the categories that were used to label the data collected.

universally available representations and constraints. While this is sometimes necessary (e.g. for constraints on lexical tone), it also seems desirable to avoid doing this when it is unnecessary—as it seems to be here. I have therefore used the simpler representation, the unlabeled grid.

However, while I have not found it necessary to label the grid layers above the word, I have found it necessary to cap the number of layers at two. This is important in what follows because there are cases in which the two-x maximum requires that some constraints be violated in order for others to be satisfied where an uncapped system would allow more constraints to remain violation-free. In short, the two-x maximum is important in motivating the absence of pitch accent from some content words.

The result of these assumptions on representation is that I have constructed a grammar that, for the most part, selects outputs like the MAE_ToBI transcriptions of the data, in particular the most common patterns transcribed for the collected data (see section 5.2.1). Note that the MAE_ToBI transcriptions distinguish only pitch accented and not pitch accented, with nuclear pitch accents identified only by their positions relative to phrase edges rather than specifically marked by the transcribers as more prominent than other pitch accents. I have assumed that the final pitch accent in an intermediate phrase is more prominent in the phonological representation than the other pitch accents in the phrase (a common assumption), although the transcription itself does not record this information.[39] What this means in terms of the grid is that the nuclear

[39] Ayers (1996) uses a phoneme monitoring task to establish that nuclear pitch accented material—even downstepped—is more prominent than non-nuclear pitch accented material. However, the downstepped nuclear pitch accented material was not as prominent as non-downstepped nuclear pitch accented material. This suggests that the final pitch accent in an intermediate phrase is "special" even when downstep has

pitch accent should have a higher grid column than other pitch accented words in the same phrase. Also in keeping with the MAE_ToBI representation, I have shown intonation phrases without heads (that is, without additional x marks on the grid).

5.2.2.2 Constraints on phrasing

Before introducing any new constraints related to pitch accent placement, we can take care of the constraints required to account for the phrasing patterns observed. For the Adj-N and NP-VP sentences as produced in Experiment 2, if exhaustive parsing and strict layering are assumed, only the economy constraint *Prosodic Phrase (defined below) is required: both of these sentence types were produced without internal phrase breaks larger than a word.

(1) *Prosodic Phrase (*PP)—There are no prosodic phrases in the output

If we use the phrasing constraints familiar from, e.g. Truckenbrodt (1999), Selkirk (1995), etc., *PP would have to outrank Align-R (XP, PP), which encourages the presence of more phrase breaks.

(2) Align-R (XP, PP)—The right edge of every syntactic XP is aligned with the right edge of some prosodic phrase

This gives a two-constraint system for phrasing, with one ranking: *PP » Align-R (XP, PP).

Truckenbrodt's Wrap-XP, which encourages phrasing a main verb and its object together, could also be added to our grammar, as it is never violated in any of the outputs. I have, however, not included it in the rankings and tableaux that follow, as it is not necessary to derive the observed outputs.

---

made it less prominent than it could potentially be. Note that the division between the two types of nuclear pitch accent (downstepped and not) in Ayers' results is minor compared to the difference between nuclear and non-nuclear pitch accent.

The single ranking *PP » Align-R (XP, PP) will give the correct phrasing for the Adj-N and NP-VP items as produced in Experiment 2. It will not, however, give the correct output for the RC-VP items, which have an ip break at the end of the restrictive relative clause (also the end of the subject). The question of why restrictive relative clauses in English tend to be produced with intermediate phrase boundaries at their right edges is larger than the scope of this project, requiring, as it does, a theory of the details of the syntactic structure of restrictive and non-restrictive relative clauses and what aspects of these syntactic structures are important for the prosodic structure.[40] I have encoded this information in the current grammar as the cover constraint RC = ip, which requires that a relative clause end with an intermediate phrase break.[41]

(3) RC = ip—The right edge of a relative clause aligns with the right edge of an intermediate phrase

Although this could be re-stated as an alignment constraint (Align-R (relative clause, ip)), I have chosen to state it more informally as a reminder that it is a cover for a part of the grammar rather than a serious claim about the phonology of English.

Adding RC = ip to the grammar and ranking it above *PP will allow us to derive all of the common phrasing patterns observed in Experiment 2: the end of a restrictive relative clause will trigger the presence of an ip break but the other sentences will be produced as single phrases (that is, without any internal breaks larger than word-size).

---

[40] A truly complete treatment would also require additional follow-up experiments to see whether these aspects of the syntax behaved as expected in other similar structures.

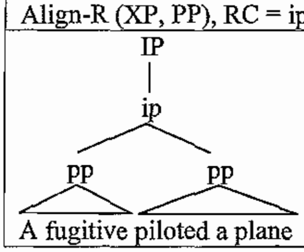
[41] Note that an additional constraint of this type might be required to ensure an even stronger break at the end of a non-restrictive relative clause; however, RC = ip can apply to both types of relative clause without causing problems, at least under the assumption of strict layering.

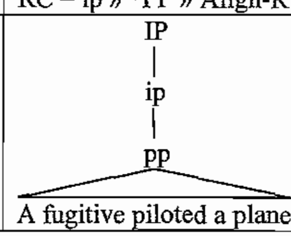
Because the productions from Experiment 1 seem to indicate that there is a degree of break between word and intermediate phrase, I have assumed that there are three levels of prosodic phrase (above the word) available to the grammar. Thus, a sentence produced as one intonation phrase will incur three violations of *PP (one each for the intonation phrase, intermediate phrase, and small prosodic phrase). The appearance of the smallest degree of prosodic boundary above the word after the subject in the Experiment 1 data (based on the increased lengthening shown in the NP-VP condition) suggests the ranking Align-R (XP, PP) » *PP. The regular use of an ip break at the end of the relative clause in the RC-VP condition indicates the ranking RC = ip » *PP.

For the Experiment 2 data, we can rank these three constraints in the order RC = ip » *PP » Align-R (XP, PP), on the grounds that only the presence of a relative clause boundary is sufficient to trigger any phrase break within the sentences; the right edge of the subject fails to trigger a phrase break (as do any other XPs with sentence-internal right edges that might be present in particular syntactic analyses of the script items). The two rankings, with their associated outputs (in prosodic tree form) are shown in Figure 31. The *fugitive pilot* NP-VP items are shown in the figure for reasons of space.

Figure 31—Rankings of RC = ip, *PP, and Align-R (XP, PP) and their outputs

Experiment 1	*Experiment 2*
Align-R (XP, PP), RC = ip » *PP	RC = ip » *PP » Align-R (XP, PP)





The difference in the rankings used to derive the two sets of data appears to be principled—that is, the variation in ranking is not random. There are relatively few factors that might be responsible for this difference in the data collected. Indeed, as we have already concluded in section 5.1 that the presence or absence of the boundary after the subject is not directly connected to the surface pitch accent pattern, there is only one consistent difference between the two sets of data: in Experiment 1, the utterances contained both new and old information, whereas all of the information in the Experiment 2 utterances was equally new. This suggests that speakers producing more pragmatically complex utterances[42] may demote *PP below Align-R (XP, PP). Assuming that the production of utterances the speaker knows to be structurally ambiguous (overall or for a considerable chunk of the utterance) also falls under the umbrella of "more pragmatically complex" than the simple, context-less reading task of Experiment 2, this explanation can also cover the findings of studies like Wightman et al. (1992).[43]

In regular conversational settings, speakers deal often with information of varying degrees of importance or newness, suggesting that the ranking used for the Experiment 1 data may be the more frequent ranking overall. However, because the lowest degree (or degrees) of boundary are not tonally marked—indeed, seem to be marked only by durational cues—the presence of such boundaries is only discernable in phonetic

---

[42] More pragmatically complex in the sense of containing a mix of new and old information rather than all new or all old information.

[43] The recording items in Wightman et al. (1992, and Price et al. 1991) were several sentences long. They consisted of a structurally ambiguous sentence preceded by a few disambiguating preceding sentences. Price et al. found that speakers were able to differentiate the two meanings when played recordings of only the ambiguous sentences, without the context. Wightman et al. found that there were four degrees of final lengthening produced in these items.

measurement in very tightly controlled experiments. This, naturally, raises the question of whether speakers and listeners can actually use this information—the lengthening observed is, after all, fairly modest in degree. While my own data do not speak to this, Price et al. (1991), shows that listeners can use prosodic information (including the presence of Wightman et al.'s degree 2 of final lengthening) to distinguish structurally ambiguous utterances. In other words, when the prosody is the only cue to a difference in meaning, as in a disambiguation task, listener judgments show us that these finer differences in duration are both perceptible and useful.

5.2.2.3 Constraints relating prominence and phrasing

If we are generating prominence and phrasing separately, we will need constraints that regulate the relationships between the two. For example, the condition that every phrase must have a head can become a violable constraint, highly ranked in languages that have a one-to-one relationship between prominences and phrases. This might be formalized as in (4), below.

(4) ∀Prosodic Phrase ∍ Head (∀PP ∍ Hd)—Every prosodic phrase has a single highest grid column (i.e. head) contained within it[44]

In English, only the ip seems to have a head, suggesting that a more specific version of (4) should be employed—that is, (4) is low-ranked but the more specific (5) is highly ranked. This constraint penalizes once for any intermediate phrase that does not have a single most prominent member. With the two-x limit on the height of any grid column, this means that in an ip with more than one prominence, one and only one should

[44] This is, in effect, a specific form of Culminativity, the requirement that some (usually metrical) constituent have one and only one peak prominence / stress.

have two x marks; in an ip with only one prominence, a single x mark is sufficient to satisfy the constraint. Note that this does leave open the possibility of creating an ip whose members are all equally prominent (or equally non-prominent) under the influence of a more highly ranked constraint.

(5) ∀intermediate phrase ∋ Head (∀ip ∋ Hd)—Every intermediate phrase has a single highest grid column contained within it

We might also introduce a constraint similar to (5) that requires every prominence above word level to be the highest in some domain, but this would be low-ranked in English and will play no role in what follows.

Although (5) can ensure that every ip has a head, it does not position the head within the phrase. For this, alignment constraints (or something like them) are still necessary.[45] The alignment constraints used in accounts such as Selkirk (1995) and Truckenbrodt (1999) are adapted from the word-stress literature, which is quite logical for a system using accent domains, which parallel feet. Moving away from accent domains suggests that it might be more logical for us to look to literature that assigns stress without the use of metrical feet, as in Gordon (2002) and Heinz et al. (2005).

---

[45] Note that alignment constraints are necessary but not sufficient: the often used Align-R (Hd, PP) requires every head to be aligned with the right edge of the phrase but is not violated when there is no head. Thus, a candidate with a perfectly aligned head and one with no head will both pass such a constraint. Using Align-R (PP, Hd) instead requires that the right edge of every PP be aligned with the right edge of its head. This constraint presents the same computational problems as its mirror image. In addition, like its mirror image, it might also be violated less (or perhaps equally) by having no head than by having a head far away from the right edge of the phrase.

In deference to the computational problems inherent in the use of alignment constraints that attempt a many-to-one alignment, such as Align-R (Stress, Word),[46] I have used constraints inspired by Heinz et al.'s FirstStressLeft and LastStressRight rather than the more recognizable alignment constraints. However, there are some properties of Heinz et al.'s stress constraints that, although beneficial for word stress, are problematic for sentence-level prominence. In particular, FirstStressLeft and LastStressRight combine in single constraints the tendency to have some prominence towards each edge and the tendency to have the maximal prominence towards each edge. For example, FirstStressLeft is violated by the number of syllable between the left edge and the first stress in the word and one additional time when the maximum stress is not on the left-most syllable. Constructing constraints with this property at the sentence level means that when focus (or some other influence) moves the greatest prominence in a phrase away from the right edge, the constraint that should be responsible for deaccenting after the focus will instead prefer to change the placement of the highest grid column only, placing other sentence-level prominence as it would in the absence of focus to decrease the number of violations resulting from an absence of sentence-level prominence close to the right edge. For this reason, it is necessary to split these constraints apart, resulting in four constraints: FirstxLeft, MaxxLeft, LastxRight, and MaxxRight, defined below. In addition, while FirstStressLeft and LastStressRight treat all syllables equally, the constraints used at the sentence level are restricted to content words.

[46] See Heinz et al. (2005) for an overview of these problems and references cited therein for more detail. Note that Align-R (XP, PP) does not present the same difficulties in the way it is used here because it does not need to be evaluated gradiently for the current data set, and the computational difficulties that arise with alignment constraints come from this particular mode of evaluation.

(6) FirstxLeft—The first content word in a phrase has prominence above the word level. Incurs one violation for every content word between the left edge of the phrase and the first x mark on the grid (if there are any grid marks)

(7) MaxxLeft—The highest grid column in a phrase is the first. Incurs one violation for every content word between the left edge of the phrase and the highest column of x marks on the grid (if there are any grid marks)

(8) LastxRight—The last content word in a phrase has prominence above the word level. Incurs one violation for every content word between the right edge of the phrase and the last x mark on the grid (if there are any grid marks)

(9) MaxxRight—The highest grid column in a phrase is the last. Incurs one violation for every content word between the right edge of the phrase and the highest column of x marks on the grid (if there are any grid marks)

Like the alignment constraints they replace, these do not actually require that there be a single highest column in a particular phrase, only that this highest column be in a particular position when it occurs. The constraints that will be important here are FirstxLeft and MaxxRight. The other two constraints defined above, MaxxLeft and LastxRight, must both be ranked below MaxxRight. With MaxxRight ranked above MaxxLeft, the main prominence in a phrase that has a single highest grid column within it will be to the right in the phrase. With MaxxRight ranked below LastxRight, post-focal deaccenting will be correctly handled by the grammar; the opposite ranking would allow for additional grid marks after the highest grid column in a phrase. Because they will not be crucial in the following account, LastxRight and MaxxLeft will not be shown in the tableaux below.

In most cases, MaxxRight and FirstxLeft will not be in conflict. The crucial case is a phrase with only one pitch accent in it, which would, by definition, be the nuclear pitch accent. In this case, where the first column of x marks is also the last, the two constraints will be in conflict. Because the nuclear pitch accent should still be right-aligned, even when it is the only pitch accent in the phrase, MaxxRight must also be ranked above FirstxLeft. Note that FirstxLeft is not inactive in longer phrases when ranked below MaxxRight—FirstxLeft can still encourage the presence of some prominence at or near the left edge of the phrase. The tendency to favor such configurations has been noted previously (e.g. Shattuck-Hufnagel & Turk 1996) and will be relevant in accounting for the data discussed here.

The constraint $\forall ip \ni Hd$, (5), is never violated in the output[47] and is thus included in the highest stratum in the grammar (along with RC = ip). MaxxRight is ranked above FirstxLeft, as discussed above, although we will see that neither of these constraints is undominated—that is, the first x mark is not always on the leftmost content word in a phrase and the last x mark is not always on the rightmost content word in a phrase. Violations of these constraints will be driven by constraints introduced in the following section.

5.2.2.4 Constraints requiring prominence

Of the constraints introduced thus far, only $\forall ip \ni Hd$ can directly force the presence of at least one x mark on the metrical grid. On a phrase at least two words long, the addition of FirstxLeft and MaxxRight can add two more x marks—one on the first content word and an additional x mark over the rightmost content word to satisfy both

[47] Recall that the greater prominence of the NPA relative to other pitch accents in these particular utterances is by hypothesis. However, this assumption does follow the findings of Ayers (1996), who finds that nuclear pitch accents, whether downstepped or otherwise, facilitate response in a phoneme monitoring task (relative to non-nuclear accented material)—that is, nuclear pitch accented material behaves as more prominent than non-nuclear accented material, regardless of how the nuelear pitch accenting is realized. (Sec also note 39, p. 88 on this topic.) The question of whether there might be ips without any pitch accents did not arise in this study.

∀ip ∋ Hd and MaxxRight. Without the addition of some other constraints, this grammar is insufficient to select a unique winner, as illustrated in Figure 32.

Figure 32—Tableau showing failure to select a unique winner (constraints on prominence only)

A fugitive piloted a plane	∀ip ∋ Hd	MaxxRight	FirstxLeft
x x x x ?☞ A fugitive piloted a plane]$_{ip}$			
x x x ?☞ A fugitive piloted a plane]$_{ip}$			
x A fugitive piloted a plane]$_{ip}$			2
A fugitive piloted a plane]$_{ip}$	1!		

Given that we are allowing prominence to be assigned (or, more accurately, permitted) independently of phrasing, we might expect to need an economy constraint—*x to go along with *PP.

(10) *x—There are no x-marks on the metrical grid

However, while we may need *x in general (see, for example, section 5.3.2 for discussion of the typological usefulness of *x), it does not select the correct winner (the first candidate in Figure 32). Instead, it selects a candidate with the fewer x marks on the grid:

Figure 33—Selection of the wrong winner

A fugitive piloted a plane	Vip ∋ Hd	MaxxRight	FirstxLeft	*x
x x x x ☹ A fugitive piloted a plane]$_{ip}$				4!
x x x 💣 A fugitive piloted a plane]$_{ip}$				3
x A fugitive piloted a plane]$_{ip}$			2!	1
A fugitive piloted a plane]$_{ip}$	1!			

Note that re-ranking can cause selection of a different wrong winner (the third or fourth candidate), but the desired winner is harmonically bounded by the second candidate. Because the example shown in Figure 32 is not unusual in having more than two pitch accents, additional constraints must be introduced that will favor the presence of additional prominences.

A constraint requiring two parts of the output to differ in degree of prominence can favor candidates with more x marks on the grid. Constraints of this type have already been introduced for syntactic positions, such as Büring & Gutiérrez-Bravo's (2001) ARGUMENT-OVER-PREDICATE (within a phonological phrase, an argument is more prominent than a predicate, similar to German et al. 2006's HEADARG, a head is less prominent than its internal argument).[48] Of course, ARGUMENT-OVER-PREDICATE (and HEADARG) will favor the incorrect winner from Figure 33, not the desired winner; thus,

[48] The original purpose of ARGUMENT-OVER-PREDICATE and HEADARG was to position heads within prosodic phrases, not to favor candidates with a greater number of prominences.

neither of these constraints will be of help to the grammar.[49] The form of these constraints, however, is of definite interest. I have used the general format X > Y, X is more prominent than Y, for such constraints. In this format, the existing examples could be re-written ARG > PRED and ARG > HEAD. More precisely, the schema for the X > Y constraints used here is

(11) X > Y—Every word in category X is more prominent than every word in category Y
X is a group of words (e.g. by syntactic category, phonological characteristic, etc.)
Y is a group of words (same range as X)
X and Y do not overlap

While it might ultimately be necessary to limit these constraints to holding within various domains, I have used the sentence as the domain for all such constraints.

The most basic constraint of this type that I will use is Content > Function (C > F).

(12) Content > Function (C > F)—Every content word is more prominent than every function word

Returning to the example in Figure 32 and Figure 33, we can see that C > F must be ranked above *x but cannot, on the basis of this example alone, be ranked with respect to the other constraints in the grammar.

[49] Note that after the constraints dealing with headedness are introduced both of these constraints will be satisfied by most of the winners. I do not omit them because the data considered here is inconsistent with them, but because they do not perform any vital function for this data set that is not already taken care of by some part of the grammar that is independently needed.

Figure 34—Tableau showing selection of the correct winner; illustrates the ranking C > F » *x

A fugitive piloted a plane	∀ip ∋ Hd	MaxxRight	FirstxLeft	C > F	*x
x x x x ☞ A fugitive piloted a plane]$_{ip}$					4
x x x A fugitive piloted a plane]$_{ip}$				2!	3
x A fugitive piloted a plane]$_{ip}$			2!	4	1
A fugitive piloted a plane]$_{ip}$	1!			6	

In general, a ranking of C > F » *x creates a system in which content words are prominent—in this case pitch accented—and function words are not.[50] While this is adequate for *a fugitive piloted a plane,* it is not sufficient for longer, more complex sentences.

For the more complex sentences that follow, three additional constraints will be needed:

(13) Noun > Adjective (N > A)—Every noun is more prominent than every adjective

(14) Noun > Verb1—(N > V1)—Every noun is more prominent than every verb in group 1 (see below for description of group 1)

(15) Noun > Verb2—(N > V2)—Every noun is more prominent than every verb in group 2 (see below for description of group 2)

The splitting of a more general Noun > Verb into two separate constraints is necessary due to a clear difference between the verbs *pilot, target,* and *partner,* which are usually pitch accented, and *fly* and *support,* which usually are not.[51]

---

[50] It is not difficult to evaluate C > F so that when *x » C > F content words can still be more prominent than function words, but in a way that does not involve adding x marks to the grid—perhaps by being phrase initial rather than pitch accented.

There is no single factor that is clearly and unequivocally responsible for the difference between these two groups of verbs, but one or more of the following factors may play a role:

- Main stress vowel height—The verbs that are usually pitch accented all have [+low] main stress vowels and those that are usually unaccented all have non-low main stress vowels. Because low vowels are inherently louder than non-low vowels, they may attract pitch accent.

- Frequency—The Corpus of American English shows that *pilot, target,* and *partner* are all less frequent than either *fly* or *support*.[52] Thus, group 1 could be the less frequent verbs and group 2 the more frequent ones. However, this is unlikely to be the sole factor, as the division between the most frequent of the group 1 verbs (*target,* 8336 instances) and the less frequent of the group 2 verbs (*fly,* 8994 instances) is not particularly striking when compared to the difference between the two group 2 verbs (*fly,* 8994 instances vs. *support,* 26,856 instances)

- Predictability—Some of these verbs are more predictable from the surrounding context than others. This is particularly striking in the case of the verb *fly,* which is probably much more expected in a sentence about pilots than, for example, the verb *target* is in a sentence with a subject whose head noun is *natural.*

- Syntactic origin—All of these verbs are related in some way to nouns. However, whether the verb or the noun is the more basic form may make a difference in how likely the verb is to be pitch accented.

These or similar factors may act, either singly or in combination, to separate the verbs into the relevant groups. Alternatively, the division may be more word-specific. Because the data clearly underdetermine which factors are responsible for this difference in how often the verbs are pitch accented, I have used the more neutral numbered group notation.

Because the verbs in group 1 were usually pitch accented and verbs in group 2 were usually not, N > V2 is usually ranked above C > F and N > V1 (or more general N >

[51] The sixth verb, *be,* is also usually not pitch accented, but this can be accounted for independently by C > F under the assumption that *be* still counts as a function word even when it is used as a main verb.

[52] All figures reported from the Corpus of American English are by lemma and from the spoken language subset only.

V) is usually ranked below C > F. The ranking C > F » N > V1 is illustrated in Figure 35. The ranking N > V2 » C > F (along with the ranking of N > A discussed below) is shown in Figure 36 (on p. 104).

Figure 35—C > F » N > V1

A fugitive piloted a plane	Vip ∍ Hd	MaxxRight	FirstxLeft	N > V2	N > A	C > F	N > V1	*x
x x x x ☞ A fugitive piloted a plane$]_{ip}$							1	4
x x x A fugitive piloted a plane$]_{ip}$						2!		3
x A fugitive piloted a plane$]_{ip}$			2!			4		1
A fugitive piloted a plane$]_{ip}$	1!					6		

In general, the adjectives *fugitive, natural,* and *radical* were coded as pitch accented,[53] while the non-initial adjectives *new, lead,* and *responsible* most often were not. The non-initial adjective *greediest* was more variable in its behavior. The non-initial modifier *novice* was consistently pitch accented; the position occupied by *novice* in the sentence under consideration is more often associated with adjectives, but *novice* may also be regarded as a noun that is acting to modify another noun in this particular case, and I have treated it as such below.[54] I have taken *new, lead,* and *responsible* to be the default cases for adjectives, with *fugitive, natural,* and *radical* pitch accented by virtue of

[53] Recall that T1 and T3 both coded these adjectives as pitch accented, although they were judged less prominent than the pitch accented nouns that followed them. T2 differed from T1 and T3 in coding these as without pitch accent on this basis. Assuming that the generative grammar is engaged in perception as well as production, this would indicate a high ranking of N > A for T2.

[54] Justification for regarding *novice* as a noun even in this position comes from the infelicity of sentences like **I'm talking about the novice one* in contrast to sentences like *I'm talking about the new one.*

their position and *greediest* and *novice* pitch accented for reasons discussed below. This is reflected in the ranking FirstxLeft » N > A » C > F, shown in Figure 36 and Figure 37.

Figure 36—N > V2 » C > F; N > A » C > F

A radical partner supported the responsible purchase	Vip ∍ Hd	MaxxRight	FirstxLeft	N > V2	N > A	C > F	N > V1	*x
x x x x x x A radical partner supported the responsible purchase]$_{ip}$				1!	2			6
x x x x x A radical partner supported the responsible purchase]$_{ip}$					2!	2		5
x x x x ☞ A radical partner supported the responsible purchase]$_{ip}$					1	4		4
x x x x A radical partner supported the responsible purchase]$_{ip}$			1!		1	4		4
x x x A radical partner supported the responsible purchase]$_{ip}$			1!			6		3

The constraint system as developed thus far, with a certain amount of variation in ranking, is capable of deriving the most common patterns for the *fugitive pilot* and *radical partner* sentences. Table 19 shows which of the patterns from Table 18 can be accounted for by the grammar in its present form. (Patterns that are accounted for have a check mark below the condition name in the second column of the table.)

Table 19—Outputs accounted for by the current grammar

fugitive pilot	Adj-N ✓	x x x x A fugitive pilot flew his plane…	
	NP-VP ✓	x x x x A fugitive piloted a plane…	
	RC-VP ✓	x x x x x a fugitive$]_{RC}$ piloted a plane…	
natural target	Adj-N	x x x x x A natural target is the greediest of the novice investors.	
	NP-VP ✓ (2)	x x x x x A natural targets the greediest new investors…	x x x x A natural targets the greediest new investors…
	RC-VP	x x x x x a natural$]_{RC}$ targeted the greediest investors…	
radical partner	Adj-N ✓	x x x x A radical partner supported the responsible purchase…	
	NP-VP ✓	x x x x A radical partnered with the lead conservative…	
	RC-VP ✓	x x x x x x a radical$]_{RC}$ partnered with the lead conservative…	

The examples in rows 2 and 7 of Table 19 (and Table 18) are successfully derived as shown in Figure 35 and Figure 36, respectively. The four remaining examples whose most frequent patterns can be accounted for at this stage are shown in Figure 37–Figure 40. In addition, one of the two most common patterns for the *natural target* NP-VP item (the one on the right) can be accounted for by the current grammar, as shown in Figure 41 on p. 110. For the Adj-N and NP-VP items, the phrasing constraints have been omitted, as in the previous examples of this type, and only candidates in which the entire sentence is a single ip (and, presumably, a single pp) have been shown. The tableaux for the RC-VP examples include the phrasing constraints; for clarity, violations of ∀ip ∋ Hd, MaxxRight, and FirstxLeft in these tableaux are assessed as though there were no other pitch accents in the ip that contains W_1. In the case of FirstxLeft, only violations from the second ip are shown—that is, violations of FirstxLeft are assessed as though the first ip perfectly satisfies this constraint. Note that violations of these constraints from material earlier in the sentence will not change the winner for the relevant portion of the RC-VP items.

As before, every pp, ip, and IP is counted as one violation of *PP. In assessing these violations in tableaux, I have assumed

- One IP per sentence
- One ip per sentence in the Adj-N and NP-VP items
- Two ips per sentence in the RC-VP items
- An equal number of pps and ips for Experiment 2 items (including Figure 37–Figure 41)
- Two pps per item from Experiment 1 (as shown in Figure 28 and Figure 31, above, and in Figure 47, below)

Figure 37—Adj-N *fugitive pilot* item, pitch accent on *fugitive, pilot, plane;* FirstxLeft » N > A[55]

A fugitive pilot flew a plane	Vip ∋ Hd	MaxxRight	FirstxLeft	N > V2	N > A	C > F	N > V1	*x
x x x x x A fugitive pilot flew a plane]$_{ip}$				1!	1			5
☞ x x x x A fugitive pilot flew a plane]$_{ip}$					1	2		4
x x x x A fugitive pilot flew a plane]$_{ip}$			1!	1		2		4
x x x A fugitive pilot flew a plane]$_{ip}$			1!			4		3

Figure 38—NP-VP *radical partner* item, pitch accent on *radical, partnered, conservative*

A radical partnered with the lead conservative	Vip ∋ Hd	MaxxRight	FirstxLeft	N > V2	N > A	C > F	N > V1	*x
x x x x x A radical partnered with the lead conservative]$_{ip}$					1!		1	5
☞ x x x x A radical partnered with the lead conservative]$_{ip}$						3	1	4
x x x A radical partnered with the lead conservative]$_{ip}$						6!		3
x x x A radical partnered with the lead conservative]$_{ip}$			1!		1	6	1	3

55 For this particular example, it would also work to rank N > V2 » N > A. However, using N > V2 » N > A and allowing FirstxLeft to be ranked below N > A would predict that adjectives in initial position would not be pitch accented if the verb belonged to group 2.

Figure 39—RC-VP *fugitive pilot* item, pitch accent on *fugitive, piloted,* and *plane*

a fugitive]$_{RC}$ piloted a plane	RC = ip	Vip ∍ Hd	*PP	Algin-R (XP, PP)	MaxxRight	FirstxLeft	N > V2	N > A	C > F	N > V1	*x
x x x x x ☞ a fugitive]$_{ip}$ piloted a plane]$_{ip}$			5								5
x x x a fugitive]$_{ip}$ piloted a plane]$_{ip}$			5						2!		3
x x x x a fugitive]$_{ip}$ piloted a plane]$_{ip}$			5							1!	4
x x x x a fugitive]$_{ip}$ piloted a plane]$_{ip}$			5						2!		4
x x x x a fugitive piloted a plane]$_{ip}$	1!		3	1						1	4

Figure 40—RC-VP *radical partner* item, pitch accent on *radical, partnered, lead,* and *conservative*

a radical]$_{RC}$ partnered with the lead conservative	RC = ip	Vip ∍ Hd	*PP	Algin-R (XP, PP)	MaxxRight	FirstxLeft	N > V2	N > A	C > F	N > V1	*x
x x x x x x ☞ a radical]$_{ip}$ partnered with the lead conservative]$_{ip}$			5								6
x x x x x a radical]$_{ip}$ partnered with the lead conservative]$_{ip}$			5						3!		5
x x x x x a radical]$_{ip}$ partnered with the lead conservative]$_{ip}$			5			1!			3		5
x x x x a radical]$_{ip}$ partnered with the lead conservative]$_{ip}$			5			1!			6		4

One interesting property of the grammar that is made apparent in Figure 40 is that the introduction of an ip break causes the grammar to favor additional pitch accenting. This effect is indirect: the ip break in the RC-VP sentences results in the subject noun (W_1) receiving a nuclear pitch accent (two grid marks) making it more prominent than pitch accented words (one grid mark) in other parts of the sentence. Because the object noun also receives a nuclear pitch accent, adjectives in the RC-VP sentences have no need to deaccent adjectives for better compliance with N > A; thus, the winner is the candidate that better satisfies C > F: the one with the sentence-medial adjective(s) pitch accented. This is the desired result in Figure 40, but there is also a less common pattern for this sentence in which *lead* is not pitch accented (just as *greediest* is not pitch accented in the most common pattern for the RC-VP *radical partner* item, shown in Table 18). The grammar cannot derive this pattern, suggesting that additional forces for deaccenting must also exist.

The last set of sentences, the *natural target* sentences, gives a couple examples of adjectives that plausibly belong to classes that receive special treatment: *greediest* (for one or more reasons discussed below) and *novice* (perhaps because it is very low frequency, especially in spoken language[56]).

First, let us consider the NP-VP *natural target* item, which is the simplest item in the set. The current grammar can derive one of the two most common outputs—the selected winner in Figure 41, below. In this pattern, only *natural, target,* and *investors*

[56] As in Mark Davies' Corpus of American English.

are pitch accented—*natural* by virtue of its position and *target* and *investors* by virtue of their syntactic categories.

Figure 41—NP-VP *natural target* item, pitch accent on *natural, target, investors*

A natural targets the greediest new investors	Vip ∍ Hd	MaxxRight	FirstxLeft	N > V2	N > A	C > F	N > V1	*x
x x x x x x A natural targets the greediest new investors]$_{ip}$					2!		1	6
x x x x x A natural targets the greediest new investors]$_{ip}$					1!	2	1	5
x x x x ☞ A natural targets the greediest new investors]$_{ip}$						4	1	4
x x x A natural targets the greediest new investors]$_{ip}$						6!		3

The other common pattern transcribed is the second candidate in Figure 41, which is not a possible winner given the current grammar. While there may be rhythmic reasons to place pitch accent on *greediest*—to prevent a lapse of two content words—it seems unlikely that rhythm is the sole deciding factor, as we have already seen that a two-word lapse is permitted in the Adj-N *radical partner* item, where the string *supported the responsible* is most commonly left entirely without pitch accent, as shown in Figure 42, below (a reproduction of one cell from Table 18):

Figure 42—Adj-N *radical partner* item, most common pattern (single)

x
x x x
A radical partner supported the responsible purchase…

If anything, *supported the responsible* should be a worse lapse than *the greediest new,* as the former is eight syllables long and the latter only five. Thus, while concerns of rhythm may favor the second candidate in Figure 41 over the third, they cannot be the deciding factor without deriving an incorrect winner for *a radical partner supported the responsible purchase.*

Instead, I hypothesize that the tendency to place prominence on *greediest* has to do with some property particular to it. Candidates for such a property include: the fact that it is not just an adjective, but a superlative; its relatively low frequency compared to *natural* (again, per the Corpus of American English); and its position in the NP-VP item as the first of the two prenominal modifiers. This can be included in the grammar as the constraint A1 > A2, requiring that every adjective in class 1 (here, only *greediest*—assigned to class 1 for one or more of the reasons mentioned above) be more prominent than every adjective in class 2 (here, any other adjective):

(16) Adjective1 > Adjective2—(A1 > A2)—Every adjective in class 1 is more prominent than every adjective in class 2

In order for A1 > A2 to have its desired effect, it must be ranked above N > A (which will still favor absence of pitch accent on *greediest).* In order for winners with and without pitch accent on *greediest* to surface, A1 > A2 and N > A must be variably ranked, which is indicated in Figure 43 with a wavy line between the two constraints.

Figure 43—Effect of A1 > A2 » N > A on the example from Figure 41

A natural targets the greediest new investors	Vip ∋ Hd	MaxxRight	FirstxLeft	N > V2	A1 > A2	N > A	C > F	N > V1	*x
x x x x x x A natural targets the greediest new investors]$_{ip}$					1!	2		1	6
x x x x x ☞ A natural targets the greediest new investors]$_{ip}$						1	2	1	5
x x x x A natural targets the greediest new investors]$_{ip}$					1!		4	1	4
x x x A natural targets the greediest new investors]$_{ip}$					1!		6		3

Reversing the ranking of A1 > A2 and N > A shown in Figure 43 will result in the winner from Figure 41 again being the preferred output, as it is in a grammar without A1 > A2.

The second most common pattern for the RC-VP *natural target* item has pitch accent on *natural, targeted, greediest,* and *investors.* The most common omits the pitch accent on *greediest.* While the current grammar can derive the former, it cannot derive the latter. With the first noun in a separate ip, both nouns can be more prominent than both the verb and the adjective, with all four content words receiving prominence above the level of the function words. Thus, all the X > Y constraints and all the prosodic well-formedness constraints can be satisfied by one candidate, as shown in Figure 44; there is no ranking that comes close to being accurate for the other items but can select the most common pattern (the second candidate in Figure 44) as the winner for the RC-VP *natural target* item.

Figure 44—RC-VP *natural target* item, pitch accent on *natural, targeted, greediest,* and *investors*

A natural]$_{RC}$ targeted the greediest investors	RC = ip	Vip ∋ Hd	*PP	Algin-R (XP, PP)	MaxxRight	FirstxLeft	N > V2	N > A	A1 > A2	C > F	N > V1	*x
x x x x x x ☞ A natural]$_{ip}$ targeted the greediest investors]$_{ip}$[57]			5									6
x x x x x A natural]$_{ip}$ targeted the greediest investors]$_{ip}$[58]			5							3!		5
x x x x x A natural]$_{ip}$ targeted the greediest investors]$_{ip}$			5			1!				3		5
x x x x A natural]$_{ip}$ targeted the greediest investors]$_{ip}$			5			1!				2		4

Note that all of the RC-VP items were sometimes produced with fewer pitch accents than the grammar predicts. However, this occurs in the most common pattern only for *a natural targeted the greediest investors.* We will return to a discussion of these patterns after seeing how this type of grammar handles patterns most commonly produced for the Adj-N *natural target* item.

The current constraint system comes close to being able to derive the most common patterns for the Adj-N *natural target* item, except for the fact that *novice* is consistently nuclear pitch accented and *investors,* unlike all the other nouns in these items, was not pitch accented in any of the common outputs. This can be captured in a constraint N1 > N2, parallel to A1 > A2. Here, the two likeliest candidates for dividing

[57] This is the second most common pattern in the output.

[58] This is the most common pattern in the output.

the two types of nouns from each other are linear order (first noun > second noun) and frequency (*novice* appears only 89 times in the spoken word portion of the Corpus of American English, while *investors* appears 2431 times).[59]

(17) Noun1 > Noun2—(N1 > N2)—Every noun in class 1 is more prominent than every noun in class 2

In assessing violations of N1 > N2, I have assumed that only *investors* is a class 2 noun, and that *target* is in neither of these classes. Note that assuming that *target* is in class 2 will not affect which candidate is selected as the winner. In this item, assuming that *target* is in class 1 will also produce the correct result, as the desired winner has both *target* and *novice* more prominent than *investors.*

If the grammar does not allow a phrase boundary to be moved to fall immediately after *novice* (and other class 1 nouns)—and I have assumed that it does not—the effect of this constraint in phrase-medial position would be to deaccent only the class 2 noun(s), without necessarily placing nuclear pitch accent on the class 1 noun: the fact that *novice* receives nuclear pitch accent here is due to its proximity to the end of the sentence, not only to the effect of N1 > N2. Note, however, that in order for *novice* to receive the nuclear pitch accent rather than *investors,* N1 > N2 must be ranked above MaxxRight, which favors the candidate with the perfectly right-aligned nuclear pitch accent (the first candidate in Figure 45).

[59] Looking at the modifiers, it would also be possible to account for part of the difference between, e.g., *novice* and *new* using a constraint Polysyllabic > Monosyllabic, but this would not distinguish between *novice* and other polysyllabic words, thus requiring some other explanation for the fact that *novice* is not just pitch accented, but nuclear pitch accented. In addition, Polysyllabic > Monosyllabic would predict that *responsible* should behave like *novice.* In fact, *responsible* behaves more like *lead* and *new.*

An additional modification to the grammar is also necessary: N > A must sometimes be ranked below C > F. Given the previous ranking N > A » C > F, other things being equal, the fifth candidate will win instead of the third. Note that A1 > A2 must still be ranked above C > F to avoid selection of an incorrect winner for the item *A natural targets the greediest new investors.* The ideal candidate according to C > F is the one in which every content word is pitch accented, and this is the candidate that would win in Figure 43 (on p. 112). This is the first candidate in Figure 43, and loses to the desired winner only because A1 > A2 prefers the winner.

Figure 45—Adj-N *natural target* item with pitch accent on *natural, target, greediest,* and *novice*

A natural target is the greediest of the novice investors	Vip ∋ Hd	N1 > N2	MaxxRight	FirstxLeft	N > V2	A1 > A2	C > F	N > A	N > V1	*x
x (investors) x x x x x A natural target is the greediest of the novice investors]$_{ip}$		1!				2		3		6
x (novice) x x x x x A natural target is the greediest of the novice investors]$_{ip}$			1!			2		6		6
☞ x (novice) x x x x A natural target is the greediest of the novice investors]$_{ip}$					1	2	5	6		5
x (greediest) x x x A natural target is the greediest of the novice investors]$_{ip}$		1!			1		10	5		4
x (novice) x x x A natural target is the greediest of the novice investors]$_{ip}$					1	2	10!	5		4
x (novice) x x x A natural target is the greediest of the novice investors]$_{ip}$				1!	2	1	10	4		4

At this point, we have a grammar that can derive many of the commonly observed patterns produced in Experiment 2, as shown in Table 20, below. Those that were already checked off in Table 19 have gray check marks; those that can be accounted for only by the modified grammar have black check marks.

Table 20—Outputs accounted for by the revised grammar

fugitive pilot	Adj-N ✓	x x x x A fugitive pilot flew his plane…	
	NP-VP ✓	x x x x A fugitive piloted a plane…	
	RC-VP ✓	x x x x x a fugitive]$_{RC}$ piloted a plane…	
natural target	Adj-N ✓	x x x x x A natural target is the greediest of the novice investors.	
	NP-VP ✓✓	x x x x x A natural targets the greediest new investors…	x x x x A natural targets the greediest new investors…
	RC-VP	x x x x x a natural]$_{RC}$ targeted the greediest investors…	
radical partner	Adj-N ✓	x x x x A radical partner supported the responsible purchase…	
	NP-VP ✓	x x x x A radical partnered with the lead conservative…	
	RC-VP ✓	x x x x x x a radical]$_{RC}$ partnered with the lead conservative…	

The patterns that we cannot derive—including the one majority pattern that we cannot derive (the *natural target* RC-VP item)—generally involve more content words produced without pitch accent than the current grammar can require. This is particularly clear in the case of the RC-VP items, which all have variants in which one content word does not have any prominence above word level; in the case of *a natural]$_{RC}$ targeted the greediest investors,* the variant with *greediest* deaccented is more common than the one the grammar derives. This is the opposite of *a radical]$_{RC}$ partnered with the lead conservative,* which behaves more as the grammar predicts, with *lead* pitch accented much more often when there is a phrase break after *radical* than when there is not. Comparing these two items does not yield a clear, convincing reason for this difference. The most likely explanation is that there are more words between *targeted* and *investors* (the closest pitch accented words before and after *greediest)* than between *partnered* and *conservative* (the closest pitch accented words before and after *lead)*. However, this explanation is not fully satisfactory, as strings of three pitch accent-less words (the length of the lapse that would be produced by failing to pitch accent *lead* in *a radical]*$_{RC}$ *partnered with the lead conservative*[60]*)* are acceptable in the examples reproduced below.

Figure 46—Three examples of commonly produced patterns with three-word lapses

natural target, NP-VP	x x x x A natural targets the greediest new investors…
radical partner, Adj-N	x x x x A radical partner supported the responsible purchase…
radical partner, NP-VP	x x x x A radical partnered with the lead conservative…

Note: These are reproduced from rows 5, 7, and 8 of Table 18/Table 19/Table 20.

Other factors would lead us to expect *greediest* to bear pitch accent more often than *lead*—even leaving aside the assertion that group 1 adjectives attract pitch accent more than group 2 adjectives (A1 > A2), *greediest* is polysyllabic and its main stress syllable is separated from the next main stress syllable by three syllables to *lead's* one. As the two words have the same vowel quality in their main stress syllables, an inherent difference in the loudness of the main stress vowel cannot explain the difference, either. In general, I suspect that the variants with fewer pitch accents per phrase than the current grammar predicts may be satisfying (or partly satisfying) some rhythmic goal that is not

[60] Assuming that the words *partnered* and *conservative* would be the closest pitch accented words.

clear from the sample under consideration, perhaps in combination with some constraints on the relative prominence on different main stress vowels relative to other vowels in their immediate vicinities.

A further remaining issue, obviously, is exactly which factors are responsible for making divisions between the various groups of words used in the X > Y constraints. For example, the idea of incorporating the [±low] difference in the main stress vowel of the verb into the N > V constraints is appealing because different vowel heights do result in differences in loudness, and thus in prominence. Using this one factor to make the division, however, is almost certainly an oversimplification and inaccurate for larger sets of data, both in assuming that main stress vowel height is always important for verbs and in assuming that it is not important for any other category. This is part of a larger issue, which is that this account, like most—perhaps all—others, does not include every factor relevant for determining the prosody of an utterance. Thus, it is not surprising that some of the observed behavior remains outside the capacity of this grammar.

Because the reasons for different frequencies of pitch accenting on each content word by item are not clear, I have not attempted to address differences in same between Experiment 1 and Experiment 2. We can, however, add a single constraint to the grammar and derive the main difference between Experiment 1 and Experiment 2—the deaccenting of W_2 in Experiment 1.

(18) New > Old—New words are more prominent than words already used[61]

[61] Obviously, this is a rough approximation of the status of W_2 relative to the other words in the items. A formal definition of the semantic / pragmatic status of W_2 was not used to construct the Experiment 1 items—they were created using trial and error so that the desired pattern on W_1 and W_2 would be produced more often than not.

The tableaux that follow show how including the goal of making new (or otherwise important) information more prominent than old (or otherwise unimportant) information can change the favored outputs for the items from Experiment 1. Furthermore, ranking this new constraint variably with respect to FirstxLeft can account for the greater difficulty of causing deaccenting of W_2 in the RC-VP condition. In the following tableaux, W_2 is considered to be the only "old" word,[62] with the rest of the content words all "new."

The effect of New > Old is shown first on *a fugitive pilot flew his plane* in Figure 47, then on *...a radical]$_{RC}$ partnered with the lead conservative* in Figure 48, and finally on *...a fugitive]$_{RC}$ piloted a plane* in Figure 49. In order to trigger deaccenting of W_2 in the Adj-N sentences, New > Old must be ranked above N > V2; note that in Figure 47, if this ranking is reversed, the intended winner will be less harmonic than the second candidate in the tableau, which violates New > Old twice but does not violate N > V2 at all.

[62] In these tableaux, W_2 has been marked as "old" in the input as a convenience to the reader, not as a claim about how the relative newness or importance of a word should be represented.

Figure 47—Effect of New > Old on the Adj-N *fugitive pilot* item in Experiment 1

	A fugitive pilot$_{old}$ flew his plane	Vip ∋ Hd·	N1 > N2	MaxxRight	FirstxLeft	New > Old	N > V2	A1 > A2	N > A	C > F	N > V1	*x
	x x x x x A fugitive pilot flew his plane]$_{ip}$					2!	1					
	x x x x A fugitive pilot flew his plane]$_{ip}$					2!				2		
☞	x x x A fugitive pilot flew his plane]$_{ip}$					1	1		1	4		
	x x x A fugitive pilot flew his plane]$_{ip}$				1!	1				4		

Note that the variable ranking of New > Old with respect to FirstxLeft is shown in Figure 47 but is not discussed (or needed) until the example in Figure 49 is treated. In Figure 47, New > Old and FirstxLeft are not in conflict; when the old material is not phrase-initial, the ranking of these two constraints with respect to each other cannot be determined.

In the RC-VP items, however, the old material is ip-initial and the effect of the ranking of these constraints is visible. In these items, W_2 did not deaccent nearly as easily; the ranking shown in Figure 48 results in W_2 with some prominence above the word level, but switching the order of these two constraints will favor a winner with no pitch accent on *partnered.*

Figure 48—Effect of New > Old on the RC-VP *radical partner* item in Experiment 1[63]

	a radical$]_{RC}$ partnered$_{old}$ with the lead conservative	RC = ip	Vip ∋ Hd	*PP	Algin-R (XP, PP)	MaxxRight	New > Old	FirstxLeft	N > V2	N > A	C > F	N > V1	*x
☞	x x x x x x a radical$]_{ip}$ partnered with the lead conservative$]_{ip}$			5			1!						6
	x x x x x a radical$]_{ip}$ partnered with the lead conservative$]_{ip}$			5				1			3		5

While the variable ranking produces the desired variation in the pitch accent status of W_2 in Figure 48, it does not produce such variation in Figure 49, below. The requirement to make old information (W_2) less prominent than new information cannot require deaccenting of the old information when the new information is marked with nuclear pitch accents for independent reasons. However, this result is actually desirable, as W_2 was almost always pitch accented in the example shown in Figure 49.[64] Thus, although some other force must be acting in the grammar to generate the few instances in which *piloted* was successfully deaccented in Experiment 1, in terms of the majority patterns with which this analysis is concerned, the grammar performs exactly as it should in this case.

---

[63] For reasons of space, the constraints N1 > N2 and A1 > A2, which are not relevant for evaluating the candidates in this tableau, have been omitted.

[64] For the remaining RC-VP item, the grammar actually does better for the Experiment 1 item than for the Experiment 2 item, since *greediest* was usually pitch accented in Experiment 1 (but not Experiment 2), and *targeted* usually deaccented, just as the grammar would predict.

Figure 49—Effect of New > Old on the RC-VP *fugitive pilot* item in Experiment 1

A fugitive]$_{RC}$ piloted$_{old}$ a plane	RC = ip	Vip ∋ Hd	*PP	Algin-R (XP, PP)	N1 > N2	MaxxRight	New > Old	FirstxLeft	N > V2	N > A	A1 > A2	C > F	N > V1	*x
x x x x x ☞ A fugitive]$_{ip}$ piloted a plane]$_{ip}$			5											5
x x x x A fugitive]$_{ip}$ piloted a plane]$_{ip}$			5					1				2		4

It is important to note that New > Old is not intended to replace the constraint(s) responsible for positioning focus. A constraint like Büring &Gutiérrez-Bravo's (2001) Focus Prominence, which requires that focused material be most prominent (compared to any other material within its phrase), will still work as intended in the proposed grammar. Depending on exactly how focus prominence is formulated, it may act only within the ip, which is independently required to have a single most prominent member, or it may act at every level of phrasing. In either case, the focused material (or word) can be assigned two x marks in the grid, which will limit the material that shares the phrase with it to a maximum of one x mark regardless of its position relative to the focus. The constraint that requires that the head be the rightmost prominence will then trigger deaccenting, just as Align-R (Hd, ip) would in a grammar in which all prominence is head assignment. As in the example shown in Figure 45 (where the cause of deaccenting is not focus), all that is required is that the constraint driving deaccenting—in the case of focus, Focus Prominence—and MaxxRight both outrank C > F.

Instead, New > Old is a way of incorporating the observation that given material may be deaccented into the grammar. As discussed in section 3.1.3, deaccenting of W_2 was encouraged by context, but the items elicited did not behave as though they contained focused material—W_1, which is both new and somewhat more salient than much of the sentence in the Experiment 1 items, is followed in the NP-VP and Adj-N by a deaccented word, but a pitch accent comes between W_1 and the closest ip boundary in almost all of the items produced.

Note that in order for the RC-VP and NP-VP items to differ in how easily (or frequently) W_2 was deaccented in Experiment 1, FirstxLeft must be split into different constraints for different phrases. In particular, the version active here seems to apply to the ip but not to the pp—thus, to W_2 in the RC-VP items but not (or not with the same force) to W_2 in the NP-VP items. (For diagrams of the relevant structures, see Figure 28 on p. 81.) Because of the nature of strict domination, claiming that FirstxLeft is violated separately for each level of phrasing at which its requirements are not met will be of no help in differentiating the two sets of items.[65] Instead, we can redefine FirstxLeft as follows (see (6) on p. 96 for the original), specifying that the phrases in question are intermediate phrases:

(19) FirstxLeft—The first content word in an intermediate phrase has prominence above the word level. Incurs one violation for every content word between the left edge of the intermediate phrase and the first x mark on the grid

[65] One solution that might allow for preservation of a single FirstxLeft is to claim that the rank of the constraint increases as its violations increase, thus allowing it to outrank New > Old more when it (FirstxLeft) is violated more, as in a Maximum Entropy OT model.

LastxRight can be modified in the same way, keeping the two constraints mirror images of each other. This means that separate constraints of this type presumably exist for each level of prosodic phrasing—the pp, the ip, and the IP. The fact that English privileges the ip in this regard, as well as in the high ranking of ∀ip ∋ Hd, is a coincidence of ranking in the current grammar; there is no single constraint that makes this choice. Thus, it would be possible to construct a grammar that ranked a constraint like ∀ip ∋ Hd high in the grammar but the constraints responsible for positioning prominence near the edges very low, allowing them to be overpowered by nearly every other requirement for the positioning of prominence. The absence of such systems, if they are found to be absent from the range of actually occurring languages, may be attributed either to chance or to other practical considerations that act on the grammar (in much the same way that relatively low rankings of *STRUC are assumed to be universal).

The current grammar does not include rhythmic constraints (e.g. against phrase-internal lapse or clash above the word level), mainly because there are often lapses and clashes at the word level (and, for that matter, at the content word level). I suspect that this is because the main stresses of the content words are often well-separated (either by unstressed syllables within the content word or by function words or both), so that it may be possible to satisfy most of the rhythmic constraints in play even with apparent "clashes" as counted by the word. More puzzling than the tendency to pitch accent most content words are the instances of deaccenting. In general, these do not fulfill any obvious rhythmic goal, giving little incentive to account for them using rhythm constraints. This is not to say, however, that rhythmic considerations do not play an important part in regulating the distribution of above-word prominence in general—just that these particular items are not ideally constructed to show such effects. In fact, the part of these items to which the most attention was devoted—W_1 and W_2—was designed to avoid such effects, thus allowing the "clash" of pitch accenting adjacent content words in Experiment 2.

It is possible that rhythmic factors, or the interaction of rhythmic and other factors, may help explain some of the outputs that the current grammar cannot account for. In particular, the current grammar can account for most of the deaccenting actually seen, but not all of it. In general, the X > Y format constraints cannot require the absence of as many pitch accents in shorter, less complex phrases as they can in longer, more complex phrases. Therefore, RC-VP items produced with unaccented content words are difficult for the current grammar to handle. In other words, there must be more constraints favoring the absence of pitch accent on content words than are included here. It seems doubtful, however, that these considerations are entirely related to rhythm, given the comparison between rows 6 and 9 of Table 18 (p. 84).

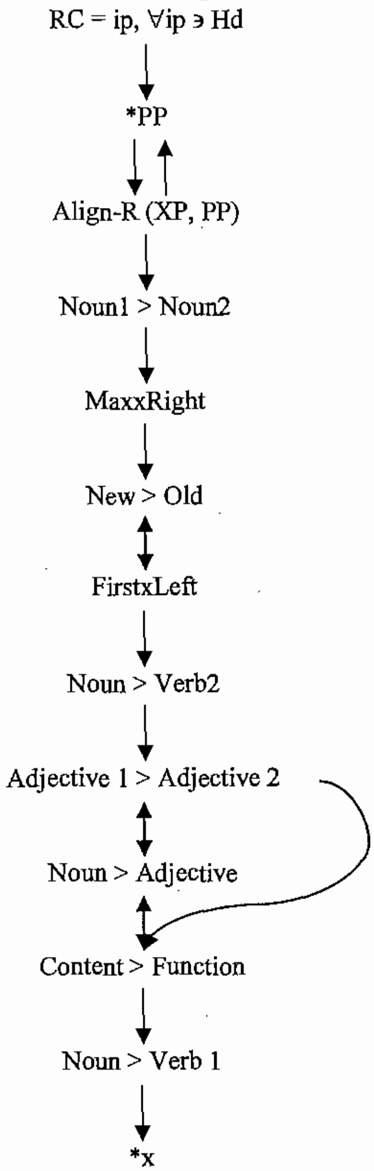
Even more generally, the way that the X > Y constraints are evaluated here means that the more instances of Y there are, the worse it is for any given X to fail to be more prominent than the Ys. Thus, the more function words in a sentence, the worse it is for any of the content words to be unaccented. In this case, the conclusion does not seem entirely unreasonable, as the function and content words are, in most situations, likely to be interleaved, so that unaccented content words will be undesirable from a rhythmic point of view, as well as for the evaluation of C > F. I am less confident, however, that the presence of large numbers of adjectives can render all of the nouns in a sentence more likely to be pitch accented, all else being equal.

Despite these apparent problems, and the certainty that this fragment cannot represent all of the constraints (or even all of the constraint types) needed to account for the prosody of English, the proposed grammar does show how above-word prominence and prosodic phrasing can be generated separately, with the grammar enforcing only some types of relations between them. Re-ranking of these constraints can yield various different types of grammar, including grammars that privilege different levels of phrasing than English does (i.e. not the ip) and those that do not use the grid (that is, pitch accent-like marking) at all. For the latter case, *x is ranked above the X > Y constraints, which may then be satisfied in other ways, as discussed in section 5.3.2

5.2.3 A summary of the grammar

The grammar described above contains fourteen constraints, ranked as shown in the Hasse diagram below. In this diagram, two separate arrows (one pointing each way) are used to represent the principled re-ranking of *PP and Align-R (XP, PP) discussed in section 5.2.2.2; one double-headed arrow is used to represent variable ranking between two constraints.

Figure 50—Hasse diagram showing proposed constraint rankings



The three variable rankings shown are between New > Old (N > O) and FirstxLeft, Adjective1 > Adjective2 (A1 > A2) and Noun > Adjective (N > A), and Noun > Adjective and Content > Function (C > F), all of which are discussed in section 5.2.2.4. Assuming that all three of these variations—plus the principled re-ranking of *PP and Align-R (XP, PP)—are independent from each other, there are twelve total rankings of the grammar.

However, the ranking between N > O and FirstxLeft will be discernable only when there is a difference between new and old information in the output, in which case Align-R (XP, PP) will be ranked above *PP according to the system laid out above (see section 5.2.2.2). Therefore, there are nine visibly different total rankings of the grammar, as shown in Table 21. (Note that Table 21 shows only the rankings of the variably ranked constraints: the rankings of the other constraints are assumed to be constant across all nine of these.)

Table 21—Possible rankings of the variably ranked constraints

*PP » Align-R (XP, PP) (all new or all old info)	Align-R (XP, PP) » *PP (new and old info)
1. N > O »« FirstxLeft, A1 > A2 » N > A » C > F	2. N > O » FirstxLeft, A1 > A2 » N > A » C > F
3. N > O »« FirstxLeft, A1 > A2 » C > F » N > A	4. N > O » FirstxLeft, A1 > A2 » C > F » N > A
5. N > O »« FirstxLeft, N >A » A1 > A2 » C > F	6. N > O » FirstxLeft, N >A » A1 > A2 » C > F
	7. FirstxLeft » N > O, A1 > A2 » N > A » C > F
	8. FirstxLeft » N > O, A1 > A2 » C > F » N > A
	9. FirstxLeft » N > O, N >A » A1 > A2 » C > F

Although there are nine such rankings, there are not nine possible outcomes for any given utterance, both because not every constraint affects every utterance and because the re-ranking of *PP and Align-R (XP, PP) is principled. In other words, there are at most three possible outcomes for the Experiment 2 utterances (the left-hand column of Table 21) and six possible outcomes for the Experiment 1 utterances (the right-hand column of Table 21). These two sets of items are considered separately in the two sections that follow.

5.2.3.1 Outcomes of variable rankings for Experiment 2 items

Because there must be both adjectives and nouns present in an item for N > A to affect the output, the *fugitive pilot* sentences will be very stable. In fact, because FirstxLeft is ranked highly enough to require pitch accent on a phrase-initial adjective regardless of the ranking of N > A and C > F, the *fugitive pilot* sentences will show no variation regardless of the variable rankings, because none of the variable rankings affect them.[66] Thus, within Experiment 2, the grammar produces only the outputs in Table 22.

Table 22—Outcomes for the *fugitive pilot* sentences, all rankings, Experiment 2

Adj-N	x x x x A fugitive pilot flew his plane…
NP-VP	x x x x A fugitive piloted a plane…
RC-VP	x x x x x a fugitive]$_{RC}$ piloted a plane…

The *radical partner* items will show an effect of the re-ranking of N > A and C > F: the non-initial adjectives in these items should be pitch accented when C > F is ranked above N > A but not with the reverse ranking. Because *greediest* is the only (discernibly) class 1 adjective in the script items and is not contained in any of the *radical partner* items, the ranking of A1 > A2 will not affect these script items at all. The possible outputs for these items are shown in Table 23.

[66] The obvious exception to this statement is the principled variation between *PP and Align-R (XP, PP), which affects all of the items.

Table 23—Outcomes for the *radical partner* sentences by ranking, Experiment 2

	N > O »« FirstxLeft, A1 > A2 » N > A » C > F N > O »« FirstxLeft, N >A » A1 > A2 » C > F	N > O »« FirstxLeft, A1 > A2 » C > F » N > A
Adj-N	x x x x A radical partner supported the resp. purchase…	x x x x x A radical partner supported the resp. purchase…
NP-VP	x x x x A radical partnered with the lead conservative…	x x x x x A radical partnered with the lead conservative…
RC-VP	x x x x x x a radical]$_{RC}$ partnered w. the lead conservative…	x x x x x x a radical]$_{RC}$ partnered w. the lead conservative…

The *natural target* items that contain *greediest* and at least one other adjective will have two variants each, not three, because A1 > A2 is always ranked above C > F. The result of this ranking is that *greediest* will always pressure the non-initial class 2 adjective *new* to deaccent in the NP-VP *natural target* sentence. In the Adj-N *natural target* sentence, A1 > A2 has no effect on the actual output because the above-word prominence of the other adjective in the sentence (*natural)* is determined by higher-ranked constraints; therefore, the only variation in ranking that causes variation in the output pattern is between N > A and C > F. The RC-VP *natural target* item is predicted to show no more variability than the *radical partner* items, because it contains only one adjective, so the constraint A1 > A2 will have no effect on it—in other words, there is only one output for this item, even with the three possible sub-rankings.[67]

[67] Note that the most common pattern for this sentence as actually uttered is the one output that cannot be derived by any ranking of the available constraints—that is, the single output that is generated for this sentence by all three rankings, with *greediest* pitch accented, is not correct. Instead, *greediest* is usually not pitch accented in the actual outputs, although the derived pattern does appear as a less common variant.

In sum, although there are three possible rankings for the Experiment 2 items—that is, three possible re-rankings of A1 > A2, N > A, and C > F given the ranking *PP » Align-R (XP, PP)—at most two different patterns are outputted for each sentence.

Table 24—Outcomes for the *natural target* sentences by ranking, Experiment 2

	A1 > A2 » C > F » N > A	A1 > A2 » N > A » C > F	N > A » A1 > A2 » C > F
Adj-N	x x x x x A nat. target is the greedst of the nov. invrs.	x x x x A nat. target is the greedst of the nov. invrs.	x x x x A nat. target is the greedst of the nov. invrs.
NP-VP	x x x x x A natural targets the greediest new investors…	x x x x x A natural targets the greediest new investors…	x x x x A natural targets the greediest new investors…
RC-VP	x x x x x x a natural$]_{RC}$ targeted the greediest investors…	x x x x x x a natural$]_{RC}$ targeted the greediest investors…	x x x x x x a natural$]_{RC}$ targeted the greediest investors…

5.2.3.2 Outcomes of variable rankings for Experiment 1 items

Just as for the Experiment 2 script items, there are actually fewer possible outcomes for each of the Experiment 1 utterances than there are possible rankings: only the RC-VP items are affected by the re-ranking of FirstxLeft and N > O and only the Adj-N and NP-VP *natural target* items may be affected by the ranking of A1 > A2 relative to other constraints (because *greediest* and another adjective must be present for A1 > A2 to have any effect). Note that in the Experiment 1 items, all forms of *pilot, target,* and *partner* are considered old information due to the influence of the preceding context sentence. These words have not been marked as old in the tables that show the different outputs for each possible total ranking for reasons of space.

As shown in Table 25, for the *fugitive pilot* sentences, the only permitted variation that seems as though it might affect the *fugitive pilot* sentences—the variation between N > O and FirstxLeft—does not, in fact result in any variation in output. For an explanation of why this variation in ranking does not affect the *fugitive pilot* RC-VP item, see the discussion above Figure 49 (p. 122). The inability of this re-ranking to affect the Adj-N and NP-VP items assumes, as the earlier discussion of the ranking variation also does (see p. 123), that the FirstxLeft that is relevant in this case must refer specifically to the ip; thus, it will not affect whether pp-initial material in the Adj-N and NP-VP items is pitch accented or not.

Table 25—Outcomes for the *fugitive pilot* sentences by ranking, Experiment 1

	N > O » FirstxLeft	FirstxLeft » N > O
Adj-N	x x x A fugitive pilot]$_{pp}$ flew his plane…	x x x A fugitive pilot]$_{pp}$ flew his plane…
NP-VP	x x x A fugitive]$_{pp}$ piloted a plane…	x x x A fugitive]$_{pp}$ piloted a plane…
RC-VP	x x x x x a fugitive]$_{RC}$ piloted a plane…	x x x x x a fugitive]$_{RC}$ piloted a plane…

As for the Experiment 2 items, A1 > A2 has no effect on the *radical partner* sentences, so there are four rankings that can reasonably be differentiated, but only two outputs per item produced by all four, as shown in Table 26. These outputs vary according to whether *partner* is pitch accented in the RC-VP item and whether the non-initial adjectives *responsible* and *lead* are pitch accented in the Adj-N and NP-VP conditions.

Table 26—Outcomes for the *radical partner* sentences by ranking, Experiment 1 (continues on next page)

	N > O » FirstxLeft, A1 > A2 » N > A » C > F N > O » FirstxLeft, N >A » A1 > A2 » C > F	FirstxLeft » N > O, A1 > A2 » N > A » C > F FirstxLeft » N > O, N > A » A1 > A2 » C > F
Adj-N	x x x A radical partner]$_{pp}$ supported the resp. purchase…	x x x A radical partner]$_{pp}$ supported the resp. purchase…
NP-VP	x x x A radical]$_{pp}$ partnered with the lead conservative…	x x x A radical]$_{pp}$ partnered with the lead conservative…
RC-VP	x x x x x a radical]$_{RC}$ partnered w. the lead conservative…	x x x x x x a radical]$_{RC}$ partnered w. the lead conservative…

	N > O » FirstxLeft, A1 > A2 » C > F » N > A	FirstxLeft » N > O, A1 > A2 » C > F » N > A
Adj-N	x x x x A radical partner]$_{pp}$ supported the resp. purchase…	x x x x A radical partner]$_{pp}$ supported the resp. purchase…
NP-VP	x x x x A radical]$_{pp}$ partnered with the lead conservative…	x x x x A radical]$_{pp}$ partnered with the lead conservative…
RC-VP	x x x x x a radical]$_{RC}$ partnered w. the lead conservative…	x x x x x x a radical]$_{RC}$ partnered w. the lead conservative…

In principle, the *natural target* sentences could show six different outputs. However, because the pitch accent pattern of most of the sentence and the phrasing of the sentence are determined by constraints whose ranking is fixed, there are actually still only two outputs per item, at most, as shown in Table 27 (see p. 137). There are two possible outputs for the RC-VP item—one with pitch accent on *targeted* and one without—depending on whether N > O is ranked above or below FirstxLeft. There are also two possible outputs for the NP-VP item: one without pitch accent on *greediest* and one with, as shown in the tableaux Figure 41, p. 110, and Figure 43, p. 112, respectively.

Given the Experiment 1 ranking Align-R (XP, PP) » *PP, there is only one output for the Adj-N *natural target* item, regardless of which of the six rankings of the freely varying constraints is selected. This is because higher ranked constraints will force two of the three varying constraints—A1 > A2 and N > A—to be violated in the winning candidate. A1 > A2, which is satisfied when *greediest* is more prominent than the other adjectives in the sentence, will be violated in the winning candidate because higher-ranked FirstxLeft and N1 > N2 require *natural* and *novice,* respectively, to bear above-word prominence. N > A, which is satisfied when all the nouns in the sentence are more prominent than all the adjectives, will also be violated in the winning output: N > O requires that *target* be deaccented. *Investors* is also without pitch accent due to the combined effect of N1 > N2 and MaxxRight (see Figure 42, p.110, for tableau). Therefore, neither of the nouns in the sentence bears any prominence above the word, which means that N > A will be equally violated regardless of whether *greediest* has any prominence above the word or not. Thus, C > F will be the deciding factor in whether *greediest* is pitch accented—and C > F favors candidates in which *greediest* is pitch accented over those in which it is not. If there were some other factor in the grammar that would place the nuclear pitch accent on *investors* rather than *novice,* the existing variable rankings in the grammar would produce two outputs with nuclear pitch accent on *investors:* one with prominence on *greediest* and one without.

In summary, because most of the variation in the grammar is in constraints that are relatively low-ranked, the actual variability in output that these rankings are capable of generating is quite limited. If anything, given the variable nature of English phrasing and pitch accent placement in general, there are probably both more constraints and more variable rankings overall than are employed in the current grammar.

Table 27—Outcomes for the *natural target* sentences by ranking, Experiment 1

	N > O » FirstxLeft, A1 > A2 » C > F » N > A	N > O » FirstxLeft, A1 > A2 » N > A » C > F	N > O » FirstxLeft, N > A » A1 > A2 » C > F
Adj-N	x x x x A nat. target]$_{pp}$ is the greedst of the nov. invrs.	x x x x A nat. target]$_{pp}$ is the greedst of the nov. invrs.	x x x x A nat. target]$_{pp}$ is the greedst of the nov. invrs.
NP-VP	x x x x A natural]$_{pp}$ targs the greedst new investors…	x x x x A natural]$_{pp}$ targs the greedst new investors…	x x x A natural]$_{pp}$ targs the greedst new investors…
RC-VP	x x x x x a natural]$_{RC}$ targeted the greedst investors…	x x x x x a natural]$_{RC}$ targeted the greedst investors…	x x x x x a natural]$_{RC}$ targeted the greedst investors…

	FirstxLeft » N > O, A1 > A2 » C > F » N > A	FirstxLeft » N > O, A1 > A2 » N > A » C > F	FirstxLeft » N > O, N > A » A1 > A2 » C > F
Adj-N	x x x x A nat. target]$_{pp}$ is the greedst of the nov. invrs.	x x x x A nat. target]$_{pp}$ is the greedst of the nov. invrs.	x x x x A nat. target]$_{pp}$ is the greedst of the nov. invrs.
NP-VP	x x x x A natural]$_{pp}$ targs the greedst new investors…	x x x x A natural]$_{pp}$ targs the greedst new investors…	x x x A natural]$_{pp}$ targs the greedst new investors…
RC-VP	x x x x x x a natural]$_{RC}$ targeted the greediest investors…	x x x x x x a natural]$_{RC}$ targeted the greediest investors…	x x x x x x a natural]$_{RC}$ targeted the greediest investors…

5.3 Properties of the X > Y constraint type

Some of the constraint revisions and additions required to make the assignment of prominence more independent from phrasing are fairly straightforward modifications to existing systems. One example of this type is the introduction of a separate constraint requiring that prosodic domains have heads, rather than assuming that this is a general property of phrases. However, the apparatus added to the grammar to create prominence that is not in the head position of a phrase is, of necessity, somewhat more of a departure from existing constraints. Although the constraints ARGUMENT-OVER-PREDICATE (Büring & Gutiérrez-Bravo 2001) and HEADARG (German et al 2006) served as models for the more generalized X > Y constraints, there are clear differences between these existing constraints and the new constraints proposed, most notably the ability of the X > Y constraints to compare parts of a sentence that have no local relation to each other. This property and its importance for capturing the data considered here are discussed in the next section. The following section, 5.3.2 compares X > Y constraints to alternative constraint types that might be used for the same purposes.

5.3.1 Long-distance effects of X > Y

A property of these constraints that may be disturbing to some linguists is their ability to compare two parts of a sentence that are arbitrarily far apart and have little or no apparent relation to each other in the syntactic tree—unlike ARGUMENT-OVER-PREDICATE and HEADARG, which rely on a basic syntactic relation between the two parts of the sentence being compared. This non-local property is actually crucial to accounting for some of the data patterns observed in the utterances collected. In particular, the pitch accenting of the subject noun (its assignment of an x mark in the grid) is what triggers the deaccenting of both class 1 verbs (*support* and *fly)* and the adjectives modifying the object noun. The object noun, because it is at the end of the sentence, receives the nuclear pitch accent. As such, it is already more prominent than anything else in the sentence and cannot trigger deaccenting of any other material. Thus, when the subject noun is not pitch accented, as in the Adj-N sentences from Experiment 1, the ability of the grammar to cause other words to deaccent is much less than when the subject noun is pitch accented, as discussed in section 5.2.3. This is because the pitch accent status of both nouns in the sentence has been determined by high-ranked constraints, fixing the number of violations of the constraints N > V1, N > V2, and N > A at a high level, nullifying any influence these constraints might otherwise have had on selecting the winner for this item.

This effect is also clearly at work in the grammar's treatment of the *radical partner* NP-VP and RC-VP sentences, reproduced in Figure 51, below.

Figure 51—NP-VP and RC-VP *radical partner* items

```
                                                 x
               x       x                         x
NP-VP:     A radical partnered with the lead conservative...
               x                                    x
               x           x                x       x
RC-VP:     a radical]RC partnered with the lead conservative...
```

In the NP-VP item, N > A prefers the candidate without above-word prominence on the adjective *lead* because this renders the subject noun *radical* more prominent than *lead.* In the RC-VP sentence, *radical* is the final noun in a relative clause modifying the subject of the larger sentence[68] and thus receives the nuclear pitch accent of the ip that ends at the

[68] The full item recorded was "Someone who was known as a radical partnered with the lead conservative on some key issues."

end of the relative clause. Because *radical* is more prominent than *lead* even when *lead* has some above-word prominence, N > A will not be violated if *lead* is pitch accented. The object noun, *conservative,* is more prominent than *lead* in both of these sentences, by virtue of being last in the sentence.[69]

Although it might prove useful to add some notion of locality to X > Y constraints in a larger, more complex grammar that treats a larger set of data, adding locality restrictions to the current grammar creates more problems than it addresses. This is true of at least syntactic tree based and string adjacency based notions of locality. Restricting the effect of the constraint N > A to a noun and its modifiers would render the constraint essentially useless in the grammar, as it would be violated in every case by the subject-modifiers (due to their initial positions) and could not be violated by object-modifiers except by leaving the object noun without any grid marks and giving the modifier the nuclear pitch accent.[70] Restricting the N > V constraints to purely local application would less problematic, assuming that both the subject and object are considered to be local to the verb. If only the object is considered local, the N > V constraints would, like locally-restricted N > A, lose most of their utility in the grammar. Locality by tree relation also causes problems for C > F: modified for locality, this can only motivate the pitch accenting of content words that are in a local relation with some function word(s). In the examples discussed here, this would predict that all of the verbs should be without

[69] Recall that this is by stipulation—both because I have followed existing practice in assuming that the last pitch accent in a domain is the most prominent and because I have stipulated that the objeet noun would still receive pitch accent if it were actually final in the utterance and thus at all levels of phrasing, even though, as note 68 makes clear, there is material following *conservative* in the actual recorded utterance.

[70] This is the actual configuration observed in the (still rather mysterious) Adj-N *natural target* item, where the modifier *novice* is nuclear piteh accented and *investors* has no above-word prominence.

prominence above the word level unless their tense morphology is considered to be some sort of local function word. While this position would help the grammar to account for the pitch accents on the verbs *pilot, target,* and *partner,* it also suggests that other inflectional morphology should be able to trigger pitch accenting. This seems likely to include the prediction that a plural noun should be more prone to pitch accenting than its singular counterpart in an identical (or near-identical) environment.

String adjacency does not create all of the same problems that tree locality does, but it creates other problems. In particular, while it does not encounter the same problem with verbs that tree locality does, it still cannot derive patterns in which object-modifiers are without pitch accent (except *new* in the NP-VP *natural target* item, which might be deaccented because it is a content word not adjacent to any function word and thus would not be subject to the effect of a string adjacent-only C > F). This modification would not allow the grammar to capture the one sentence it does not currently account for and would add an additional three cases, bringing the number of common patterns that the grammar could not account for from one to four (out of a total of ten).[71] In sum, the non-local quality of the X > Y constraints is somewhat unusual, but it helps the grammar to capture more data than it would with X > Y constraints with locality restrictions, a criterion that merits at least some attention. In addition, it should be noted that while this sort of non-local effect in constraint evaluation may be unusual, it is not, in itself, necessarily bad.

[71] The items that the grammar would lose the ability to account for under this modification are the version of the *natural target* NP-VP item with *greediest* unaccented, the Adj-N and NP-VP *radical partner* items.

5.3.2 X > Y versus alternatives

The main alternatives to X > Y constraints are constraints that either prohibit pitch accent (or grid marks) on words of a particular category, like German et al's (2006) *ACCPREP (do not accent a preposition, p. 165) or require pitch accent on words of a particular category. For convenience, I will use the generic forms *AccY (words in category Y have no marks on the grid) and AccX (words in category X have at least one mark on the grid). Using constraints of the *AccY or AccX form would crucially require rhythmic and other constraints to interact with them. In the case of *AccY constraints, some constraint(s) encouraging the presence of pitch accents would be necessary.[72] In the case of AccX constraints, the other constraints would be devoted to deaccenting. For the latter purpose particularly, rhythmic constraints banning plateaus—sequences of two or more columns of identical heights in a row) of various lengths at various levels of the grid seem like the most likely candidates.[73]

For example, a grammar that uses AccX constraints (AccN, AccV1, AccV2, AccAdj, etc.), *x, and constraints banning long plateaus (more than two columns in a row) at the first level of the grid is a fairly successful alternative to the proposed grammar in terms of data coverage. Given some variation in the ranking of AccAdj and the ban on long plateaus, such a grammar can get most of what the proposed grammar gets, although it cannot capture the one data point that the current grammar misses. In addition, it has difficulty with the pattern observed for the NP-VP *natural target* item, *a natural targets*

[72] German et al's (2006) proposal is concerned with focus, so the constraints used to motivate the presence of accent in their paper are focus-related.

[73] Note that such constraints could also be incorporated into a grammar that uses X > Y constraints.

the greediest new investors, in which both *greediest* and *new* are deaccented. In order to capture this pattern, such a grammar must allow AccAdj to sink below *x, predicting deaccenting of all (non-initial) adjectives. AccAdj will also have to be able to float up above the constraint against long plateaus in order to account for the *natural target* Adj-N item's most common pattern, reproduced in Figure 52 for ease of reference.

Figure 52—Most common pattern for the *natural target* Adj-N item

```
                                          x
    x       x             x               x
A natural target is the greediest of the novice investors.
```

Because the use of the AccX constraints renders the relative prominence of various parts of the sentence more independent from each other, the variation in outputs that is produced by this variable ranking of AccAdj will be greater than the variation in output that results from the variable rankings in the proposed grammar. Thus, an apparent advantage of the AccX constraints—the locality of their evaluation—arguably has a downside. For example, this variation in ranking will allow three outputs for an all-new (or all-old) rendition of the NP-VP *natural target* item, as shown in Figure 53, below. This problem can be alleviated by introducing multiple constraints that distinguish between different lengths of long plateau (e.g. three columns versus four columns versus five, and so on), but this has the unfortunate effect of producing a potentially infinite number of additional anti-plateau constraints, one for each plateau length.

Figure 53—The three possible outputs for the NP-VP *natural target* item, AccX-based analysis

Ranking	*Output*
AccAdj over the ban on long plateaus and *x	x x x x x x A natural targets the greediest new investors…
AccAdj between the ban on long plateaus and *x	x x x x x A natural targets the greediest new investors…
AccAdj under the ban on long plateaus and *x	x x x x A natural targets the greediest new investors…

It should also be noted that accounting for the pattern in Figure 52 (p. 143), some additional apparatus would have to be introduced to move the final pitch accent off of *investors* and onto *novice.* While the proposed grammar also requires an additional constraint to accomplish this, the AccX-based alternative would require a separate type of constraint. (In fairness, it should be noted that this pattern is somewhat odd given the other patterns observed and is therefore likely to create some problems for any grammar that captures the other patterns well.)

Another concern that played a role in the selection of the X > Y format rather than the AccX (or *AccY) format is less driven by data coverage and more by larger concerns about how the grammar relates to our understanding of what prominence is: X > Y explicitly incorporates our understanding of prominence as relative into the formal grammar, which *AccY and AccX do not. Conversely, *AccY and AccX make explicit reference to the grid, while X > Y does not. This is certainly true for the definitions that I have given for these three constraint types, all repeated below for ease of reference.

(20) X > Y—Every word in category X is more prominent than every word in category Y
X is a group of words (e.g. by syntactic category, phonological characteristic, etc.)
Y is a group of words (same range as X)
X and Y do not overlap

(21) *AccY—Words in category Y have no marks on the grid

(22) AccX—Words in category X have at least one mark on the grid

While it would be possible to re-write the schema for X > Y to make explicit reference to the grid, it is not clear that *AccY and AccX can be successfully re-written to refer more generally to prominence without becoming simply Y > Y or X > X. For example, a version of AccX that did not make explicit reference to the grid might be called PromX, defined as in (23).

(23) PromX—Words in category X are prominent

Because prominence is only established relative to something else, it seems that PromX should contain an additional term to be interpreted—either a Y to go along with our X, or an implicit additional term, which will make PromX into X > X (all words in category X are more prominent than words that are not in category X). The only one of the X > Y constraints used that could actually be re-written this way is C > F, since the total set of all words can plausibly be split into either *content* or *function*. Differentiation based on most of the other characteristics that seem to be of potential interest may be difficult to achieve unless we either have two explicit terms (X > Y) or refer to the metrical grid (*AccY or AccX).

In contrast, the schema for the X > Y constraints refers to prominence but not to the metrical grid. While prominence is defined exclusively in terms of the grid for the

English data treated here, there is nothing that explicitly requires this. Thus, in a grammar where *x is ranked above all of the X > Y constraints, the grammar may minimize violations of the X > Y constraints using some means that does not involve the metrical grid—for example, by placing phrase breaks before (or after) words in category X but not before words in category Y, thus placing category X words in more prominent positions. In this way, the X > Y constraints leave open the possibility of capturing in the phonology the observation that "the function of postlexical pitch accent in English and other West Germanic languages is [performed] by prosodic phrasing in 'edge' prominence languages" such as Korean and Japanese (Jun 2005, p. 441).

The possibility of using the X > Y constraints in this way raises the question of whether position at a phrase edge competes directly with grid marking for which confers greater prominence. I have clearly assumed in evaluation of the X > Y constraints in the analysis of the English data that grid marks confer more prominence than phrasal positions. In fact, I have gone further than this and treated English as a system in which only the grid marks are used to determine prominence. This decision, like the decision to regard the final pitch accent in an intermediate phrase as more prominent than other pitch accents in the phrase, has more to do with the representations of the sentences available from coding than with a strong personal or theoretical conviction that it must be so. As a first approximation, we might say that when C > F is ranked above *x, the grid is used in establishing relative prominence and when the reverse is true, phrasal position is used.[74]

[74] Alternatively, we might say simply that phrase-initial position confers less prominence than any degree of grid marking. Although this would make a phrase-initial function word more prominent than a phrase-medial unaccented content word, there would be no change in the number of C > F violations incurred for the items considered here because C > F is violated equally when a function word is greater than or equal to

Using X > Y constraints in this way creates two types of constraints requiring or regulating prominence: those that refer more generally to prominence, whose effects we expect to see regardless of whether a language uses post-lexical pitch accenting or not, and those whose effects we expect to see only in languages that use post-lexical pitch accenting. Examples of this latter type from the above grammar are FirstxLeft and MaxxRight which, with their explicit references to the grid, would have no effect in languages that do not use post-lexical pitch accents. Assuming strict layering and exhaustive parsing, it would be unnecessary to introduce constraints that would ensure the alignment of an accentual phrase with the left or right edge of an intermediate or intonation phrase, so restricting FirstxLeft and MaxxRight to languages that make vital reference to the grid does not predict that accentual phrase languages without post-lexical pitch accenting should have gaps or misalignments at the edges of their larger prosodic units.

The question of exactly which effects should be captured by constraints referring more generally to prominence and which should refer explicitly to the grid is a question beyond the scope of this proposal. However, given the understanding that variations in prominence are accomplished using different prosodic mechanisms in different languages, there is a definite appeal to having constraints that allow us to unify these effects in the grammar. Otherwise, we are in the position of having two sets of constraints (both putatively universal in the most traditional versions of OT) devoted to accomplishing the same goals, one acting on the pitch accent–related representation (the

---

a content word. The other phrase-initial unaccented words in the data considered here are the class 1 verbs (*fly* and *support*), which would not change the winning outputs for the relevant sentences; if anything, this would render these verbs slightly more prominent than the function words in these sentences, decreasing C > F violations and making the winners more harmonic.

grid) used by English and other West Germanic languages and the other acting on phrasing to accomplish the same goals in edge-prominence languages like Korean and Japanese. The fact that the X > Y constraint type presents a possibility for avoiding this situation without either claiming that there are covert accentual phrases in some languages (e.g. English) or covert pitch accents in other languages (e.g. Korean) is at least as much a reason to give them serious consideration as the fact that they make use of our intuitive understanding of prominence as a relative quality.

5.4 Summary

The results of the two experiments presented in chapters 3 and 4 (compared with each other in section 5.1.1) show that it is very unlikely that English has any level of prosodic phrasing whose head is the pitch accent (or some similar degree of prominence). It does, however, appear that English has some level of phrasing larger than a word but smaller than an intermediate phrase. In addition, the appearance of this level of phrasing (and of phrasal boundary after the subject) in Experiment 1 but not Experiment 2 suggests that speakers may elect to use this level of phrasing—indeed, to form more phrases and more levels of phrasing overall—when they are handling more pragmatically complex utterances.

We have also seen that it is possible to form an OT grammar that generates phrasing and prominence separately, while still retaining the ability to relate prominence and phrasing. Such a grammar can still make use of many traditional notions, including headedness and alignment, but includes additional constraints requiring differences in prominence. These prominence-requiring constraints may reference phrasing, as in the case of constraints that require that particular prosodic domains have heads (that is, single most prominent members) or they may be entirely independent of phrasing, as the X > Y constraints are. The X > Y constraints also present the possibility of formally unifying the functions served by pitch accent in head-prominence languages and by phrasing in edge-prominence languages. The question of exactly which parts of the grammar should be used in this way and which parts should be specific to the grid representation or to the tree-based phrasal representation remains a question for future inquiry, as do the ways in which segmental, rhythmic, and even word-specific factors interact with some or all of the proposed constraints.

Appendices

6.1 Full script items from Experiment 1

1. Everyone who operates out of the small, remote air-strip is a certified pilot, but only one of the pilots is a fugitive. Fleeing from the police, the fugitive pilot flew his plane into a mountainous area.

2. Creative piloting of a small plane was involved in a daring escape just last week: Fleeing from the police, a fugitive piloted a plane into a mountainous area.

3. Creative piloting of a small plane was involved in a daring escape just last week: Someone who had become a fugitive piloted a plane into a mountainous area.

4. Picking the right targets can make all the difference in whether a stock scam is successful. For almost all swindles, the natural targets are the greediest of the novice investors.

5. Who a con artist targets can make all the difference in whether a swindle is successful. In almost every case, a natural targets the greediest new investors for his stock scams.

6. Targeting the right person can make all the difference in whether a swindle is successful. A con artist who was a natural targeted the greediest investors for his stock scam.

7. Before the last meeting, only the more traditional of the partners was in favor of the firm buying the new subsidiary. After the last meeting, the radical partner supported the responsible purchase, as well.

8. Unlikely alliances and partnerships among the more extreme members of the council have become more common in the past months. After the last meeting, the radical partnered with the lead conservative on some key issues.

9. Unlikely alliances and partnerships among the more extreme members of the council have become more common in the past months. Someone who was known as a radical partnered with the lead conservative on some key issues.

6.2 Full script items from Experiment 2

1. Fleeing from the police, a fugitive pilot flew his plane into a mountainous area.
2. Fleeing from the police, a fugitive piloted a plane into a mountainous area.
3. Someone who had become a fugitive piloted a plane into a mountainous area.
4. For almost all swindles, a natural target is the greediest of the novice investors.
5. In almost every case, a natural targets the greediest new investors for his stock scams.
6. A con artist who was a natural targeted the greediest investors for his stock scam.
7. After the last meeting, a radical partner supported the responsible purchase, as well.
8. After the last meeting, a radical partnered with the lead conservative on some key issues.
9. Someone who was known as a radical partnered with the lead conservative on some key issues.

6.3 Segmentation criteria

Criteria for sentences 6–3 in preceding Appendices: (ð)ə fjúdʒɪɾɪv pʰáɪlɪˀt

Transition from [ə] to [f]: End of voicing; mostly from spectrogram but waveform and pulse display in Praat also consulted. Additionally, beginning of frication energy.

Transition from [ɾ] to [ɪ]: Increase in amplitude, [ɪ] starts after any evidence of "burst" or frication after the [ɾ].

Transition from [ɪ] to [v]: End or clear perturbation of stable F1, F2, F3 (spectrogram), beginning of frication energy (spectrogram), drop in amplitude (mostly waveform).

Transition from [v] to [p] closure: End of frication energy; both waveform and spectrogram consulted.

Transition from [p] closure to [p] aspiration: Closure ends at the beginning of noise, either as a sharp burst or as the more gradual onset of aspiration when not preceded by a strong ballistic release. For any token with no full closure, aspiration begins at the sharpest increase in amplitude.

Transition from [p] aspiration to [a]: The end of aspiration (and the start of the vowel) is marked at the beginning of voicing and clear F2. When these do not correlate, beginning of regular formant structure is usually the preferred criterion. When F2 itself is not well-defined until well into the vowel, the beginning of stable F1 and F3 is considered an acceptable substitute (to avoid considering the entire first half of the vowel to be part of the stop release).

Criteria for sentences 4–6 in preceding Appendices: (ð)ə nætʃɚl̩ tʰɑɹgə̣ˀt

Transition from [ə] to [n]: Sharp drop in amplitude. Also discontinuity in formant structure, appearance of anti-formants or smearing of formants when [n] is produced without closure (needed for only some of the speakers).

Transition from [ʃ] to [ɚl̩]: End of voicing, end of frication energy, start of formants. The division between [ɚ] and [l̩] is unreliable for segmentation.

Transition from [l] to [t] closure: End of voicing and end of energy in formants (especially F2). When voicing continues after formants, the end of second formant energy is considered the beginning of the stop closure. In some cases, the end of the second formant in the [l] is also the start of some weak frication energy, which is then considered part of the [t] closure. Some of the tokens have frication in the [t]; in these cases, there is generally still a clear division between the frication and the aspiration, and the frication in the [t] is considered part of (or a replacement for) the closure, and is marked as 't-fric' rather than 't-clos' in the annotation of the file. When there is a mixture of closure and frication (or very weak frication, rather than the typical strong frication) between the end of the [l]'s F2 and the beginning of the [t]'s aspiration, this section of the sound file is marked as 't-clos-part'.

Transition from [t] closure to [t] aspiration: Marked at the beginning of the burst for almost all tokens, at the sharp drop in amplitude for the tokens with [t]-frication rather than closure.

Transition from [t] aspiration to [a]: (Same as above.) The end of aspiration (and the start of the vowel) is marked at the beginning of voicing and clear F2. When these do not correlate, beginning of regular formant structure is usually the preferred criterion. When F2 itself is not well-defined until well into the vowel, the beginning of stable F1 and F3 is considered an acceptable substitute (to avoid considering the entire first half of the vowel to be part of the stop release).

Criteria for sentences 7–9 in preceding Appendices: (ð)ə ɹærɪkḷ phɑɹˀtnɚ

Transition from [k] to [ḷ]: Marked at the start of voicing and the start of strong formant energy in the [l] (these generally correlate and, when they do not, the start of formant energy is taken to be the beginning of the [l]).

Transition from [ḷ] to [p] closure: Marked at the end of strong F2 for the [l] (often precedes the end of voicing by a few pulses).

Transition from [p] closure to [p] aspiration: (Same as above.) Closure ends at the beginning of noise, either as a sharp burst or as the more gradual onset of aspiration when not preceded by a strong ballistic release. For any token with no full closure, aspiration begins at the sharpest increase in amplitude.

Transition from [p] aspiration to [ɑ]: (Same as above.) The end of aspiration (and the start of the vowel) is marked at the beginning of voicing and clear F2. When these do not correlate, beginning of regular formant structure is usually the preferred criterion. When F2 itself is not well-defined until well into the vowel, the beginning of stable F1 and F3 is considered an acceptable substitute (to avoid considering the entire first half of the vowel to be part of the stop release).

Bibliography

Ayers, Gayle (1996) "Nuclear Accent Types and Prominence: Some Psycholinguistic Experiments" unpublished Ph.D. dissertation, The Ohio State University, Columbus ,Ohio.

Beckman, Mary (1996) "The parsing of prosody" in Language and Cognitive Processes 11, 17–67.

Beckman, Mary and Gayle Ayers Elam (1997) "Guidelines for ToBI Labeling" Available online at http://www.ling.ohio-state.edu/~tobi/ame_tobi/labelling_guide_v3.pdf as of June 2006.

Beckman, Mary and Jan Edwards (1990) "Lengthening and shortening and the nature of prosodic constituency" in Laboratory Phonology I, J. Kingston and M. E. Beckman (eds.), Cambridge University Press, Cambridge, 152–178.

Beckman, Mary and Jan Edwards (1994) "Articulatory evidence for differentiating stress categories" in Laboratory Phonology III, Patricia Keating (ed.), Cambridge University Press, Cambridge, 7–42.

Beckman, Mary and Janet Pierrehumbert (1986) "Intonational Structure in Japanese and English" in Phonology Yearbook 3, 15–70.

Boersma, Paul and David Weenink (2005) Praat: doing phonetics by computer [Version 4.3]. Downloaded from http://www.praat.org/.

Büring, Daniel, and Rodrigo Gutiérrez-Bravo (2001) "Focus-related word order variation without the NSR: A prosody-based crosslinguistic analysis" in Syntax at Santa Cruz 3, Séamas Mac Bhloscaidh (ed.), 41–58.

Cambier-Langeveld, Tina and Alice Turk (1999) "A cross-linguistic study of accentual lengthening: Dutch vs. English" in Journal of Phonetics 27, 171–206.

Cho, Taehong (2004) "Prosodically-conditioned strengthening and vowel-to-vowel coarticulation in English" in Journal of Phonetics 32 (2), 141–176.

Cho, Taehong (2005) "Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /□, i/ in English" in JASA 117(6), 3867–3878.

Chomsky, Noam and Morris Halle (1968) The Sound Pattern of English, Harper and Row: New York.

Cole, Jennifer, Heejin Choi, Hansook Kim and Mark Hasegawa-Johnson (2003), "The effect of accent on the acoustic cues to stop voicing in Radio News speech" in M.J. Solé, D. Recasens and J. Romero, Editors, *Proceedings of the 15th international congress of phonetic sciences,* Causal Productions, Adelaide (2003).

Davies, Mark. (2008-) Corpus of American English: 360 million words, 1990-present. Available online at http://www.americancorpus.org.

de Jong, Kenneth (1995) "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation" in Journal of the Acoustical Society of America 91 (1), 491–504.

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

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

Fry, Dennis B. (1955) "Duration and Intensity as Physical Correlates of Linguistic Stress" in JASA 35, 765–769.

Fry, Dennis B. (1958) "Experiments in the Perception of Stress" in Language and Speech 1, 120–152.

Gordon, M. (2002) "A factorial typology of quantity insensitive stress" in Natural Language and Linguistic Theory 20, 491–552.

Hayes, Bruce (1989) "The Prosodic Hierarchy in Meter" in Rhythm and Meter, Paul Kiparsky and Gilbert Youmans, (eds.), Academic Press, Orlando, FL, 201–260.

Hayes, Bruce (1995) Metrical Stress Theory: Principles and Case Studies, University of Chicago Press.

Hayes, Bruce and Aditi Lahiri (1991) "Bengali Intonational Phonology" in Natural Language and Linguistic Theory 9, 47–96.

Heldner, Mattias (2001) "Spectral Emphasis as an Additional Source of Information in Accent Detection" in Prosody 2001: ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding, M. Bacchiani, J. Hirschberg, D. Litman & M. Ostendorf (eds.), Red Bank, NJ: ISCA, 57–60.

Jacobs, Joachim (1991/1992) "Neutral stress and the position of heads" in Informationsstruktur und Grammatik (also cited as Linguistische Berichte Sonderheft 4), Joachim Jacobs (ed.), Westdeutscher Verlag, Opladen, 220–244.

Jun, Sun-Ah (1993) "The Phonetics and Phonology of Korean Prosody" unpublished Ph.D. dissertation, The Ohio State University, Columbus ,Ohio.

Jun, Sun-Ah (1998) "The Accentual Phrase in the Korean prosodic hierarchy" in Phonology 15, Cambridge University Press, 189–226.

Jun, Sun-Ah (2005) "Prosodic Typology" in Prosodic Typology: The Phonology of Intonation and Phrasing Sun-Ah Jun, (ed.), Oxford University Press, Oxford, 431–458.

Jun, Sun-Ah and Cécile Fougeron (2002) "Realizations of accentual phrase in French intonation" in Probus 14, 147–172.

Keating, Patricia, Taehong Cho, Cécile Fougeron, and Chai-Shune Hsu (2003) "Domain-initial strengthening in four languages" in Phonetic Interpretation: Papers in Laboratory Phonology VI, J. Local, R. Ogden, and R. Temple (eds.), Cambridge University Press, Cambridge, 143–161.

Khan, Sameer (2006) "Bengali Intonational Structure" Poster presented at the 4th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, Honolulu, HI, 29 November 2006.

Ladd, Robert (1986) "Intonation phrasing: the case for recursive structure" in Phonology Yearbook 3, 311–340.

Nespor, Marina and Irene Vogel (1986) Prosodic Phonology, Foris: Dordrecht.

Norcliffe, Elizabeth and T. Florian Jaeger (2005) "Accent-free prosodic phrases? Accents and phrasing in the post-nuclear domain" in Proceedings of Interspeech, to appear. Downloaded from: http://www.bcs.rochester.edu/people/fjaeger/articles.html 3/4/07.

Okobi, Anthony (2006) "Acoustic Correlates of Word Stress in American English" unpublished Ph.D. dissertation, MIT, Cambridge, Massachusetts.

Oller, D. Kimbrough (1973) "The Effect of Position in Utterance on Speech Segment Duration in English" in Journal of the Acoustic Society of America 54, 1235–1247.

Pierrehumbert, Janet (1980) "The phonology and phonetics of English intonation" Ph.D. dissertation, Department of Linguistics and Philosophy, MIT, Cambride, Massachusetts.

Pierrehumbert, Janet and David Talkin (1992) "Lenition of /h/ and glottal stop" in Laboratory Phonology II, G. Docherty and D. R. Ladd (eds.), Cambridge University Press, Cambridge, 90–117.

Price, Patti, Mari Ostendorf, Stefani Shattuck-Hufnagel, and Cynthia Fong (1991) "The use of prosody in syntactic disambiguation" in Journal of the Acoustical Society of America 90(6), 2956–2970.

Prince, Alan (1983) "Relating to the Grid" in Linguistic Inquiry 14, 19–100.

Prince, Alan and Paul Smolensky (1993) "Optimality Theory: Constraint Interaction in Generative Grammar," Rutgers University Center for Cognitive Science Technical Report 2.

Selkirk, Elizabeth (1980) "On Prosodic Structure and its Relation to Syntactic Structure", Indiana University Linguistics Society, Bloomington Indiana. (Reproduction of a paper presented at the Sloan Workshop of the Mental Representation of Phonology, University of Massachusetts, November 1978.)

Selkirk, Elisabeth (1984) Phonology and syntax: The relation between sound and structure. Cambridge, Mass.: MIT Press.

Selkirk, Elisabeth (1986) "On derived domains in sentence phonology" in Phonology Yearbook 3, 371–405.

Selkirk, Elisabeth (1995) "The prosodic structure of function words", in University of Massachusetts Occasional Papers 18: Papers in Optimality Theory, 439–469. GLSA, University of Massachusetts, Amherst.

Shattuck-Hufnagel, Stephanie and Alice Turk (1996) "A prosody tutorial for investigators of auditory sentence processing" in the Journal of Psycholinguistic Research 25(2), 193–247.

Sluijter, Agaath and Vincent van Heuven (1996) "Acoustic correlates of linguistic stress and accent in Dutch and American English" in Journal of the Acoustical Society of America 100, 2471–2474.

Sluijter, Agaath, Vincent van Heuven, and Jos Pacilly (1997) "Spectral balance as a cue in the perception of linguistic stress" in Journal of the Acoustical Society of America 101, 503–513.

Snedecker, J. and J. Trueswell (2003) "Using prosody to avoid ambiguity: Effects of speaker awareness and referential context" in Journal of Memory and Language 48, 103–130.

Taglicht, Josef (1998) "Constraints on intonational phrasing in English" in the Journal of Linguistics 34, 181–211.

Truckenbrodt, Hubert (1998) "Phrasal stress". Rutgers ms.

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

Truckenbrodt, Hubert (2006) "Phrasal Stress" in The Encyclopedia of Languages and Linguistics, 2nd Edition Vol. 9, Keith Brown (ed.), Oxford: Elsevier.

Turk, Alice and James R. Sawusch, (1997) "The domain of accentual lengthening in American English" in Journal of Phonetics 25, 25–41.

Turk, Alice and Laurence White (1999) "Structural effects on accentual lengthening in English" in Journal of Phonetics 27, 171–206.

Turk, Alice and Stephanie Shattuck-Hufnagel (2000) "Word-boundary-related durational patterns in English" in Journal of Phonetics 28, 397–440.

Venditti, Jennifer (1995) Japanese ToBI Labeling Guidelines. Manuscript with examples, Ohio State University.

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