UNIVERSITY OF CALIFORNIA

Los Angeles

Prenuclear Accentuation in English:

Phonetics, Phonology, Information Structure

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Linguistics

by

Jason Brandon Bishop

2013

ABSTRACT OF THE DISSERTATION

Prenuclear Accentuation in English:

Phonetics, Phonology, and Information Structure

by

Jason Brandon Bishop

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 2013

Professor Sun-Ah Jun, Chair

A primary function of prosody in many languages is to convey information structure—the "packaging" of a sentence's content into categories such as "focus", "given" and "topic". In English and other West Germanic languages it is widely assumed that focus is signaled prosodically by the location of a nuclear pitch accent. As a result, prenuclear, or “secondary” accents are standardly regarded as optional, phonological objects that are unrelated to the information structural representation. This dissertation investigates, from the perspective of the listener, how valid this claim about prenuclear accents is.

As a case study, I consider a putative prosodic ambiguity: the size of the focus constituent in English SVO constructions (i.e., "broad focus" on a VP versus "narrow focus" on an object). My approach to this issue is essentially a three-pronged one, considering the production, perception and processing of prenuclear accents in relation to this contrast. Recent phonetic evidence from

production studies is the starting point for a set of perception experiments (Chapter 2) and a pair of cross-modal priming experiments (Chapter 3). Both sets of experiments provide evidence that listeners have expectations about focus-prenuclear accent correspondences that mirror patterns reported in speakers' productions, suggesting that the broad versus narrow focus contrast is not a genuine prosodic ambiguity. An additional matter that is investigated is the extent to which individual differences in "cognitive processing styles" (autistic traits and verbal working memory) contribute to variation among listeners.

To account for the experimental findings, I argue that the prosodic realization of the size of the focus constituent in English SVOs represents conventionalized, phonological behavior. The variation, it is shown, can be captured by an Autosegmental Metrical model of prosodic structure that includes syntagmatic relations of tonal prominence along the lines proposed by Ladd (1990). This level of "tonal metrical structure" represents linguistically-specified pitch range, and it is demonstrated that such structure is independently needed to account for the prosodic realization of several other information structural contrasts.

The dissertation of Jason Brandon Bishop is approved.

Robert Daland

Maria Luisa Zubizarreta

Patricia Keating

Sun-Ah Jun, Committee Chair

University of California, Los Angeles

2013

To Diana

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Acknowledgements

My time at UCLA has been exhilarating, and for this I owe the greatest thanks to my teachers. I came to the department six years ago, connecting with Sun-Ah Jun right away. Like several before me, I was drawn in by her excitement and genuine, child-like curiosity in all things prosodic. After nearly every single meeting with Sun-Ah—of what must amount to nearly a hundred meetings—I have walked out of her office feeling rejuvenated and more determined than ever. She has never failed to breathe new life into every question, every problem, every project that I have come to her with, somehow convincing me every time that what I am doing is important and fascinating. She will probably never fully know how much I have benefited from her faith in me, and from her interest in my work.

Pat Keating has also been my teacher, and I believe her to be very consciously leading her students by example. She does this to perfection. She and Sun-Ah have taught me essentially everything I know about my field, and to this Pat adds the image of the model scientist and of the model academic. I harbor no illusions that I will be able to replicate her in this regard, but I now begin my own academic career with a highly salient model. I have certainly benefited from her interest in my interests, and frankly I can conceive of no better person to hold together and inspire our lab.

I have had the good fortune to be influenced by a host of other linguists at UCLA as well. I happened to discover information structure the summer before my first year, and had Daniel Büring here to teach me more when I arrived. My experience working with him and Sun-Ah on my MA is in no small way the origin of this dissertation, most of my other graduate work, and many more projects yet to come. In my second year Robert Daland arrived, and I suspect that he

has been one of my greatest advocates in the department, something for which I am quite grateful. Over the past few years he has also made a few remarkably well-timed comments to me, and for this I am also grateful. Numerous other inspirational characters have influenced me in less direct but nonetheless significant ways, including Bruce Hayes, Kie Zuraw, Russ Schuh, and Megha Sundara. I am also eternally grateful to Anoop Mahajan and Nina Hyams, who, as Department Head and DGS, made sure that all of us were well supported all of the time, and that any disasters that I happened to create were swiftly neutralized. Outside of UCLA, and an important member of my committee, is Maria Luisa Zubizarreta. While concentration on a niche-interface area is surely dangerous to one's prospects on the job market (of course I was warned about this), I might never have met Maria Luisa had I been pursuing some more traditional research topic in phonetics. Maria Luisa is a superb human being and linguist, and this dissertation is surely the beginning of my work with her, not the end. Finally, the list teachers ends with the first of them. Beverley Goodman and Daniel Seely at Eastern Michigan University are the source of all this, and I'll never forget their passion for linguistics, or their encouragement and support of me as an undergraduate. They share this distinction with Carla Damiano, also at EMU, and also an early inspiration.

I shared the linguistics graduate student experience with many fine colleagues. First, I thank those whose time in the Phonetics Lab overlapped with mine, all of whom influenced my work in some way at some point: Roy Becker-Kristal, Adam Chong, Marc Garellek, Sameer Khan, Jianjing Kuang, Grace Kuo, Ingvar Löfstedt, Kuniko Nielsen, Nancy Ward, James White, Yu Tanaka, Chad Vicenik, and Kristine Yu. My gratitude also to Henry Tehrani, for providing technical help and various scripts over the years, often on short notice. Elsewhere in the department, I am fortunate to have had my time here overlap with that of Melanie Bervoets,

Dustin Bowers, Philippe Cote-Boucher, Niki Foster, Anne Shisko, and Jun Yashima. Certainly not least of all among my colleagues is Natasha Abner—friend, roommate, officemate, co-conspirator, co-complainer, unintended provider of hard liquor, buyer of bread, lender of interest free loans, and long-time lover of the long shot.

Finally, I am grateful to Diana Hill, without whom I shudder to imagine the last few years—or the future—and to my parents. Their support has allowed me to do all of this, and I have not forgotten that at any point.

VITA

2006	B.A., with honors in Linguistics Eastern Michigan University Linguistics, Psychology, German Language and Literature
2006-2007	Fulbright Fellowship Leipzig, Germany
2006	Research Assistant Neuropsychology Department Max Planck Institute for Cognitive and Brain Sciences
2009-2013	Teaching Assistant Department of Linguistics, University of California, Los Angeles
2008-2009	Graduate Summer Research Mentorship Grant University of California, Los Angeles
2009	Linguistic Society of America Institute Fellowship University of California, Berkeley
2010-2011	Teaching Assistant Coordinator Department of Linguistics University of California, Los Angeles
2011	Dissertation Year Fellowship University of California, Los Angeles
2011	Ladefoged Scholarship Department of Linguistics University of California, Los Angeles

PUBLICATIONS AND PRESENTATIONS

Bishop, J. & Toda, K. (to appear). Syllabication, sonority, and segmentation: evidence from word-spotting. *Proceedings of the 47th Meeting of the Chicago Linguistics Society*.

Bishop, J. (2013). “Individual differences in the perception and processing of prosody”. Invited talk given to the Hearing Seminar, Center for Computer Research in Music and Acoustics, Stanford University, Spring 2013.

Jun, S-A. & Bishop, J. (2013). Implicit phrasing and prominence influence relative clause attachment: evidence from prosodic priming". Paper presented at the West Coast Conference on Formal Linguistics 31, Arizona State University, 8-10 February 2013.

Jun, S-A. & Bishop, J. (2013). “Implicit prosodic priming in relative clause attachment”. Paper presented at The 87th Annual Meeting of the Linguistic Society of America, Boston, MA, 3-6 January 2013.

Bishop, J. & Keating, P. (2012). Perception of pitch location within a speaker's range: Fundamental frequency, voice quality and gender. *Journal of the Acoustical Society of America* 132(2), 1100-1112.

Bishop, J. (2012). “Information structure, prosodic prominence and individual differences: evidence from on-line processing". Paper presented at Generative Linguistics in the Old World 35 *(Workshop on the Production and Perception of Prosodically-Encoded Information Structure)*. University of Potsdam, Germany, 27 March 2012.

(2012) Bishop, J. “The role of listeners' autistic traits in the interpretation of prosody and contrastive Focus”. Poster presented at the Conference on Sources of Individual Linguistic Differences, University of Ottawa, Canada, 2-4 March 2012.

(2012) Bishop, J. “The role of prosody and information structural interpretation in lexical processing: evidence from cross-modal associative priming”. Paper presented at the 86th Meeting of the Linguistic Society of America, Portland, Oregon, 5-8 January 2012.

Bishop, J. (2012). Information structural expectations in the perception of prosodic prominence. In G. Elordieta and P. Prieto (eds.) *Prosody and Meaning* (Interface Explorations). Berlin: Mouton de Gruyter.

Bishop, J. (2012). Focus, prosody, and individual differences in “autistic" traits: evidence from cross-modal semantic priming. *UCLA Working Papers in Phonetics* 111, 1-26.

Bishop, J. (2011). English listeners' knowledge of the broad versus narrow focus contrast. *Proceedings of the International Congress of Phonetic Sciences XVII*, 312-315.

Bishop, J. (2011). “Prosodic prominence and information structural interpretation in on-line lexical processing”. Poster presented at Experimental and Theoretical Advances in Prosody 2, McGill University, 23-25 Sept. 2011.

Bishop, J. (2011). “English listeners' knowledge of the broad versus narrow focus contrast”. Paper presented at the International Congress of Phonetic Sciences XVII, Hong Kong, 17-21 August, 2011.

Bishop, J. and Toda, K. (2011). “Syllabification, sonority, and spoken word segmentation: evidence from word-spotting”. Paper presented at the 47th Meeting of the Chicago Linguistics Soc., University of Chicago, 7-9 April 2011.

Bishop, J. (2010). “Voice quality and the perception of f0 range location”. Poster presented at the 159th Meeting of the Acoustical Society of America, Baltimore, Maryland, 19-23 April 2010.

Bishop, J. (2010). “Narrow focus as prominence lending”. Paper presented at the West Coast Conference on Formal Linguistics (Special Session on Prosody in Grammar, Acquisition and Processing), University of Southern California, 19-21 February, 2010.

Abner, N. & Bishop, J. (Eds). (2009). *Proceedings of the 27th West Coast Conference on Formal Linguistics*. Cascadilla Press.

Jun, S.-A. and Bishop, J. (2009). “An acoustic study of the /w/ approximant in Korean”. Poster presented at the 157th Meeting of the Acoustical Society of America, Portland, Oregon, 18-22 May 2009.

Bishop, J. (2008). The effect of position on the realization of second occurrence focus. *Proceedings of Interspeech 2008*, 869-872.

Bishop, J. (2008). “Effects of coarticulatory timing on vowel nasalization in English: an aerodynamic study”. Poster presented at the 156th Meeting of the Acoustical Society of America, Miami, 10-14 November 2008.

Wake, M., Bishop, J., & Panganiban, C. (2008). “The production and perception of second occurrence focus”. Poster presented at the 156th Meeting of the Acoustical Society of America, Miami, Florida, 10-14 November 2008.

Bishop, J. (2008). “The effect of position on the realization of second occurrence focus”. Paper presented at Interspeech 2008, Brisbane, Australia, 22-26 September 2008.

Lao, J. and Bishop, J. (2008). “A cross-dialectal perceptual study of Eastern Andalusian coda neutralization”. Poster presented at the Linguistic Society of America Summer Meeting. Columbus, Ohio, 10-13 July 2008.

Bishop, J. (2007). Incomplete neutralization in Eastern Andalusian Spanish: perceptual consequences of durational differences involved in s-aspiration. *Proceedings of the International Congress of Phonetic Sciences XVI*, 1765-1768.

Bishop, J. (2007). “Incomplete neutralization in Eastern Andalusian Spanish: perceptual consequences of durational differences involved in s-aspiration”. Paper presented at the International Congress of Phonetic Sciences XVI, Saarbrücken, Germany, 6-10 August 2007.

Bishop, J. (2007). “Phonetic explanation in phonology”. Poster presented at the 15th Manchester Phonology Meeting". Manchester, United Kingdom, 24-26 May 2007.

Bishop, J. (2007). “A perceptual study of incomplete neutralization in Andalusian Spanish”. Poster presented at the Old World Conference in Phonology 4. Rhodes, Greece, 18-21 January 2007.

Bishop, J. & Goodman, B. (2006). “The production and perception of American English /p,t,k/”. Paper presented at the 80th Meeting of the Linguistic Society of America, Albuquerque, New Mexico, USA, 4-6 January 2006.

Bishop, J. & Goodman, B. (2006). “Auditory distinctiveness and the perception of voiceless stop consonants in continuous speech”. Poster presented at the 150th Meeting of the Acoustical Society of America, Providence, Rhode Island, USA, 5-9 June 2006.

CHAPTER 1

Prosody, prominence, and focus marking in English

1.0 Introduction to the dissertation

One of the functions of prosody is to express how the information in a sentence relates to a larger discourse context. For example, a sentence's focus, or the pragmatically important or informative part of an utterance, can be signaled by the location of a prosodic prominence, such as an intonational pitch accent, as in (1):

(1) a. JOHN bought a motorcycle.

b. John BOUGHT a motorcycle.

In (1), where capital letters indicate the "nuclear", or "primary" accent in each sentence, *John* is unambiguously the focus in (a) and *bought* is the focus in (b). In this dissertation, I will be exploring how prosodic prominence is able to signal information structural meaning such as focus in this way.

One of the more specific and central questions I will be investigating is how focus is conveyed by different levels of prominence. In intonational phonological models of English, such as that developed by Pierrehumbert (1980) and Beckman and Pierrehumbert (1986), as well as related ones in Ladd (1996), and Gussenhoven (1984), there is a fundamental distinction made between nuclear accents and prenuclear accents. A nuclear accent is the last pitch accent in an intermediate phrase, and roughly corresponds to the nuclear stress of Chomsky and Halle (1968). The nuclear accent is sometimes referred to as the "primary" accent, both because it is considered to be the highest level of structural, phonological prominence, but also because its function in English is closely tied to expressing the location of the sentence's focus constituent. The examples in (1), above, demonstrate the close relation between nuclear accent location and focus location.

Much less well understood, and more tenuous, is the relation between prenuclear accents, sometimes referred to as "secondary accents", and focus. For example, consider the possible pronunciations of an answer to the question in (2) when focus is on *motorcycle* (nuclear accents are shown in capitals, prenuclear accents in italics):

(2) Q: What did John buy?

A1: *John* bought a MOTORCYCLE.

A2: John *bought* a MOTORCYCLE.

A3: *John bought* a MOTORCYCLE.

Here it does not seem that the reliable alignment of focus and pitch accent that holds for nuclear accents holds for prenuclear accents, since they all seem, at least impressionistically, to be felicitous pronunciations of sentences with focus on *motorcycle*. That is, focus seems to be marked by nuclear accents, and the presence and location of prenuclear accents is optional.

In this dissertation, I will be considering just how optional prenuclear accents really are. My goal will be to better understand when speakers produce them, and, especially, how listeners interpret them. In so doing, I will consider the contrasts in the size of the focus constituent, such as (3):

(3) Q1: What did John do?

Q2: What did John buy?

A: John bought a MOTORCYCLE.

The case in (3) has been of great interest in studies of English intonation and focus prosody because the answer sentence (3A) has the same syntactic structure and the same nuclear accent placement, but is assigned a different information structure depending on whether it is produced in response to Q1 versus Q2. As an answer to Q1, which asks about an event carried out by *John*, the focus is the entire verb phrase (VP) *bought a motorcycle*. In response to Q2, however, which asks about an object of a buying event, focus is on only the object, *a motorcycle*. This (pragmatically dependent) semantic contrast will be the case study in this dissertation, and the primary question that I will ask is whether prenuclear prominence correlates with this distinction. While the question seems to be a simple one, it turns out to present problems for phonological models of English intonation, as well as theoretical models of the prosody-information structure interface. For these reasons, we stand to learn from investigating more carefully speakers' use of, and listeners' expectations for, prenuclear accentuation.

In the first sections (1.1-1.2) I describe the preliminaries for studying prosodic prominence, and the basic model of English intonational phonology that I am assuming. In section 1.2.3 I also discuss some ways in which individual differences might influence how prosodic prominence is perceived and processed. In Section 1.3 I describe basic notions concerning information structure and focus, and how focus is claimed to be encoded in prosodic structure. I will then consider more detailed experimental evidence for the relation between focus and prosody in Section 1.4, and describe how this issue will be explored further in the rest of the dissertation in Section 1.5.

1.1 Prominence in Autosegmental Metrical theory

Prominence is a complex and multifaceted concept, conceived of in different ways by different researchers at different times. However, the concept of a primary "sentence stress" and additional "secondary" prominences has a long history in the study of English prosody. In earlier models in the British tradition (e.g., Palmer 1922, O'Connor and Arnold 1973), an utterance was defined minimally by a "nucleus", or a most prominent stress and pitch movement, which was optionally preceded by a prenuclear stretch of syllables, or the "head"[1], and a postnuclear part of the contour called the "tail". A similar division of the contour into separate regions was eventually also utilized by researchers at the Institute for Perception Research (IPO; t'Hart and Collier (1975), 't Hart, Collier, and Cohen (1990), who distinguished the root, prefix and suffix. For both these British and Dutch (at least after 't Hart, Collier, and Cohen 1990) approaches, the nucleus marked the most prominent syllable, usually the last (lexically) stressed syllable in a sentence. Thus, there is a long history for the intuitive notion of a "most prominent" sentence accent, one that in some sense serves as the "center" for a more complex unit in which the prenuclear stretch is an independent part. Although not making reference to the intonational contour in the way of the British tradition, the concept of a maximally prominent "nuclear" stress is also central to theories of metrical phonology (notably Chomsky and Halle 1968).

The basic model I assume in this dissertation represents a significant departure from these earlier conceptions of the structuring of sentence prosody, namely the Autosegmental Metrical model of Pierrehumbert (1980), and more specifically the revised model for English intonational phonology in Beckman and Pierrehumbert (1986). According to this basic approach (see also Ladd 2008 and Gussenhoven 1984), the intonational contour is not a single unit divided into

[1] Subsequently to Palmer (1922) it was also commonly assumed that the head was further divided so as to include a "prehead", which served to explain the early pitch movement/secondary sentence stress frequently found in very early in an utterance in which the nucleus occurred on a relatively late syllable (see Kingdon 1958).

Table 1.1. Inventory of boundary tones, phrase accents and pitch accents in the current intonational phonological model for English based on Pierrehumbert (1980) and Beckman and Pierrehumbert (1986).

Boundary Tone	Pitch Accents	Phrase Accent	Boundary Tone
H%	H*	H-	H%
L%	L*	L-	L%
	L*+H		
	L+H*		
	!H*		
	H+!H*		

regions, but is derived from a sequence of level low (L) and high (H) tones, aligned with syllables and connected by simple interpolation. In the phonological model for English, there is a relatively small inventory of tonal categories in paradigmatic opposition to one another, as shown in the Table 1.1. The tones fall into three basic larger categories: pitch accents, which align with lexically stressed syllables and mark prominence, and two kinds of edge marking tones, which can align with either stressed or unstressed syllables, but do not mark prominence. The edge-marking tones are phrase accents (marking the intermediate phrase (ip) level of phrasing), and boundary tones (marking the higher-level intonational phrase, (IP) level of phrasing). This basic model is also the basis for the Tones and Break Indices (ToBI) conventions for the prosodic annotation of Mainstream American English (Beckman and Ayers Elam 1997; Beckman, Hirschberg, and Shattuck-Hufnagel 2005).

In this AM model, however, a distinction between nuclear and prenuclear material is still made, although it is understood somewhat differently. Rather than referring to particular regions, the nuclear accent is the last accent in an intermediate phrase, and a prenuclear accent is any pitch accent preceding the nuclear accent in the same intermediate phrase. Although it is defined strictly by position in structure, i.e., by its phrase-finality, it is recognized that the nuclear accent has a special status in structure, and represents the highest level of phonological prominence, and

```
                              x                         nuclear pitch accent
           x                  x                         pitch accent
           x        x         x                         (primary) lexical stress
       x x x    x   x     x   x   x x                   syllable nucleus
     M a r i a n n a  made  the  marmalade )ip )IP
           |                  |          |   |
           H*                 H*         L-  L%         tonal tier
```

Figure 1.1. Prominence structure and the intonational contour in the AM model.

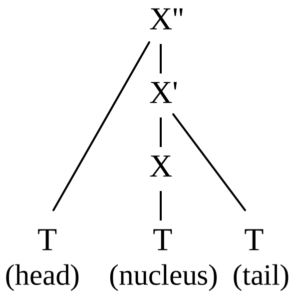
often phonetic prominence as well. Thus, whereas in the earlier British and Dutch traditions described above the nuclear was defined by its prominence, and its position in the sentence was largely unexplained, in the AM models assumed here, the nuclear accent is defined by position in structure, and the source of its special prominent status is less clear. Notably, its prominence representation does not in any way relate to the tonal tier.

The basic assumption about levels of prominence, and crucially which aspect of structure encodes this phonological prominence is sketched out in Figure 1.1. An important aspect of this phonological model of prominence structure is that it is intended to predict phonetic realization. The most important distinction is perhaps that between the lexically-specified level of prominence occurring at the word level, namely lexical stress. In English, the primary cues to word-level prominence are (a) duration and intensity (together contributing to loudness) and (b) vowel quality (full or reduced) (Beckman 1986). Above the level of the word, the primary cue is f0; not a specific level of f0, but the alignment of an f0 target with a syllable that is prominent at the previous, lexical level of structure. Then, as stated above, status as the last pitch accent in a linear sequence within an ip defines the highest level of structural prominence, the nuclear accent.

What is also important to note about the structuring of phonological prominence is that there is no place in the prominence structure for the tonal representation itself (which also distinguishes it from the earlier whole-contour approaches described above). That is, there is nothing inherent in the phonological structure that assigns a particular prominence value to a H versus a L tone associated with an accented syllable, nor is there any modification of the pitch range of a H target that has phonological prominence status. That is, although the model does contain the downstepped !H* accent, which is defined by a lower pitch range than a preceding H target, this does not enter into its structural prominence in any way. This is not to say, however, that downstepped accents are equally *phonetically* prominent to non-downstepped ones; by definition they have a reduced pitch range, and it has also been shown that they tend to have less of the ancillary durational enhancement enjoyed by full-fledged accents, leading downstepped accent to be perceptually less prominent (Ayers 1996). However, as far as the phonological model itself is concerned, these are just properties that accompany the paradigmatic "!H*" accent; they do not have a specified place in the metrical prominence structure in Figure 1.1, and pitch range itself does not have any phonological status in this AM model.

However, one of the questions we will address later is whether or not this is a desirable aspect of the model, and it has been suggested that it is not. Ladd (1990; see also Ladd 1993 and 2008), for example, has proposed two revisions to the basic Pierrehumbert/Beckman and Pierrehumbert model, both of which allow us to maintain the advantage of a well-defined phonological structuring of phonetic cues while taking from the earlier whole-contour approaches some of their more intuitive properties. Both of these involve adding structure to the tonal tier. The first organizes the nuclear, prenuclear, and postnuclear regions of the contour—although still conceived of as their AM level tones or tone sequences—into a structure that is very much in the spirit of the British tradition. Ladd's conception of this organization is shown in (4):

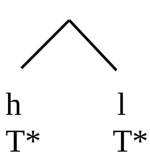
(4)



The obvious consequence of this structure is to make explicit that the nucleus has some central prominence status in addition to being part of a linear sequence. Additionally, it also makes explicit that prenuclear accents (making up the head) are more structurally related to each other than they are to the nucleus—a desirable result given that, when multiple prenuclear accents are present in a sentence, they are usually all of the same type (see Ladd 1986, and the discussion in Ladd 2008: 283). While it is not possible (or of interest) to discuss all of Ladd's motivations for identifying prenuclear and nuclear constituents in this way, suffice it to say that it is not an inevitability of the AM model that a linear ordering of level tones precludes any such additional hierarchical organization.

The second proposal of Ladd's is to let the tonal tier contribute to phonological prominence. He proposes that (in addition to their paradigmatic properties) pitch accent type categories should be permitted to have syntagmatic relations. Just like the strong-weak relations that characterize prominence at lower levels of structure (i.e., the lexical and phrasal prominence levels in Figure 1.1), tones have high (h) and low (l) relations with one another, where h and l make reference to pitch range. This kind of "tonal prominence" is represented in a binary-branching metrical tree:

(5)



The principal function of this structure in the model is to represent downstepped accents, and do so in a way that, first, raises pitch range to a linguistic status, and, second, gives to pitch range an explicitly syntagmatic nature as well. The h-l structure can be thought of as a kind of prominence, since pitch range is well known to be an effective prominence lending cue (e.g., Gussenhoven and Rietveld 1988, Dilley 2005). However, Ladd makes clear that this level of structure is in addition to the lower level weak-strong relations. A number of questions appear to be left unspecified in this theory, however, such as what the domain of this structure is, i.e., what level of phrasing does it define. It is clear that it is not the intermediate or intonational phrases, since a primary purpose of this level of structure is in fact to phonologically encode relations across these phrases. While this is a matter that would need to be resolved, Ladd's theory represents a serious attempt to combine the AM model with a phonological model of pitch range, allowing, in a sense, for the tonal tier to contribute a level of phonological prominence.

To summarize, prominence here is assumed to be an aspect of phonological structure; at the phrase level, nuclear accents are the highest level of phrasal prominence, and prenuclear accents are structurally less prominent. The distinction is one of linear order in an intermediate phrase. Tonal targets align with these two levels of prominent syllables, i.e., pitch accents, and also with the edge of phrases to mark ip and IP phrase boundaries. In standard AM theory based on Pierrehumbert's (1980) model, and the revisions in Beckman and Pierrehumbert (1986), the version on which the ToBI conventions are based, the tonal tier does not represent any kind of phonological prominence, although such a notion has been proposed by Ladd (1990).

1.2 Perceiving prominence

1.2.1 Predicting prominence from the signal

It has long been of interest to know which factors contribute to a listener's perception of prosodic prominence, and considerable effort has been devoted to understanding how structural prominence might be cued by acoustic prominence. In early literature on this subject in English, the question was often approached in such a way that emphasized prominence at the word level, confounding it with sentence/phrase level prominence. Thus, for example, Fry (1955) demonstrated that f0 is an extremely reliably cue to identifying lexical stress in pairs such as *pérmit* and *permít*, and duration and intensity paled in comparison. While this is true when such words are pronounced in isolation, in the present AM model, it is understood as due to the fact that the lexically stressed syllable will also carry the phrase-level prominence of a nuclear accent, marked by an f0 target. That is, while f0 is marking prominence in such cases, it is not the direct acoustic correlate of word-level prominence, but of higher-level prosodic organization.

Further evidence that f0 is not a primary perceptual correlate of word-level prominence per se comes from Beckman (1986), who showed that native English speakers relied primarily on intensity and duration (crucially, a combination of the two), and very little on f0 when presented with manipulated acoustic versions of *pérmit* and *permít*-type pairs. Native speakers of Japanese, a pitch accent language, on the other hand, showed the inverse pattern. Additionally, there was some evidence from English-Japanese bilinguals that showed that the importance of f0 increased in L1 native English speakers only when they had extensive L2 Japanese experience, suggesting that a heightened use of f0 for word-level prominence f0 was dependent on L2 interference. It has also been shown that duration and intensity (again, a combination of the two, although duration was found to be the most important) are predictors of syllable stress when f0 is flat across a disyllable such as *mama* (Turk and Sawusch 1996). Thus the phonological structuring of phonetic prominence is a fundamental aspect of the AM model, and is supported by previous research. The most important assumption that I take from it is that f0 is the primary cue to prominence at level of prosodic structure above the word. Duration and intensity, although they may occur as ancillary enhancement of prominence at higher levels, primarily define lower, lexically-specified, word-level prominence.

In fact, the chief interest of the present dissertation is higher, phrase/sentence level prominence, and how it is perceived by listeners. One experimental approach to this problem has examined the role of f0 in cuing the relative prominence of two separate f0 peaks in a string of syllables. These types of studies all demonstrate that listeners' judgments of prominence vary along with f0 levels for the individual peaks such that higher f0 peaks tend to correspond to greater perceived prominence (Rietveld and Gussenhoven 1985), with some important but predictable qualifications, such as the listeners' correction for declination (Pierrehumbert 1979) and certain complex interactions between the relative prominence of the two peaks (Gussenhoven and Rietveld 1988; Ladd 1993) and their ordering (Jagdfeld and Baumann 2011). Thus listeners are shown to be quite sensitive to f0 when making prominence judgments (especially relative prominence judgments) that is probably best regarded as prominence above the word-level.

In stark contrast to the sensitivity that these previous studies report are the findings of a more recent line of research utilizing corpus methodology. For example, Kochanski, Grabe, Coleman, and Rosner (2005) modeled the binary prominent/nonprominent judgments of native English-speaking listeners, most of whom were phonetically-trained, using materials from the IViE (Intonational Variation in English) corpus as stimulus materials (Grabe, Post, and Nolan

2001). Testing five acoustic phonetic parameters, including approximations of perceptual loudness (based on intensity and amplitude), segmental duration, periodicity of voicing, spectral slope (cf Sluijter and van Hueven 1996) and f0. Their results showed that all measures were relevant to predicting prominence judgments at a statistically significant level, but loudness was found to be most important, and f0 accounted for very little in comparison.

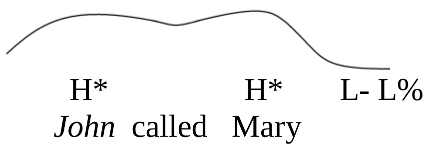
This finding is echoed in a subsequent study by Cole, Mo, and Hasegawa-Johnson (2010), who tested only phonetically/linguistically naïve listeners (see also Mo 2008). These listeners also made prominent/non-prominent decisions, and stimulus materials consisted of excerpts from the Buckeye Corpus (Pitt et al. 2007). Using a similar set of acoustic predictors similar to Kochanski et al's, Cole and colleagues, also found f0 to be a quite poor predictor of prominence judgments in relation to the other predictors, although they find duration, rather than intensity to be the most relevant. Notably, these studies seem to echo, in stronger form, findings from similar studies reported in the automatic speech recognition literature, which suggest that, although f0 is relevant, factors such as duration and intensity increase prominence prediction considerably (e.g., Conkie, Ricardi, and Rose 1999; Chen and Hasegawa-Johnson 2004; Sridhar, Narayanan, Nenkova, and Jurafsky 2008).

When considering previous studies of prominence perception by human listeners, it is apparent that methodologies are often quite different. While it is clear that f0 can be used by listeners to make prominence judgments, the question that remains is why f0 plays such a weaker role in corpus-based studies, which often make use of spontaneous (or perhaps more accurately, "unscripted") speech. One difficulty may have to do with the role that relative measures are permitted to play; while this is of clear importance to measures of intensity/loudness, which have strictly relative importance, it should also be of importance to the interpretation of f0 as well, particularly for speakers who utilize a rather compressed overall range. Thus, f0 may become a more useful predictor if a relative measure of a target syllable against a background global or (utterance-wide) pitch range is used. This is somewhat supported by Rosenberg, Hirschberg and Manis (2010), who included such measures, although their study tested non-native English-speaking listeners, making it difficult to compare their results directly with the others.

Another methodological issue, however, is likely a much deeper one. As noted above, f0 seems to be less relevant when modeling binary prominent/non-prominent decisions rather than explicitly or implicitly comparing two f0 peaks, and using a gradient rating scale to do so. It is also possible that f0 may be more useful in deciding *levels* of prominence above some minimal level of accentuation, a possibility that is pointed out by Calhoun (2006; 91). This would help explain why f0 seems to be most robust when comparing two f0 peaks; presumably, both peaks represent categorical prominence, i.e., accents, to listeners, and f0 height then serves as difference in phonetic prominence on top of those categories. Such a comparison also makes more intuitive sense than judgments on f0 for two syllables for which the structural prominence is differs: i.e., comparing a syllable that is non-prominent at the word-level (i.e., a reduced vowel) with a syllable that is prominent at the level of nuclear accentuation. In such a case, reminiscent of Fry's paradigm, a specific f0 value may or may not be most important, depending on (a) what f0 target the accented syllable was assigned, H or L, and (b) the value of the unaccented syllable's f0, which could also be high or low—not because it is specified for prominence, but due to the previous f0 target's value, which is interpolated through the unaccented syllable in question.

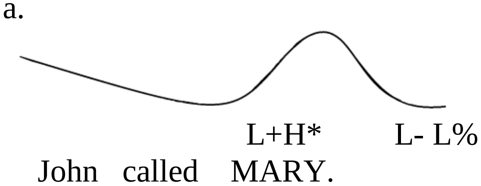
One way to see this kind of problem is to consider a sequence of two H* accents with an unaccented word in-between them. In this case, the f0 on the middle word is relatively high for the speaker, and similar in absolute value to the adjacent accented syllables' f0. For example, in (6), *John called Mary*, each word has a f0 value:

(6)

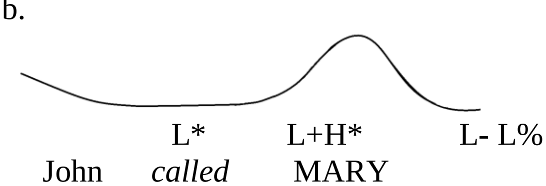


Nonetheless, *called* it is not predicted by the intonational phonological model to be prominent, since it lacks an identifiable f0 target of its own. A similar situation arises when an unaccented syllable precedes a syllable bearing a L+H*, as in (7a); in that case, the unaccented syllable will have a relatively low value, differing only slightly if at all from a case where it is accented with a L*, as in (7b):

(7) a.



b.



Examples like these are ones that Beckman (1996) points out as ambiguities for human ToBI annotators, who in such cases must make use of only ancillary durational and intensity cues to make decisions about the accent status of words in question (see also Beckman and Ayers 1997 and Ladd 2008: 261). However, for human annotators (and presumably ordinary perceivers), the ambiguity is only for the syllable which lacks an identifiable turning point, while for a statistical model, the fact that unaccented and accented syllables can have the same absolute mean/max/min f0 value reduces the predictive power of f0 for all syllables. Thus, studies like some of those just described do not accurately model what the listener is doing in most cases, which is not just identifying f0 values, or even H or L targets, but identifying those targets that align with lexically stressed syllables.

One final comment here is to note that, although our analytical tools require us to decompose the signal into such independent measures as f0, duration, intensity, spectral cues, and so forth, listeners do not hear them this way. Rather, they hear all cues to prominence in unison, and so in this sense, all approaches to testing prominence perception discussed fall short of modeling what the listener is truly doing. This is, of course, not a problem specific to modeling the perception of prominence, but one general to modeling the perception of any multidimensional phonological contrast. However, given how many phonetic dimensions there seems to be for prosodic prominence, and the additional (non-phonological) factors that they are sensitive to, it is surely a more complex and difficult problem than most segmental contrasts. The matter is complicated yet further when we consider that prominence perception is also a function of factors that are completely absent from the signal itself.

1.2.2 Top-down prominence perception

Like other aspects of human perception, there is now considerable evidence that the perception of prominence emerges from factors not found directly in the signal. In a study of prominence perception in Swedish, for example, Eriksson, Thunberg, and Traunmüller (2001) compared two types of models of gradient prominence judgments for each word in a single sentence (*Jag tog ett violett, åtta svarta och sex vita.;* "I took one purple, eight black and six white.") that was produced by multiple speakers at different levels of "vocal effort". The first model contained only acoustic predictors, including measures of f0 height and excursion, and duration. The second model contained only top-down factors, such as whether the syllable was in a word that was contrastive in the sentence, and whether the syllable was capable of bearing a lexical accent. The top-down model was found to account for 57% of the variance in listeners' judgments, while the acoustics-only model accounted for only 48%. While it is surely the case that the two models accounted for overlapping portions of variance, and that a more sophisticated acoustic model might have performed better, the fact that the top-down model was relatively successful is suggestive that linguistic factors might actually contribute something on their own.

In fact, this is the conclusion reached by Cole, Mo, and Hasegawa-Johnson in their (2010) study, discussed in Section 1.2.1. In addition to the acoustic factors they tested, Cole and colleagues also examined correlations between listeners' prominence ratings and lexical and discourse variables, which included (a) the lexical frequency of the word and (b) the number of previous occurrences of that word in the corpus excerpt heard by the subject. The authors found that both of these factors were negatively correlated with the probability of the word's being judged as prominent. As for Eriksson et al's study, we would reasonably assume that these non-signal-based variables have the same relation to the acoustic variables; that is, it is well known that words with lower lexical frequency (Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, and Gildea 2003) and words with previous discourse mentions (Fowler and Housum 1987) tend to be pronounced with less phonetic prominence. However, in addition to the correlational analyses, Cole and colleagues show in regression models that these non-signal-based factors are significant predictors even when the acoustic measures are included in the models. In fact, the top-down factors (when combined into a single factor) contribute more to the model than do the acoustic predictors. Thus, the lexical frequency and discourse properties of the words were contributing something independent and important to the model of listeners' judgments of prominence.

The finding that lexical statistics and discourse variables influence prominence perception in a top-down manner raises the question of what the mechanism(s) for such effects might be. One clear possibility would be that these effects are analogous to top-down effects of linguistic knowledge on the perception of segmental information. For example, the well-established phoneme restoration effect (Warren 1970, Samuel 1981) demonstrates that listeners use their knowledge of a word's phonological structure to perceive phonetic information that has been masked or entirely removed from the signal. For example, a listener reports hearing an intact word like "*legislatures*" even when the /s/ has been removed, replaced, or masked in the signal. Another well-replicated effect is the perceptual "repair" of illegal phonotactic sequences. For example, Japanese speakers tend to hear illegal **ebzo* as legal *ebuzo* (Dupoux, Kazohiko, Yuki, Pallier, and Mehler 1999), English-speakers hear **lbif* as *lebif* (Berent, Steriade, Lennertz, and Vaknin 2007), and Spanish speakers hear illegal **stib* as legal *estib* (Theodore and Schmidt 2003, Cuetos, Hallé, Domínguez, Segui 2011). It is known that other language-specific phonological and phonetic knowledge has similar effects (e.g., Pitt 1998, Hallé and Best 2007, Davidson and Shaw 2012). Thus, it is known that (mis)perception can occur as the result listeners' perceptually "filling in" information that is "missing" based on their expectations. It is therefore reasonable to assume that listeners' expectations based on levels of prominence for certain words or certain contexts similarly influences their perception of prosodic aspects of the signal; that is, low frequency words and first-occurrence words, based on experience with speech, *should* be pronounced with more prominence, and so the listener tends to perceive them that way. This "restorative" top-down perceptual phenomenon is one that will be exploited in Chapter 2 to learn about listeners' expectations about the focus size contrast.

In the analysis of their top-down findings, however, Cole and colleagues do not appeal to such a restorative phenomenon, offering instead a processing-based explanation. They propose that rather than listeners being indirectly influenced by acoustics (via expectations), the effects of lexical frequency and discourse directly reflect ease of lexical access. In this case, the processing of words with low resting activation levels (i.e., low frequency or less predictable words) demands more effortful processing—assumed to mean "requiring more attentional resources"—and one of the consequences of that more effortful processing is the sensation of prominence. Thus prominence perception in their account arises in part as a top-down, processing-based epiphenomenon. The authors also suggest that both bottom-up acoustic cues and top-down processing cues are ultimately related to processing; if it is assumed that we can construe processing-resources as a kind of attention, then prominence perception ultimately results from anything demanding of attention. This can come from the speaker's efforts, by drawing the listener's attention to a particular word using acoustic prominence, or from the listener, who may need more attention to process certain unexpected kinds of information. This sort of explanation might be plausible given recent fMRI research, which shows that hearing focus-marking pitch accents draws upon the same cortical regions that are recruited during auditory spatial attention tasks.

While it is not yet completely clear how best to understand Cole et al.'s (2010) lexical and discourse-related findings, there is another, particularly relevant example of top-down prominence perception that seems best explained by restorative mechanisms. This evidence comes from Finnish, and so does not directly inform us on English prosody, but also serves as a proof of concept for top-down effects related to information structure, most pertinent to the present goals. This study is reported by Vainio and Järvikivi (2006), who tested the perception of the relative prominence of two f0 peaks in Finnish sentences. These two peaks occurred on two different NPs (which were part of adverbial expressions) in sentences such as (8a) and (8b):

(8) a. Menemme laivalla Lemille
We go by boat to Lemi

b. Menemme Lemille laivalla
We go to Lemi by boat

In Finnish, word order is, strictly speaking, free, but is exploited in highly predictable ways to express pragmatic meanings. In the present case, sentences like (8a) and (8b), which order the adverbials of place and manner differently, have different preferred focus interpretations. The order in (8a) is the unmarked order, and is appropriate for either broad focus on the sentence or narrow focus on the second NP, *Lemille*; it is a possible but dispreferred structure for expressing focus on the first NP *lavailla*. The configuration in (8b), a marked order, has preferred narrow focus on the second NP, *laivalla*. Vainio and Järvikivi found that prominence perception depended in part on which of the two configurations was presented. The primary result with respect to this top-down effect was that the f0 peak on the second NP in (8b) was perceived as more prominent even when the acoustic information (at least with respect to f0 and intensity) was the same as in (8a). Additionally, in order for the two peaks to be perceived as equally prominent in (8b), the first peak had to be increased by 10Hz more than it did in the unmarked version in (8a). Thus information structure, in this case being in a focused syntactic position, was itself prominence lending, independent of its acoustic properties. This is an important finding for the present purposes, because it can in fact be understood as listeners' expectations "filling in" the prosodic patterns that correspond to such structures; in a production study, Vainio and Järvikivi (2007) found that a sentence-final NP is pronounced with a higher f0 peak when it is interpreted as narrowly focused compared with when it is part of a broader sentence focus. Thus, these results for Finnish, although not directly applicable to English, suggest that listeners can make prominence judgments based on information structure that are consistent with the typical productions of such structures.

1.2.3 Individual differences: autistic traits

As can be seen from the last section, factors such as discourse context and lexical frequency provide a basis for prominence perception in a top-down way; they bias listeners to perceive information that is not in the signal. There is also a reason to believe, however, that listeners may not all be equally sensitive to such factors. While the emphasis on variation in phonetics and speech science has generally focused on speakers rather than listeners, recent research is moving towards closing this gap (e.g., Surprenant and Watson 2001; Makashay 2003, Yu 2010, Kong and Edwards 2011; Ladd, Turnbull, Browne, Caldwell-Harris, Ganushchak, Swoboda, Woodfield, and Dediu in press), and one finding that is of interest to the current study is that some variation among listeners can be predicted based on their "autistic traits".

Autistic traits are behaviors and patterns of information processing associated with a clinical diagnosis with an Autism Spectrum Disorder. However, such traits—for example, non-holistic attentional focus, lack of social engagement, and poor communication skills—are known to occur to varying degrees in the neurotypical population as well. These traits are measured in non-clinical populations using the Autism Spectrum Quotient (AQ; Baron-Cohen et al. 2001), a non-diagnostic, self-administered questionnaire that divides autistic traits into five separate dimensions pertaining to *social skills*, *attention to detail*, *attention switching abilities*, *communication skills,* and *imagination*. Studies have shown the instrument, which is scored such that higher scores indicate more autistic traits, to have a high level of cross-cultural validity (Wakabayashi et al. 2006; Hoekstra et al 2008; Ruta et al. 2011; Sonié et al. 2012), although there may be some variation related to culture on the Imagination and Attention Switching subscales (Freeth, Sheppard, Ramachandran and Milne 2013). It is also known that males generally score higher than females, scientists and mathematicians score higher than humanists (Baron-Cohen et al. 2001, Freeth et al. 2013), and musicians with absolute pitch discrimination score higher than those without (Dohn, Garza-Villarreal, Heaton, Vuust 2012).

Most relevant to the present purposes, autistic traits have been shown relevant to predicting the use of top-down knowledge in speech perception experiments. For example, Stewart and Ota (2008) found that high total AQ scores (i.e., the sum of all five subscales) were associated with less perceptual shifting of segment identifications in the direction of real words compared with nonce words (i.e., the “Ganong effect”; Ganong 1980), suggesting that autistic traits are associated with less reliance on top-down information from the lexicon. A similar effect was found by Yu, Grove, Martinović, and Sonderegger (2011). In their study, they tested for the effect of phonotactic context on the perception of an ambiguous sibilant fricative, and found weaker top-down effects for listeners with either high total AQ scores or high working memory capacity (as measured by the Reading Span task; Daneman and Carpenter 1980). Thus, autistic traits predict to what extent listeners are sensitive to the signal rather than to their (lexically-based) expectations about it.

Another way autistic traits might influence speech perception, although perhaps less directly, is by inhibiting the use of, or attention to, a more global pragmatic context. For example, it has been shown that at least the Communication subscale of the AQ predicts sensitivity to pragmatic “violations”. In an ERP investigation, Nieuwland et al. (2010) asked subjects to read sentences that were either informative (*Some people have pets, which require good care.*) or uninformative (*Some people have lungs, which require good care.*), depending on whether the target word (underlined in these examples) should be trivially true via pragmatic implicature. Readers with low to mid-range AQ-Communication scores (i.e., very good to average communication skills) exhibited the expected adverse brain response (the N400) following the target word in the uninformative sentences, while those with high AQ-Communication scores showed no such effect. Interestingly, neither group showed an effect when the target word was not placed before a comma, (e.g., *Some people have* <u>*lungs*</u> *that require good care.*), suggesting that the violation may have depended on the target words being phrase-final (leading to non-restrictive interpretation of the relative clause).

A similar lack of pragmatic influence on sentence interpretation was recently demonstrated by Xiang, Grove and Giannakidou (2011), who investigated whether sentences with illicit negative polarity items (i.e., those lacking a c-commanding licensor) such as *ever* can be "rescued" by way of pragmatic inference. They compared grammaticality judgments for sentences such as (9a-c), which contained either a fully c-commanding licensor (9a), a non-c-commanding but pragmatically inferable licensor (9b), or no licensor (9c):

(9) a. <u>*Only*</u> documentaries that the network TV stations have played during prime time have ***ever*** been very popular. **(grammatical)**

b. The documentaries that <u>*only*</u> network TV stations have played during prime time have ***ever*** been very popular. **(ungrammatical but pragmatically rescuable)**

c. The documentaries that the network TV stations have played during prime time have ***ever*** been very popular. **(ungrammatical)**

Xiang and colleagues found acceptability of sentences like (9b) to be dependent on subjects' scores on AQ-Communication, such that those with lower scores (i.e., good communication skills) were more likely to judge them as grammatical. Interestingly, this AQ-dependent pragmatic rescuing was possible when the licensor was either *only* or *no*, but not when it was *every*, which the authors cite as being the result of the former two licensors' association with negation (which the meaning of *every* lacks). Another possibility is that the difference is more related to (contrastive) focus than negation. Some evidence for this comes from a second study, testing ERPs, that compared the licensors "*no*", "*few*", and "*only*". In that study, it was found that the interaction with AQ-Communication was replicated only for sentences in which "*only*" was the NPI licensor. Thus, from the ERP study reported by Xiang and colleagues, and the previous one reported in Nieuwland et al. (2010), there is evidence that at least the AQ-Communication subscale may be a predictor of the use of a specific kind of pragmatic information, namely that related to information structural representations. This is potentially relevant to the goals of the present study, which will use perceptual and processing mechanisms to probe listeners' knowledge of the prosody-information structure relation.

1.3 Focus and Focus marking

1.3.1 Basic overview for English

The goal of this section is to define what is meant by information structure and by focus, and to outline how they might relate to the prosodic structure described in Section 1.1. This will bring us closer to understanding what the basic issues are, and highlight the particular problem to be addressed in this dissertation.

In the broadest and most basic sense, information structure is the linguistic mechanism that relates the content of a sentence to a larger discourse. Perhaps more specifically, information structure relates such content to a discourse *model* constructed by interlocutors. This model is meant to include the set of beliefs that the speaker holds to be mutual with the listener. Information structure has also been conceived of as a kind of "packaging" of an utterance's content into independent categories (Halliday 1967; Chafe 1974); although there is still not widespread agreement regarding how many categories there are, or how to define them, most researchers recognize some version of the categories "focus", "background", and "topic" (Chomsky 1971; Jackendoff 1972; Szabolsci 1981; Krifka 1984; Rochemont 1986; Vallduví 1990; Rooth 1992; Vallduví and Engdahl 1996; Vallduví and Vilkuna 1998; Lambrecht 1996; Kiss 1998; Schwarzschild 1999; Steedman 2000; Büring 2007, Beaver and Clark 2008, Féry and Krifka 2008, Neeleman, Titov, van de Koot, and Vermeulen 2009). Among these, the focus/background distinction is perhaps the most well-studied distinction, and it is also the one of interest in this dissertation.

As Büring (2007) notes, least controversial about focus is that it corresponds to (a) the new information in a sentence, and (b) the information required to answer a WH-question. In this sense, the focus is the "informative" part of the sentence. For identifying the focus of a sentence, I will be appealing to these widely-held assumptions, without committing to the particular details of focus's interpretive properties—particularly the issue of whether there are (grammatically) distinct focus types, an area of study still rife with disagreement.[2]

For the present purposes, then, I will utilize question-answer exchanges to manipulate focus, as in (10a) and (10b), which have focus on the subject and object, respectively:

[2] For recent discussions of the matter, see Kratzer (2004), Gussenhoven (2007); Beaver and Velleman (2011); and Katz and Selkirk (2011). My impression is that consensus is moving in the direction of recognizing distinct focus types, at least a contrastive versus non-contrastive distinction, due mostly to an accumulation of experimental phonetic evidence (see Katz and Selkirk 2011 or Bishop 2012 for a review).

(10) a. Q: Who called Mary?

A: $[\text{JOHN}]_{\text{Foc}}$ called Mary

b. Q: Who did John call?

A: John called $[\text{MARY}]_{\text{Foc}}$

Important to the questions we will want to ask is the fact that the focus of the sentence can also be a larger constituent, such as the entire VP or sentence, as in (11):

(11) a. Q: What did John do?

A: John $[\text{called MARY}]_{\text{Foc}}$

b. Q: What happened?

A: $[\text{John called MARY}]_{\text{Foc}}$

(11a) and (11b) contrast the *size of the focus constituent*, sometimes called the “domain” or “breadth” of focus, and contrasts along the dimension of the size of a focus are what I will be most concerned with in this study. I will refer to cases where focus is on a single, relatively small constituent such as a noun object, as *narrow focus*, and cases where the focus constituent is on larger constituents (e.g., on the VP or sentence) as *broad focus*. The terms “broad” and “narrow” have their origins in Ladd (1980), but note that they are being used here (and elsewhere in the recent literature) in a slightly different way; whereas Ladd used broad focus to refer to the sentence-wide focus that was said to be marked by “normal stress”, it is now common to talk about the relative size of the constituent under focus, which can vary depending on the size of the syntactic constituent.

Comparing the contrast for the location of focus as in (10), versus that of the size of the focus as in (11), highlights the fact that in English (as well as the other West Germanic languages) there is a correlation between the presence of focus and the location of a nuclear accent. It is also clear that it is not an optional correspondence, as a nuclear accent on a non-focused part of the previous examples would result in a clearly infelicitous response. What is more, it is clear from (10b) and (11), repeated in (12), that the nuclear accent does not distinguish the size of several different focus interpretations in simple SVO sentences:

(12) a. Q: What happened?

A: [John called MARY]$_{Foc}$

b. Q: What did John do?

A: John [called MARY]$_{Foc}$

c. Q: Who did John call?

A: John called [MARY]$_{Foc}$

Predicting the prosody-focus correspondences illustrated in these examples is a primary goal of a highly influential kind of theory called Focus Projection, as proposed in Selkirk (1995) (see also Selkirk 1984) and Gussenhoven (1999) (see also Gussenhoven 1984), which is the subject of the next section.

1.3.2 Focus Projection and prenuclear accents

Focus Projection is a grammatical mechanism that allows a single accent on the internal argument of a verb to "project" a focus feature up to larger constituents, allowing for focal interpretations of, for example, broad focus on a VP, to result from a nuclear accented object, even though the verb itself may contain no pitch accent. That is, it is a way of addressing the apparent facts in (12). The two major versions of Focus Projection, however, differ in non-trivial ways. For example, Selkirk's theory claims the domain of focus marking to be syntactic, while Gussenhoven's theory makes reference to semantic constituents. For the purposes here, however, the most relevant difference between the two theories has to do with their treatment of prenuclear accents, which I will briefly describe.

In Gussenhoven's theory of Focus Projection, a sentence contains one or more domains for focus marking—defined as a semantic constituent that requires only a single pitch accent in order to be interpreted as focused. One such constituent is the predicate-argument structure, which, in the case of simple SVO constructions, is the verb and the object. In such a structure, a pitch accent on the object is sufficient to mark the verb as part of the focus, although a separate rule states that accents *may* occur to the left of the nuclear accent—i.e., prenuclear accents—by way of an optional phonological rule. As a result, nuclear accents are the vehicle for focus, and prenuclear accents are completely optional with respect to meaning.[3]

In Selkirk's (1995) theory, which, again, makes reference to syntactic structures, accenting a word results in its being assigned a focus feature—F-marking—in the syntactic representation. Then, by way of two basic rules of Focus Projection, F-marks may percolate up and through the syntactic tree, marking larger constituents as follows. First, F-marking of a head licenses F-

[3] This is not to say that Gussenhoven claims that *all* prenuclear accents are optional; indeed, one of the key functions of his theory is to predict when prenuclear accents will be optional and when they will be obligatory. For example, he predicts that in constructions containing an adjunct, a prenuclear accent on the verb will be required to mark it as part of the focus constituent (e.g., contained within a VP focus). That English-speaking listeners prefer this has been supported experimentally (Gussenhoven 1983; Birch and Clifton 1995). Interestingly, Gussenhoven (1999: 46) suggests that there may be phonetic differences between optional prenuclear accents and obligatory ones, in the direction of optional ones being less prominent. I do not know of this being explored further in the context of complement versus adjunct structures, but we will see further below that speakers (at least in English and German) manipulate the phonetic prominence of prenuclear accents in narrower focus conditions.

marking of the larger syntactic phrase; second, and most important for our purposes, F-marking of the internal argument of a head licenses F-marking of the head. Any constituent that is not dominated by F-marks (i.e., the highest F-marked node) is then interpreted as the sentence's Focus. Importantly, the second rule allows for the facts in (12), since F-marking the object with an accent allows either the object or a larger syntactic phrase containing that object to be F-marked as well. It is clear, then, that in the kind of the simple SVO constructions under consideration, the emphasis is on the nuclear accent, as in Gussenhoven's model. However, Selkirk also includes the stipulation that a word that is accented but not the Focus (i.e., is F-marked, but dominated by other F-marks) is interpreted as "discourse new". The result of Selkirk's theory, then, is that a prenuclear accent on the verb, although it does not mark the Focus per se, should nonetheless be infelicitous if focus is narrow on the object. This is because in contexts requiring narrow focus on the object (e.g., *"Who did John call?"*), the verb will be given—not new—information. Therefore, prenuclear accents are optional in broad focus contexts, due to Focus Projection, but predicted to be absent in narrow focus contexts.

1.3.3 Calhoun (2006): a probabilistic model

One final kind of model to describe differs considerably from the Focus Projection models. This is the probabilistic theory proposed in Calhoun (2006) and Calhoun (2010). Calhoun's model is most interesting for the present study because it recognizes an information structural status for prenuclear accents, albeit a weak one. As I will describe further below in Section 1.4, this is a desirable property for a theory to have.

Calhoun's assumption about prosodic prominence is that it is derived from a metrical-prosodic structure of binary branching weak-strong nodes. From the perspective of the speaker, the decision to make a word prominent amounts to a decision to assign that word to a prominent position in the structure. That decision, however, is probabilistically influenced by two basic kinds of factors: information structural factors (e.g., focus) and non-information structural factors (e.g., lexical frequency, predictability, rhythm). It is also assumed, in order to account for facts of English like the ones discussed in the previous section, that information structure makes strong demands on nuclear accents and weaker ones on prenuclear accents. Prenuclear accents, while not unrelated to information structure, thus encode it less reliably.

What is most important for the present purposes is what this all means from the perspective of the listener, who must decode this information. From Calhoun's model of speaker's behavior (for which she provides corpus evidence), it is predicted that the likelihood of a listener interpreting a prenuclear accent as information structurally-meaningful should be dependent on, for example, the word's lexical frequency. Thus, in a sentence such as "*I pawned the stereo.*", a prenuclear accent on the verb is unlikely to be perceived as marking information structure, since *pawned*, being a low frequency word, is likely to be prominent simply due to frequency (e.g., Bell et al. 2003). In order for the listener to interpret *pawned* as information structurally relevant, it would have to be considerably more prominent than expected to, based on its status as a relatively low frequency word of English.

Calhoun's theory is an interesting one, since it offers a way to account for prenuclear accents, while also explaining some of the variation in people's intuitions about when prenuclear accents are felicitous, which will be discussed below in Section 1.4. One final aspect of her theory has to do with its assumptions about prosodic prominence. In her metrical prosodic model, prominence is a relative notion, and one of its key linguistic functions is to mark certain kinds of information—for example, focused information or *rhemes*—as more phonetically prominent than other kinds of information—for example, salient information, or *themes*. My understanding of her model is that this prominence can be marked in many of the ways discussed in Section 1.1: by using duration, intensity, f0, spectral cues, or “a combination of all these measures” (Calhoun 2012:332). While it is true that the phonetic cues to prominence seldom come one at a time, this view of the speakers’ use of prominence is much less phonologically organized than the one assumed here, which is essentially that of Beckman (1986). The result is that Calhoun’s model is powerful enough to predict what the signal often contains, but most likely not restricted enough to make detailed predictions about how the listener is structuring these cues phonologically. Nonetheless, I believe Calhoun’s model represents an advance, since it offers us a way to think about how prenuclear accents relate to information structure, in a way that is much more dynamic than the Focus Projection theories allow. In Chapter 2, I will take very seriously the idea that how listeners interpret prenuclear accents depends crucially on their non-information structural properties, such as lexical frequency.

1.3.4 Summary of focus marking in English

To summarize, focus realization in English exploits prosodic structure, and the most important aspect of that structure is prominence. However, the primary relation is arguably with prominence at the level of nuclear accentuation rather than with accentuation in general, since prenuclear accents, at least impressionistically, seem to be optional. Indeed, although there are some differences between two influential accounts of focus marking, namely Gussenhoven’s and Selkirk’s Focus Projection models, suffice it to say that the information structural function of prenuclear accents is generally de-emphasized in formal theories. The question I wish to ask is whether this is justified experimentally, i.e., based on phonetic data and also on perceptual data, and literature on this topic is to be reviewed next. Specifically, it is of interest to know whether the information structural contrast represented by size of the focus constituent is truly ambiguous. As will be shown, this is not clear, and in fact prenuclear prominence appears to be relevant.

1.4 The phonetic realization of broad and narrow focus

1.4.1 Speakers' encoding of the contrast

An accumulation of phonetic evidence in English (and closely related Dutch and German) suggests that broad and narrow foci are not entirely ambiguous at the phonetic level. One of the first indications of this was reported by Gussenhoven (1983), who acquired productions of sentences with broad VP or narrow object focus to present to listeners in a perception experiment. Although Gussenhoven did not provide an acoustical analysis of these production data, differences are inferable from the fact that a group of listeners reported perceptible differences between the sentences. These differences had to do with the prominence of the verbs, such that verbs that had been produced in VP focus contexts sounded more prominent. Soon after Gussenhoven's study, multi-speaker production studies of the contrast took place, such as Eady and Cooper (1986) and Eady, Cooper, Klouda, Mueller, and Lotts (1986), and these experiments also revealed differences, although they had to do with the properties of the objects. In particular, significant differences between broad and narrow focus sentences were found that showed f0 to be slightly higher on final objects that were narrowly focused compared with those that were embedded in larger sentence foci.

A number of later studies report similar findings. Because these studies are often small in number of speakers (and therefore tend to report individual subject data), they also draw attention to important interspeaker variation. First, Sityaev and House (2003) conducted two experiments with speakers of Southern British English, testing broad sentence versus narrow object focus. The first experiment (6 speakers) tested sentences such as *"I broke my neck.*" and revealed no significant effects of focus on the height of f0 peaks on the object or the verb. There was, however, a consistent pattern across subjects, significant for two of them, involving the relative scaling of these peaks. Although three of the four speakers always produced the peak on the object as significantly lower than the peak on the verb (i.e., a "downstepped" final accent), this was less dramatic in the narrow focus context than in the broad focus context. A fourth subject showed the opposite pattern in all conditions: the height of the second peak was always significantly higher than the first, but that speaker did this more dramatically in the narrow focus condition than in the broad focus condition. An additional finding is that objects under narrow focus were also pronounced with slightly longer duration than those under broader focus. Thus, these speakers seem to be employing a strategy whereby, given some speaker-specific baseline, the object's relative prominence is increased under narrow focus, and decreased under broad focus.

Patterns like these were also reported for German by Baumann, Grice, and Steindamm (2006), with differences somewhat more statistically reliable (see also Baumann, Becker, Grice, and Mücke 2007). Baumann and colleagues' study (6 speakers) also compared qualitative measures of prosodic structure, reporting intonational phonology categories using the ToBI conventions for the prosodic transcription of Standard German (GToBI; Grice et al. 2005). In the German equivalent to the English SVO constructions, annotators labeled nuclear accented objects as downstepped, or categorically lower than a preceding prenuclear accent, in 42% of the sentences produced with broad focus, but in only 25% of narrow object focus productions. This pattern held for 4 of the 6 speakers. Of the two remaining speakers who did not make use of a downstepped accent, one of them gradiently varied f0 of the nuclear accent in the direction of a relatively lower f0 in the broad focus condition, much like in Sityaev and House's study. Thus, while it is not easy to draw firm conclusions from studies with so few speakers, it does allow for examination of variation that seems quite informative. First, there is interspeaker variation with respect to what acoustic parameters speakers make use of to signal focus size; speakers have been shown to use duration and intensity as well as f0, although f0 seems to be the most reliable correlate across studies. Second, there is variation with respect to how speakers employ these acoustic cues; at least in the case of f0, it is clear that speakers can either *directly* increase f0 on a narrowly focused object, or *indirectly* increase f0 on the object, by way of suppressing prenuclear f0—or possibly some combination of the two.

In fact, recent larger studies have subsequently confirmed the basic findings of these smaller ones. Notably, Xu and Xu (2005) report f0 data for 8 American English speakers showing that focus size is systematically related to f0. The authors claim the most robust manipulation by speakers to be the suppression of f0 following the nuclear syllable, a region not considered in other phonetic studies examining the focus size contrast (but see Hannsen et al. 2008 for similar findings for Dutch). These authors also find, however, that speakers produced a higher peak on narrowly focused objects than the same objects under broad focus and some, but not all, speakers additionally suppressed prenuclear peaks. Thus, as in the previous studies, speakers seem to be manipulating cues for both the nuclear accented word as well as words in the prenuclear stretch.

The importance of such relative prominence patterns is particularly emphasized by Breen, Fedorenko, Wagner, and Gibson (2010), who present perhaps the most thorough study on this topic to date (although it is strictly quantitative). Breen and colleagues show persuasively that

Table 1.2. Rate (above chance level) of correct classification for focus size based on prosody across the three Experiments in Breen et al. (2010). Values in bold are classification by discriminant analysis; values in italics are for classification by human listeners (carried out in Experiments 2 and 3 only).

	Expt 1		Expt 2		Expt 3	
Sentence Focus	**67%**	*n/a*	**77%**	*46%*	**79%**	*80%*
Object Focus	**74%**	*n/a*	**88%**	*69%*	**92%**	*80%*

native English speakers can in fact disambiguate focus size, and that they do so using multiple phonetic cues to prominence—f0, word duration and especially intensity—and that the distribution of these cues across the sentence is crucial. They found discriminant analysis of their production data to be successful in classifying sentences as having the intended focus size above chance level. They also found such classification to be even more accurate for productions made by speakers who were aware of the potential ambiguity between broad and narrow focus, further indicating that listeners have knowledge about how to communicate the distinction. This later finding can be seen in Table 1.2, which reproduces Breen and colleague's correct classification rates for each of the three experiments in their study (as well as classification by human listeners, discussed further below). In all three experiments, classification rates are significantly above chance level for both broad and narrow focus; an increase in classification accuracy occurs in Experiments 2 and 3, where speakers were deliberately trying to disambiguate their productions for the benefit of a listener who was not told the intended size of the focus constituent in the sentences. There was also a modest increase in Experiment 3, where additional prefocal material was added to the SVO sentences, namely the lead-in phrase "I heard that…", which would seem to indicate that more material to the left of the focus was exploited for disambiguation. Finally, though no statistical comparison is made, it is also evident from the classification rates that, at least numerically, narrow focus is more reliably classified across experiments than broad sentence focus.

As can be seen from reviewing phonetic studies, close inspection of speakers' productions of sentences with different focus sizes contradicts the claim that focus size is inexpressible prosodically. As noted above, however, there seems to be at least two types of variation: that regarding the particular cues involved in the distinction (duration, intensity, f0), and with respect to how speakers apply them (directly or indirectly to the nuclear accented word). The next question to ask is whether these patterns that speakers provide are successfully used by listeners to extract the intended meaning. While it has been shown by Breen et al. (2010) that a statistical model is quite capable of classifying based on acoustics, it is not yet clear how well human listeners do this. In fact, as is discussed below, the results of studies with human listeners have been quite mixed, in some cases supporting the ambiguous behavior generally predicted by theoretical models.

1.4.2 Listeners' decoding of the contrast

As noted above, and consistent with much of the production evidence, Gussenhoven (1983) found there to be some perceptible differences between sentences produced in broad and narrow focus context, in the direction of verbs being perceived as more prominent if they had been produced with broad focus. This was found when listeners were asked to make subjective ratings of prominence when the sentences were presented out of context. However, in an experiment where listeners were asked to use such differences to match up an answer sentence with its intended broad or narrow focus question context, listeners' responses did not correlate with any prosodic differences.

This finding is to a large extent replicated by Birch and Clifton (1995), who used similar methodology. These authors presented listeners with SVO answer sentences that either had or did not have prenuclear accents on the verb, and these appeared to listeners following VP focus questions. In one experiment, the task for participants was to listen to the question-answer pairs, and to rate (from 1 to 5) how "appropriate" the answer pronunciation sounded; in a second experiment with the same materials, the task was instead to judge how well the question-answer exchanges "made sense". Thus in the case of the first experiment, listeners were told explicitly to attend to prosody, and in the second were arguably more focused on interpreting meaning. In the first case, results indicated that listeners had a small but statistically significant preference for VP foci to contain a prenuclearly accented verb, inconsistent with a genuine ambiguity. In Experiment 2, there was no significant difference in listeners' responses, although there was a trend towards listeners taking longer to respond when broad foci lacked a prenuclear accent. Thus, at least for broad focus, listeners showed some signs of a preference for the production patterns reviewed: broad VP focus sentences are judged as better, and are possibly easier to process, when they contain a prenuclear accent on the verb. Possibly, when listeners are asked specifically to judge prosodic appropriateness, they are simply more attentive to the signal than they are when judging sentence meaning. It is also possible that Gussenhoven's listeners were more focused on meaning than on prosody, which would explain their ambivalence.

However, appropriateness ratings of prosody were also used in a subsequent study by Welby (2003), and her results better resemble Gussenhoven's earlier, completely ambivalent listeners. Welby tested preferences for a prenuclear accent on the verb in sentences with either broad or narrow focus, and found no evidence for a preference either way. Worth noting is that her results are consistent with Gussenhoven's theory, but not Selkirk's. This is because Selkirk predicts a prenuclear accent on a verb to signal that the verb is new, and thus part of the focus constituent. To Gussenhoven, however, prenuclear accents (for these constructions) are always optional.

Additionally, Welby showed that there was no effect related to the type of nuclear pitch accent used, as sentences with the ToBI H* and those with the L+H* on the object were rated as equally appropriate to listeners in the broad focus context. (Pitch accent type was not tested for narrow focus).

Thus, previous studies, which have utilized context matching and appropriateness ratings—and generally concentrated on broad focus—have produced results that are somewhat mixed. Note, however, that using these sorts of metalinguistic tasks, and using productions produced in uncommunicative contexts, is one of several criticisms of previous work cited by Breen et al. (2010), reviewed above. Breen and colleagues also carried out a perception experiment, but used as stimuli the productions that speakers produced with the intent to disambiguate. For these productions, they found that broad and narrow focus sentences were correctly paired with the correct question contexts by most listeners, shown in italics in Table 1.2. Of the 13 listeners in their Experiment 2, for example, six (46.2%) were able to correctly identify broad focus sentences above chance, and nine (69.2%) were able to correctly identify narrow focus sentences above chance level. Of the 10 listeners in their Experiment 3, this increased to 8 out of 10 for both focus sizes. While the authors do not provide modeling of listeners' identification as a function of individual acoustic predictors, we do know from their detailed production study that the narrow object focus sentences were generally produced with greater intensity, somewhat longer durations, and higher f0 on the object (even more so in their Experiment 3) relative to prenuclear words. Thus, Breen and colleagues show that listeners can use speakers' productions to extract the intended focus structure, although the elicitation methods for the stimuli are relevant. Additionally, however, individual differences are apparent, since not all listeners performed above chance level, even with highly expressive stimuli.

To summarize, there is a considerable amount of evidence from speakers that the size of the focus constituent in English (and some closely related languages) is distinguished prosodically by speakers. First, it is clear that this is not done by manipulating nuclear accent location; in all cases it must fall on the object in SVO constructions. Rather, the appropriate generalization seems to be that, in order to indicate narrow focus on a sentence-final object, speakers will increase the prominence of that nuclear accent using a possibly speaker-specific combination of intensity, duration and f0. Further, this can be accomplished by either increasing the prosodic prominence of nuclear material, or by suppressing the prominence of surrounding, particularly prenuclear, words. Additionally, there is some evidence that this may be done more reliably for narrow focus compared with broad focus (Breen et al. 2010). Finally, results from perception studies are rather mixed with respect to whether or not listeners can use this information to decode the intended focus interpretation.

1.5 Goals for the rest of the dissertation

As I have described in this chapter, there is a clear relation between prosody and focus in English, and the nuclear accent is a key part of this relation. This is something that theoretical work has sought to capture, and one result of this is that the role of prenuclear accents has been de-emphasized. This state of affairs is particularly apparent when we look at how theories have handled the size of the focus constituent in English SVO sentences. Whereas such work regarded prenuclear accents as largely unrelated to focus, we have seen evidence from production studies, and to some extent perception studies, that indicates theory may not be on solid empirical ground in this regard. In this rest of this dissertation, my goal will be to contribute further to this issue, concentrating on the listener's perspective. Whereas previous research has largely concentrated on the listener's ability to detect mismatches between focus and prosody in a fairly explicit way, I will probe listeners more indirectly, by way of their expectations.

In the second chapter of this work, this will be done by exploiting the fact that perception, as described in Section 1.2.2, takes place in part as a top-down process. To the extent that listeners' expectations serve as the basis for top-down effects, we can deduce the character of those expectations by observing them in perception. As will be shown in the experiments in Chapter 2, listeners are unable to make prominence judgments that reflect only the signal's properties; they are influenced by information structure in a way that teaches us how they expect the two to relate.

In Chapter 3, this same question is approached, but from the perspective of on-line processing mechanisms. Using the cross-modal priming paradigm, it will be shown that English speaker listeners are able to process sentences and lexical information more easily when sentence prosody and information structural interpretation correspond in a well-defined way. As is demonstrated in Chapter 1, the experiments in Chapter 2 show that prenuclear prominence is relevant to characterizing what it is that listeners know about how prosody should express the size of the focus constituent. Further, it will also be shown that the use of prosody on-line in the cross-modal priming task is particularly sensitive to individual differences in "autistic"-like personality traits.

Finally, in Chapter 4, I consider the findings from the two experimental chapters, and reconsider what has been reviewed in the present chapter. I then propose a revision to the theory of prosodic structure in English that can account for what has been learned. I argue that listeners represent the focus size contrast in the grammar in a way that requires a significant, yet conceptually simple, addition to the basic AM model described in Section 1.1. The proposal will be along the lines of Ladd’s (1990) theory of tonal metrical structure, which I also show to be independently needed for a number of other information structural contrasts in English.

CHAPTER 2

Information structural influences on prominence perception

2.1 Introduction

Listeners' perception of prosodic prominence, like other aspects of speech, has an important top-down component to it. As discussed in Section 1.2.2, one source of this top-down knowledge is previous experience with speakers' productions. In Section 1.4, I also reviewed production studies that provided a compelling picture of what that experience is likely to look like with respect to the realization of focus size in English. Important to the present purposes, a crucial aspect of how speakers disambiguate broad focus from narrow focus involved prenuclear prominence, something not predicted by grammatical models of the information structure-prosody interface. In particular, it is reported that speakers are more likely to produce prenuclear accentuation on a verb in an SVO sentence if that verb is part of the focus constituent, i.e., under broad sentence or VP focus. However, some of the studies reviewed also indicated that prenuclear prominence is not the only aspect of sentence prosody that speakers manipulate. Rather, they also reduce the prominence of the nuclear accent on the object in broader focus conditions, either categorically (by using a different accent, namely a downstepped one), or gradiently (by using a full-fledged, non-downstepped accent, but a phonetically less prominent one). Moreover, speakers may manipulate both: increasing prenuclear prominence and reducing nuclear prominence in tandem. However, it is still an open question to what extent listeners internalize their experience with such production patterns; to the extent that they do, it suggests that the prosodic representation of the focus size contrast is ambiguous for neither speakers nor listeners, which a theory of the prosodic expression of focus should accommodate. As discussed in Section 1.4.2, experimental evidence thus far has been inconsistent on this point.

In the experiments presented below, the goal is to probe native English-speaking listeners for expectations that we can relate to the patterns reported in production studies. The prediction is that any such expectations should be apparent in the form of top-down effects on prominence perception. I will employ a relatively simple prominence rating task to test this hypothesis with linguistically/phonetically-untrained listeners who were presented with simple SVO sentences with different focus structures. Experiments 1a and 1b are preliminary experiments, which use the prominence rating task with naturally produced stimuli, and establish the existence of a top-down, auditory illusion based on information structural interpretation. In Experiment 2, this auditory illusion is replicated and used to explore prominence perception in more detail. In particular, Experiment 2 probes for interactive effects between expectations based on focus and on lexical statistics. As discussed in Section 1.3.4, Calhoun's (2006) proposal regarding the interpretation of prominence predicts that information structural factors should interact with other such factors, a possibility that, if confirmed, might account for some of the variation across studies on the use of prosodic cues to decode focus size.

2.2 Experiment 1a

2.2.1 Method

2.2.1.1 Materials

Materials were designed to serve as auditory stimuli for a prominence rating task. Recorded sets of question-answer pairs were designed to serve as mini-dialogues to present to listeners. A dialogue set included one version of an SVO answer sentence such as *I bought a motorcycle*, and

Table 2.1: Example set of question-answer dialogues used in the Experiment 1a.

Focus Condition	**Question Context**	**Answer Sentence**
Sentence focus	What happened yesterday?	I bought a motorcycle.
VP focus	What did you do yesterday?	
Object focus	What did you buy yesterday?	

three different set-up questions, all WH-questions such as *What happened yesterday?*, *What did you do yesterday?*, or *What did you buy yesterday?* The answer sentence was to be used as the test sentence for which linguistically naïve listeners would provide prominence ratings. The set-up questions allowed for the pragmatic manipulation of the size of the focus in test sentences, resulting in three experimental conditions: one in which the entire answer sentence was the focus (sentence focus), one in which the verb-phrase was the focus (VP focus), and one in which only the object was the focus (object focus) (Table 2.1; see Appendix A for the full list of sentences and contexts).

The dialogues were recorded and used to create stimuli as follows. Two native speakers of American English (both linguistics graduate students) read 17 sets of question-answer exchanges from a printed booklet. The printed dialogues contained no intonational annotations and neither the speaker of the questions (a female), nor the speaker of the answers (a male) was instructed on how to produce the sentences, beyond being told to read the exchanges as naturally as possible. Consequently, no control was exerted over how the test sentences might have been produced within the limitations of being felicitous to the speakers. The speakers were recorded over two separate channels while speaking into head-mounted microphones in a sound-attenuated booth. Recordings of questions and answers were digitized at 22.05 kHz, saved and stored as separate wav files.

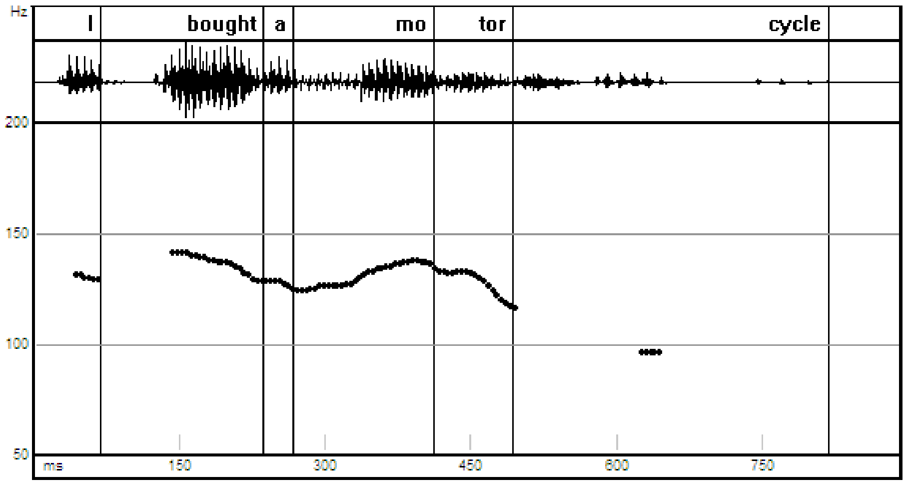


Figure 2.1: Example waveform and pitch track for the sentence *I bought a motorcycle.*, recorded as answer to the VP focus question *What did you do yesterday?*. This production (ToBI transcribed with H* on both the verb and object) was presented as an answer to each of the questions in the three focus conditions.

Because the purpose of the experiments was to test the independent effect of information structure on the perception of prominence for words in the answer sentences, it was necessary to hold all acoustic information in those sentences constant across the conditions in which they would be presented. This was accomplished by extracting the recordings of answer sentences produced in response to VP focus questions in the original recordings and using them as the test sentences for all three focus conditions. Thus, for example, the production of the answer sentence *I bought a motorcycle* (Figure 2.1), originally produced as an answer to the VP focus question *What did you do yesterday?*, was made to follow each of the three different questions recorded in that set. Based on previous research (reviewed in Section 1.4), it was expected that each of the VP focus answer sentences would be pronounced as a single prosodic phrase with the nuclear accent on the object. However, other details of the realization of an "appropriate" answer to these kinds of questions are less predictable for an individual utterance. As reviewed above, the presence or absence of a prenuclear accent, and also the type of nuclear accent are said to vary (e.g. Gussenhoven 1983; Selkirk 1995; Ladd 1996; Jun 2008). It was therefore necessary to identify the intonational pattern of the sentences that were to be used as stimuli.

Those data are reported here in the form of ToBI annotations. Two linguists trained in using the model for Mainstream American English independently transcribed tones (not break indices) for the verb phrase of the 17 answer sentences used as test stimuli. Labeling of these test sentences was done without any question context included in the sound file. Here I will report the tones assigned by the first labeler (Table 2.2), and describe rates of agreement below.

It is often noted that one of the least reliable distinctions in prosodic transcription of English is that between what ToBI represents as H* and L+H* (e.g., Syrdal and McGory 2000; Calhoun 2006; Breen, Dilley, Kraemer, & Gibson 2012), and for this reason, the two categories are sometimes collapsed in calculating transcriber agreement (e.g., Pitrelli, Beckman, and Hirschberg 1994). The distinction was maintained here, but the phonetically more subtle downstepped categories !H* and the L+!H* were collapsed.

Agreement between the two labelers for verbs in the test sentences was 100%, all being transcribed with a prenuclear H*. Agreement for tones on objects was 76.4%. While the two labelers agreed that all of these objects carried a phrase-final pitch accent, the disagreements that arose regarded whether that pitch accent was a H* or !H*, and Table 2.2 shows the labeler who tended to use H*.[4] Agreement for boundary tones was 100% (for both intermediate and intonational phrases). The ends of sentences were usually marked by low targets, although in five cases the speaker's productions showed a fall from the nuclear accent (i.e., an L-, associated with the intermediate phrase), followed by a slight rise (H%, associated with the intonational phrase).

[4] In one of the cases where the phrase edge was marked as L-H%, the second labeler annotated the object as ambiguous between a !H* and L* (an unsurprising ambiguity in the reduced pitch range near the end of the utterance). The !H* annotation was used here to calculate agreement between raters.

Table 2.2: Intonational structure of the 17 test sentences, described in the form of ToBI transcriptions.

Verb	Object	ip-Boundary	IP-Boundary	# of items
H*	H*	L-	L%	7
H*	!H*	L-	L%	4
H*	!H*	L-	H%	3
H*	H*	L-	H%	2
H*	L+H*	L-	L%	1

The completed 51 recorded dialogues (17 test sentences, each occurring in three question contexts) were arranged in three different pseudorandomizations, intermixed in each with 37 non-experimental filler dialogues. The fillers closely resembled the experimental dialogues in most respects (including the focus conditions in which they appeared) but differed from them in that (a) fillers varied syntactic structure (some contained adjuncts instead of simple argument objects) and (b) they were subject to additional focus conditions (some occurred in double (subjects/objects) focus contexts, and some had narrow focus on the verb). In each ordering of the stimuli, members of a crucial set (e.g., Table 2.1) were separated by at least 6 question-answer dialogue items.

2.2.1.2 Participants

Thirty native speakers of American English were recruited from the University of California, Los Angeles, to participate in the experiment. All were undergraduate students or (non-academic) employees at the university. Many of the student participants were linguistics or psychology majors, although none had any training in intonational phonology or the transcription of prosody. All participants confirmed they had no previous diagnosis of a hearing or communication disorder, and all were paid for their participation.

2.2.1.3 Procedure

Listeners participated in a prominence rating task. PowerPoint presentations were used to present the question-answer stimuli. Each slide contained only an item number and a play button, which participants used to listen to the items; no orthographic representation of the dialogues or any visualization of the prosody/acoustics appeared to listeners. They were able to proceed through experimental items at their own pace, listening to each recorded dialogue as many times as they wished, although listeners were discouraged from listening more than two or three times to a single item. As they played each question-answer exchange, they were to listen for how "stressed" words in the male speaker's answers sounded. The experimenter emphasized to participants that their task was to listen to how his answer sentences were pronounced, and this was described in the following way:

> "This experiment is about how speakers pronounce words in a sentence. Your task is to tell us as accurately as possible how stressed the underlined words sound relative to other words in the sentence. By "*stressed*" we mean "*how much did the speaker use his voice to make the word stand out*".

Participants were provided with printed transcripts of the dialogues they heard, ordered and numbered as they appeared on the PowerPoint lists. They were instructed to follow along on the transcript and to provide ratings of "stress" from 1 ("not at all stressed") to 5 ("very stressed") for words that were underlined on that transcript. These words were the verbs and the objects in the answer sentences, and they were to write in their ratings above the word. An example of how items appeared on the transcript is shown in (1); the numbers appearing above the underlined words are hypothetical examples of how listeners provided their judgments.

(1) a. Q: What did you do yesterday?

2 *4*

A: I bought a motorcycle.

b. Q: What did you eat at the picnic?

2 *5*

A: I ate a hamburger.

Before beginning the experiment, participants completed a short practice session of three dialogue items to familiarize themselves with the style of the dialogues, the speakers and the general set-up for the task. After completing the practice session and asking questions, participants listened to a pseudorandomized list of 88 dialogues (51 test + 37 filler dialogues) binaurally over Sony MDR-V500 closed, dynamic headphones at a comfortable listening volume (held constant across participants) in a sound attenuated booth in the UCLA Phonetics Laboratory. They provided prominence ratings as above for verbs and objects in each sentence (30 listeners × 17 test sentences × 2 words (verbs and objects) × 3 focus conditions (sentence focus, VP focus, object focus) = 3,060 ratings). These ratings served as the outcome variable in a mixed-effects linear model using the *lmer* function in the *lme4* package (Bates and Maechler 2009) for R Statistics (R Development Core Team 2012). The predictors in the model included *listener* and *item* as random factors and the following fixed effects factors: the *word* that was rated (verb or object)[5], the experimental manipulation *focus size* (sentence, VP, or object), and the interaction of these two factors.

2.2.2 Results

Average listener ratings for objects and verbs are shown for each of the focus size conditions in Figure 2.2. Table 2.3 shows the results of the model when "*verb*" and "*object focus*" are the

[5] Note that in this experiment, the factor *word* (verb versus object) represents both lexical category and accent status, since all verbs were prenuclear accented, and all object were nuclear accented.

default values (i.e., comparison groups) for *word* and *focus size*, respectively. A comparison of the fit of the full model (i.e., the model containing the three fixed effects parameters) to the data was shown to be superior to that of a baseline model containing only random effects parameters ($LogLik_{full}$ = -4235 vs. $LogLik_{baseline}$ = -4244; *anova* function in R: $p < .0001$). According to the model, there was a significant effect for *word*, such that the verbs were associated with lower overall prominence ratings than the objects. There was also a smaller but significant main effect for the focus manipulation, *focus size*, which indicated that words were rated as less prominent under the two broad focus conditions compared with the narrow object focus condition.

Both of these effects, however, are best understood in terms of the significant interaction between *word* and *focus size*. As is easily seen in Figure 2.2, although objects were judged as more prominent than verbs in each of the focus conditions, the difference was greatest in the object focus condition, due to both object as well as the verb being significantly different from the other two conditions. With respect to the two broad focus conditions, the results of the model indicate they had the same distance from the object focus condition, suggesting no significant differences between sentence and VP focus were likely. To confirm this, the default value for *focus size* in the (same) model was reset to “VP focus” so that a direct comparison could be made with sentence focus; the groups were found to be statistically the same (VP vs. sentence focus: β = .0480, *SE* = .0834, $t = .58$, $p = .5646$).

2.2.3 Discussion

The results of Experiment 1a demonstrate two important things. The first is that listeners’ judgments of prosodic prominence are significantly and independently affected by their interpretation of the utterances’ information structure. More specifically, prenuclear verbs were

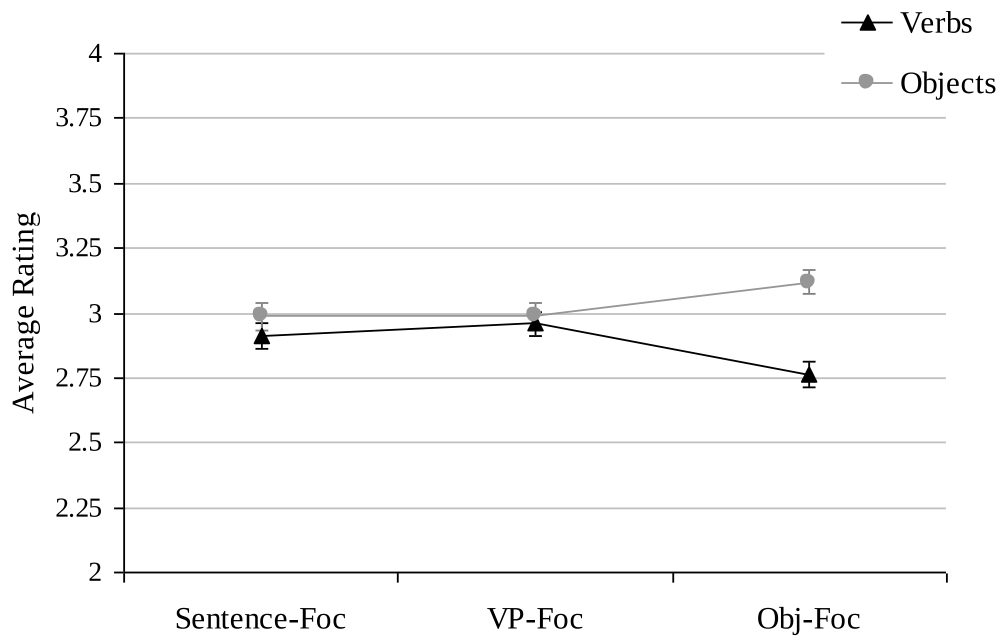


Figure 2.2: Average prominence ratings for verbs and objects in test sentences in the three focus conditions in Experiment 1. '1' is lowest in prominence, '5' is highest. Error bars show standard error.

Table 2.3 Results for fixed-effect factors in the model of listeners' prominence ratings in Experiment 1a.

	β	*SE (β)*	*t-value*	*p-value*
(Intercept)	3.1191	0.1201	25.97	< **.0001**
Word (verb)	-0.3578	0.0590	-6.07	< **.0001**
Focus Size (Sen)	-0.1304	0.0590	-2.21	**.0271**
Focus Size (VP)	-0.1304	0.0590	-2.21	**.0271**
Word(verb)*Focus Size (Sen)	0.2824	0.0834	3.39	**.0007**
Word(verb)*Focus Size (VP)	0.3284	0.0834	3.94	< **.0001**

heard as more prominent when they were part of the focus constituent (VP and sentence focus) than when they were excluded from it (object focus). However, the significant interaction between *word* and *focus size* indicated that it was not only the case that prenuclear verbs were heard as more prominent under broad focus, but objects were also heard as less prominent. Notably, this pattern closely resembles that of speakers in the studies reviewed in Section 1.4.1.

Having found evidence that listeners' ratings of prominence were influenced by information structure, it should be pointed out that it is unlikely that listeners were relying only on information structure. As can be seen in Figure 2.2, and indicated in the model, verbs were overall less prominent than objects, and this is generally consistent with the phonological structure of the sentences, since objects were always nuclear accented and verbs were always prenuclear accented. However, the ratings probably reflect the phonetic properties of the sentences to some extent as well, since the numerical differences between objects and verbs are remarkably small in the two broad focus conditions. As reported in Table 2.2, while all of the verbs in Experiment 1a contained a prenuclear H*, only about 60% of the objects bore a full-fledged (i.e., non-downstepped) nuclear accent, and so, though phonologically strong, objects were often phonetically rather weak. This may be evidence that listeners were in fact attending to properties of the signal, and that their expectations based on phonological structure and information structure served to modify their perception of it. Further evidence for interpretation will be seen in Experiment 1b, which used a different set of stimuli, elicited from, and presented in, contrastive focus contexts.

2.3 Experiment 1b

The purpose of Experiment 1b was to test whether the results of Experiment 1a could be replicated for a different kind of focus, namely "corrective" focus. Narrow corrective focus is often said to correspond most reliably to an increase in prominence compared with non-contrastive WH-focus (e.g., Bartels & Kingston 1994; Ito, Speer, and Beckman 2004; Breen et al. 2010), and so we might expect the effects to actually be larger than in Experiment 1a, which used WH-focus. Experiment 1b is also an opportunity to test another set of sentences that have accent patterns

Table 2.4: Example set of stimuli used in Experiment 1b.

Focus Condition	Question Context	Answer Sentence
Sentence-Foc	Why's your wife mad? …Because the roof's leaking?	No, because I bought a motorcycle.
VP-Foc	Why's your wife mad? …Because you lost your job?	
Obj-Foc	Why's your wife mad? …Because you bought a car?	

different from those in Experiment 1, where the phonetic prominence of verbs and objects might have been more in favor of verbs being more phonetically (though less phonologically) prominent.

2.3.1 Method

2.3.1.1 Materials

A second set of 51 short dialogues were prepared for Experiment 1b, and were similar to those in Experiment 1a; the same two speakers were recorded reading question and answer sentences, and those sentence were of the same basic structure as in Experiment 1a. However, in order to test for the effect of focus size for contrastive focus, the dialogues differed in the following ways. First, the set-up questions read by the female speaker were followed by complementizer phrases headed by "*because*", which themselves contained a set-up question (see Table 2.4; complete list shown in Appendix A). For example the female speaker read questions such as "*Why aren't you hungry?",* and offered a possible reason which was intended as a set-up to the interpretation of the focus structure of the answer. That proposition was then corrected in the answer sentence read by the male speaker. Thus, in the context of "*Why's your wife mad?... because you lost your job?"*, the answer "*No, because I bought a motorcycle."* is assumed to be a correction to the VP,

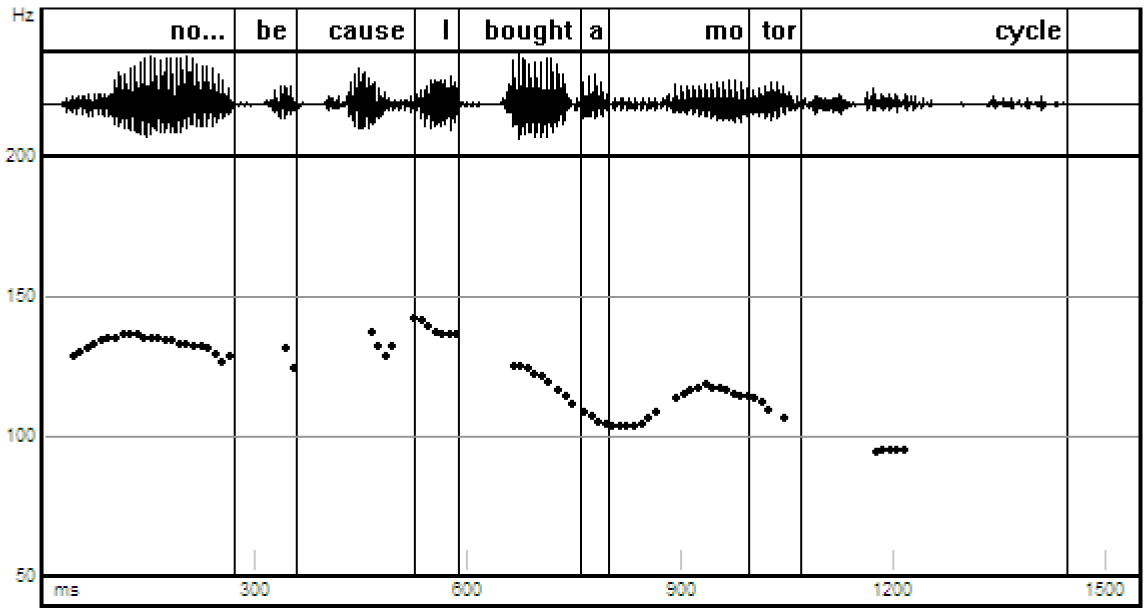


Figure 2.3: Example waveform and pitch track for the sentence *No, because I bought a motorcycle*, recorded as an answer to the VP focus question *Why's your wife mad…because you lost your job?.* This production (ToBI transcribed with an unaccented verb and L+!H* on the object) was presented as an answer to each of the three focus conditions.

"lost your job". In the context of "*Why's your wife mad?... because you bought a car?"* that same answer is assumed to be a correction only to the object, "*motorcycle"*. These materials were recorded as in Experiment 1a, and productions of the answer sentence spoken in a VP context (e.g. Figure 2.3) were saved and paired with the three different questions in the same way. ToBI annotations were again assigned to the test sentences by the same two labelers.

The first labeler's transcriptions are shown in Table 2.5; agreement for accents on verbs was 76.5%, accents on objects 76.5%, boundary tones 100% (all sentences being transcribed with low boundary tones after the object). Disagreements in assignments for accents on verbs regarded the presence or absence of an accent and whether an accent, if present, was H* or !H*. Disagreements for objects involved whether they bore a H* rather than L+H*; rate of agreement as to the presence of a nuclear accent on the object, however, was 100%. Worth noting are the overall differences from the production elicited for Experiment 1a; although objects were produced with downstepped nuclear

Table 2.5: Intonational structure of the 17 test sentences, described as ToBI transcriptions. 'Ø' indicates the absence of a pitch accent.

Verb	Object	ip-Boundary	IP-Boundary	# of items
Ø	L+H*	L–	L%	6
H*	H*	L–	L%	3
H*	L+!H*	L–	L%	3
Ø	!H*	L–	L%	2
H*	!H	L–	L%	1
H*	L+H*	L–	L%	1
!H*	L+H*	L–	L%	1

accents at a similar rate as in Experiments 1a, unaccented verbs were much more likely in Experiment 1b's sentences, as were objects with the L+H* accent rather than the simple H* accent. Thus, the baseline relative prominence between prenuclear verbs and nuclear objects was rather different across the two experiments, with the balance tipped more in the direction of objects being prominent in the contrastive focus stimuli for Experiment 1b; this is consistent with some of the production work cited above.

2.3.1.2 Participants

Thirty native speakers of American English were recruited from the University of California, Los Angeles as in Experiment 1a (none had participated in Experiment 1a). No participant reported any previous diagnosis or knowledge of a hearing or communication disorder; all were paid for their participation.

2.3.1.3 Procedure

The procedure for Experiment 1b was carried out as for Experiment 1a.

2.3.2 Results

It was discovered that two of the participants in Experiment 1b had also participated in Experiment 1a, and so data from these two participants were removed from the analysis. Average listener ratings across conditions for the remaining twenty-eight subjects are shown in Figure 2.4. A linear mixed-effects model was constructed as in Experiment 1a, using the same parameters. The fit of that full model (LogLik = -3594) to the data was significantly better ($p < .0001$) than that of the base-line model containing only random effects factors (LogLik = -3810) according to a log likelihood ratio test. The outcome of the full model is shown in Table 2.6 and indicated the following. There was a significant effect for *word*, such that verbs were overall judged as less prominent than objects, although the difference was not as pronounced as in Experiment 1a, presumably due to the character of the stimuli. As in Experiment 1a, there was also an effect for *focus size* that indicated words were rated as less prominent under the two broad focus conditions compared with the narrow object focus condition. However, also as in Experiment 1a, both of these effects are best understood in terms of their interaction in the model; the highly significant interaction between *word* and *focus size* indicated that the experimental manipulation did not influence both verbs and objects in the same manner; rather, verb in VP focus contexts were rated as more prominent than verbs in narrow object focus contexts. Object ratings of prominence, however, increased under narrow object focus. A second model was constructed to directly compare the two broad focus groups, and it was found that there were no differences between them (VP vs. sentence focus: $\beta = .0047$, $SE = .0073$, $t = .641$, $p = .5212$).

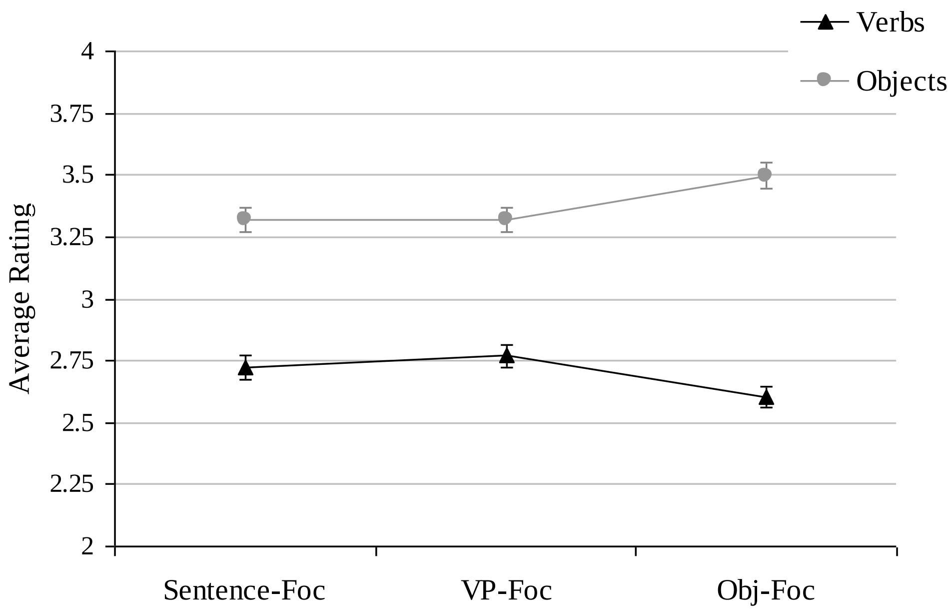


Figure 2.4: Average prominence ratings for verbs and objects in test sentences in the three focus conditions in Experiment 1b. '1' was lowest in prominence, '5' was highest. Error bars show standard error.

Table 2.6: Results for fixed-effect factor in the mixed-effects model of listeners' prominence ratings in Experiment 1b.

Fixed effects	*β*	*SE (β)*	*t-value*	*p-value*
(Intercept)	3.4517	0.1465	23.55	< **.0001**
Word (verb)	-0.8446	0.0535	-15.78	< **.0001**
Focus Size (Sen)	-0.1324	0.0535	-2.47	**.0135**
Focus Size (VP)	-0.12	0.0535	-2.20	**.0280**
Word(verb)*Focus Size (Sen)	0.2416	0.0757	3.19	**.0014**
Word(verb)*Focus Size (VP)	0.2710	0.0757	3.58	**.0003**

2.3.3 Discussion

In relation to focus size, the pattern of results in Experiment 1b did not differ in any significant way from those in Experiment 1a; verbs were perceived as more prominent under broad VP or sentence focus than they were under narrow object focus, and this boosting in prominence ratings for verbs under broad focus was accompanied by a suppression of prominence ratings for nuclear accented objects. Also as in Experiment 1a, there is evidence that listeners were attending to bottom-up information about the stimuli as well. As noted above, the baseline difference in relative prominence between verbs and objects in Experiment 1b was such that verbs were much less prominent overall than objects in the experiment, which listeners' ratings reflect. Thus, listeners' interpretations of the sentence's focus structure, and presumably their expectations about the prosody associated with that structure, were modulating their perception of prominence, but not overriding it.

The results of Experiments 1a and 1b, then, have establish that listeners' interpretation of the size of a sentence's focus constituent can produce a type of auditory illusion, warping the perception of prominence for words in that sentence, notably affecting both prenuclear verbs and nuclear accented objects. The goal of Experiment 2 was to investigate this phenomenon in the context of a larger and more sophisticated model of prominence perception.

2.4 Experiment 2

The primary interest of Experiment 2 was to examine whether the focus effect established in Experiments 1a and 1b interacts with another factor that is associated with prominence, namely a word's lexical frequency. Recall from the discussion in Section 1.3.4 that Calhoun's (2006) model of prominence marking would predict that such interactions should be present, since any factor that affects baseline expected prominence should modify the amount of prominence needed to indicate focus marking. Additionally, this is predicted to be most relevant for prenuclear words, since prominence in the prenuclear domain is most sensitive to non-information structural factors. However, it still unknown whether the interactions which Calhoun find to characterize production (as indicated by a model of prosodic transcriptions of productions) also characterize the perceptual processes of listeners. Therefore, it was explored whether a prenuclear verb's lexical frequency served as a predictor of the size of the focus effect found in my first two experiments.

Another goal, however, was to explore whether there might be any individual differences related to listeners' autistic traits influencing the patterns of prominence perception. As discussed in Section 1.2.3, autistic traits have been found to predict listeners' sensitivity to (a) lexical statistics, and (b) pragmatic context, both of which are crucial to present subject matter. Therefore, listeners' AQ profiles were also collected for use in the model of results from Experiment 2.

Finally, in order to allow listeners' expectations to have maximal top-down influence on their perception of the signal, the stimuli used in Experiment 2 were intended to be more ambiguous with respect to the relative prominence of the verb and object. Rather than the uncontrolled productions of questions and test sentences that were elicited as stimuli in the first those used in Experiment 2 were therefore controlled, and the accent pattern was also held constant across all test items.

2.4.1 Method

2.4.1.1 Materials

The design of materials to be used in Experiment 2 was similar to that for Experiments 1a and 1b, with three primary differences. First, because there were no significant differences between the two broad focus conditions in Experiment 1a, only VP was used as the broad focus condition here. Second, an additional manipulation in Experiment 2 was the frequency of the verb in the test sentences, and so verbs were selected to have different lexical frequencies. Third, the same accentual pattern was used for all test sentences, and the purpose of the one chosen was to render

Table 2.7: Example set of stimuli used in Experiment 2.

Focus Condition	Question Context	Answer Sentence (Low Freq Verb)
VP focus	What did you do just now? …Did you break something?	No, I latched the gate.
Object focus	What did you latch just now? …Did you latch the car door?	

Focus Condition	Question Context	Answer Sentence (Mid Freq Verb)
VP focus	What did you do? …Did you cut yourself?	No, I touched the stove.
Object focus	What did you touch? …Did you touch the grill?	

Focus Condition	Question Context	Answer Sentence (High Freq Verb)
VP focus	What did you do in the workshop? …Did you drill?	No, I used the wrench.
Object focus	What did you use in the workshop? …Did you use a hammer?	

the relative phonetic prominence of the verb and object in each sentence more ambiguous. How this was carried out is described in more detail below.

A set of 36 SVO sentences was designed with verbs that fell into one of three groups, based on CELEX log frequency: low frequency (log frequency = 0), mid frequency (log frequency = .47 – 1.14), or high frequency (1.2 – 2.4). Nouns were then selected to serve as objects of the verbs in those sentences, and were chosen to be relatively semantically/pragmatically predictable based on the verb. Although this restricted the possibility of controlling the lexical frequency of the objects, it resulted in objects with a range of frequencies. Further, and importantly, there was no correlation between the verb's frequency and the object's frequency in the sentences (r=.047). An example of a test sentence in each of the three frequency conditions is shown below in Table 2.7 (full list shown in Appendix B).

To produce recorded versions of the 36 test sentences and their two focus contexts, two speakers (the same two as for Experiments 1a and 1b) were again recorded. This time, the male speaker, who was recorded separately from the female speaker, read each sentence with the same intonational structure. The accent pattern chosen was one intended to produce a "falling hat pattern", and the goal was to have the verb and object be of roughly equal phonetic prominence. To create these, the male speaker was instructed (and sometimes coached using an imitation method of elicitation) to produce a H* accent on the verb and an accent on the object that was phonetically ambiguous between a !H* and H*. In most cases this was not particularly difficult with some practice,[6] although it was observed that, impressionistically, placing a prenuclear accent on verbs in the high frequency verb condition sometimes required more concentration. Many repetitions of each of the test sentences were produced in this way, and the most impressionistically natural, fluent production of each item with the desired accent pattern was selected as the answer sentence to be used. These were then made to follow as answers to the question recordings as in Experiments 1a and 1b.

Because the test sentences in Experiment 2 were made to have the same accentual structure, it was desirable to have fillers that would produce a wider range of range of prosodic prominence across experimental items. Therefore, unlike in Experiments 1a and 1b, where the fillers varied syntactic structure (recall that some contained adjuncts), the 36 filler sentences for Experiment 2 varied nuclear accent location (and, necessarily, focus structure). Eighteen of the filler sentences appeared in narrow verb focus contexts (and so had a nuclear accent on the verb), while the other 18 appeared in narrow subject focus contexts (and so had the nuclear accent on the subject). The structure of the sentences was otherwise analogous to those of the test sentences (i.e., SVO), although the verbs in the filler sentences all had frequencies that fell within the mid frequency range (CELEX log frequencies between .47 – 1.14).

[6] Subsequent to Experiments 1a and 1b, the speaker who read the answer sentences had completed (and excelled in) a course in intonational phonology (Ling 111) at UCLA, facilitating the task of eliciting the desired productions.

Unlike the first two experiments in this chapter, Experiment 2 made use of a between-subjects design, in part because it was a concern that multiple exposures to verbs (especially lower frequency verbs) might weaken any frequency-related differences. Finally, two lists were formed, containing all fillers and all 36 test items, with the test sentences counterbalanced across the two lists, so as to appear equally in the two focus conditions. One pseudorandomization was used for the two lists, and an additional version of each list was created that had the items appear in the inverse order, to help reduce any possible effect of trial order on listeners' ratings.

2.4.1.2 Participants

Seventy-four native speakers of American English were recruited from UCLA. None had participated in Experiments 1a or 1b, and none reported any previous hearing or communication disorder. Participants received either monetary compensation or course credit.

2.4.1.3 Procedure

All aspects of the prominence rating task itself were carried out as in Experiments 1a and 1b. Subsequent to prominence rating, however, participants also completed the Autism Spectrum Quotient (AQ). The self-report questionnaire, administered to participants electronically, consists of 50 items measuring autistic-like personality traits along five dimensions: Social Skills (e.g., ‘‘*I would rather go to a library than a party*.”), Attention to Detail (e.g., ‘‘*I don't usually notice small changes in a situation, or a person's appearance*.”), Attention Switching (e.g., ‘‘*I frequently get so absorbed in one thing that I lose sight of other things*.”), Imagination (e.g., “*I find it difficult to imagine what it would be like to be someone else*.”), and Communication (e.g., ‘‘*I frequently find that I don't know how to keep a conversation going*.”). Participants provide

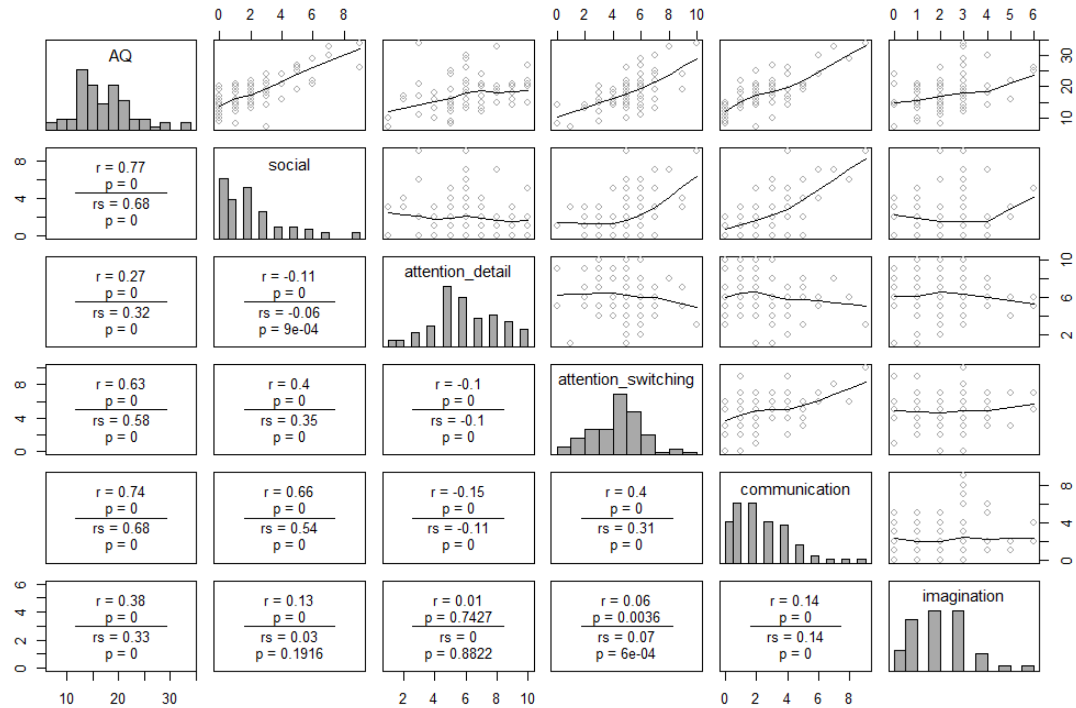


Figure 2.5. Matrix of scatterplots and Pearson (r) and Spearman (rs) correlations (with p values for each) for listeners' AQ scores in Experiment 2. Scores are shown for each of the individual subscales, as well as for the total AQ scores.

"definitely agree", "slightly agree", "slightly disagree" or "definitely disagree" responses to such statements, and receive a point for each autistic-like response (e.g., a "definitely agree" or "slightly agree" response would earn a point in each of the examples just given).[7] Although the AQ is not a diagnostic tool, it has been reported that scores of 32 and higher represent the low end of the distribution of scores for the clinical population (specifically, those with a diagnosis of High Functioning Autism), although there is some overlap with the non-clinical population (Baron-Cohen et al. 2001). Of the participants sampled in Experiment 2, two had scores above this (one scored 33 and the other 34). Figure 2.5 shows the distribution of scores (histograms on the diagonal) for each of the AQ subscale and for total AQ scores, as well as the correlations

[7] Some recent studies (e.g., Yu 2010,, Jun and Bishop in progress) have used a scoring method that treats the possible responses as a 4-point Likert scale. In the present study, both the binary and Likert scale scoring methods were calculated and tested during modeling, but only the binary scoring method was used for the final analyses.

between them (graphically as scatterplots above the diagonal, with correlations shown in the lower triangle).

2.4.2 Results

2.4.2.1 Comparison with Experiments 1a and 1b

Before exploring a more sophisticated model of listeners' prominence ratings, a simple analysis was carried out in order to compare any overall effect of focus with the effects found in Experiments 1a and 1b, since additional control was exerted over the form of the stimuli in Experiment 2.

Average prominence ratings are plotted in Figure 2.6, where it is evident that the focus effect was not only replicated, but was more robust, in that whether it was the verb or the object that was more prominent was (on average) entirely dependent on focus. Unsurprisingly, this interaction was highly significant, as shown in Table 2.8. Thus the test sentences in Experiment 2, which were purposefully more ambiguous with respect to the relative prominence of verbs and objects, allowed information structure to play a larger role.

It is also apparent from Figure 2.6 is the fact that average ratings for all test words were lower than those in Experiments 1a and 1b. While this may reflect the acoustic properties of the words themselves, it is more likely due to the additional prosodic variability across items in Experiment 2. Recall that in Experiments 1a and 1b, the location of the nuclear accent in the fillers deviated from the object only about half of the time. In Experiment 2, the nuclear accent was always in a marked position in the fillers, occurring on either the verb or the subject. It is possible that these accents were perceived as more prominent by listeners than the accents in the test sentences, causing the accents in test sentences to sound less prominent in comparison. This

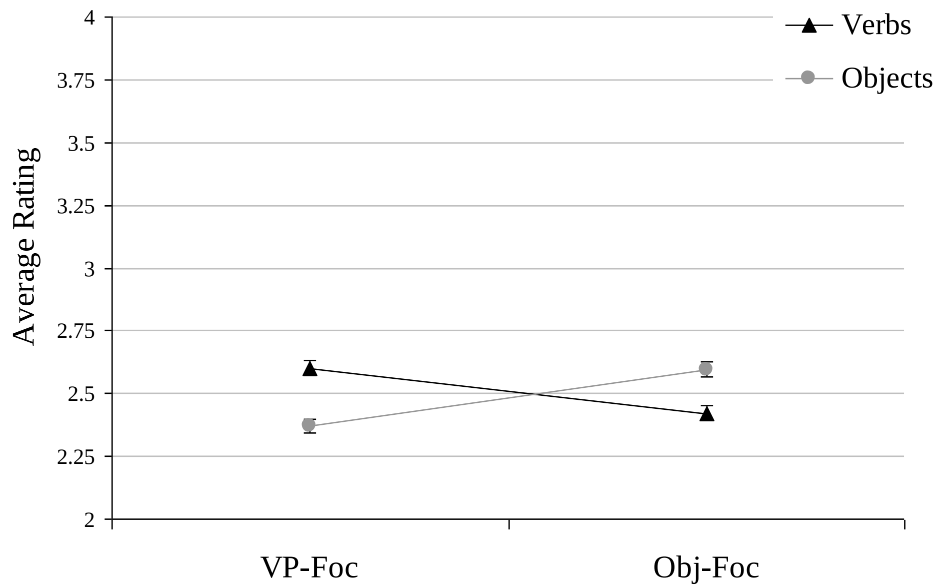


Figure 2.6 Average prominence ratings for verbs and objects in test sentences in the three focus conditions in Experiment 2. '1' is lowest in prominence, '5' is highest. Error bars show standard error.

Table 2.8 Results for fixed-effect factor in the mixed-effects model of listeners' prominence ratings in Experiment 2.

Fixed effects	*Estimate*	*Std. Error*	*t value*	*p value*
(Intercept)	2.5843	0.0838	30.85	**< .0001**
Word (verb)	-0.2222	0.0314	-7.08	**< .0001**
Focus Size (VP)	-0.1760	0.0314	-5.60	**< .0001**
Word(verb)*Focus Size (VP)	0.4017	0.04441	9.04	**< .0001**

could be due to their acoustic properties, although listeners have been shown to judge accents in marked positions as more prominent even when controlling for their acoustic prominence (e.g., Rosenberg, Hirschberg & Manis 2010, at least for speakers with bilingual experience).

To explore this possibility further, the average rating was calculated for verbs in the 18 fillers that had narrow verb focus[8], and plotted in Figure 2.7 along with average ratings for the

[8] Recall that, since subjects rated only verbs and objects for all sentences, there were no ratings for nuclear accented subjects, which represented the other 18 filler items.

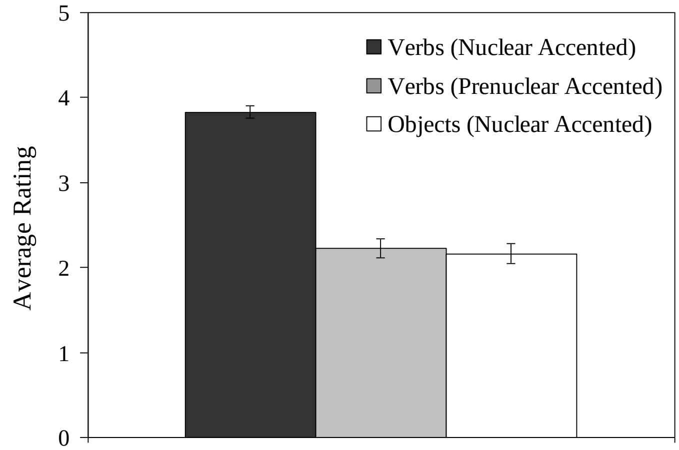


Figure 2.7. Average ratings for words bearing accents of different status. For prenuclear accented verbs and nuclear accented objects, the values shown collapse across the two focus contexts they occurred in. Nuclear accented verbs were fillers and occurred only in narrow verb focus contexts.

verbs and objects in the test sentences (collapsing across focus conditions). As can be seen, nuclear-accented verbs were rated as nearly twice as prominent as the less marked prenuclear accented verbs or nuclear accented objects. Whether this is due to the acoustic properties, their marked status, or both, it seems likely that having more such cases throughout the course of the experiment might have had a lowering effect on the perceived prominence of the test words.

Thus, despite some overall differences, which likely result from details of the experimental design differences between Experiment 2 and the previous experiments, it is clear that the basic focus effect was replicated. The next sections explore whether the focus effect occurred equally across the lexical frequency conditions.

2.4.2.2 Testing for interactive effects on prominence ratings

An additional round of modeling was carried out to understand the focus effect in a larger context of prominence perception. To keep the models maximally interpretable, verbs and objects were modeled separately. Mixed effects linear models were constructed as above, but tested for a larger variety of factors. For both verbs and objects, preliminary models contained the primary experimental variables: (a) *focus size*, (b) *verb frequency* (low, mid, or high), as well as (c) *object frequency* (the CELEX log frequency, as a continuous variable), listeners' scores on each of the five *AQ* subscales[9], and the control variables (d) *trial* and (e) experimental *list*. In the model, the primary experimental variables were permitted to enter in to two-way interactions with each other, and also with *trial* and each of the *AQ* subscales. Random effects factors included intercepts for subject and item, and a by-trial slope for subject. From this preliminary model, a factor was dropped if doing so did not significantly deteriorate the fit of the model. Results of verb ratings and object ratings are discussed separately.

2.4.2.2.1 Verbs

Mean ratings for verbs in the two focus conditions are shown by *verb frequency* group in Figure 2.8. A comparison of the full model (LogLik = -3165) with a model containing only random effects (LogLik = -3188) showed the full model to be superior (p <.0001). According to the full model, reported in Table 2.9, there was a significant main effect for *trial*, such that verbs occurring in later trials were rated higher, a likely artifact of listeners' learning to use the rating scale as a wider range of stimuli were encountered as the experiment progressed. There was no significant effect for the *object frequency*, although it contributed to the overall fit of the model. There was a significant effect for one of the *AQ* measures, namely AQ-Attention Switching; higher AQ-Attention Switching scores (indicating worse attention switching abilities) were associated with higher prominence rating for verbs, although there was no interaction with any other factor in the model.

[9] A separate round of modeling carried out this way but including only the total AQ scores (i.e., all subscales combined) determined that total AQ did not contribute significantly to any model.

Unsurprisingly, given the results just reported in 2.4.2.1, there was a highly significant effect of *focus size*, such that verb ratings were more prominent under broad focus than under object focus. There was also an effect for *verb frequency*, with verbs being rated as more prominent if they came from the low frequency group than from the mid frequency group. However, the effect of the verb's frequency on its perceived prominence was not consistent, since verbs from the high frequency group were also rated as higher than verbs from the mid frequency group, although the effect was only marginally significant. Further, a version of the model with the high frequency group used as comparison showed no difference between it and the low frequency group (High vs. Low: β=.0768, *SE*=.0982, t=.078, p=.4342). Thus the three frequency groups, as a baseline, do not correlate reliably with prominence ratings, although the lowest group was numerically lowest in ratings.

Of most interest however, was any interaction between the *verb frequency* and the focus conditions. As can be seen in Figure 2.8, numerically, the increase in prominence ratings for verbs going from object focus to VP focus was smallest for high frequency verbs. However, the difference is largest for mid frequency verbs rather than for low frequency verbs. Indeed, as can be seen in the output of the model, the effect for focus was only marginally smaller for high frequency verbs compared with mid frequency verbs, and a second model with the high frequency condition as default determined that there was no significant difference between the high and low frequency conditions (β=.0503, *SE*=.0701, t=.718, p=.4728). Thus, a verb's lexical frequency was not a particularly good predictor of its perceived prominence overall, nor did it appear to be modulating the focus effect in a consistent way.

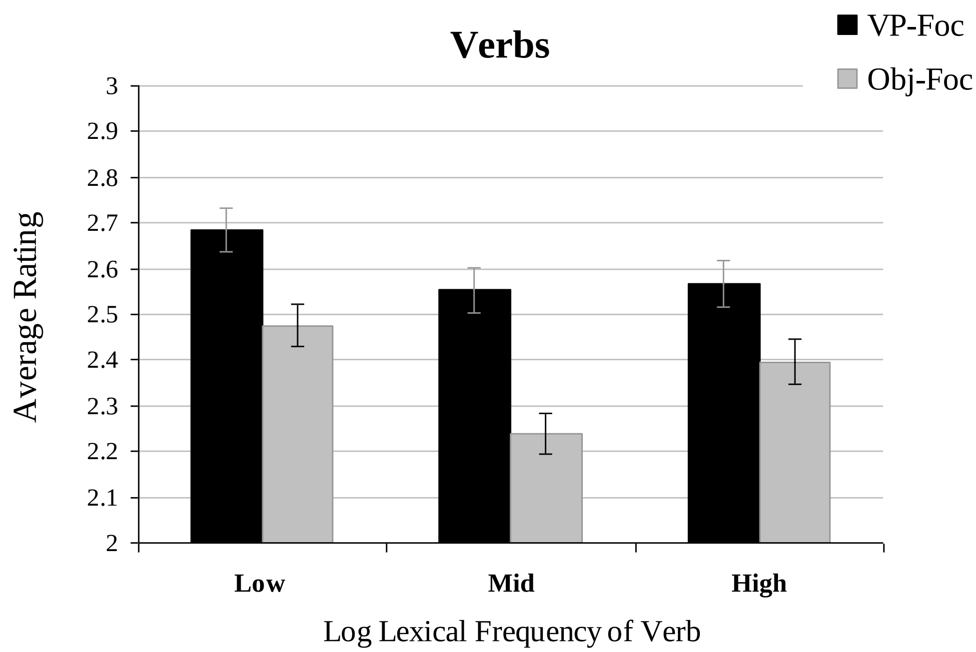


Figure 2.8. Average prominence ratings for verbs across all frequency and focus conditions in Experiment 2. Error bars show standard error.

Table 2.9 Results for fixed-effect factors for the model of listeners' ratings of verbs in Experiment 2.

Fixed effects	*Estimate*	*Std. Error*	*t value*	*p value*
(Intercept)	1.5448	0.2267	6.82	< **.0001**
Trial	0.0028	0.0011	2.64	**.0083**
AQ-Attention Switching	0.1056	0.0383	2.76	**.0058**
LogFrequency of Object	0.0595	0.0589	1.01	.3126
Focus(VP)	0.2999	0.0496	6.05	< **.0001**
LogFrequency of Verb(High)	0.1636	0.0987	1.66	.0931
LogFrequency of Verb(Low)	0.2404	0.0995	2.42	**.0158**
LogFrequency of Verb(High)*Focus(VP)	-0.1313	0.0701	-1.87	.0612
LogFrequency of Verb(Low)*Focus(VP)	-0.0806	0.0702	-1.15	.2486

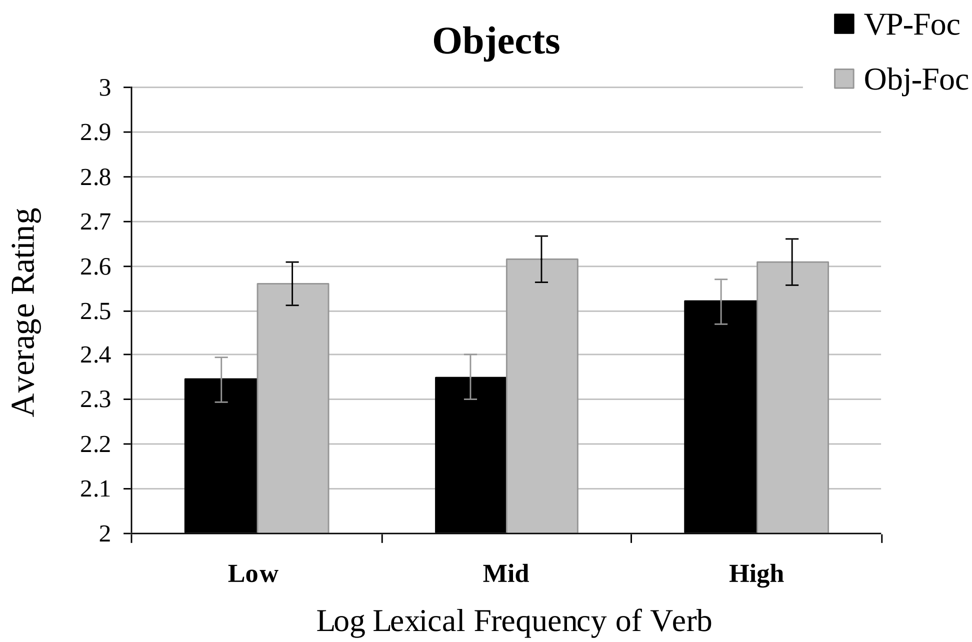


Figure 2.9 Average prominence ratings for objects across all frequency and focus conditions in Experiment 2. Error bars show standard error.

Table 2.10 Results for fixed-effect factors for the model of listeners' ratings of objects in Experiment 2.

Fixed effects	*Estimate*	*Std. Error*	*t value*	*p value*
(Intercept)	1.9393	0.2370	8.18	< **.0001**
Trial	0.0042	0.0014	3.08	**.0021**
AQ-Attention Switching	0.1210	0.0408	3.19	**.0014**
LogFrequency of Object	-0.0681	0.0569	-1.20	.2314
Focus(VP)	-0.2788	0.0496	-5.62	< **.0001**
LogFrequency of Verb(High)	-0.0252	0.0951	-0.26	.7918
LogFrequency of Verb(Low)	-0.0785	0.0959	-0.82	.4129
LogFrequency of Verb(High)*Focus(VP)	0.1910	0.0701	2.72	**.0065**
LogFrequency of Verb(Low)*Focus(VP)	0.0729	0.0703	1.04	.2998

2.4.2.2.2 Objects

Turning now to the perceived prominence of objects, Figure 2.9 plots average ratings in the same manner as was done for verbs, including grouping by *verb frequency* (i.e., not the object's own frequency). The output of the model is reported in Table 2.10. There was an effect for *trial,* and there was also an effect for *AQ-Attention Switching*—both of these factors were positively related to perceived prominence. Just as it did not have a significant effect on verb ratings, o*bject frequency* did not have a significant effect on object ratings (although, as evident from the coefficient in the model's output, the trend was in the opposite direction for objects and verbs).

The effect for *focus size* was highly significant, with objects being rated less prominent under broad focus compared with object focus. And while there was no main effect for *verb frequency* on object ratings, there was a significant interaction between *focus size* and *verb frequency*. In particular, the increase in perceived prominence of objects going from VP focus to object focus was less when the object followed a high frequency verb compared to a mid frequency verb. However, in a second version of the model, which compared the high and low frequency conditions, the effect was shown to be only marginally significant (β = -.1181, *SE* = .0702, t = -1.68, p =.0927). Still, somewhat surprisingly, this suggests that the relation between verb frequency and the size of the focus effect was more consistent for object ratings than it was found to be for the verb's own ratings in the previous section. Additionally, as can be seen clearly in Figure 2.9, the interaction between *verb frequency* and *focus size* is being driven primarily by changes in object ratings in the VP focus condition, not the object focus condition. This indicates that, to the extent that a prenuclear verb's frequency was able to influence the perception of a nuclear accented object's prominence, it is dependent on that verb being part of the focus constituent.

Finally, it is important to note that it is not the case that any of the effects related to the verb's frequency is actually reflecting the object's own frequency, since the objects' frequencies were also accounted for in the model.

2.4.2.2.3 Relative prominence

The fact that focus condition affected both verbs and objects, and also that object ratings were influenced by the frequency of preceding verbs, suggests that listeners were attending to a relative prominence relation. To explore this matter more specifically, the relative prominence of verbs was calculated by subtracting the object's rating from the verb's rating for each trial, and that value was then to be modeled using the same factors as above.

Average relative verb prominence ratings are shown in Figure 2.10, plotted by focus condition. Numerically, there appears to be a consistent relationship between the relative prominence of the verb and the verb's lexical frequency under VP focus only, and so a model was run on that condition only. According to the model, reported in Table 2.11, the only significant difference was between the high frequency and low frequency groups. However, a second model was run which included the verb's frequency as a continuous rather than a categorical variable, and in this model the verb's frequency was highly significant (β = -0.1551, *SE* = .059, t= -2.63, p =.0086). Thus there is some evidence that the frequency of the verb is a predictor of prominence verb's perceived, but primarily when calculated relatively, and only when focus is on the entire VP . Finally, it is also noteworthy that when focus was calculated relatively for the verb, the object's lexical frequency also became a more important predictor of ratings. There was no evidence that this effect interacted in any way with focus, however.

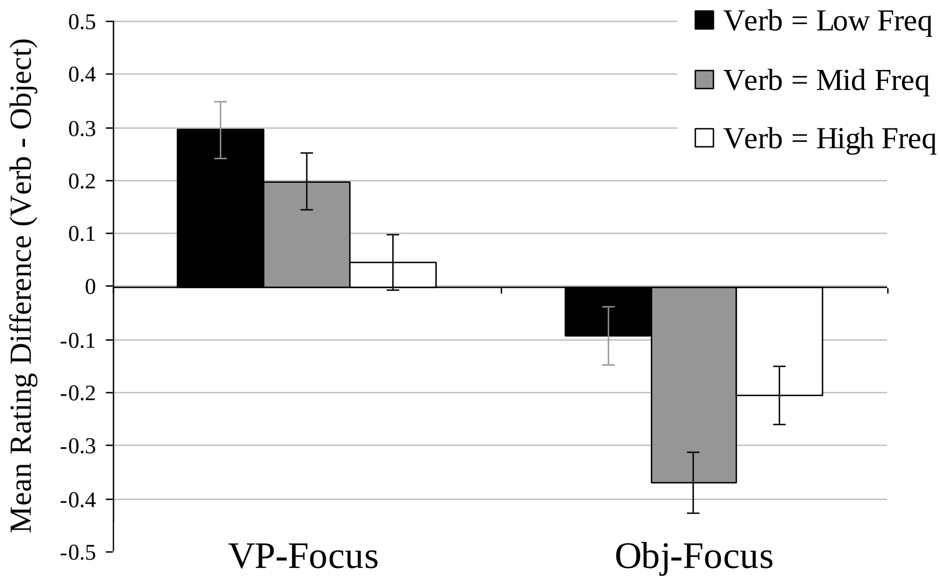


Figure 2.10. Mean relative prominence ratings for verbs (calculated as verb rating – object rating for each trial) in each of the verb frequency conditions. Error bars show standard error.

Table 2.11 Results for fixed-effect factors for the model of listeners' ratings of relative verb prominence (calculated as the verb's rating - the object's rating) in Experiment 2

Fixed effects	*Estimate*	*Std. Error*	*t value*	*p value*
(Intercept)	-0.0512	0.1450	-0.353	.7241
LogFrequency of Object	0.1734	0.0718	2.433	**.0150**
VerbFrequency: Low vs. Mid	0.1790	0.1130	1.584	.1133
VerbFrequency: High vs. Mid	-0.1255	0.1120	-1.12	.2628
VerbFrequency: High vs. Low	0.2121	0.0951	2.23	**.0064**

2.4.3 Discussion

The primary result of Experiment 2, as for Experiments 1a and 1b, was that listeners' perception of prominence was influenced by information structure. The effect of the size of the focus constituent was somewhat more robust in Experiment 2, in that there was a crossover effect, indicating that whether it was the verb or object that was more prominent depended primarily on whether focus was narrow or broad. This was likely the result of the fact that the accent pattern of the stimuli was designed so as to make the bottom-up information about their relative prominence more ambiguous. The fact that the sentence's information structure was such a crucial predictor when the physical stimulus was most ambiguous is the hallmark of an auditory illusion.

One of the more specific goals of Experiment 2, however, was to examine whether this focus effect was systematically dependent on listeners' knowledge about other factors that condition prosodic prominence. As discussed in Section 1.2.2, English-speaking listeners have been shown to have expectations that low frequency words are more prominent than higher frequency words, even when they are not necessarily pronounced that way. It was predicted based on Calhoun's (2006) work that listeners would expect a larger "boost" in prominence for lower frequency verbs, since their baseline (i.e., non-focus) expected prominence would be higher. This was expected to hold true primarily for verbs, since prenuclear prominence is said to be more sensitive to non-information structural factors, while nuclear prominence is primarily related to information structure.

The frequency-related results were rather mixed, however. Although there was an interaction such that the highest frequency verb group did show numerically the smallest focus effect, as Calhoun's (2006) theory would predict, there was otherwise not a reliable relationship between verb frequency and the size of the focus effect for verbs. Interestingly, however, the verb's frequency was shown to have some influence on the perceived prominence of objects. It was found that the perceived prominence of the object increased when the preceding verb's frequency was high, at least when focus was on the VP. Presumably, listeners expect high frequency verbs to be low in prominence, but this was perceptually achieved by an increase in the object's prominence. A similar finding was that the verb's frequency was a more reliable predictor of the difference between verb and object ratings than of verb ratings alone. Again, however, this depended on the focus being broad on the VP. It therefore seems that relative prominence calculations may be more complicated when both the verb and the object share the focus constituent, while their prominence may be manipulated more independently when only the object occupies the focus constituent.[10]

Another interesting finding from Experiment 2 was that individual differences related to autistic traits were relevant to predicting prominence ratings. As discussed in Section 1.2.3, individuals with high scores on the Communication subscale of the AQ (indicating poorer communication skills) rely less on pragmatic information in sentence processing (Nieuwland et al. 2010, Xiang et al. 2011). Thus, there was the possibility that the effect of discourse context, which was used to manipulate focus structure in the test sentences, might have been attenuated for such individuals in Experiment 2. Similarly, individuals with high AQ have been shown to have weaker top-down effects from lexical knowledge in speech perception (Stewart and Ota 2008, Yu et al. 2011), and so autistic traits could have conceivably modulated the effects of verb frequency in Experiment 2 as well.

However, total AQ scores and AQ-Communication scores were predictors of neither of these effects. Instead, it was found that the Attention Switching subscale of the AQ was the relevant measure of autistic traits, and was positively associated with prominence ratings for both prenuclear verbs and nuclear accented objects. While it does not bear on my particular research questions, this finding may help us better understand how prominence perception takes place by human listeners, and so it is worth considering it in light of some of the literature reviewed in Section 1.2.2, in particular the distinction made between restorative prominence perception, or

[10] Possibly, the added complexity in calculating relative prominence may help explain why listeners seem to be less accurate at recovering speakers' intended VP focus productions compared with intended object focus productions (see discussion of Breen et al.'s (2010) results in Section 1.4.1). Another possibility is that the results may be influenced by the frequency or predictability of the verb+object sequences used in the experiment (i.e., the entire VPs in the test sentences); this possibility will have to remain an issue for future research.

perceptual repair of the signal, and processing-based prominence perception of the sort proposed by Cole et al. (2010).

Recall that under restorative processes, listeners perceptually “repair” the signal so as to make it conform to their linguistic knowledge and experience-based expectations. Thus, when listeners exhibit top-down perception, it is expected that there is an observable basis for the illusion in, for example, speakers’ productions of the relevant structure. It seems unclear how the effect for AQ-Attention Switching can be explained by a restorative process, since it would predict an interaction with the linguistic and lexical variables rather than an across-the-board main effect.

Alternatively, the basic idea behind the processing-based theory was that the sensation of prominence occurs when attentional resources are being taxed, such as when lexical access is attempted for a word with a low resting activation level. This would seem to make the prediction that individuals with lower baseline attentional allocation abilities should, other things being equal, be more likely to experience a given stimulus as prominent. This seems to better explain the result found, assuming that the AQ-Attention Switching subscale is in fact measuring this attentional-processing capacity. This hypothesis could be more directly tested in a study that attempted to predict individual differences in prominence perception as a function of performance on an auditory selective attention task, or possibly within-subject as a function of task demands. Particularly in light of the links between prominence and selective attention reported by Kristensen, Wang, Petersson, & Hagoort (2012), a processing based theory along the lines of that proposed by Cole and colleagues seems promising.

The prominence-as-resource-exertion account would also explain the example of syntactically-driven prominence perception in Finnish discussed in Section 1.2.2 In their study, Vainio and Järvikivi (2006) found that moving a noun into a focused position resulted in its being perceived as prominent, even when acoustic information was controlled. It seems that simply being focused comes with its own prominence, which would be predicted by processing-based theory, since focused information is known to undergo deeper or additional processing compared with non-focused information (Rooth 1992, Cutler & Fodor 1979, Birch & Rayner 1997, Ward & Sturt, 2007, Braun & Tagliapietra 2010, Wang, Bastiaansen, Yang, & Hagoort 2011). Importantly, since Vainio and Järvikivi (2007) found that speakers do not produce focused nouns with added prominence in relevant constructions (indeed, they pronounce them with *lower* prominence), it does not seem that restorative processes can account for that finding either. Thus, one insight from the experiments in this chapter is that prominence perception is complex in a way that other aspects of speech perception might not be, in that perceived prominence can emerge as a result of either a knowledge-driven restorative process, or as an epiphenomenon of effortful processing.

2.5 Local General Discussion

In the experiments presented in this chapter, the basic question asked was whether English-speaking listeners have expectations regarding how prosody can be used to disambiguate the size of the focus constituent in simple SVO sentences. It was found that listeners have very clear expectations, ones that are strong enough to serve as the basis for an auditory illusion. Moreover, these expectations involved prenuclear prominence, indicating that prenuclear prominence is not free of information structural meaning. Interestingly, however, listeners also had expectations about nuclear prominence in a way that suggests a prenuclear/nuclear relation is crucial to the disambiguation. Not only does this contradict the claim that prenuclear accents lack information structural relevance, it also seems particularly irreconcilable with a theory in which focus is expressed by the presence versus absence of pitch accents. This is especially troubling since listeners' expectations in Experiment 2 closely mirror what speakers have been shown to do in the production studies reviewed in Section 1.4.1, suggesting that the pattern in question represents conventionalized knowledge.

Given that the link between production and perception patterns was established so clearly in the present study, before concluding this chapter, it is worthwhile to consider why results of previous studies have been so mixed. For example, as reviewed in Section 1.4.2, previous studies have attempted to elicit expectations about prenuclear accents by presenting question contexts and answer sentences to be "correctly" paired by listeners (Gussenhoven 1984, Breen et al. 2010), or presenting listeners with question-answer pairs to be rated for appropriateness (Welby 2003, Birch & Clifton 1995). Possibly this sort of semantic judgment task simply does not encourage listeners to attend to prosody beyond its coarsest features, such as nuclear accent location. For example, it has long been noted that speakers, when reading aloud, tend to go into a mode of "reading what it says" (Brazil, Coulthart, and Johns 1980, as cited in Gussenhoven 1983:68), i.e., speaking for fluency rather than the encoding of structure and context-sensitive meaning.[11] Indeed, it has recently been shown that speakers do not disambiguate certain syntactic structures reliably unless they are explicitly aware of the need to do so (Snedecker and Trueswell 2003; see also Jun 2010 and the discussion therein). It seems reasonable that listeners may similarly engage in a sort of "listening for comprehension" when asked to make semantic judgments, largely ignoring anything more subtle than the placement of nuclear accent. Indeed,

[11] "Fluency" in this case most likely means "reading without making an error", and relying on default phrasing and accentuation patterns. However, it may to some extent also mean "reading rhythmically". This seems to be the case in English, as, in addition to the common observation that read speech contains more pitch accents than conversational speech, it has also been reported to contain remarkably evenly-spaced accentuation (Ostendorf, Price, Shattuck-Hufnagel 1995).

the very thing that the existence of restorative perceptual processes implies is that listeners will be overly sympathetic when asked to give judgments about appropriateness, since they tend to perceive the signal as more contextually appropriate than it truly is. Thus, a reliance on context, and vulnerability of perception to expectations may combine to cause listeners to be more ambivalent in making acceptability judgments in an off-line task. This interpretation finds some support in the experiments presented in Chapter 3, which approached listeners' expectations using a more on-line methodology.

CHAPTER 3

Prosodic expectations in processing: cross-modal priming

3.1 Introduction

The experiments in Chapter 2 found that English-speaking listeners had strong expectations about how prosody disambiguates the focus size contrast, since these expectations were evident in their perceptual restoration of the signal. In the present chapter I will approach this same matter using a more on-line measure of listeners' expectations: the cross-modal lexical decision. As is discussed below, this paradigm is especially well-suited to exploring questions about what listeners expect from the signal, because priming effects in word recognition are not automatic, but dependent on a good match between expected input and actual input. In order to understand the details of listeners' prosodic expectations, I will again rely on what is known from production studies, and my primary interest is in how prenuclear prominence might contribute to the disambiguation of the information structural contrast in the minds of listeners.

3.2 Cross-Modal Associative Priming

In the cross-modal associative priming paradigm, a listener is auditorily presented with a word (e.g., *dog*), either in isolation or embedded in a sentence, and must then respond in some way to a related target word (e.g., CAT), presented visually. [12] In the lexical decision version of the task, which will be used here, listeners' response to the target word must be a "yes" or "no" to indicate whether they recognize the target as a real word. "Priming" occurs when the target word is

[12] Throughout this chapter, I will adhere to the convention of representing auditory prime words in italics and visual target words in capitals.

recognized more quickly in comparison to an unrelated control target (e.g., *dog* → CAT vs. *dog* → BUS).

The use of cross-modal priming in the present study is motivated by recent work demonstrating that priming effects are highly dependent on semantic and prosodic contextual factors. This basic idea was emphasized in early work on lexical activation by Foss & Ross (1983), who argued that priming of one word by another is not an automatic process based just on the first word's meaning; rather it results from a prime word's sentence level meaning, what they called the "effective context". In a thorough review, Norris, Cutler, McQueen & Butterfield (2006) discuss evidence that in fact priming seems to fail in just those cases where the associative relationship that the prime has with the target is not supported by a more global context (see also Tabossi 1996). In particular, priming is actually less reliable when the prime is embedded in a sentence than when it occurs in isolation, which they argue is because when a word is in isolation, it effectively provides its context for itself. However, when embedded in a sentence, there is an increased chance that the larger sentence-level context might be less supportive.

Consistent with this, and particularly important for the present purposes, is a finding of Norris and colleagues, which is that, while isolated words prime consistently, priming effects for words embedded in sentences are crucially dependent on prosody. In their (2006) study, the authors demonstrate that primes embedded in sentences did not facilitate recognition of targets unless the sentence contained a highly prominent ("contrastive") nuclear accent somewhere in the sentence. For example, in sentences such as *"The rebels were expected to be reasonable and to <u>come</u> and surrender to the government forces."* priming of the related target GO by the prime *come* did not occur if the sentence had what Norris et al. called "neutral" prosody. Although the authors do not speak explicitly in terms of information structure, they conclude that upon hearing the word produced with a prominent accent, listeners were encouraged “not only to consider the meaning of the sentence, but also to construct the alternative scenario implicit in the utterance” (Norris et al. 2006:174). That is, these words were interpreted as focused, and thus have meanings that are implicitly contrastive. This interpretation would also help explain why, across experiments, isolated primes are more reliable primes than sentence-embedded primes: assuming that such primes are also produced as isolated, single-phrase utterances, they will always be marked by a nuclear accent, increasing the probability of evoking contrastive meaning. Importantly, there is further evidence from priming that prominent accentuation evokes a contrastive interpretation, reported by Braun and Tagliapietra (2010). In their study, in Dutch, listeners heard primes such as *flamingo,* which occurred as objects in sentences such as *“In Florida he photographed a* <u>*flamingo*</u>*”*. These sentences were presented to listeners with either “neutral” prosody (a prenuclear accent on the subject followed by a downstepped nuclear accent on the object) or with “contrastive” prosody (a highly prominent H*L in the ToDI conventions for transcribing Dutch intonation; Gussenhoven 2005). They found that priming of contrastively related targets such as PELICAN occurred only when the prime was contrastively accented. However, if the target was semantically related, but in a non-contrastive way (e.g., PINK), priming occurred regardless of prosody, and was weaker.

3.3 Present investigation

Braun and Tagliapietra’s findings are important because they show both that there is psychological reality to the notion of focus alternatives (i.e., to be contrastive is to activate alternatives), and that a particular prosody cues that contrastive meaning in the absence of explicit discourse context. The experiments in this chapter were intended to similarly probe listeners for such knowledge, but to do so by providing them with both the prosody and an explicit context. This allows us to pair sentences with an unambiguous focus interpretation (in our case broad VP or narrow object focus) with different prosodic structures (sentences with or without a prenuclear accented verb) and observe the resulting priming. If listeners have internalized the patterns reported in production studies (as the listeners in Experiments 1a, 1b, and 2 were shown to have done), we would expect priming to occur only when listeners encounter those patterns. For example, a narrowly focused prime word like *blonde* in the sentence in (1a) should facilitate recognition of related target BRUNETTE only if the sentence lacks a prenuclear accent, since narrow focus corresponds to low prenuclear prominence. Similarly, priming of BRUNETTE by *blonde* should be weaker (or fail to occur) if the sentence has VP focus, but lacks a prenuclear accent, since expectations should be for broad focus sentences to contain higher prenuclear prominence.

(1) a. He kissed a [*blonde*]$_{Foc}$

b. He [kissed a *blonde*]$_{Foc}$

These predictions are also summarized in Table 3.1, with the subset of questions that will be tested (and which experiments tests them) shown in the shaded cells. Specifically, in Experiment 3a, I test the basic prediction that an SVO sentence without a prenuclear accent on the verb is appropriate in both broad VP and narrow object focus sentences. In Experiment 3b, I test whether SVOs with a nuclear accent on the object and a prenuclear accent on the verb are appropriate for the expression of narrow focus on an object.

Table 3.1. Predictions of priming patterns, based on listeners preferences in production studies. # Indicates a contextually inappropriate prosodic realization.

	+ Prenuclear Accent	- Prenuclear Accent
VP focus	√	# (3a)
Object Focus	# (3b)	√ (3a, 3b)

At this point it is also worth noting that, as discussed in Section 1.3, Focus Projection theories do not all agree on this matter. In particular, Gussenhoven's (1999) theory predicts all conditions to be equally felicitous, since prenuclear accents are always added optionally. In Selkirk's (1995) theory, however, only VP focus is truly ambiguous, since the presence of a non-focus-marking accent (i.e., a prenuclear accent in our case) is interpreted as marking the information as new. Therefore, for Selkirk, a prenuclear accent on a verb should be infelicitous for a narrow focus sentence, since the verb is given. In any case, a hypothesis based on production studies clearly predicts there to be no ambiguity; higher prenuclear prominence is preferred for broad focus, and dispreferred for narrow focus.

3.4 Experiment 3a

3.4.1 Methods

3.4.1.1 Materials

The basic design of primes, targets and sentences for Experiments 3a and 3b was similar to that used by Braun and Tagliapietra (2010). The materials consisted of target words (e.g., PEPPER), and primes that were either related to the target or were unrelated control primes (e.g., *salt* and *tape*, respectively). Because it has been shown that contrastive associative relationships are most likely to facilitate priming in contrastive contexts (Braun and Tagliapietra 2010), all related primes were contrastive with the targets. The prime-target pairs were selected as follows. First, 32 English nouns, mostly monosyllables or disyllabic words with a strong-weak stress pattern were chosen to serve as the primes. These primes were then used in a web experiment to elicit contrastively related associates from 80 native English-speakers. These participants were presented with the 32 primes in frames such as "He didn't say "X", he said ___", to which they responded with the first word that came to mind. For each of the primes, the most frequent response was selected and used as the target for that prime (the mean association rate was 47.7% of responses; range 26.2% – 87.7%). Thirty-two simple SVO sentences were then constructed, in which the primes were to serve as the sentence-final objects. Care was taken so that, for each sentence, the object prime was the only word semantically related to the target. Thus 64 SVO sentences were prepared (one version of all 32 sentences containing the related prime, a second version containing the unrelated control prime).

In order to create the two prosodic conditions, the 64 sentences were then produced with two different accent patterns by a male speaker of American English trained in intonational phonology. First, a production was recorded in which the verb bore a prenuclear H* pitch accent with a following nuclear accent that I will call "H*", but which was intended to be ambiguous between a H* and a !H*, much like in Experiment 2. This production was used as the prenuclear accented condition (see Figure 3.1A), which would be used in Experiment 3b. A second version of each of the sentences was then read without a prenuclear accent on the verb (or on the subject, which was in most cases a pronoun). The object was produced with a prominent L+H* nuclear accent (see Figure 3.1B); this accent pattern was intended to be used for all answer sentences in Experiment 3a, and the prenuclear unaccented condition in Experiment 3b.

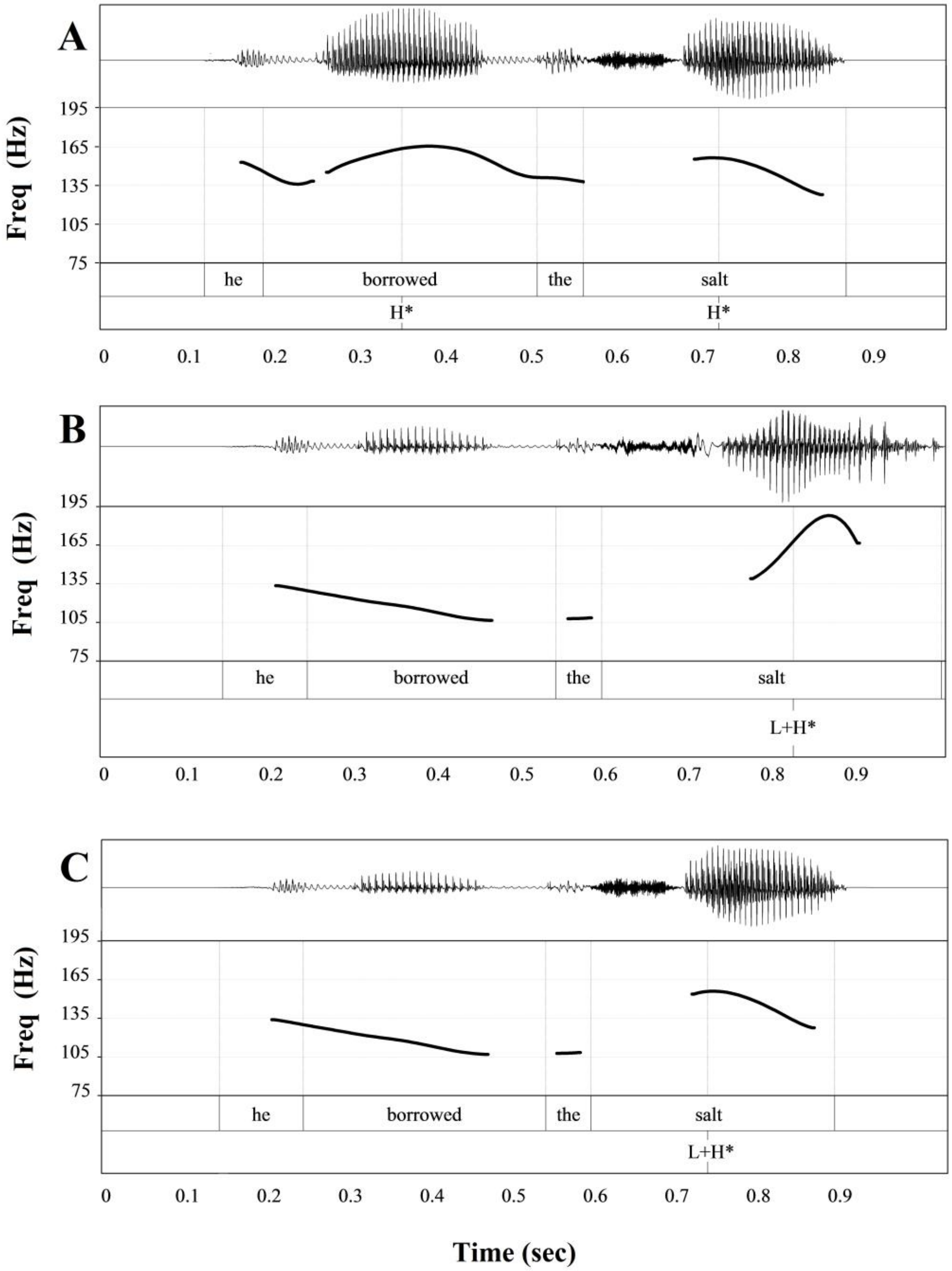


Figure 3.1. Example of an SVO test sentence in the two prosodic conditions. (A) shows the sentence produced with a prenuclear H* on the verb, used in the +prenuclear accent condition; (C) shows the sentence used for the –prenuclear accent condition, which was created by splicing the object "*salt*" from (A) into the production of the same sentence in (B).

Table 3.2. Acoustic properties of the verb and object in the two prosodic conditions for test and control sentences. The same object was used in both conditions, and so only one object is shown for each item type. Values shown are means with standard deviations in parentheses.

Test Items	**+ Accent Verb**	**– Accent Verb**	**Object**
Dur (ms)	294 (87)	257 (88)	466 (99)
Intensity (dB)	76.1 (2.9)	66.2 (3.3)	73.5 (3.1)
f0 min (Hz)	138 (12)	112 (7)	106 (9)
f0 max (Hz)	175 (19)	126 (10)	158 (13)
f0 range (Hz)	35 (14)	15 (8)	52 (14)

Control Items	**+ Accent Verb**	**– Accent Verb**	**Object**
Dur (ms)	304 (92)	260 (86)	520 (96)
Intensity (dB)	75.5 (3.4)	67 (4.9)	73.9 (3.6)
f0 min (Hz)	146 (13)	112 (9)	108 (12)
f0 max (Hz)	177 (16)	132 (14)	160 (12)
f0 range (Hz)	31 (15)	19 (9)	52 (15)

Although two accent patterns would be used in Experiment 3b, it was desirable to hold the acoustic properties of the primes themselves constant. This was accomplished by excising the production of the prime (i.e., the object) from the prenuclear accented condition (e.g., Figure 3.1A) and splicing it into the unaccented condition, replacing the original production of the L+H* object (Figure 3.1B). Thus, the sentences in the prenuclear accented condition were original, unedited productions of a H* H* pattern, and the final versions of sentences in the prenuclear unaccented condition (e.g, Figure 3.1C) were manipulated versions. These manipulated versions were still of the phonological structure L+H* on the object, although the accent was generally phonetically less prominent than the one it replaced. All experimental sentences (i.e., those containing related primes, and those containing unrelated control primes) were created in this way and saved as wav files.

To demonstrate that the prosodic manipulation resulted in the intended accentuation contrast, acoustic measurements were carried out for verbs and objects in the final stimuli.

Measurement criteria followed the recommendations in Turk, Nakai and Sugahara (2006). The mean values across items for the most common acoustic correlates of phonetic prominence are shown in Table 3.2. As can be seen, f0 max values for verbs in the +prenuclear accent condition were significantly higher than those of the objects for both the test and control conditions (on average 110% the f0 height of the object for each type of prime, consistent with a "falling hat pattern", or H* H* sequence). In addition, verbs were considerably longer and had higher intensity when they were produced with a prenuclear accent than when they were unaccented. Thus, while the primes themselves were the same productions with the same absolute acoustic properties, their prominence relative to a preceding verb was considerably different across the two prosodic conditions. Again, since Experiment 3a was to vary information structure and not accent pattern (see Table 4.1), only one prosodic condition (prenuclear unaccented) would be used for Experiment 3a.

Finally, in order to create the two information structural conditions that these sentences would occur in for Experiment 3a, lead-in questions were created. These were WH-questions such as *"What did Robert do?"* (in the case of VP focus) and *"What did Robert borrow?"* (in the case of object focus). In order to produce maximally contrastive contexts for the focused constituents, yes/no questions were additionally made to follow the WH questions, much like for Experiments 1b and 2. For example, the full question contexts were of the form *"What did Robert do after the party?...Did he leave?"* for VP focus, or *"Who did Robert kiss after the party?...Did he kiss Mary?"* for object focus. Question contexts, like the test sentences, were constrained such that only words unrelated to the targets could be used. All question contexts were produced and recorded by a female speaker of American English and were made to precede the SVO sentences in the stimuli, so that the SVO test sentences appeared as corrective answers to them. The full list of all materials used (sentences, test primes, control primes, and targets) is shown in Appendices C and D. In addition to these materials, there were also 96 filler sentences with filler primes and filler targets. 64 of the filler trials contained non-word targets; of the remaining 32 filler trials, half contained primes that were semantically unrelated to the target words, and half were related. In other respects, filler trials were the same as the experimental trials, with the same two prosodic versions of each. (Fillers were all original productions, with no splicing being done). An additional set of 6 filler sentences was also created to be used as items in a brief practice session to familiarize participants with the experimental task.

3.4.1.2 Participants

Ninety-two native English speakers, most of them members of the university community, participated as listeners in the lexical decision task in Experiment 3a. None of these participants had taken part in web-based association experiment used for stimulus design, and none reported any history of a hearing, speech or communication disorder. All received either monetary compensation or course credit.

3.4.1.3 Procedure

Participants took part in a cross-modal lexical decision task, individually in a sound-attenuated booth. The auditory stimuli were played binaurally over Sony MDM headphones at a comfortable listening volume (held constant across participants). Visual targets appeared on a computer screen directly in front of the participant (in 72pt white font on a black background), immediately at the offset of the sentence-final primes. Participants were to push a ‘yes’ or ‘no’ key as quickly as possible to indicate that they recognized the word on the screen (the ‘yes’ key corresponded to the dominant hand for each participant). Four lists were formed by taking all 32 test sentences and corresponding visual targets and rotating them through the two prime type conditions (related and unrelated prime) and the two information structural conditions (broad VP and narrow object focus); all sentences contained prenuclearly accented verbs. Thus there were 8 items per condition in each list, and participants were assigned (randomly) to one of the lists. A MATLAB script presented the stimuli in random order (different for each participant), and recorded responses and reaction times (RT). Following the lexical decision task, participants also completed (a) the Reading Span Task (Daneman and Carpenter 1980; Unsworth et al. 2005) as a measure of verbal working memory capacity, and (b) the Autism Spectrum Quotient. Participation in the entire experiment took approximately 40 minutes.

3.4.2 Results

Reponses to experimental targets were considered errors if the participant failed to hit the 'yes' key, or their response was slower than 1800 ms. For the 92 subjects, this resulted in 63 errors (approximately 2.1% of experimental trials), which were evenly distributed across the conditions and of no further interest. RTs for all correct responses falling within 2 standard deviations of the mean RT (511 ms; SD=103), a total of 2819 observations, were analyzed using mixed-effects linear regression using the *lmer* function in the *lme4* package (Bates and Maechler 2009) for R Statistics (R Development Core Team 2012).

The predictors that were of primary interest were (a) the linguistic predictor, the size of the focus constituent (VP versus object), (b) the prime type (related prime versus unrelated control prime), and, particularly, (c) their interaction. Also included were a number of stimulus-level variables known to be relevant to the lexical decision task: the CELEX log frequency of the visual target word and its length (in characters), as well as the reaction time to the preceding trial. Finally, the participant-level predictors were RSPAN score, scores on each of the AQ subscales, listener sex, and the interaction of these predictors with each of the primary predictors (a-c). The initial model included all predictors as fixed-effect terms, as well as random intercepts for participant and item, and a by-participant random slope for trial. From this initial model, factors were removed if doing so did not result in a significant decrease in model fit as assessed by a log-likelihood ratio test. After removing non-contributing predictors in this way, the simplest model was refitted.

The resulting model is shown in Table 3.3. As expected from previous lexical decision studies, several stimulus-based predictors had a significant effect on RTs. In particular, RTs to targets were slower when the RT in the preceding trial was slower, when the target was longer or of low lexical frequency, and for trials that occurred earlier in the experiment. There were also main effects for (a) prime type, such that RTs were faster for targets following related primes, and (b) for focus, such that RTs were faster for targets that were interpreted as narrow object foci. Conspicuously absent from the model (because it was found to contribute nothing to model fit) was an interaction between prime type and focus, indicating that the priming effect observed was statistically equal for broad VP and narrow object focus.

The overall mean RTs for each of the experimental conditions is shown in Figure 3.2. Although there was no significant interaction between prime type and focus size, a tendency for priming of targets to be more effective when the target was narrowly focused was apparent (an average priming difference of 11.3 ms for object focus compared with 7.8 ms for VP focus). Nonetheless, the overall finding that priming occurred regardless of the size of the focus

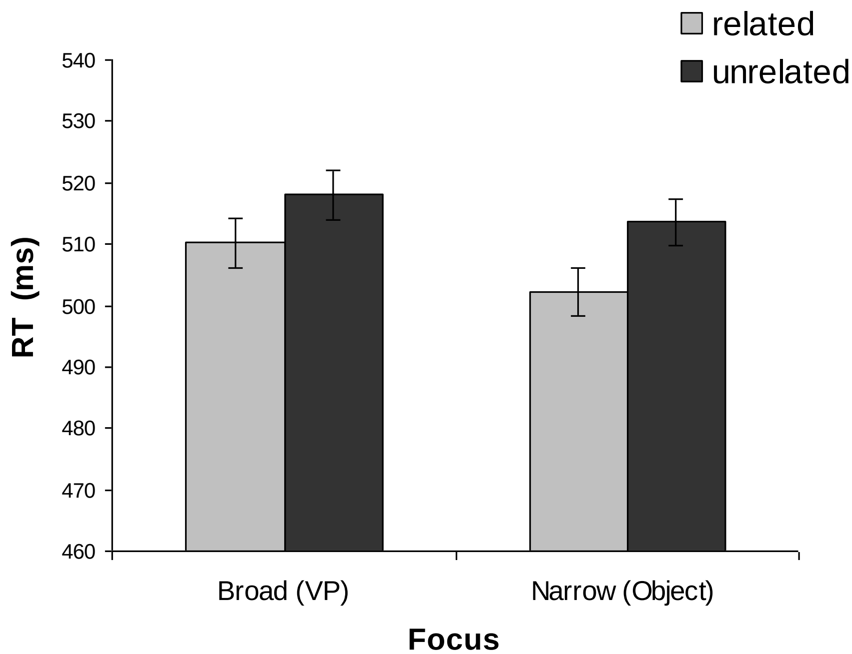


Figure 3.2. Mean reaction times for each of the information structural (focus) and prosodic conditions in Experiment 3a. Error bars show standard error.

Table 3.3. Results for fixed-effects factors for the model of reaction times in Experiment 3a. Default level for the binary categorical factors is shown in italics.

	β	*SE (β)*	*t-value*	*p-value*
(Intercept)	523.740	15.334	34.15	< **.0001**
Trial	-0.298	0.058	-5.09	< **.0001**
Previous RT	0.045	0.008	5.41	< **.0001**
LogFreq of Target	-20.058	3.848	-5.21	< **.0001**
Target Length	3.664	1.795	2.04	< **.041**
Focus (*Narrow*)	-6.844	3.000	-2.28	< **.022**
Prime type (*Related*)	-7.744	2.996	-2.58	< **.009**

constituent indicates that sentences with a prenuclear accent on the verb were acceptable in both conditions.

3.4.3 Discussion

Two important findings resulted from the Experiment 3a. The first was the significant main effect for focus size, which indicated that listeners were able to make decisions faster to targets that followed narrowly focused primes, regardless of whether or not they were related to those primes. This is the pattern we might expect if broader focus constituents required more processing resources. It seems reasonable that, other things being equal, the set of possible alternatives to broader constituents such as a VP should be larger than the set of alternatives to a single object, simply by virtue of the combinatorial potential of verbs and arguments. Larger alternative sets should, in principle, impose a greater processing cost if they are actually activating lexical and conceptual representations. It may also be that processing larger chunks of new information simply requires more processing resources than do smaller chunks, regardless of any additional focus semantic value that is calculated. In either case, the main effect found for focus size provides evidence for the psychological reality of the discourse-dependent semantic difference between broad and narrow focus.

The second result, however, was more pertinent to the primary goal of Experiment 3a, which was to probe listeners' processing mechanisms for expectations for particular prosody-information structure correspondences. It was tested whether sentences with VP focus and sentences with object focus were equally well processed when they bore a single pitch accent on the object; they were predicted not to be. Listeners should, based on production evidence (and the results of the study in Chapter 2), have expectations for VP focus to have prenuclear accentuation, and so processing should have been more difficult in the absence of a prenuclear pitch accent, resulting in less reliable priming in Experiment 3a. However, this turned out, statistically speaking, not to be the case, since there was no significant interaction between prosody and focus; priming occurred in both conditions. Thus, the results of the cross-modal priming experiment look statistically very much like the results of context matching studies such as those reported by Gussenhoven (1984) and Welby (2003), and which are predicted by both Gussenhoven's (1984/1999) and Selkirk's (1995) focus projection theories.

Still, since there was in fact a trend in the direction I predicted, i.e., numerically more robust priming occurred when focus was narrow on the object, and so there is some suggestion that a preference might exist. At this point it should be noted that negative evidence (finding no disruption to priming) is not as strong as positive evidence (finding priming to fail) in the present experiment. Since all prime words were highly prominent in the stimuli, it may have been easier for listeners to ignore the infelicitous context. To the extent that they were able to do this, listeners could have treated them as individual words, allowing the more reliable word-to-word priming to occur. Thus, we might expect a different result when the prime is less prominent, or at least relatively less prominent. This is what was tested in Experiment 3b, where focus was held constant as narrow object focus, and the relative prominence of the object prime varied.

3.5 Experiment 3b

3.5.1 Methods

3.5.1.1 Material

The materials for Experiment 3b were the same materials used in Experiment 3a, with the addition of the second prosodic condition, and only the narrow focus question contexts were used.

3.5.1.2 Participants

Eighty-eight native English speakers took part in Experiment 3b, none of whom had participated in Experiment 3a or the web-based association experiment used for stimulus design. None reported any history of a hearing, speech or communication disorder. All received either monetary compensation or course credit.

3.5.1.3 Procedure

For Experiment 3b, 4 lists were created by rotating the same 32 sentences/visual targets through the two prime conditions and the two prosodic conditions (with or without a prenuclear accent on the verb); all sentences were presented in the narrow object focus contexts. Fillers also occurred equally in the two prosodic conditions. The procedure was otherwise as for Experiment 3a.

3.5.2 Results

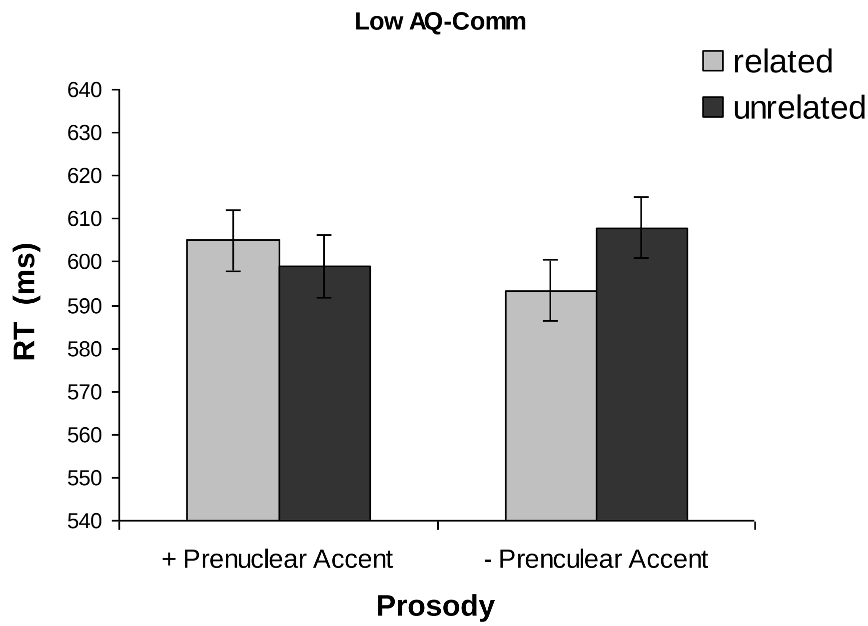
Data from one participant were excluded due to very slow overall RT (fewer than 20% of responses were below 1800 ms). For the remaining subjects, error rate was determined as in Experiment 3a, and was similar (2.5%, resulting in 2601 usable observations for the model). The average RT was 607 ms (SD=161). The modeling procedure and predictors were as in Experiment 3a, with the exception that the linguistic predictor of primary interest was not the information structural status of the prime, but the sentence's prosodic structure (+/– prenuclear accent on the verb). This variable was included in the same two and three-way interactions as *focus size* was in Experiment 3a.

The resulting model is shown in Table 3.4. The stimulus-based predictors (i.e., previous reaction time, log lexical frequency of the target, target length, and trial) all had the same effect on reaction times as in Experiment 3a, significantly, with the exception of target length. Additionally, RSPAN scores were a robustly significant predictor, such that higher scores (reflecting greater working memory capacity) were associated with shorter RTs. There were also non-significant trends for both prime type and the prosodic manipulation, although both were in the opposite direction predicted; RTs were faster following primes that were either unrelated to the target, or were presented with low prosodic prominence.

Table 3.4. Results for fixed-effects factors for the model of reaction times of participants' reaction times in the lexical decision experiment in Experiment 3b. Default factor for the binary categorical factors is shown in italics.

	β	*SE (β)*	*t-value*	*p-value*
(Intercept)	741.809	42.690	17.38	< **.0001**
Trial	-0.324	0.010	-3.25	**.001**
Prev RT	0.014	0.006	2.22	**.027**
LogFreq of Target	-28.448	7.243	-3.93	< **.0001**
Target Length	4.281	3.385	1.27	.206
RSPAN	-2.489	0.723	-3.44	< **.001**
Prime Type(*Related*)	17.093	9.327	1.83	*.067*
Prominence(*High*)	10.796	9.307	1.16	.246
AQ-Comm	2.697	7.326	0.37	.713
Prime Type(*Related*)*Prominence(*High*)	-34.474	13.145	-2.62	**.009**
Prime Type(*Related*)*AQ-Comm	-6.876	3.444	-1.99	**.046**
Prominence(*High*) *AQ-Comm	-5.701	3.419	-1.67	*.096*
Prime Type(*Related*)*Prominence(*High*)*AQ-Comm	10.477	4.879	2.15	**.039**

The effects of these factors, however, are better understood in terms of their interactions with each other, and with participants' AQ-Comm scores. First and most important for our purposes, there was a significant two-way interaction between prime type and prosodic prominence, such that the facilitation of targets by primes (i.e., reduction in RTs relative to targets following controls) occurred primarily when the sentence lacked a prenuclear accent on the verb. However, this pattern is further qualified by the significant three-way interaction between prime type, prosody, and AQ-Comm. The consequence of this interaction can be seen in Figure 3.3, which divides participants into two groups based on their AQ-Comm score. For participants that fall into the lower end of the spectrum of AQ-Comm scores (indicating good communication skills), we see the expected pattern: there is significant facilitation of targets following related primes relative to unrelated primes when the sentence lacked any prenuclear accent, and no effect (even a trend towards inhibition, or increase in RTs relative to unrelated control primes) when the sentence contained a prenuclear accented verb. However, for those subjects on the higher end of the AQ-Comm spectrum (indicating poorer, more autistic-like comm-



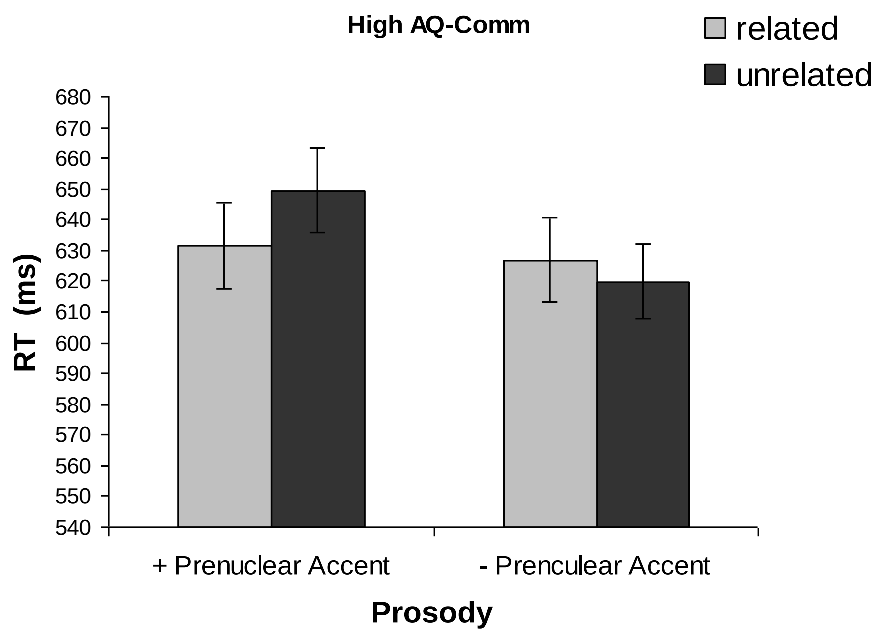


Figure 3.3. Mean reaction times for each of the experimental conditions in Experiment 3b for participants with low AQ-Comm scores (top) and those with high AQ-Comm scores (bottom). The "high AQ-Comm" group represents subjects with scores higher than one standard deviation above the group's mean, the low group all others. Error bars show standard error.

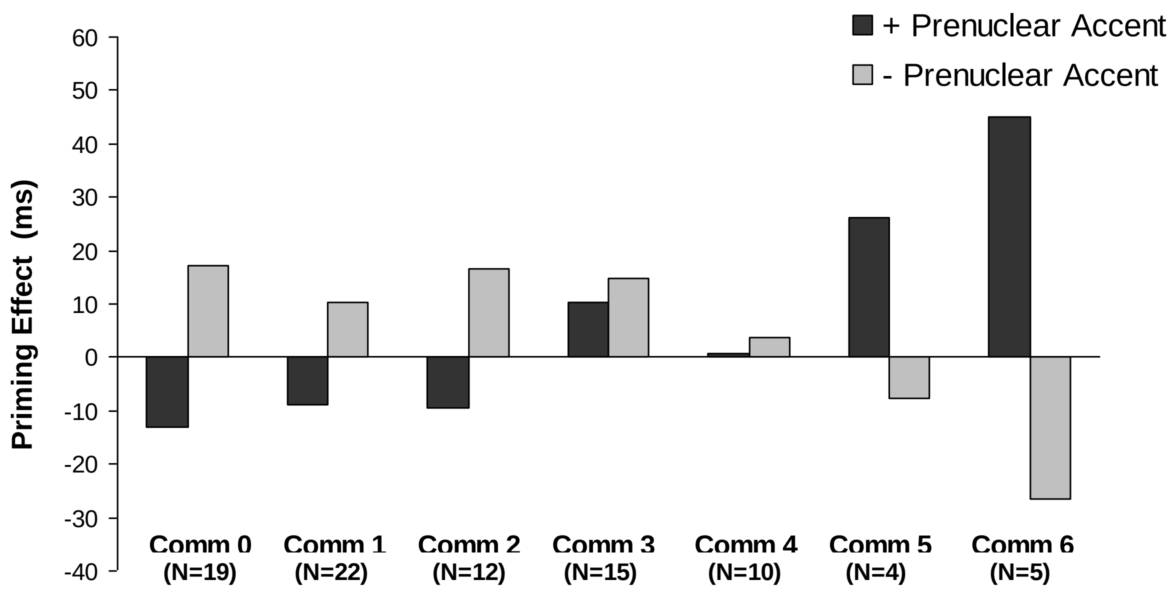


Figure. 3.4. Priming effects for the two prosodic conditions in Experiment 3b, binned by participants' AQ-Comm score (ranging from 0 to 6). Priming effects are mean reaction time differences (in milliseconds) between responses to targets after unrelated control primes versus after related primes (positive numbers reflect facilitation relative to the control condition; negative numbers reflect inhibition). Higher AQ-Comm scores indicate more "autistic"-like communication skills. (Number of subjects in each bin is shown in parentheses).

unication skills), there are no clear differences between conditions, and indeed there is a trend in just the opposite direction: slight facilitation when sentences contained a prenuclear accent and slight inhibition when they did not.

This inverse pattern is even more apparent in Figure 3.4, which shows priming effects binned by listeners' actual AQ-Comm scores (which ranged from 0 to 6 for these subjects, out of 10 possible points). Priming effects in the figure are determined by taking the average difference in RTs between targets following related and unrelated primes for each prosodic condition. As can be seen, both facilitation of targets and inhibition of targets covaries with AQ-Comm; participants at the low end of the spectrum (e.g., AQ-Comm=0-2) are very sensitive in the direction expected (based on production studies), showing robust facilitation of related targets in sentences with no prenuclear accent, and inhibition when there is a prenuclear accent. In the mid-range of AQ-Comm scores (AQ-Comm 3-4), however, facilitation is less reliable, and there is no inhibition. Finally, at the high end of the AQ-Comm spectrum (AQ-Comm 5-6), we find a pattern that is the inverse of those on the low end of the spectrum. While it must be noted that the average differences are least reliable at this higher end of the range of scores (because the distribution is such that fewer participants scored in this region), the relationship is clear.

Finally, there was also a significant two-way interaction between prime type and AQ-Comm, indicating that the relatedness of the prime was a better predictor of priming overall for participants with higher AQ-Comm scores, regardless of prosody. There was also a marginally significant trend in the direction of high relative prominence on a prime being associated with faster reaction times overall, regardless of the relatedness between prime and target, or AQ-Comm score.

3.5.3 Discussion

3.5.3.1 Basic findings

In Experiment 3b, listeners heard sentences with narrow focus on an object with two different prosodic structures. It was shown that priming patterns for listeners as a group were sensitive to which of the two prosodic structures was heard, very much in line with my predictions: facilitation was observed only if the sentence lacked a prenuclear accent. In fact, there was even an overall tendency towards inhibition of lexical decisions when narrow foci were presented with a prenuclear accent on the verb. These results demonstrate a dispreference that is consistent with reported production patterns, as well as the results of Chapter 2; listeners seem to expect prenuclear prominence to be higher under broad focus VP focus, but low when focus is narrow on the object.

3.5.3.2 Individual differences

In addition to the primary, overall finding, the results also make it clear that it is somewhat inaccurate to talk about listeners "overall", since there were individual differences that were systematically related to verbal working memory and to autistic traits.

3.5.3.2.1 RSPAN

First, it was found that listeners with lower RSPAN scores (indicating lower working memory capacity) were on average slower to respond. This may have been due to the fact that in Experiment 3b, unlike in Experiment 3a, the prosody of the test sentence varied from trial to trial. Having to integrate an unpredictable (and sometimes dispreferred) prosodic structure with the previous discourse context seemed to be more difficult, as evidenced by the approximately 110ms slower average RT in Experiment 3b compared with 3a. These more difficult conditions would be expected to slow down participants with lower working memory resources more than those with higher working memory resources.

3.5.3.2.2 Autistic Traits

Unlike working memory capacity, which exerted general effects on RTs, listeners' autistic traits interacted crucially with the prosodic manipulation. However, these effects were complex, in that they did not simply distinguish listeners who were sensitive to prosody from those who were not. Instead, as is apparent in Figure 4.4, there are two separate processes at play related to AQ-Communication scores: facilitation and inhibition; I will attempt to interpret them separately.

Considering first the patterns of facilitation, recall from the discussion of Norris et al. (2006) in Section 3.2 that isolated words serve as more reliable primes than those embedded in sentences. This was said to be due to the fact that priming is largely dependent on whether sentence-level semantic meaning supports the relationship between prime and target; this is necessarily the case for single-word utterances, but may or may not be so in larger sentence contexts. Since listeners with higher AQ-Communication scores in Experiment 3b showed more robust overall priming (as indicated by the significant two-way interaction between prime type and AQ-Communication), I propose that these listeners were attending primarily to the simple prime-target lexical relationships, largely ignoring sentence/discourse context. This sort of "shallow" processing, which is the kind typically used for non-focused information (Sanford & Garrod 1998, Sanford & Sturt 2002), allowed listeners with poorer, more autistic-like communication skills to exhibit the robust priming typical of primes presented in isolation. Thus facilitation effects for these individuals was largely independent of factors such as discourse context and prosodic context.

It is somewhat less clear how best to explain the patterns of inhibition observed in Experiment 3b, as they relate to AQ-Communication scores. One possibility would be to again appeal to individual differences in the propensity to use sentence-level interpretation. For example, we might assume that, when the object's prominence is high (i.e., the [–prenuclear accent] condition), a contrastive interpretation, i.e., a set of alternatives, is evoked, causing two things to then happen. First, lexical activation of words in that alternative set occurs, and, second, inhibition of words not in that alternative set occurs.[13] The AQ-related individual differences could then arise in the construction of the alternative set, with higher-AQ listeners generating the "wrong" alternatives and thus suppressing the "right" ones. According to such an account, then, listeners with higher AQ-Communication scores would differ from their low-AQ counterparts in

[13] This is consistent with Husband and Ferreira (2012), who recently proposed that the generation of alternative sets is a two-step process that proceeds sequentially in time, with the inhibition mechanism being the later process.

terms of interpretative mechanisms, but not in a basic use of prosody. One attractive and fairly intuitive property of this interpretation-based explanation is that it provides a unified account of when we find facilitation and when we find inhibition, since each process is inversely—and presumably, equally—predictable based on the contents of a given alternative set.

One problem is immediately apparent for this explanation, however. Since it assumes that facilitation and inhibition are complementary processes triggered by the same semantic interpretation—and AQ only predicts the details of semantic interpretation—we do not expect to find listeners who lack one or the other process. However, we did find such listeners in Experiment 3b, namely those with mid-range AQ-Communication scores, for whom there was no inhibition, only facilitation for both prosodic conditions. Importantly, this cannot be explained away by assuming that the process of inhibition itself is systematically weakened, in a general sort of way, by autistic traits. This is because robust inhibition occurs at both ends of the AQ-Communication spectrum, albeit in opposite prosodic conditions. It therefore seems that the AQ-based effect cannot be understood as the result of an equal sensitivity to, but different interpretation of, patterns of prosodic prominence.

In light of the discussion in Section 2.5, and especially the results of Experiment 2, we might instead consider a more processing-based explanation. Recall that Kristensen et al. (2012) found that processing prosodic prominence was taxing on selective attentional resources, and also Cole and colleagues' (2010) related claim that prominence perception occurs when such resources are overly taxed. Assuming that individuals with high AQ-Communication scores might also have low attentional resources, this would suggest that processing prominence would be particularly taxing for them. The question would then be how to connect this attention-intensive processing of prominence specifically to the process of inhibition.

In fact, it has been argued that inhibition in cross-modal associative priming can sometimes occur as the result of attentional suppression. For example, inhibition seems to occur due to the need to suppress word-level representations in order to attend to the construction of a larger, sentence-level semantic interpretation (Marí-Beffa, Houghton, Estévez, and Fuentes 2000). Similarly, it has been found that inhibition to targets occurs when participants are made to actively ignore a target's related prime (Tipper 1985, Marí-Beffa et al. 2000). Thus, there is independent evidence that inhibition and attention are closely linked in cross-modal priming tasks. If listeners with more autistic traits are more easily overwhelmed by the difficult task of processing highly prominent accentuation, this may explain their inhibitory patterns in Experiment 3b. Left to be explained are details such as whether this inhibition occurs as a direct result of overwhelmed attention (cf Cole et al. 2010, and my Experiment 2 discussion), or whether under such circumstances, listeners actively shift their attention away from (i.e., ignore) such prosodic events. While this matter clearly requires further study, it seems likely that a processing-based story along these lines is more promising than an explanation that assumes that the difference between listeners with high and low autistic traits is merely a matter of semantic/pragmatic interpretation.

3.6 **Local General Discussion**

The experiments in this chapter were intended to test the same basic hypothesis that the experiments in Chapter 2 tested, namely that listeners do not regard broad focus on a VP and narrow focus on an object as prosodically ambiguous. Unlike in Experiment 2, which made use of an off-line measure of listeners' preferences, the cross-modal priming experiments here allowed us to get at whether listeners preferences were also evident on-line, in lexical processing.

The primary finding was, again, that listeners show a preference, such that broad focus should be pronounced with high prenuclear prominence, and narrow object focus with lower prenuclear prominence, although the processing evidence for the former was only in the form of a numerical trend. These patterns, however, look very much like listeners' productions of broad and narrow focus meaning, and counter a theory that assumes prenuclear accentuation does not contribute to this contrast in particular (Ladd 1996) or to information structural meaning in general (e.g., Büring 2007). It should be noted also, however, that while the results of Experiment 3b are inconsistent with Gussenhoven's model of focus projection, which assumes prenuclear accents are always assigned randomly (with respect to semantic meaning) to these syntactic structures, they are consistent with Selkirk's (1995) model. This is because Selkirk's theory of focus projection contains the stipulation that prenuclear accents mark words as "new". Since verbs in sentences with narrow object focus must be given, such sentences are predicted to be infelicitous with a prenuclear accent on the verb.

It is also worth considering how the results of the present chapter compare with similar experimental attempts to reveal listeners' knowledge of how prosody relates to broad and narrow focus. In particular, the present study might be regarded as an implicit, on-line version of the appropriateness rating experiments of Birch and Clifton (1995) and Welby (2003). In the case of Birch and Clifton's study, which tested prosodic expectations for broad VP focus only, recall that there were two experiments. The first used a task meant to draw attention to prosodic acceptability, the second more to semantic acceptability. They found mixed results, in that the former experiment revealed expectations in line with our predictions, and the second did not. Thus, their conclusion was that a single accent on an object did not fully "project" focus to the entire VP. Results of Experiment 3a were similarly inconclusive on this point; statistically, priming was shown not to be disrupted when VP focus sentences lacked prenuclear accent verbs, but there was also a trend for priming to be less robust. In my discussion of Experiment 1a in Section 3.4.2, I suggested the possibility that the weak but observable priming despite what should be the dispreferred prosody might have to do with the experimental design. However, it is also possible that there may be something inherently more ambiguous about broad focus prosody than object focus prosody, an observation made in Section 1.4.2.

In Welby's study, however, the presence versus absence of prenuclear accents was tested for both VP focus and narrow object focus, and no evidence of a preference was found in either case. We might assume, that, overall, the off-line appropriateness rating task will be more prone to listeners' perceptual corrections, and attempts to construct a basic semantic interpretation for the sentence. While this would account for the difference between my narrow focus results and hers, it does not account for why her results for VP focus differ from Birch and Clifton's. Possibly, despite being given similar instructions in both author's studies, Welby's listeners nonetheless were less effectively attending to the signal, and so tended to perform more of a semantic acceptability task. Again, the inconsistency within and across studies using off-line, explicit rating of appropriateness further highlights the problematic nature of this task.

CHAPTER 4

Summary and General Discussion

4.1 Summary of the primary experimental findings

A longstanding and widely-held assumption in research on the prosody-information structure interface in English is that focus is conveyed by the sentence's nuclear accent. Prenuclear accents, on the other hand, are assumed not to contribute to information structural interpretation, and are thus considered optional, unpredictable, or "ornamental" in that they serve only some other non-information structural purposes. In this dissertation, I have explored to what extent this assumption about the linguistic knowledge of English speakers is justified by probing them for expectations regarding the correlation between prenuclear prominence and focus meaning.

The test case for this exploration was the contrast between broad focus on a VP and narrow focus on an object in English SVO sentences. This information structural contrast has been of particular interest to linguists, since, on the surface, it seems to especially highlight the disparity between the ability of nuclear and prenuclear accents to encode focus. That is, since the nuclear accent falls on the object in both cases, and because prenuclear accents are said not to relate to the meaning, the contrast should be neutralized. This is in fact a basic prediction built into focus projections theories such as those of Selkirk (1995) and Gussenhoven (1984/1999).

In Chapters 2 and 3, experimental methodology was used to determine (a) whether listeners have knowledge about how prenuclear prominence relates to the size of the focus constituent, and (b) how best to characterize its contribution, if any. In Chapter 2, Experiments 1a, 1b, and 2 approached these questions using a prominence rating task to investigate whether listeners had expectations for prenuclear prominence in relation to focus structure that could be seen in their top-down use of those expectations in prominence perception. In fact, it was determined in each of those experiments that the interpretation of focus structure was a crucial predictor of perceived prominence of prenuclear material, completely independent of the signal. However, it was found that listeners' expectations were equally for nuclear prominence, such that nuclear accents were perceived as less prominent in broad VP focus sentences, and prenuclear verbs were perceived as more prominent; the opposite pattern held for narrow object focus. This finding is important because it closely mirrors the pattern we see in production studies that show speakers to manipulate the relative prominence of nuclear and prenuclear material in relation to focus. The fact that both speakers and listeners show these patterns supports the notion that prosody disambiguates the size of a focus constituent in these structures in the conventionalized sort of way that should be included in the grammar.

Further evidence was found in Chapter 3, where listeners' expectations were tested using a more on-line, less explicit comprehension-based measure, namely the cross-modal associative priming task. Listeners were presented with sentences with either broad or narrow focus interpretations, and only the presence or absence of a prenuclear accent was manipulated. It was shown, overall, that processing was more difficult (i.e., priming effects were disrupted) when narrow object focus sentences contained a prenuclear accent on the verb. A non-significant trend was also found, suggesting that broad VP focus sentences without a prenuclear accent tended to be somewhat less easily processed than narrow object focus sentences without a prenuclear accent. These results are quite unexpected if listeners do not have any expectations about how focus size and prenuclear prominence relate.

Thus, the experiments in this dissertation reveal very clearly that listeners harbor expectations about how focus size is expressed prosodically, and those expectations are ones that we can relate in a detailed way to the behavior of speakers. In both cases, relative prominence disambiguates focus size.

This concludes the basic summary of results, as they pertain to the theoretical matter of whether these information structural meanings are truly ambiguous. In the following sections, I will consider the implications of my findings for the theory of prosodic structure, and, in particular, how prosodic structure encodes discourse-sensitive sentence meaning, such as focus.

4.2 Implications for the prosodic encoding of information structure

4.2.1 Characterizing the realization of focus size: metrical relations

The fact that speakers and listeners agree so closely on the details about how prosody expresses the focus size contrast is evidence that it is highly conventionalized; at the same time, however, it seems that any attempt to characterize this knowledge as straightforwardly part of the phonology is bound to fail, since listeners' behavior is mostly phonetically gradient.

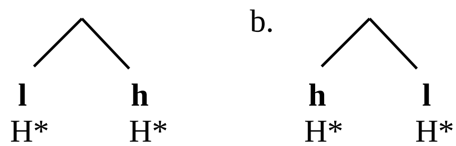
To illustrate, consider first the determined analyst who might seize upon some of the categorical aspects that do happen to be present. This could be, for instance, the now well-established correlation between the presence of broad focus on a VP and the presence of a prenuclear pitch accent—uncontroversially a phonological entity in AM models—on the verb. A second kind of phonological analysis might instead attempt to characterize the contrast as making reference to the tonal string, since accent types are also distributed differently across broad and narrow focus sentences; a !H* on a sentence-final object is more likely for broader foci, and a H* or L+H* is more likely for narrow object focus.

The two problems that immediately arise for both phonological characterizations, however, are that (a) speakers fail to produce the correct categorical structures at a rather non-trivial rate, and (b) speakers sometimes produce the wrong categorical structure, but phonetically adjust in the direction of the "right" structure. While a probabilistic grammar might be able to explain away some of the kind of variation represented by (a), it would treat as random the highly predictable variation represented by (b).

I propose that both the categorical and gradient patterns that characterize the focus size contrast in English can be understood as restricted by a linguistic grammar if we make certain assumptions about the nature of prosodic structure. In particular, if we treat the intonational contour as reflecting a layer of metrical structure, along the lines proposed by Ladd (1990), discussed in Section 1.1, then relative prominence marking of the type speakers produce and listeners expect is structurally available for encoding the focus size contrasts.

Recall Ladd's basic proposal that some pitch range variation, namely that which is characterized as downstep, is controlled linguistically, although orthogonally to the tonal string. That is, rather than being a property of a paradigmatic pitch accent type contrast such as "!H*" downstepping is the result of a l-h syntagmatic relation between two accents, illustrated as in (1), where downstepping of a final accent occurs in (1b) but not (1a).

(1) a. b.



Importantly for our purposes, it is possible to re-characterize pitch accent types as conforming to one or the other of these tonal relational structures, greatly reducing the amount of possible variation to two basic categories: right tonal prominence (2a) and left tonal prominence (2b).

(2)

	a.		b.	
	l	**h**	**h**	**l**
	Ø	H*	H*	Ø
	L*	H*	H*	L*
	H*	H*	H*	!H*

To account for the details of focus size, some additional stipulations are necessary regarding how these accent patterns are interpreted. The first is that the H*~L+H* contrast does not have a structural role in tonal metrical structure, although I would suggest adding a leading low target to the H* is likely useful in perceptually reinforcing the intended pitch range of a H*. The second, however, is that the most phonetically ambiguous of all the tonal sequences in (2) is likely the case of H* accents (as in 2a). I assume that, in such cases, there is a bias towards interpreting the second of the two as having the larger pitch range. Thus, any phonetic value for the second tone not low enough to be perceived as downstepped is regarded as more tonally prominent than the preceding accent. This is reminiscent of the well known correction for declination that occurs in perception, but may also be a linguistic bias, which Calhoun (2006) calls the right-branching bias. Assuming such a bias also makes clear there is no need for an upstepped category.

Given these basic assumptions, this partitioning of possible intonational contours captures the broad and narrow focus marking behavior of speakers very clearly; whereas broad focus requires the structure in (2b), narrow focus requires the structure in (2a). This conception of the prosodic structure predicts that speakers' realizations of the contrast will sometimes show categorical behavior of a certain asymmetry; however, it also predicts the details of phonetic variation. Namely, if a speaker chooses the accent pattern that is most phonetically ambiguous with respect to tonal metrical structure (i.e., the H* H* sequence), they have the option of phonetically disambiguating it, and, crucially, doing so by modifying the two accents *in tandem*. If we do not build syntagmatic relations of this sort into the basic prosodic structure, we are forced to assume that all of these facts are unrelated to one another. This seems highly improbable, particularly given the results of Chapter 2, which demonstrated how vivid a perceptual object the relative prominence relation is.

Thus, to summarize, I propose that the best understanding of the prosodic realization of the focus size contrast reflects the linguistic system; this requires the tonal string, in addition to its paradigmatic nature, to be viewed as having syntagmatic structure. When such metrical structure, as has been previously argued for by Ladd (1990), is applied, it is more readily apparent that the behavior of speakers and listeners does not vary freely, and in this way should be distinguished from paralanguage. Because broad and narrow focus meanings correspond to a complementary exploitation of a small number of categories, and also a complementary implementation of gradient modifications to those categories, both these phonological and phonetic patterns are part of grammar.

As Ladd (2008:6) emphasizes, and I have stated above, syntagmatic l-h relations at the level of tonal structure exist in addition to, not in place of, syntagmatic s-w relations at lower-level levels of metrical structure. The most important property of this appeal is that it limits the use of tonal metrical structure in making linguistics contrasts, since contrasts such as +/- pitch accent or nuclear/prenuclear accent status will be determined at a lower level of metrical structure. However, even given this basic limitation, my proposed use of Ladd's metrical theory would seem to predict, at first glance, more variation for the marking of focus size than is actually observed. I will therefore briefly address two additional, mostly theory-internal issues.

4.2.2 Additional implications

4.2.2.1 Asymmetrical de-accentuation

One aspect of the structures in (2) is that they suggest the possibility of a symmetrical distribution of a null accent, when in fact the null accent is not symmetrically observed in nuclear versus prenuclear position. That is, in considering SVO sentences, the Ø H* pattern corresponds to an unaccented prenuclear verb, while the H* Ø pattern corresponds to a nuclear accent on the verb and a deaccented object; while the former should be dispreferred for broad focus sentences, the latter pattern is categorically impossible and unambiguously marks narrow focus on the verb.

The reason for this disparity highlights the point just made above, which is that tonal metrical structure does not replace strong-weak relations at lower levels of metrical organization. In this case, both broad and narrow focus have the same requirement on that lower level structure, such that the object is parsed into a strong position, and so requires an accent. Hence the structural ambiguity that represents the reason for studying this contrast in the first place. The implication, however, is that the h-l tonal metrical structure can only place requirements on a word's local (but relative) pitch range; it cannot specify accentuation, as this is determined at a different level of metrical structure. Thus, tonal metrical relations will be useful as the primary basis for encoding a linguistic contrast only when the lower-level metrical structure is ambiguous on the matter. This will often be the case across phrases, which is discussed further below.

4.2.2.2 *The role of L**

Another apparent prediction derived from reference to tonal metrical structure in marking focus structure involves the occurrence of pitch accents with low targets (i.e. L*). The first prediction is that L* should be among the relatively preferred realizations for the left node of the l-h structure, which is preferred for narrow object focus. While I have not tested this issue in the present undertaking (considering only prenuclear accents of the H* type), [14] it has been widely assumed since Pierrehumbert and Hirschberg's (1990) compositional analysis of intonational contours that the L* marks information as relatively uninformative or given. In their words, one of the functions of L* is to "convey that the accented item already figures in to what H[earer] currently believes to be mutually believed" (1990:294). This interpretation is supported by Röhr and Baumann (2011) for German, who found that listeners perceived as more "known" (rather than "new") words that were pronounced with a L*. Thus, it seems likely that, although not handled here (and possibly the L* H* L-L% is a less pattern overall), the L* does not seem to pose any problem with the proposal that focus size is constrains tonal metrical relations.

I also have not addressed L* as a possible realization for the l-node in the h-l structure; however, this seems even less likely to pose a problem for the present proposal than the case just discussed. For example, it is often observed in prosodic transcription that a !H* and a L* (and the absence of an accent) are very difficult to distinguish in the reduced pitch range typical near the end of a phrase marked by the L- phrase accent (recall the transcribers in Experiment 2, for example, who disagreed on exactly this point; see footnote 1). However, the L* in nuclear position presents other complicated issues (e.g., its interaction with rising boundary tones for

[14] Although it can only be regarded as anecdotal, a pilot version of Experiment 3b did contain some items that with prenuclear L*. Among the findings from a small sample of listeners was that prenuclear L* on the verb in narrow object focus contexts did not result in the same disruption to priming that prenuclear H* was ultimately found to, suggesting it may have been more appropriate.

Yes/No questions), and is overall very rare in the declarative sentences considered, and so is not discussed further here.

4.2.3 Applications to other phenomena

So far I have proposed that an appeal to a syntagmatic level of prominence relations based on pitch range provides an insightful analysis of the phenomenon I have investigated experimentally in this dissertation. However, I believe that an appeal to tonal metrical structure of this sort can help us to understand the prosodic encoding of other linguistic contrasts that have become of much interest since Ladd's original proposal. Here I discuss three basic cases, involving accessibility, (contrastive) topics, and special types of double focus construction.

4.2.3.1 Accessibility

It is often acknowledged that information that is 'old', or given in the discourse, may sometimes be marked by a pitch accent, yet nonetheless distinguishable from new or focused information, which is also pitch accented (e.g., Halliday 1967). For example, in Pierrehumbert and Hirschberg's (1990) analysis of intonational meaning, they discuss the speaker's tendency to use the !H* accent in order to mark information as inferable (usually pragmatically inferable) to the listener. Pierrehumbert and Hirschberg note that, presumably related to the similarity between given and inferable information, the !H* accent seems able to alternate with no accent, allowing sentences like both (3B) and (3C) to follow sentence (3A) felicitously:

```
                H*                 H*      H*       L-H%
(3)     A:      I know you have great credentials…

                     H*                            !H*  !H*    !H*        L-L%
        B:      And I'm looking for someone with just such credentials.

                     H*                                                   L-L%
        C:      And I'm looking for someone with just such credentials.
```

(Pierrehumbert and Hirschberg 1990: 298)[15]

Patterning of this sort for inferable or "accessible" information has also been reported for German by Baumann and Grice (2006), who investigated how several potentially distinct types of accessibility relate to pitch accent type preferences. In their study, listeners heard sentences such as (4), where *Kellner* is assumed to be accessible based on the topic introduced by *Tischnachbarn*:

```
              H*                          H* / H+!H* / Ø
(4)     Unsere Tischnachbarn reifen den Kellner
          our   table-neighbors called  the  waiter
```

"*The people at the next table called the waiter*."   Baumann and Grice (2006)

Baumann and Grice presented versions of sentences such as (4) to listeners, who were to asked to rate the contextual appropriateness of their intonation, which varied across three conditions: one in which the accessible word had a nuclear H* accent, one in which it bore a H+!H* accent, and one in which it was deaccented.

The authors in fact found that listeners have different expectations about the various sorts of accessibility that were tested: some types corresponded to a preference for no accent; some types of accessibility were equally appropriate with either no accent or a downstepped accent,

[15] This is a modification of the authors' example, which is attributed to Gregory Ward. The intonational annotations also reflect the current ToBI symbols rather than the original Pierrehumbert (1980) equivalents.

consistent with Pierrehumbert and Hirschberg's observation, discussed above. Finally, other types of accessibility showed a significant preference for the downstepped accent over deaccenting. Thus, although there may be some internal complexity to accessibility, accessible information is consistently dispreferred with a full-fledged H* accent. Important to the present purposes is the fact that this sort of variation is accommodated if accessible information is marked by the h-l tonal metrical configuration, assigning a subordinate pitch range to the accessible item.

Bauman and Grice's study also draws attention to the prediction made by the theory of syntagmatic tonal contrasts, which is that different downstepped pitch accent types should have a unified function. Indeed, in a subsequent production study, Grice, Baumann, and Jagdfeld (2009) found that downstep within accents (a H+!H*) alternated with downstep across accents (a H* !H* sequence) for information that was referentially accessible. Thus, in both cases, it seems that tonal metrical structure, which assumes downstep should be largely orthogonal to the tonal string, is able to assign a common structure to a set of related meanings. Interestingly, the authors also reanalyzed the phonetic data from Baumann et al. (2007), reviewed in Section 1.4.1, in terms of pitch accent type, and found that this same within/across accent downstep alternation occurred for German broad focus productions. While it does not seem that broad focus can be construed as semantically related to accessibility, it seems to be expressed by the same structure.[16] Thus in German there exist alternations that must be regarded as happening

[16] Note that this is a general property of intonational meaning. That is, there are numerous examples in English where the same prosodic structure has different and unrelated interpretations in different contexts. Examples include simple declaratives and WH-Questions (both being realized by H* L-L%), the "uncertainty" and "incredulity" contour (both said to be marked by L+H*L-H%; but see Barnes, Veilleux, Brugos, and Shattuck-Hufnagel (2012) for a reanalysis of the latter case). Indeed, a key property of prosodic structure that distinguishes it from syntactic structure is its context-dependent interpretation (see also Hirschberg, Gravano, Nenkova, Sneed, and Ward (2007) on this point).

completely by accident if downstep does not reflect some structural property orthogonal to paradigmatic pitch accent type contrasts.

Although patterns reported for German do not necessarily imply the same patterns for English, English does seem to behave similarly. For example, it was noted that for Experiments 1a and 1b in this dissertation that the speaker of the answer sentences used for stimuli produced VP focus sentences in which the object sometimes had !H* and sometimes had L+!H*, which were combined for the purposes of calculating labeler agreement (see Section 2.2.1.1). More systematically, Hirschberg, Gravano, Nenkova, Sneed, and Ward (2007) explored the occurrence of English downstepped accents in read and spontaneous speech in the Boston Directions Corpus (Nakatani, Grosz, and Hirschberg 1995; Hirschberg and Nakatani 1996). Testing for the influence of a number of factors, Hirschberg and colleagues found that several discourse-related features (e.g., whether the word was "given" or "inferable") predicted not only the occurrence of !H*, but other downstepped accents as well. This suggests that, as in German, there is some functional unity to different pitch accent type categories that contain downstep.

It is interesting to note a particular case of anaphoric reference (a kind of "referential accessibility") discussed by Ladd (1980; see also 2008). Although at this point these are only impressionistic observations, they all seems to relate to a use of downstep,:

(5) Q: Everything OK after your operation?

A: Don't talk to me about it…

(i) I'd like to STRANGLE the butcher.

(ii) I'd like to strangle the BUTCHER. Ladd (1980)

In the context of the question in (5), the deaccented pronunciation of *butcher* (with nuclear accent on *strangle*) in (5Ai) is interpreted as referring to the implied doctor (the "epithet" reading in Ladd's words). This is contrasted with the non-referential, literal interpretation of *butcher* when *butcher* bears a (nuclear) accent.

For cases like (5), it seems that the relevant level of metrical structure is at a lower level than tonal metrical structure, since the contrast is a matter of accent versus no accent on *butcher*. That is, it seems clear that any accent on *butcher,* including a downstepped one, renders the epithet reading just as unavailable. However, Ladd also notes that when *butcher* appears earlier in the sentence, as in (6), things change:

(6) Q: Everything OK after your operation?

A: Don't talk to me about it…

H* H* L-L%
(i) The butcher charged me a *thousand* BUCKS.

H* H* H* L-L%
(ii) The *butcher* charged me a *thousand* BUCKS.

H* L- H* H* L-L%
(iii) The BUTCHER charged me a *thousand* BUCKS.

The relevant observation about these sentences is that in the first two cases, (6Ai) and (6Aii), in which *butcher* is either prenuclear unaccented or prenuclear accented, the epithet reading is available. When it bears a nuclear accent, however, as in (6Aiii), the literal meaning seems to be the preferred reading of *butcher*. This would seem to suggest that the referential meaning depends on the status of the accent on *butcher* being prenuclear rather than nuclear.

However, Ladd points to another possible pronunciation of the answer sentence in (6), an emphatic one in which each content word is set in its own intermediate phrase, as in (6Aiv). In this case, although *butcher* bears a nuclear accent, the epithet reading seems to be nonetheless preferred again:

```
                          H*       L-     H*      L-        H*       H*   L-L%
(6)    A:   (iv)  The BUTCHER CHARGED me a thousand BUCKS.
```

Ladd points out that the generalization seems to be that, when not at the end of the sentence, the epithet reading for *butcher* requires a prominence subordination of the utterance-final nuclear accent. This is seen in (6Ai) and (6Aii), and also in (6Aiv), since there is a sense in which the prominence of all the non-final accents is reduced in these cases.

Ladd argues we can understand these examples, not by appealing to his theory of tonal metrical structure, but rather to a separate hypothesis of compound, or recursive, intermediate phrases. According to this view, (6Ai) and (6Aii) are both single prosodic phrases, and so *butcher* is straightforwardly prenuclear in these cases. Then, it is simply assumed that *butcher*, when nuclear accented in (6Aiv), is actually also prenuclear, albeit in the larger complex intermediate phrase. This therefore renders (6Aiii) as the only case in which *butcher* is truly nuclear, and so it is also the only case where the epithet reading is unavailable.

Although not discussed by Ladd, it seems to me that tonal metrical structure can be appealed to directly in the understanding of these sorts of contrasts. To see this, we hold both phrasing and accent placement constant, and manipulate only the presence of downstep on the nuclear accent. That is, we contrast a l-h tonal structure with a h-l tonal structure:

(7) Q: What happened at the doctor's office today?

A: Don't talk to me about it…

H* H* L-L%
(i) The *butcher* charged me a thousand BUCKS.

H* !H* !H* L-L%
(ii) The *butcher* charged me a *thousand* BUCKS.

H* H+!H* !H* L-L%
(iii) The *butcher* charged me a *thousand* BUCKS.

These judgments are impressionistically difficult, but my intuition is that the epithet reading is preferred in (7Ai), but less available in either of the two downstepped cases (7Aii) or (7Aiii) (although there may be something else additionally interfering with the felicity of the H+!H* in (7Aiii)). These sorts of matters would likely be fruitfully investigated using on-line methods of interpretation, such as eye-tracking in the visual world paradigm. However, if these asymmetries in the epithet reading's apparent availability are valid, it indicates that Ladd's observations are better explained by his tonal metrical structure than his recursive phrase structure.

What we have seen from considering these cases, then, is that the literature has recognized a relation between pitch accent type (downstepped accents versus non-downstepped accents) and pitch accent status (nuclear versus prenuclear) in marking information as accessible (in some inferential or referential sense). I have argued that all of these cases can instead be recast insightfully as representing prominence structure at a particular level, namely tonal prominence. While notions of accessibility are unrelated to the broad versus narrow focus contrast that was my particular object of investigation, both accessibility and the focus size contrast can be argued to exploit this level of syntagmatic tonal structure in their prosodic expression. To some extent, accessibility may also be related to particular cases of contrastive topic marking, which I consider in the next section.

4.2.3.2 Contrastive topics

Considerations of the cases above, particularly (6Aiv), also raise one of the most important issues for a theory of syntagmatic tonal relations, and that is the domain of downstep. For German, it has been claimed that downstep patterns hold across phrases (van den Berg, Gussenhoven, and Rietveld 1992; Truckenbrodt 2002; Féry and Truckenbrodt 2005), although this can sometimes be difficult to assess, depending on how phrases are defined/indentified/derived in the prosodic model (e.g., Katz and Selkirk 2011, discussed below). According to the original AM analyses of English in Pierrehumbert (1980) and Beckman and Pierrehumbert (1986), the application of downstep is limited to the intermediate phrase. Thus any tonal dependencies across intermediate phrases is largely unpredicted from linguistic structure, and instead has to be explained by a lowering effect, or by non-phonological paralinguistic behavior. However, Ladd's observation about (6Aiv) suggests that there may be a clear example of a linguistic contrast that is expressed by exactly such across-phrase tonal dependencies.

Incidentally, there is another, much more widely familiar case of inter-phrasal tonal dependency that we have already discussed, and that is the marking of contrastive topics in English. This realization, sometimes characterized as a contrast of pitch accent and boundary tone type, is shown again in form of Pierrehumbert's (1980) example in (8):

(8) **Background-Answer:**

Q: What about Manny? Who did he come with?

L+H* L-H% H* L-L%

A: MANNY came with ANNA.

Answer Background:

Q: What about Manny? Who did he come with?

H* L-L% L+H* L-H%

A: ANNA came with MANNY.

Although Liberman and Pierrehumbert (1984) investigate the pitch range scaling patterns of the Answer-Background contrast right alongside that of downstep (and find remarkable within-speaker "constants" in the scaling of each), only the within-phrase pitch range difference (i.e., downstep) is given phonological status. Indeed, in Pierrehumbert's model, and the subsequent related AM model on which ToBI is based, there does not seem to be any way to assign a similar status to the Answer-Background/Background-Answer patterns.

Clearly, however, the pattern is captured by allowing the semantic contrast to be encoded in the phonology if the phonological object is the syntagmatic relation between (inter-phrase) pitch accents. In that case, the l-h structure is used to mark the topic-focus (i.e., Background-Answer) ordering, and h-l structure for the focus-topic (Answer-Background) configuration. This particular case is pointed out by Ladd as a potential use for the tonal metrical contrast, and, compared with some of the single-phrase cases of accessibility discussed above, the intuitions are very strong and categorical.

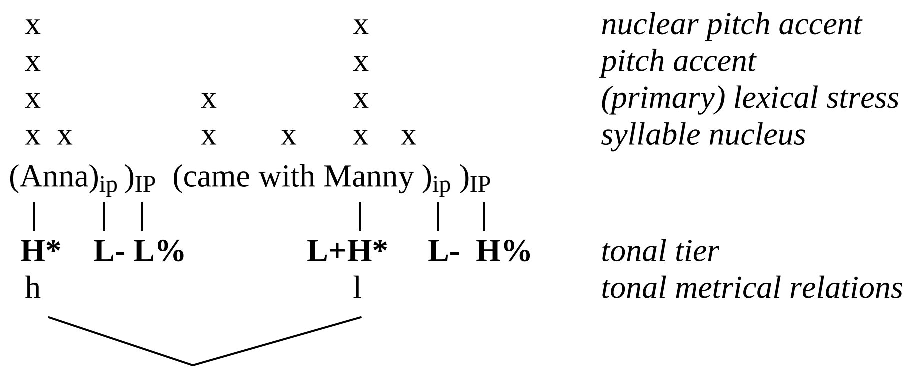


Figure 4.1. Example metrical structure (including tonal metrical structure for the Answer-Background (i.e., focus-topic) construction from Pierrehumbert (1980). The tonal metrical relations exist separately from the stress and accentual structure.

Here it is again important to point out that the tonal contrast does not replace the lower level metrical one, and in topic-marking constructions like the present one, it is very clear that there can be no changes to the lower level metrical structure, which determines the presence versus absence of accent. That is, both the topic (when it is contrastive) and the focus must be (nuclear) pitch accented, and so both are required to be metrically strong at a lower level of structure. We therefore assume a structure that is like that in Figure 4.1, which separates the accentual and tonal relations. This feature of the theory is important because it validates tonal metrical structure as a phonological organization of phonetic parameters, because it predicts that pitch range cannot be replaced by another phonetic cue to prominence, such as duration or intensity (although we might assume that these can occur as ancillary markers, enhancing the prominence at the tonal level). As I see it, this is the key difference between the kind of structure Ladd proposes, and the kind of analysis that Calhoun (2006) argues for. While Calhoun's theory has predictive power about which semantic phenomena will be less phonetically prominent (indeed, capturing all of the phenomena discussed so far, and also those below), it does not make specific predictions about which phonetic parameters will be relevant in which cases, because it is not explicit about the phonological structuring of these phonetic parameters. As Katz and Selkirk (2011) point out, this is also a property of the more semantically-sophisticated models of focus-sensitive phenomena. What is more, there is recent phonetic data for these sentences (elicited in very ecologically valid contexts), which show (a) that the pitch scaling asymmetry is the most consistent phonetic correlate of the topic-focus/focus-topic alternation, and (b) although other correlates of prominence seem to accompany pitch scaling, they are smaller and less reliable (Meyer, Fedorenko, and Gibson 2011; Calhoun 2012,).[17] Thus, the prosodic realization of topic *is not simply a matter of relative prominence*; rather, it is a matter of relative *pitch prominence*—which can be modeled, after Ladd, as a syntagmatic contrast relating the level of pitch accents.

While the link between tonal prominence and contrastive topic marking has received some attention in the literature (as has the tonal prominence-accessibility link), this is not the case for sentences with multiple foci. These are the final cases I consider here, and I argue that the theory advanced thus far can be easily extended to them as well.

4.2.3.3 Double focus constructions

Here I consider a number of kinds of cases that come up in Büring's (2008) study of accentless foci and some related work by subsequent authors. As will be apparent, these cases may indeed be semantically related to the phenomena of both accessibility and contrastive topic, discussed in the previous sections. In the very active literature on sentences with multiple foci, currently at least three configurational distinctions seem necessary, based on intonational behavior and semantic interpretation. These are double focus constructions where (a) both foci are new and

[17] The relative importance of the rising boundary (H%) is also at issue, although traditionally its role has been emphasized (e.g., Jackendoff 1972, Büring 2003). Possibly, however, it might also be regarded as an ancillary cue to marking the topic, or at least highly dependent on topic-focus ordering, as both Calhoun (2012) and Meyer, Fedorenko, and Gibson (2011) present experimental evidence that it is not a completely required part of the prosodic realization.

contrastive, (b) where one of the foci is a second occurrence focus and the other a new focus, and (c) cases where one of the foci is contrastive and the other non-contrastive.

<u>*New, contrastive foci*</u>

Consider first the generic sort of sentence with two separate foci, where each of the two foci in the rejoinder in (9B) are both new and explicitly contrastive with something in the prior discourse:

(9) A: I heard Mary drank water…

(L+)H* L- (L+)H* L-L%
B: …and JOHN drank SOYMILK.

In sentences, such as (9B), it is clear that both of the foci *John* and *soymilk* must be marked by pitch accents.[18] Büring's chief concern is deriving the relative prominence between such foci in a model of the prosody-information structure interface that relies upon potentially violable constraints. In particular, he derives a wide array of focus marking phenomena by assuming a highly-ranked constraint, Focus Prominence, which requires a focus to be maximally prosodically prominent in a relevant domain (cf Truckenbrodt 1995[19]). In the present case, however, it is unclear which of the two foci, which share a common domain—the utterance, the sentences, and the intonational phrase—should be assigned higher prominence. In Büring's proposal, an independent phonological constraint that requires that the rightmost accent in an intonational phrase be most prominent—a commonly appealed to alignment device (e.g., Hayes

[18] As is typically the case with focus in English, each of the two foci must be *nuclear* accented, although this does not fall out automatically from Büring's model in an obvious way. Possibly he is assuming this is handled independently by principles for prosodic phrasing, along the lines of Truckenbrodt 1995 (see also Selkirk 2004).

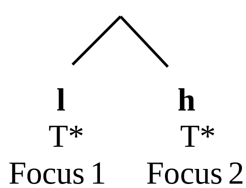
[19] Truckenbrodt (1995) is a contemporary, constraint-based implementation of the requirement that foci align with prominence, reminiscent of earlier work (e.g., Chomsky 1971 and Jackendoff 1972).

and Lahiri 1991, Selkirk 2004, Féry and Samek-Lodovici 2006)—essentially does the work of the traditional nuclear stress rule (Chomsky and Halle 1968, Halle and Vergnaud 1987, Zubizarreta 1998; see also Zubizarreta and Vergnaud 2005 and Zubizarreta to appear). That is, assuming that the foci have semantic properties that are equivalent in the right sort of ways (that is, are both "free" or both "bound" foci in Büring's terms), the rightmost one will have intonational phrase-level prominence.

Büring is thus able to capture his intuition, which I share, that *soymilk* in double focus constructions equivalent to (9) should be prosodically more prominent than *John*. However, it is unclear just what the phonetic correlates of this rightmost prominence should be, at least within AM models of intonational phonology, since the highest level of prominence is the nuclear accent, which heads the intermediate phrase.

It seems attractive to postulate that the primary difference is in the pitch range utilized for the two foci, which allows us to appeal to tonal metrical structure. That is, the relative prominence of the two foci is determined by a requirement for the l-h metrical structure at a tonal level:

(10)



Possibly, then, this example could help address a problem that has been unmentioned up until now, which is how to add tonal prominence to the existing prominence grid. That is, for what phrase level does tonal prominence serve as head? This example suggests the possibility of claiming it is the intonational phrase, which would also allow Büring's analysis go through with some added phonetic and phonological explicitness.

Consider, however, cases such as 11, also discussed by Büring, where focus sensitive operators are present.[20]

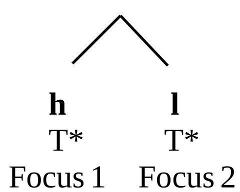
(11) A: John's party was a real drag…Hardly anyone had an alcoholic drink.

(L+)H* L- (L+)H* L-L%
B1: # Yeah… even JOHN only drank SOYMILK.

(L+)H* L- (L+)!H* L-L%
B2: Yeah… even JOHN only drank SOYMILK.

Whereas the fact remains that both foci must be (nuclear) accented in cases like (11B), it is conspicuously the case that the pitch accent marking *soymilk* must be lower than the one on *John*.[21] That is, in tonal metrical structure, (11B2), much like the Answer-Background ordering, requires the following structure:

(12)



Thus, to the extent that Büring's alignment constraint seemed to capture an intuition about relative prominence in (9B), it fails in (11B2), since it predicts exactly the same prosodic structure for both cases. Since both foci must be accented, and because both foci are similar semantically (in this case, both are "bound" by operators[22]), the alignment constraint once again would assign higher prominence to the rightmost accented focus.

---

[20] Rooth (1996) considers the following examples, which are similar in syntactic, and I suspect prosodic, structure:

(i) Even the mayor's closest supporters$_{\mathbf{FOC}}$ are saying that only his opponents$_{\mathbf{FOC}}$ have any chance of winning.

[21] Note that a deaccented soymilk would be appropriate only in a situation in which it was given that *soymilk* was the only available drink.

[22] Büring (2008:8) notes that it is unlikely that these effects can be explained by asymmetrical scopal relations between the operators. While the semantic scope of *even* contains that of *only* in the *soymilk*-example, Büring offers

However, there is an additional problem with assuming that the pitch range difference represents the marking of the head of an intonational phrase, and that is the fact that phrasing *John* and *soymilk* in completely separate intonational phrases does not seem to make the prominence relation any less important, just as it does not in the Answer-Background/Background Answer cases:

(13) A: I heard Mary drank water…

(L+)H* L-L% (L+)H* L-L%
B: …and JOHN drank SOYMILK.

(14) A: John's party was a real drag…Hardly anyone had an alcoholic drink.

(L+)H* L-L% (L+)H* L-L%
B: # Yeah… even JOHN only drank SOYMILK.

There are at least two ways in which we might assume, however, that the tonal differences we are observing are unrelated to focus structure; this would absolve Büring from having to account for them, although it renders mysterious what it might mean, in phonetic or phonological terms, to be the head of IP in his system. The first is to assume that *soymilk* in (11B2) is actually not a focus, but a topic, since it does seem that (11B2) provides only a partial answer to the larger question (which was not explicitly asked) "*Who (even) drank (only) what?*". Although it

---

cases such as (i) as evidence against the possibility that scopal relations explain the prosodic pattern (The subscripted focus markers link a focus with its associated operator):

(i) Context: (We all agreed that Sam's mother loves Ernie.)

A: … Only$_{\mathbf{F1}}$ John$_{\mathbf{F1}}$ suggested that Sam's mother also$_{\mathbf{F2}}$ loves SUE$_{\mathbf{F2}}$.

B: … And only$_{\mathbf{F1}}$ John$_{\mathbf{F1}}$ suggested that Sam's mother also$_{\mathbf{F3}}$ loves SAM$_{\mathbf{F3}}$.

Here it is clear that, in the rejoinder of B, *only* outscopes *also*, just as *even* outscopes *only* in the *soymilk*-example, and yet the downstepping does not seem appropriate on the focus of also, *Sam*. One thing about this sentence, however, is that *John* is a repeated focus, and seems suspiciously like a second occurrence focus, which, as is argued below to require a l-h tonal structure. In any case, however, this example does demonstrate that the matter cannot be so simple as to depend on the semantic scope of focus sensitive operators.

lacks the rise usually associated with contrastive topic marking, experimental evidence was noted above that indicates that this rise may be optional, especially in the focus-topic ordering (Calhoun 2012).

However, another, and I think more likely, possibility is that *soymilk* in (11B2) is in fact a contrastive focus, but is a highly *accessible* one. Although *soymilk* is new information, it is a member of (and contrasted with other members of) the highly salient set of "non-alcoholic drinks", mentioned in the context. As discussed above, when accentuated, accessible information tends to be produced in a reduced pitch range. Whatever the best semantic or pragmatic characterization of this construction turns out to be, it seems clear that it phonological encoding should make reference to pitch range.

Second occurrence focus

The theory of tonal metrical structure also has the potential to treat a certain case of second occurrence focus (SOF) (Partee, 1999; Beaver, Clark, Flemming, Jaeger, Wolters 2007; Büring 2008; Rooth 2010 inter alia). The basic phenomenon of SOF is shown in (15) from Partee (1999):

(15) A: Everyone already knew that Mary only eats VEGETABLES.

B: If even PAUL knew that Mary only eats vegetables, then he should have suggested a different RESTAURANT.

Partee (1999)

In (15A), *vegetables* is the focus of *only*, and, like most foci in English, it receives a nuclear accent. In (15B), *vegetables* is again the focus of *only*, but this time it does not seem to require a pitch accent. The relevant generalization about such pitch-accentless foci is that they (1) are foci that are repeated in the discourse, and (2), they follow, and are prosodically phrased with, a new and unrepeated focus (which receives the nuclear accent). It has been noted, however, that such second occurrence foci do show some phonetic reflexes of their focal status, in the direction of being more prominent in terms of duration and intensity (Beaver et al. 2007, Howell 2011), although these differences are usually extremely numerically small.

Although it has received less discussion, second occurrence focus can also occur before the nuclear-accented first occurrence focus, as the examples in (16) and (17) show:

(16) A: I hear that John only gave a $[\text{BOOK}]_{\textbf{Foc}}$ to Mary.
B: True, but John only gave a $\text{book}_{\textbf{SOF}}$ to $[\text{MANY people}]_{\textbf{Foc}}$.

(Dryer 1994)

(17) A: Only $[\text{MANNY}]_{\textbf{Foc}}$ likes John.
B: …and only $\text{Manny}_{\textbf{SOF}}$ likes $[\text{ANNA}]_{\textbf{Foc}}$ (too).

(Rooth 1996)

In both (16A) and (16B) of Dryer's (1994)[23] example, the focus of *only* is *book*; however, in (16B), it precedes the new focus *many people*. In Rooth's (1996) example, *Manny* in (17B) is similarly repeated and focused, and occurs before the new focus on *Anna*. While it is clear that the new focus in both cases receives a nuclear pitch accent, it is unclear in both of the (B)-sentences whether a pitch accent is required for the second occurrence focus.

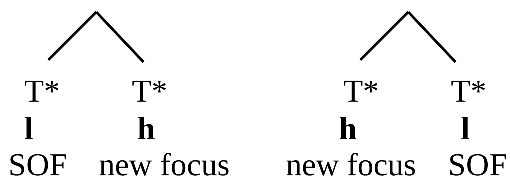
In a production study (Bishop 2008), I examined the acoustic properties of SOF like those in (16) and (17) and found that f0 did not distinguish them from non-foci in similar sentences. They were, however, distinguished by very small differences in duration and intensity, much like postnuclear SOF. Bishop (2009), however, showed that this was not necessarily because these prenuclear cases were always unaccented, but because they were accented at about the same rate as non-foci in the same position (with which they were compared). What is relevant is that these

[23] This example is attributed to Dryer by Beaver et al. (2007) .

accents that occurred in prenuclear SOF sentences tended not to be any higher in f0 than the nuclear accent marking the following new focus. [24]

This pattern, now quite familiar, lends itself well to an analysis as a tonal metrical requirement. That is, we may assume that SOF-new focus orderings require the l-h structure, which allow SOF to be pitch accented so long as they are not more tonally prominent than the following new focus:

(18)



Impressionistically, I think there is a clear preference in sentences like (16B) and (17B) that can be observed if one tries to put extra emphasis on the SOF; this results in the need to put yet more emphasis on the following new focus. Assuming tonal metrical structure is part of the contrast accounts for the optional accenting in both SOF-new focus orderings; when the SOF precedes the focus, a l-h structure is satisfied by any tonal realization on the SOF ranging from no pitch accent to a H* of equal or lesser value to the accent marking the new focus. In the new focus-SOF ordering, any realization ranging from no accent on the SOF, to a H* of significantly lower height than the H* marking the new focus will be sufficient. Since when the SOF follows the new focus, the SOF is typically deaccented, however, it is likely the case that lower-level metrical structure makes its own requirements on post-nuclear SOF.[25]

[24] See Féry and Ishihara (2009) for results of this sort for German. In their data, prenuclear SOF are far more likely to be prenuclear accented than non-foci, however, whereas Bishop (2009) did not find such a large disparity for a small set of English speakers. A possible reason for this may be the syntactic structures used; whereas prenuclear SOF in Bishop (2009) were in object position (like Dryer's example in Z), Féry and Ishihara's SOF were in subject position, and thus occurred far closer to the beginning of the utterance.

[25] But, again, see Beaver et al. (2007) for evidence that speakers do occasionally pitch accent SOF that follow new foci, and that, when they do, it is significantly lower in f0.

Thus, although a more focused production study may be needed, it seems likely that the realization of second occurrence focus can be understood in terms of a theory of the phonology that makes reference to syntagmatic relations of tonal prominence like the ones proposed.

Contrastive versus non-contrastive focus

The last example I will consider is the prosodic realization of contrastive versus non-contrastive focus. Whether or not there is a special prosodic marking for contrast in English has been of an issue of debate for some time. It has been reported that when speakers mark foci as contrastive, they do so using somewhat more acoustic prominence (e.g., Hanssen et al. 2008; Baumann et al. 2007; Bartels and Kingston 1994; Ito, Speer and Beckman 2004). However, studies are quite mixed with respect to how reliably speakers are in their marking of contrastive focus, and which phonetic parameters are relevant—making any phonological claims very tenuous.

This issue has recently been approached by Katz and Selkirk (2011), who point out that much of the uncertainty regarding the results of production studies may be due to the fact that previous studies have investigated sentences in which there is only one focus, and context is used to indicate its new versus contrastive status. In their study, the authors test sentences that contain two foci, and manipulate their contrastive (in their terms "Focused") versus non-contrastive (i.e., simply "new") status, allowing them to compare the two directly to examine any consistent differences between them.

They compared the two kinds of foci in sentences such as those in Table 4.1, which manipulated the ordering of foci as: (A) focus-new order, (B) new-focus order, and (C) "all-new" (i.e., two new items, neither of which is contrastive). Katz and Selkirk reason that an all-new sentence should serve as a kind of baseline condition for prominence, and the effect of being

Table 4.1 Example set of stimuli used in Katz and Selkirk (2011). Manny and yellow occur in three contexts, manipulating their status as focused (i.e., contrastive focus) and discourse-new (non-contrastive focus).

Condition A: *Focus*-New	
Context:	The Red Sox had an exhibition game for charity, and they gave the players various bright-colored uniforms. Bill Mueller and Nomar Garciaparra have really played well this year.
Sentence:	**But they only gave [Manny]**$_{\text{Foc}}$ **[the yellow one]**$_{\text{New}}$ (…That's the one that's reserved for the most valuable player.)
Condition B: New-*Focus*	
Context:	The Red Sox had an exhibition game for charity, and they had special bright-colored uniforms made for the occasion. There were a lot of different colors; a couple of the jerseys were orange, one was purple.
Sentence:	**But they only gave [Manny]**$_{\text{New}}$ **[that yellow one]**$_{\text{Foc}}$ (…That was a lousy color.)
Condition C: New-New (*no Focus*)	
Context:	The Red Sox had an exhibition game for charity, and they gave all the players crazy bright-colored uniforms to wear for the occasion. The whole thing was pretty funny to watch.
Sentence:	**They gave [Manny]**$_{\text{New}}$ **[the yellow one]**$_{\text{New}}$ (…It was so ugly.)

contrastive (i.e., "focus" status) should increase prominence relative to the non-contrastive new item. In a multi-speaker production study, they perform both an analyses of pitch accent choices, as well a phonetic analyses of pitch and duration, and intensity of the foci in their three conditions, with the aim of providing a phonological analysis of the prominence structure that will relate the phonetic cues to the information structural meanings.

Katz and Selkirk's results show that, like most double focus constructions, both contrastive foci and discourse-new words are realized with a H* pitch accent, and in the vast majority of cases a *nuclear* pitch accent according to their transcriptions (with the first focus being followed by an L-, marking the edge of a phonological phrase boundary).[26] Thus, it is not the case that any distinction that the grammar might make between focus and discourse-newness could rest on a categorical +/- accent contrast. In their words:

> …it is not a contrast in the tonal representation or the phonetic interpretation of individual pitch accents that correlates with a difference in Focus/new status, but rather a difference in the phonological prominence (stress) of the pitch-accented words, which is reflected phonetically in the within-sentence relation between the phonetic prominence of these words. (2011:788)

Their claims are supported by the phonetic results, which show a three-way distinction corresponding to their conditions in terms of (a) duration, (b) intensity, and (c) pitch excursion—all in the direction of focus being more prominent than discourse-new. Thus, in the present terms, Katz and Selkirk argue that the distinction is one of lower-level metrical structure, with the acoustic correlates of stress accent, and not any syntagmatic relation between tonal structure, although they do find pitch excursion to mark foci as more prominent.

Interestingly, however, the authors also measured f0 height of the accent—which is the primary predictor of tonal metrical relations—and those results do not support a three-way contrast for their conditions. In particular, they did not find a difference between the baseline "all-new" condition, which did not contain any focus, and the sentences that contained a focus-new ordering; in both cases, there was a tendency for the accent on the first test word to be higher than the second. Instead, only the new-focus ordering differed from the base-line, with the second accent in this case being approximately equal in height to the first accent. Exemplary pitch accents are offered by the authors, and reproduced here in Figure 4.2, along with Katz and

[26] The authors explicitly do not attempt to make a H*/L+H* distinction.

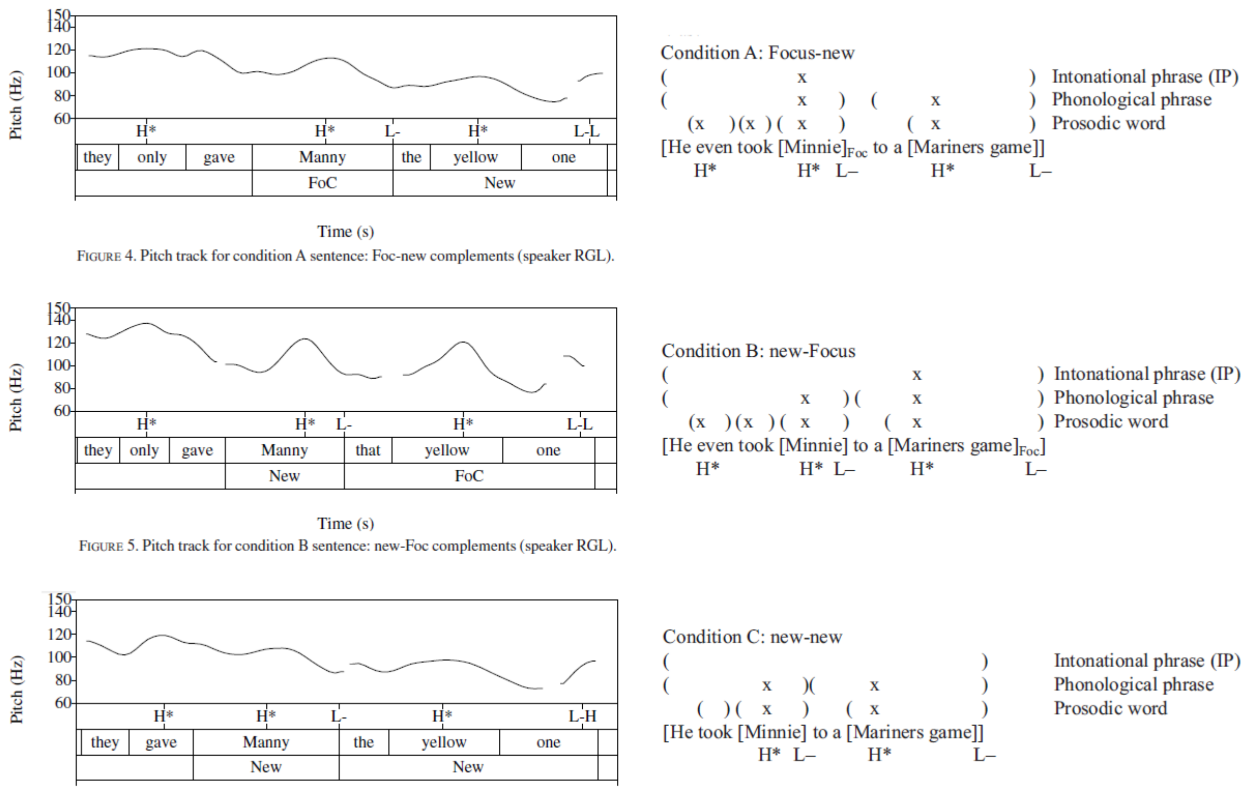
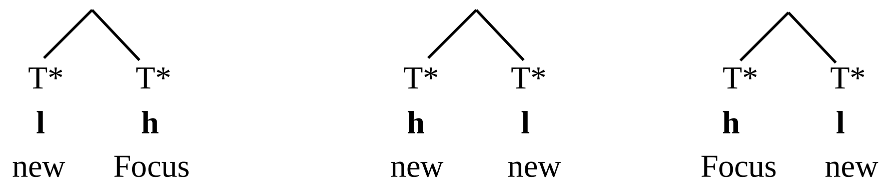


Figure 4.2 Example pitch tracks and proposed metrical phonological representations for different configurations of "focus" (i.e., contrastive focus) and "discourse new" (i.e., non-contrastive focus) described in Katz and Selkirk (2011). Reprinted with permission from the Linguistics Society of America.

Selkirk's proposed three-way phonological contrast as metrical representations. Having found such reliable phonetic correlates of prominence in relation to the location of focus, the authors are puzzled regarding why pitch height should not distinguish the new-new and focus-new conditions (C and A in the Figure, respectively).

However, this pattern is captured very neatly by a theory of tonal metrical structure; as I have already proposed for broad focus sentences (the equivalent of Katz and Selkirk's new-new constructions), there is a preference for the h-l structure, which favors a range of variation that keeps the second accent lower than the first (downstepped, within or across phrases). We thus simply assume that contrastive focus require alignment with the h-node (indicating tonal prominence), and the pattern of results falls out:

(19)



The consequence of this structure is that we predict only a two-way contrast for pitch range to be possible, and so ambiguity will have to occur for at least two of the three orderings; in fact this is what we find. Katz and Selkirk's A and C conditions do not differ in terms of f0 height because they make the same requirements on tonal metrical structure. Additionally, considering the discussion of Büring's *soymilk* example from (11B2), it seems that when both foci are new and contrastive (and neither is *accessible*), such an ordering should be ambiguous with Selkirk's new-Focus ordering, both requiring the l-h structure.

Thus, while the appeal to tonal metrical structure does not disprove Katz and Selkirk's claim (indeed it supports their basic claim that we need a distinction between focus and discourse-new), it does raise questions about what the most relevant perceptual object would be to listeners in a perception experiment. Based on Katz and Selkirk's interpretation of the phonological relevance to their findings, we would expect listeners to easily make a 3-way contrast; under the tonal metrical analysis, ambiguity would be more likely between new-new and focus-new. This is no small matter for Katz and Selkirk, however, who advance their theory as one that explicitly mediates between phonetics and syntax via phonological representations. While this is a question for future study, it demonstrates the promise of the theory of tonal metrical structure for guiding future research questions.

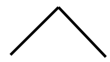
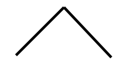
4.2.4 Summary of the proposal

To summarize, it has been proposed that syntagmatic tonal relations control certain types of local pitch range variation. An appeal to this representation as part of AM theory of intonational phonology was first made by Ladd (1990), and the subsequent years have brought attention to a number of information structural phenomena, some of which were discussed above, that can be understood as making reference to this structure in their prosodic realization. To be explicit, Figure 4.2 shows the examples we have considered, and some glosses of the possible information structural representations.

To reiterate the approach taken here, I will end this matter as follows. The theory of tonal metrical structure is particularly useful because standard AM theory includes no level of phonological representation that can encode systematic relations between the pitch scaling of different accents. Within phrases, asymmetrical patterns can be captured by a paradigmatic contrast, namely downstep; however, we have seen evidence above that different downstepped accents in English (e.g., !H*, L+!H*, and possibly H+!H*) seem to have closely related pragmatic functions, suggesting downstep is orthogonal to paradigmatic contrasts. Even more problematic, there are inter-phrasal dependencies in the pitch range of accents that looks very much like downstep, but are, in theoretical terms, regarded as completely unrelated to downstep. One such case is the Answer-Background contrast, but I have shown above that the pattern seems more productive than just that case.

Finally, it should be noted that many if not all of the prominence asymmetries I have discussed above would be predicted by Calhoun's (2006) theory, because she assumes that most phenomenon such as accessibility, contrastive topic, second occurrence focus—all of the information I have assigned to the l-node in tonal metrical structure—is thematic (in the sense of

Table 4.2 Summary of information structure-tonal metrical correspondences, intended to explain pitch range asymmetries for sentences with double foci, a focus and topic, and VP versus narrow object focus sentences.

Requires Right-Prominence			Requires Left-Prominence		
	 T* T* **l** **h**			 T* T* **h** **l**	
(1)	**Focus 1** [+contrastive]	**Focus 2** [+contrastive]	**(1)**	**Focus 1** [+ new] [-contrastive]	**Focus 2** [+ new] [-contrastive]
(2)	**Focus 1** [SOF]	**Focus 2** [new focus]	**(2)**	**Focus 1** [+new] [+contrastive]	**Focus 2** [+new] [-contrastive]
(3)	**Verb** [given]	**Object** [focus]	**(3)**	**Focus 1** [new focus]	**Focus 2** [SOF]
(4)	**Topic**	**Focus**	**(4)**	**Focus**	**Topic**
			(5)	**Verb** [+ new]	**Object** [+ new]
			(6)	**Focus 1** [+new] [+contrastive]	**Focus 2** [+new] [+contrastive] [+accessible]

Steedman 2000), and themes are made by the grammar to be less phonetically less prominent than rhemes. One problem with her particular implementation of this, however, is that it does not provide a specific phonological structuring of prominence sufficient to make predictions about what phonetic cues should be relevant, and when. As I have argued, it seems there are a number of contrasts for which phonetic cues do not seem equally exchangeable, but require the adjustment of pitch range of adjacent accents. I therefore suggest that prominence at the level of syntagmatic tonal relations, as proposed by Ladd, allow us to make optimal use of the insights of Calhoun's observations about the role of relative prominence in marking of information structure.

4.3 Concluding remarks

In this dissertation I have examined the role that prenuclear accents might play in the marking of focus, and the case study for this investigation was the broad versus narrow focus contrast in English. As we have seen in Chapters 2 and 3, prenuclear accents are not as optional as many theorists have claimed. The findings in those chapters, and the consideration of other information structural phenomena in this Chapter, led me to propose that the role prenuclear accents play in marking the size of the focus constituent is modulated by syntagmatic tonal relations. These tonal relations represent the abstract, phonological control of pitch range, first proposed by Ladd (1990), and not part of the standard Autosegmental Metrical models assumed currently by most researchers. However, as I have also shown, it seems such structure is needed to account for a number of other independent phenomena, which were summarized in Table 4.2.

The results of the present study therefore represent a starting point for future work. To begin with, it is clear that there are a number of questions that should be asked more carefully in production studies, such as whether speakers adhere to certain phonetic relations, or "constants" for pitch range relations for these contrasts, the way that Liberman and Pierrehumbert (1984) show is the case for the Answer-Background construction. In perception, it will be necessary to understand just how asymmetrical a pitch range relation needs to be in order to be interpreted as the l-h rather than the h-l structure, especially across phrases. As mentioned during the discussion of accessibility in Section 4.2.3.1, it is likely that many such questions of perception and interpretation can be fruitfully addressed using on-line experimental methodology.

Finally, before concluding, it is important to state that the findings in this study only directly pertain to English, although it is likely that many aspects of the findings will turn out also to apply to the other West Germanic languages as well. However, if the marking of information structure in English is correctly analyzed as making reference to syntagmatic tonal relations, as I believe it to be, the connections to the larger typological situation are clear. In particular, it aligns English with the large number of languages, most without lexical stress, that show the phenomena of post-focal f0 compression. English post-nuclear deaccenting is often thought to be the stress-language analogue to post-focal compression. However, post-nuclear deaccenting makes primary reference to lower-level, stress-based metrical structure. The implication of the present study is that linguistically-determined pitch range relations exist independently of such metrical structure, aligning English more closely with languages lacking lexical stress that exhibit post-focal compression. This thus suggests typological similarities that were previously underestimated, and sets a new tone for future research on the cross-linguistic prosodic realization of focus and information structure.

Appendix A

Test Sentences (Experiments 1a and 1b)

1.) ***(No… because) I bought a motorcycle.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened yesterday?	Why's your wife mad? Because your roof's leaking?
VP-Foc	What did you do yesterday?	Why's your wife mad? Because you lost your job?
Obj-Foc	What did you buy yesterday?	Why's your wife mad? Because you bought a car?

2.) ***(No… because) I lost my wallet.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you upset? Because of the economy?
VP-Foc	What did you do?	Why are you upset? Because you overslept?
Obj-Foc	What did you lose?	Why are you upset? Because you lost your keys?

3.) ***(No… because) I failed my midterm.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you so worried? Because of the GREs?
VP-Foc	What did you do?	Why are you so upset? Because you're running late?
Obj-Foc	Your grade is really low... what did you fail?	Why are you so upset? Because you failed your quiz?

4.) ***(No… because) I met a girl.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you so happy? Because school's out?
VP-Foc	What did you do at the party last night?	Why's your mom so excited? Because you graduated?
Obj-Foc	Who did you meet at the party last night?	Why are you so happy? Because you met a movie star?

5.) ***(No… because) I read a book.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened at home?	Why were you up so late? Because it was noisy?
VP-Foc	What did you do at home?	Why were you up so late? Because you were doing homework?
Obj-Foc	What did you read at home?	Why were you up so late? Because you read a magazine?

6.) ***(No… because) I passed the final.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened in class?	Why are you so happy? Because it's Friday?
VP-Foc	What did you do in class?	Why are you so happy? Because you finished reading?
Obj-Foc	What did you pass?	Why are you so happy? Because you passed the quiz?

7.) ***(No… because) I bought a car.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened while I was gone?	Why are you so broke all of the sudden? Because of the economy?
VP-Foc	What did you do with all your money?	Why are you so broke? Because you started gambling?
Obj-Foc	What did you buy with all your money?	Why are you so broke? Because you bought a house?

8.) ***(No… because) I rode a Harley.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened today?	Why are you so excited? Because of the game?
VP-Foc	What did you do?	Why are you so excited? Because you went jogging?
Obj-Foc	What did you ride?	Why are you so excited? Because you rode a pony?

9.) ***(No… because) I bought a watch.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What's up?	Why are you so happy? Because of the weather?
VP-Foc	What did you do?	Why are you so happy? Because you talked to Suzie?
Obj-Foc	What did you buy?	Why are you so happy? Because you bought a hat?

10.) ***(No… because) I drove a Porsche.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you so happy? Because of the party today?
VP-Foc	What did you do?	Why are you so happy? Because you went shopping?
Obj-Foc	What did you drive?	Why are you so happy? Because you drove a Mercedes?

11.) ***(No… because) I finished my paper.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you so happy? Because class was cancelled?
VP-Foc	What did you do?	Why are you so happy? Because you went on a date?
Obj-Foc	What did you finish?	Why are you so happy? Because you finished your homework?

12.) ***(No… because) I ate a hamburger.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened at the picnic?	Why aren't you hungry? Because of the medication?
VP-Foc	What did you do at the picnic?	Why aren't you coming to lunch? Because you're dieting?
Obj-Foc	What did you eat at the picnic?	Why aren't you hungry? Because you ate a hot dog?

13.) ***(No… because) I called the doctor.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you feeling so much better? Because of the weather?
VP-Foc	What did you do?	Why are you feeling so much better? Because you slept in?
Obj-Foc	Who did you call?	Why are you feeling better? Because you called the nurse?

14.) ***(No… because) I pawned the stereo.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you so rich all of the sudden? Because of the stimulus check?
VP-Foc	What did you do?	Why are you so rich all of the sudden? Because you worked overtime?
Obj-Foc	What did you pawn?	Why are you so rich all of the sudden? Because you pawned the T.V.?

15.) ***(No… because) I fixed the roof.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why's your wife so happy? Because of the vacation?
VP-Foc	What did you do?	Why's your wife so happy? Because you took her out to dinner?
Obj-Foc	What did you fix?	Why's your wife so happy? Because you fixed the fence?

16.) ***(No… because) I painted the kitchen.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why's your wife so happy? Because it's her birthday?
VP-Foc	What did you do?	Why were you busy all day? Because you were working out?
Obj-Foc	What did you paint?	Why's your wife so happy? Because you painted the fence?

17.) ***(No… because) I kissed another cheerleader.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened after the game?	Why are you smiling like that? Because of the game?
VP-Foc	What did you do after the game?	Why are you smiling like that? Because you played well?
Obj-Foc	Who did you kiss after the game?	Why are you smiling like that? Because you kissed another pompom girl?

Appendix B

Test Sentences (Experiment 2)

	Item	Test Sentence
Low Frequency Verb	1.)	I braised a lamb.
	2.)	I ditched the truck.
	3.)	I latched the gate.
	4.)	I pawned the stereo.
	5.)	I snagged some tickets.
	6.)	I spooked the visitors.
	7.)	I stashed the coins.
	8.)	They annexed the township.
	9.)	He pestered the kids.
	10.)	She swaddled the infant.
	11.)	They voided the payment.
	12.)	I wagered my savings.
Mid Frequency Verb	13.)	I cleaned the windows.
	14.)	I climbed a mountain.
	15.)	I cooked the lentils.
	16.)	I counted the apples.
	17.)	I earned a check
	18.)	I hit the ball.
	19.)	I kissed a blonde.
	20.)	I touched the stove.
	21.)	They destroyed the city.
	22.)	I ordered some fish.
	23.)	He published a novel.
	24.)	They removed the paint.
High Frequency Verb	25.)	They saw a film.
	26.)	He bought a motorcycle.
	27.)	I read a book.
	28.)	I wrote a memo.
	29.)	I used the wrench.
	30.)	I called a doctor.
	31.)	I found a hat.
	32.)	I took a pencil.
	33.)	I got a soda.
	34.)	I made a pizza.
	35.)	I lost my wallet.
	36.)	I finished my lunch.

Appendix C

Experimental Materials (Experiments 3a and 3b)

Test Sentences, Primes and Targets

Item	Test Sentence with sentence-final prime	Control-prime	Visual Target
1.	He amused the butler.	bachelor	maid
2.	He borrowed a screwdriver.	newspaper	hammer
3.	She broke the table.	teapot	chair
4.	He called a doctor.	donor	nurse
5.	I dropped my wallet.	bible	keys
6.	He fixed the window.	cable	door
7.	I found a penny.	bottle	nickel
8.	He hurt his arm.	horse	leg
9.	She identified the artist.	burglar	painting
10.	She impressed the tutor.	locals	teacher
11.	He kissed a blonde.	guest	brunette
12.	He made a cake.	scarf	pie
13.	He observed a dolphin.	stars	whale
14.	He offended an employee.	attendant	boss
15.	She ordered a whiskey.	movie	beer
16.	He bought a motorcycle.	supermarket	car
17.	He requested a pillow.	toothbrush	blanket
18.	He smelt coffee.	smog	tea
19.	He stole a brownie.	dollar	cookie
20.	She took my fork.	hat	spoon
21.	He wanted a pancake	flashlight	waffle
22.	He washed the dog.	bike	cat
23.	He won an apple.	trophy	pear
24.	He loves spinach.	cities	broccoli
25.	He annoyed a girl.	bird	boy
26.	He borrowed the salt	tape	pepper
27.	I bought boots.	books	shoes
28.	He broke the toaster.	picture	oven
29.	He called his sister.	student	mother
30.	He finished his soup.	chores	salad
31.	I fixed the sink.	fence	toilet
32.	I found a pencil.	cell phone	pen

Question Contexts

Item	VP Focus Context	Object Focus Context
1.	What did Robert do at the party? Did he keep to himself?	Who did Robert amuse at the party? Did he amuse the children?
2.	What did John do when he was here? Did he chat?	What did John borrow when he was here? Did he borrow a book?
3.	What did Natasha do when she found out? Did she scream?	What did Natasha break when she found out? Did she break the dishes?
4.	What John you do when he found out? Did he cry?	Who did John call when he found out? Did he call his mother?
5.	What did you do? Did you trip and fall?	What did you drop? Did you drop your hat?
6.	What did John do when he came home? Did he sit around?	What did John fix when he came home? Did he fix the refrigerator?
7.	What did you do? Did you forget something?	What did you find? Did you find someone's keys?
8.	What did John do? Did he forget about today?	What did John hurt? Did he hurt a person?
9.	What did Natasha do when she arrived? Did she teach?	Who did Natasha identify when she arrived? Did she identify the owner?
10.	What did Natasha do? Did she give up?	Who did Natasha impress? Did she impress the visitor?
11.	What did Robert do after the party? Did he leave?	Who did Robert kiss after the party? Did he kiss Mary?
12.	What did Robert do? Did he clean?	What did Robert make? Did he make something healthy?
13.	What did Robert do while he was there? Did he walk around?	What did Robert observe while he was there? Did he observe people?
14.	What did John do at the restaurant? Did he behave?	Who did John offend at the restaurant? Did he offend your friend?
15.	What did Natasha do? Did she get up and leave?	What did Natasha order? Did she order a pizza?
16.	What did Robert do with his money? Did he invest?	What did Robert buy with his money? Did he buy a Rolex?
17.	What did the executive do when he called up? Did he complain?	What did the executive request when he called up? Did he request a refund?
18.	When the executive came downstairs, what did he do? Did he read?	When the executive came downstairs, what did he smell? Did he smell smoke?
19.	What did John do when he came in? Did he clean up?	What did John steal when he came in? Did he steal the remote?
20.	What did Natasha just do? Did she ask for something?	What did Natasha just take? Did she take your seat?
21.	What did John do when you woke him up? Did he go back to sleep?	What did John want when you woke him up? Did he want a glass of water?
22.	What did John do this afternoon? Did he paint?	What did John wash this afternoon? Did he wash the car?
23.	What did the student do? Did he give up?	What did the student win? Did he win a ribbon?
24.	What did you hear about the actor? Does he work out?	What does the actor love? Does he love exercise?
25.	What did John do at the Anderson's party? Did he dance?	Who did John annoy at the Anderson's party? Did he annoy Mr. Anderson?
26.	What did Robert do when he came by? Did he snoop around?	What did Robert borrow when he came by? Did he borrow the broom?
27.	What did you do this afternoon? Did you work?	What did you buy this afternoon? Did you buy groceries?
28.	What did John do? Did he help out?	What did John break? Did he break a mirror?
29.	What did Robert do? Did he go to bed?	Who did Robert call? Did he call his friend?
30.	What did Robert do before he left the house? Did he clean up?	What did Robert finish before he left the house? Did he finish his homework?
31.	What did you do all day? Did you sleep in?	What were you fixing all day? Did you fix the roof?
32.	What did just do? Did you trip on something?	What did you find? Did you find a dollar

BIBLIOGRAPHY

Ayers, G. (1996). Nuclear pitch accent types and prominence: some psycholinguistic experiments. Ph.D. dissertation, The Ohio State University.

Barnes, J., Veilleux, N., Brugos, A., and Shattuck-Hufnagel, S. (2012). Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology* 3(2), 337–383.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): evidence from Asperger Syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders* 31, 5-17.

Bartels, C. and Kingston, J. (1994). Salient pitch cues in the perception of contrastive focus. In: P. Bosch and R. van der Sandt (Eds.), *Focus and natural language processing, Volume 1: Intonation and Syntax*, 1-10. IBM Deutschland Informationssysteme GmbH Scientific Center, Institute for Logic and Linguistics.

Bates, D. and Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. Available at: *http://CRAN.R-project.org/package=lme4*.

Baumann, S., Grice., Steindamm, S. (2006). Prosodic marking of focus domains: categorical or gradient? *Proceedings of Speech Prosody 2006*, 301-304.

Baumann, S., Becker, J., Grice, M., Mücke, D. (2007). Tonal and articulatory marking of focus in German. *Proc. of the XVIth International Congress of Phonetic Sciences*, 1029-1032.

Beaver, D., Clark, B., Flemming, E., Jaeger, T.F., Wolters, M. (2007). When semantics meets phonetics: acoustical studies of second occurrence focus. *Language* 83, 245-276.

Beaver, D. and Clark, B. (2008). *Sense and sensitivity: how focus determines meaning*. Oxford: Wiley-Blackwell.

Beckman, M. (1986). *Stress and non-stress accent.* (Netherlands Phonetic Archives 7). Dordrecht: Foris.

Beckman, M. (1996). The parsing of prosody. *Language and Cognitive Processes* 11, 17- 67.

Beckman, M., and Hirschberg, J. (1992). The ToBI annotation conventions. Ms. The Ohio State University.

Beckman, M. and Ayers Elam, G. (1997). Guidelines for ToBI labeling (version 3.0, March 1997). MS. The Ohio State University Research Foundation.

Beckman, M., Hirschberg, J., and Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed): *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Beckman, M., and Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook* 3, 255-309.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., and Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113(2), 1001-1024.

Berent, I., Steriade, D., Lennertz, T., and Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104, 591-630.

van den Berg, R., Gussenhoven, C., and Rietveld, T. (1992). Downstep in Dutch: implications for a model. In G. Docherty and R. Ladd (Eds): *Papers in laboratory phonology II*. Cambridge: Cambridge University Press.

Birch, S. and Clifton, C. (1995). Focus, accent and argument structure: effects on language comprehension. *Language and Speech* 38, 365-391.

Birch, S. and Rayner, K. (1997). Linguistic focus affects eye movements during reading. *Memory and Cognition* 25, 653-660.

Bishop, J. (2008). The effect of position on the realization of second occurrence focus. *Proceedings of Interspeech 2008*, 869-872.

Bishop, J. (2009). On the prenuclear realization of second occurrence focus in English. Unpublished MA thesis, University of California, Los Angeles.

Bishop, J. (2012). Information structural expectations in the perception of prosodic prominence. In G. Elordieta and P. Prieto (Eds.): Prosody and Meaning (Trends in Linguistics), 239-270. Berlin: Mouton de Gruyter.

Blutner, R. and Sommer, R. (1988). Sentence processing and lexical access: the influence of the focus identifying task. *Journal of Memory and Language* 27, 359-367.

Braun, B. and Tagliapietra, L. (2010). The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes* 25, 1024-1043.

Brazil, D., Coulthart, M., and Johns, C. (1980). *Discourse intonation and language teaching*. London: Longman.

Breen, M., Dilley, L., Kraemer, J., and Gibson, E. (2012). Inter-transcriber reliability for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch). *Corpus Linguistics and Linguistic Theory* 8, 277-312.

Breen, M., Fedorenko E., Wagner., M., and Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes* 25, 1044-1098.

Büring, D. (1997). *The meaning of topic and focus—the 59th Street Bridge Accent*. London: Routledge.

Büring, D. (2003). On D-trees, beans, and B-accents. *Linguistics and Philosophy* 26, 511-545.

Büring, D. (2007). Intonation, semantics and information structure. In G. Ramchand and C. Reiss (Eds.): *The Oxford Handbook of Linguistic Interfaces.* Oxford: Oxford University Press.

Büring, D. (2008). Been there, marked that—a theory of second occurrence focus. Ms, University of California, Los Angeles.

Calhoun, S. (2006). Information structure and the prosodic structure of English: a probabilistic relationship. Ph.D. dissertation, University of Edinburgh.

Calhoun, S. (2010). The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language* 86, 1-42.

Calhoun, S. (2012). The theme/rheme distinction: Accent type or relative prominence? *Journal of Phonetics* 40, 329-349.

Chafe, W. (1974). Language and Consciousness. *Language* 50, 111-133.

Chen, K. and Hasegawa-Johnson, M. (2004). An automatic prosody labeling system using ANN-based syntactic-prosodic model and GMM-based acoustic-prosodic model. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing.*

Chomsky, N. (1971). Deep structure, surface structure, and semantic interpretation. In D Steinberg and L. Jakobovits (Eds): *Semantics: an interdisciplinary reader in philosophy, linguistics, and psychology*. Cambridge: Cambridge University Press.

Chomsky, N. and Halle, M. (1968). *The sound pattern of English*. New York: Harper Row.

Cole, J., Mo, Y., and Baek, S. (2010). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes* 25, 1141-1177.

Cole, J., Mo, Y., Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology* 1, 425-452.

Conkie, A., Riccardi, G., and Rose, R. (1999). Prosody recognition from speech utterances using acoustic and linguistic based models of prosody events. In *Proceedings of Eurospeech*. 523-526.

Constant, N. (2012). English rise-fall-rise: A study in the semantics and pragmatics of intonation. *Linguistics and Philosophy* 35, 407–442.

Cuetos, F. Hallé, P., Domínguez, A., Segui, J. (2011). Perception of prothetic /e/ in #sC utterances: Gating data. *Proceedings of the XVIIth International Congress of Phonetic Sciences*, 540-543.

Cutler, A. and Fodor, J.A. (1979). Semantic focus and sentence comprehension. *Cognition* 7, 49-59.

Daneman, M. and Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19, 450-466.

Davidson, L and Shaw, J. (2012). Sources of illusion in consonant cluster perception. *Journal of Phonetics* 40, 234-248.

Dilley, L. (2005). *The phonetics and phonology of tonal systems*. Ph.D. dissertation, MIT.

Dohn A., Garza-Villarreal E., Heaton P., and Vuust P. (2012). Do musicians with perfect pitch have more autism traits than musicians without perfect pitch? An empirical study. *PLoS ONE* 7(5): e37961. doi:10.1371/journal.pone.0037961.

Dryer, R. (1994). The pragmatics of association with *only*. Paper presented at the annual meeting of the Linguistic Society of America, Boston MA. 6-9 January 1994.

Dupoux, E., Kaheki, K., Hirose, Y., Pallier, C., and Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25: 1568-1578.

Eady, S. and Cooper, W. (1986). Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America* 80, 402-415.

Eady, S., Cooper, W., Klouda, G., Mueller, P., and Lotts, D. (1986). Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech* 29, 233-251.

Eriksson, A., Thunberg, G., and Traunmüller, H. (2001). Syllable prominence: a matter of vocal effort, phonetic distinctness and top-down processing. *Proceedings of the European Conference on Speech Communication and Technology*, 399-402.

Fedorenko, E., Gibson, E., and Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: evidence against domain-specific working memory resources. *Journal of Memory and Language* 54, 541-553.

Féry, C. and Ishihara, S. (2009). The phonology of second occurrence focus. *Journal of Linguistics* 45, 285-313.

Féry, C. and Krifka, M. (2008). Information structure Notional distinctions, ways of expression. In Piet van Sterkenburg (Eds.): *Unity and diversity of languages*, 123-136. Amsterdam: John Benjamins.

Féry, C. and Samek-Lodovici, V. (2006). Focus projection and prosodic prominence in nested foci. *Language* 82, 131-150.

Féry, C. and Truckenbrodt, H. (2005). Sisterhood and tonal scaling. *Studia Linguistica* 59 223-243.

Foss, D. and Ross, J. (1983). Great expectations: context effects during sentence processing. In G. Flores D'Arcais and R. Jarvella (Eds.): *The process of language understanding*. Chichester: Wiley. 169-191.

Fowler, C. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception and Psychophysics* 68, 161-177.

Fowler, C. and Housum, J. (1987). Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26, 489-504.

Freeth M, Sheppard E, Ramachandran R, Milne E. (2013). A cross-cultural comparison of autistic traits in the UK, India and Malaysia. *Journal of Autism and Developmental Disorders*. [On-line version ahead of print publication DOI: http://dx.doi.org/10.1007/s10803-013-1808-9].

Fry, D. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of* America 27, 765-768.

Ganong, W. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6, 110-125.

Grabe, E., Post, B., and Nolan, F. (2001). Modelling intonational variation in English: the IViE system. In S. Puppel and G. Demenko (Eds.): *Proceedings of Prosody 2000*, 51-57.

Grice, M., Baumann, S., and Benzmüller, R. (2005). German intonation in autosegmental-metrical phonology. In S.-A. Jun (Ed.): *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 55-83.

Grice, M., Baumann, S., and Jagdfeld, N. (2009). Tonal association and derived nuclear accents—the case of downstepping contours in German. *Lingua* 119, 881-905.

Gussenhoven, C. (1983). Testing the reality of focus domains. *Language and Speech* 26, 61-80.

Gussenhoven, C. (1984). *On the grammar and semantics of sentence accents*. Dordrecht: Foris.

Gussenhoven, C. (1999). On the limits of focus projection in English. In Bosch, P., and van der Sandt, R. (Eds.): *Focus: Linguistic, Cognitive, and Computational Perspectives*. Cambridge: Cambridge University Press.

Gussenhoven, C. (2005). Transcription of Dutch intonation. In S.-A. Jun (Ed.): *Prosodic typology: the phonology of intonation and phrasing*, 118-145. Oxford: Oxford University Press.

Gussenhoven, C. (2007). Types of Focus in English. In C. Lee, M. Gordon, and D. Büring (Eds.): *Topic and focus—Cross-linguistic perspectives on meaning and intonation*. (No. 82 in *Studies in Linguistics and Philosophy*). Dordrecht: Springer.

Gussenhoven, C. and Rietveld, T. (1988). Fundamental frequency declination in Dutch: testing three hypotheses. *Journal of Phonetics* 16, 355-369.

Hallé, P., Best, C. (2007). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters. *Journal of the Acoustical Society of America* 121, 2899-2914.

Halle, M. and Vergnaud, J-R. (1987). *An essay on stress*. Cambridge, MA: MIT Press.

Halliday, M. (1967). Notes on transitivity and theme in English, Part II. *Journal of Linguistics* 3, 199-244.

Hanssen, J., Peters, J., and Gussenhoven, C. (2008). Prosodic effects of focus in Dutch declaratives. *Proceedings of Speech Prosody 2008,* 609-612.

Hayes, B. and Lahiri, A. (1991). Bengali intonational phonology. *Natural Language and Linguistic Theory* 9, 47-96.

Hendriks, H. (2002). Information packaging: from cards to boxes. In K. van Deemter and R. Kibble (Eds.): *Information sharing: reference and presupposition in language generation and interpretation*. Stanford: Csli.

Hirschberg, J. and Nakatani, C. (1996). A prosodic analysis of discourse segments in direction-giving monologues. *Proceedings of the Association for Computational Linguistics*, 286-293.

Hirschberg, J., Gravano, A., Nenkova, A., Sneed, E., and Ward, G. (2007). Intonational overload: Uses of the downstepped (H* !H* L- L%) contour in read and spontaneous speech. In J. Cole and J.I. Hualde (Eds.): *Laboratory Phonology 9.* Berlin: Mouton de Gruyter.

Hoekstra, R., Bartels, M., Cath, D., and Boomsma, D. (2008). Factor structure, reliability and criterion validity of the Autism-Spectrum Quotient (AQ): A study in Dutch population and patient groups. *Journal of Autism and Developmental Disorders* 38, 1555-1566.

Howell, J. (2011). Second occurrence focus and the acoustics of prominence. In R. Folli and C. Ulbrich (Eds): *Interfaces in Linguistics: New Research Perspectives*. Oxford: Oxford University Press.

Husband, M. and Ferreira, F. (2012) Generating contrastive alternatives: Activation and suppression mechanisms. Paper presented at the CUNY Conference on Human Sentence Processing, The Graduate Center of the City University of New York.

Ito, K., Speer, S., Beckman, M. (2004). Informational Status and Pitch Accent Distribution in Spontaneous Dialogues in English. *Proceedings of Speech Prosody 2004*, 279-282.

Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.

Jagdfeld, N. and Baumann, S. (2011). Order effects on the perception of relative prominence. *Proceedings of the 17th International Congress of Phonetic Sciences,* 958-961.

Jun, S.-A. (2008). Focus: domains, types, and realizations. Talk given at the Yale Linguistics Department Colloquium Series, Yale University, New Haven, CT.

Jun, S.-A. (2010). The Implicit Prosody Hypothesis and overt prosody in English. *Language and Cognitive Processes* 25, 1201-1233.

Jun, S.-A. and Bishop, J. (in progress). Implicit phrasing and prominence influence relative clause attachment. Ms, University of California, Los Angeles.

Katz, J. and Selkirk, E. (2011). Contrastive focus vs. discourse-new: evidence from phonetic prominence in English. *Language* 87, 771-816.

Kingdon, R. (1958). *The groundwork of English intonation*. London: Longman.

Kiss, K. (1998). Identificational focus versus informational focus. *Language* 74, 245-273.

Kochanski, G., Grabe, E., Coleman, J., and Rosner, B. (2005). Loudness predicts prominence; fundamental frequency lends little. *Journal of the Acoustical Society of America* 11, 1038-1054.

Kong, E.-J. and Edwards, J. (2011). Individual differences in speech perception: Evidence from visual analogue scaling and eye-tracking. *Proceedings of the XVIIth International Congress of Phonetic Sciences*, 1126-1129.

Kratzer, A. (2004). Interpreting focus: presupposed or expressive meanings? In *Theoretical Linguistics* 30, 123-136

Krifka, M. (1984). Focus, Topic, syntactische Struktur und semantische Interpretation. Ms, Universität Tübignen.

Krifka, M. (1999). Additive particles under stress. *Proceedings of Semantics and Linguistics Theory (SALT) 8*, 111-128.

Kristensen, L., Wang, L., Petersson, K., and Hagoort, P. (2012). The interface between language and attention: Prosodic focus marking recruits a general attention network in spoken language comprehension. *Cerebral Cortex*. Advance online publication: doi:10.1093/cercor/bhs164.

Ladd, R. (1980). *The structure of intonational meaning: Evidence from English*. Bloomington: Indiana University Press.

Ladd, R. (1986). Intonational phrasing: the case for recursive prosodic structure. *Phonology Yearbook* 3, 311-340.

Ladd, R. (1990) Metrical representation of pitch register. In J. Kingston and M. Beckman (Eds.): *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge University Press.

Ladd. R. (1993). Constraints on the gradient variability of pitch range, or, Pitch Level 4 lives!. In P. Keating (Ed): *Papers in laboratory phonology 3: phonological structure and phonetic form*. Cambridge: Cambridge University Press.

Ladd, R. (1996). *Intonational Phonology*. [Cambridge Studies in Linguistics 79]. Cambridge: Cambridge University Press.

Ladd, R. (2008). *Intonational Phonology*. [Second Edition]. Cambridge: Cambridge University Press.

Ladd, R., Turnbull, R. Browne, C., Caldwell-Harris, C., Ganushchak, L., Swoboda, K., Woodfield, V., and Dediu, D. (in press). Patterns of individual differences in the perception of missing-fundamental tones. *Journal of Experimental Psychology: Human Perception and Performance.*

Lambrecht, K. (1996). *Information structure and sentence form: topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Liberman, M. and Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff and R. Oerhle (Eds): *Language sound structure*. Cambridge, MA: MIT Press.

Makashay, M. (2003). Individual differences in speech and non-speech perception of frequency and duration. Ph.D. dissertation, The Ohio State University.

Mann, V. (1980). Influence of preceding liquid on stop consonant perception. *Perception and Psychophysics* 28: 407-12.

Mann, V. and Repp, B. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America* 69, 548-558.

Marí-Beffa, P., Houghton, G., Estévez, A., and Fuentes, L. (2000). Word-based grouping affects the prime-task effect on semantic priming. *Journal of Experimental Psychology: Human Perception and Performance 26*, 469-479.

Meyer, M-C., Fedorenko, E., and Gibson, E. (2011). Contrastive topic intonation: an empirical evaluation. Poster presented at Experimental and Theoretical Advanced in Prosody 2 (ETAP2), McGill University.

Mitterer, H. (2006). On the causes of compensation for coarticulation: Evidence for phonological mediation. *Perception and Psychophysics* 68, 1227–1240.

Mo, Y. (2008). Duration and intensity as perceptual cues for naïve listeners' prominence and boundary perception. In P. Barbosa, S. Madureira, and C. Reis (Eds.): *Proceedings of the Fourth International Conference on Speech Prosody*, 739-742.

Nakatani, C., Grosz, B., and Hirschberg, J. (1995). Discourse structure in spoken language: studies on speech corpora. *Proceedings of the AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*.

Neeleman, A., Titov, T., van de Koot H., and Vermeulen, R. (2009). A syntactic typology of topic, focus and contrast. In J. van Craenenbroeck (Ed.): *Alternatives to cartography*. Berlin: Mouton de Gruyter.

Nieuwland, M., Ditman, T., and Kuperberg, G. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language* 63, 324-346.

Norris, D., Cutler, A., McQueen, J., Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology* 53, 146-193.

O'Connor, J. and Arnold, D. (1973). *Intonation of colloquial English*. [Second Edition]. London: Longman.

Ostendorf, M., Price, P.J., and Shattuck-Hufnagel, S. (1995). The Boston University radio news corpus. *Boston University Technical Report No. ECS-95-001*.

Palmer, H. (1922). *English intonation, with systematic exercises*. Cambridge: Heffer.

Partee, B. (1999). Focus, quantification, and semantics-pragmatics issues. In Bosch, P., and van der Sandt, R. (Eds.): *Focus: Linguistic, Cognitive, and Computational Perspectives*. Cambridge: Cambridge University Press.

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America* 66, 363-369

Pierrehumbert, J. (1980). The phonology and phonetics of English intonation. Ph.D. dissertation, MIT.

Pierrehumbert, J. and Hirschberg, J. (1990) The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, and M. Pollack (Eds.): *Intentions in Communication*. Cambridge, MA: MIT Press.

Pitrelli, J., Beckman, M., and Hirschberg, J. (1994). Evaluation of prosodic transcription labeling reliability in the ToBI framework. In *Proceedings of the International Conference on Spoken Language Processing*, 123-126.

Pitt, M (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception and Psychophysics* 60, 941-951.

Pitt, M., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. (2007). The Buckeye corpus of conversational speech. Columbus, OH: Department of Psychology, Ohio State University.

R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, *http://www.R-project.org.*

Rietveld, A. and Gussenhoven, C. (1985). On the relation between pitch excursion size and prominence. *Journal of Phonetics* 13, 299-308.

Rochemont, M. (1986). *Focus in generative grammar*. Amsterdam: John Benjamins.

Röhr, C. and Baumann, S. (2011). Decoding Information Status by Type and Position of Accent in German. *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, China. 1706-1709.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics* 1, 75-116.

Rooth, M. (1996). On the interface principles for intonational focus. In T. Galloway and J. Spence (Eds): *Proceedings of SALT VI*, 202-226. Ithaca, NY: Cornell University.

Rooth, M. (2010). Second occurrence focus and relativized stress F. In: C. Féry and M. Zimmerman (Eds.): *Information Structure: Theoretical, Typological and Experimental Perspectives*. Oxford: Oxford University Press.

Rosenberg, A., Hirschberg, J., and Manis, K. (2010). Perception of English prominence by native mandarin Chinese speakers. *Proceedings of Speech Prosody 2010, available at http://speechprosody2010.illinois.edu/program.php.*

Ruta, L., Mazzone, D., Mazzone, L., Wheelwright, S., and Baron-Cohen, S. (2011). The Autism-Spectrum Quotient: Italian version: a cross-cultural confirmation of the broader autism phenotype. *Journal of Autism and Developmental Disorders* 42, 625-633.

Samuel, A. (1981). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General* 110: 474-494.

Sanford, A. and Garrod, S. (1998). The role of scenario mapping in text comprehension. *Discourse Processes* 26, 159-190.

Sanford, A. and Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Sciences* 6, 382-386.

Sanford, A.J.S., Sanford, A.J., Molle, J., and Emmott, C. (2006). Shallow processing and attention capture in written and spoken discourse. *Discourse Processes* 42, 109-130.

Schwarzschild, Roger (1999). "GIVENness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics* 7, 141-177.

Selkirk, E. (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.

Selkirk, E. (1995). Sentence prosody: intonation, stress, and phrasing. In J. Goldsmith (Ed.). *The Handbook of Phonological Theory*. Oxford: Blackwell, 550-569.

Selkirk, E. (2004). Bengali intonation revisited: an optimality theoretic analysis in which FOCUS stress prominence drives FOCUS phrasing. In C-M. Lee, M. Gordon and D. Büring (Eds.): *Topic and Focus: A Cross-Linguistic Perspective*, 217-246. Dordrecht: Kluwer.

Sityaev, D. and House, J. (2003). Phonetic and phonological correlates of broad, narrow and contrastive focus in English. *Proceedings of the XVth International Congress of Phonetic Sciences*, 1819-1822.

Sluijter, A. and van Hueven, V. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100, 2471-2485.

Snedeker, J. and Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language* 48: 103-130.

Sonié, S., Kassai, B., Pirat, E., Bain, P., Robinson, J., Gomot, M., Barthélémy, C., Charvet, D., Rochet, T., Tatou, M., Assouline, B., Cabrol, S., Chabane, N., Arnaud, V., Faure, P., and Manificat, S. (2012). The French version of the Autism-Spectrum Quotient in adolescents: A cross-cultural validation study. *Journal of Autism and Developmental Disorders*, 1-6 (online version accessed: Doi: 10.1007/s10803-012-1663-0).

Sridhar, V.K.R., Nenkova, A., Narayanan, S., and Jurafsky, D. (2008). Detecting prominence in conversational speech: pitch accent, givenness and focus. *Proceedings of Speech Prosody 2008*, 453-456.

Steedman, M. (1991). Structure and Intonation. *Language* 67, 260-296.

Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry* 31, 649-689.

Stewart, M. and Ota, M. (2008). Lexical effects on speech perception in individuals with "autistic" traits. *Cognition* 109, 157-162.

Surprenant, A., and Watson, C. (2001). Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *Journal of the Acoustical Society of America* 110, 2085-2095.

Swets B., Desmet T., Hambrick D., and Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: a psychometric approach. *Journal of Experimental Psychology: General* 136, 64-81.

Swinney, D. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior* 18, 645-659.

Syrdal, A. and McGory, J. (2000). Inter-transcriber reliability of ToBI prosodic labeling. *Proceeding of ICSLP 2000*, 235-238. Beijing.

Szabolsci, A. (1981). The semantics of topic-focus articulation. In T. Jansen and M. Stokhof (Eds.): Formal methods in the study of language. Amsterdam: Mathematisch Centrum.

Tabossi, P. (1988). Effects of context on the immediate interpretation of unambiguous nouns. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14, 153-162.

Tabossi, P. (1996). Cross-modal semantic priming. *Language and Cognitive Processes* 11, 569-576.

't Hart, J. and Collier, R. (1975). Integrating different levels of intonation analysis. *Journal of Phonetics* 3, 235-255.

't Hart, J., Collier, R., and Cohen, A. (1990). *A perceptual study of intonation: an experimental phonetic approach*. Cambridge: Cambridge University Press.

Theodore, R., Schmidt, A. (2003). Perceptual prothesis in native Spanish speakers [Abstract]. *Journal of the Acoustical Society of America* 113, 2256.

Tipper, S. (1985). The negative priming effect: inhibitory effects of ignored primes. *Quarterly Journal of Experimental Psychology* 37A, 571-590.

Truckenbrodt, H. (2002). Upstep and embedded register levels. *Phonology* 19, 77-120.

Turk, A., Nakai, S., and Sugahara, M. (2006). Acoustic segment durations in prosodic research: a practical guide. In S. Sudhoff, D. Lenertova, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, and J. Schließer (Eds.): *Methods in Empirical Prosody Research (Language, Context and Cognition 3)*. Berlin: Walter de Gruyter.

Turk, A. and Sawusch, J. (1996). The processing of duration and intensity cues to prominence. *Journal of the Acoustical Society of America* 99, 3782-3790.

Unsworth, N., Heitz, R., Schrock, J., and Engle, R. (2005). An automated version of the operation span task. *Behavior Research Methods* 37, 498-505.

Vainio, M. and Järvikivi, J. (2006). Tonal features, intensity, and word order in the perception of prominence. *Journal of Phonetics* 34, 319-342.

Vainio, M. and Järvikivi, J. (2007). Focus in production: tonal shape, intensity, and word order. *Journal of the Acoustical Society of America* 121, EL55-61.

Vallduví, E. (1990). The informational component. Ph.D. dissertation, University of Pennsylvania.

Vallduví, E. (1993). Information packaging: a survey. *Technical Report, Human Communication Research Centre*, University of Edinburgh.

Vallduví, E. and Engdahl, E. (1996). The linguistic realization of information packaging. *Linguistics* 34(3), 459-520.

Vallduví, E., and Vilkuna, M. (1998). On rheme and kontrast, in Culicover, Peter and Louise McNally (Eds.), *Syntax and Semantics Vol. 29*: *The Limits of Syntax*. San Diego, CA: Academic Press.

Vallduví, E. and Zacharski, R. (1993). Accenting phenomena, association with focus, and the recursiveness of focus-ground. In P. Dekker and M. Stokhof (Eds.): *Proceedings of the Ninth Amsterdam Colloquium*.

Wagner, M. (2012). Contrastive topics decomposed. *Semantics and Pragmatics* 5, 1-54.

Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., and Tojo, Y. (2006). The Autism-Spectrum Quotient (AQ) in Japan: A cross-cultural comparison. *Journal of Autism and Developmental Disorders* 36, 263-270.

Wang, L., Bastiaansen, M., Yang, Y., and Hagoort, P. (2011). The influence of information structure on the depth of semantic processing: how focus and pitch accent determine the size of the N400 effect. *Neuropsychologia* 49, 813-820.

Ward, P. and Sturt, P. (2007). Linguistic focus and memory: an eye movement study. *Memory and Cognition* 35, 73–86.

Warren, R. (1970). Perceptual restoration of missing speech sounds. *Science* 167: 392-393.

Watson, D., Arnold, J., and Tanenhaus, M. (2008). Tic Tac TOE: effects of predictability and importance on acoustic prominence in language production. *Cognition*, 106, 1548-1557.

Welby, P. (2003). Effects of pitch accent position, type, and status on focus projection. *Language and Speech* 46, 53-81.

Xiang, M., Grove, J. and Giannakidou, A. (2011). Interference “licensing” of NPIs: pragmatic reasoning and individual differences. Poster presented at the 24th annual meeting of the CUNY Conference of Human Sentence Processing, March, Stanford.

Xiang., M., Giannakidou, A., and Grove, J. (2012). Two stages of NPI licensing: An ERP study. Paper presented at the CUNY Conference on Human Sentence Processing, The Graduate Center of the City University of New York.

Xu, Y. and Xu, C. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33, 159-197.

Yu, A. (2010). Perceptual compensation is correlated with individuals' “autistic” traits: implications for models of sound change. *PLoS ONE*, *5*, doi:10.1371/journal.pone.0011950.

Yu, A., Grove, J., Martinović, M., Sonderegger, M. (2011). Effects of working memory capacity and “autistic” traits on phonotactic effects in speech perception. *Proceedings of the XVIIth International Congress of Phonetic Sciences*, 2236-2239.

Zubizarreta, M.L (1998). Prosody, focus, and word order. *Linguistic Inquiry Monographs* 33. Cambridge, MA: MIT Press.

Zubizarreta, M.L. and Vergnaud, J-R**.** (2005). Phrasal stress, focus, and syntax. In M. Everaert and H. van Riemsdijk (Eds.): *The Blackwell Companion to Syntax*. Malden, MA: Blackwell Publishing.

Zubizarreta, M.L (to appear). Nuclear stress and information structure. In C. Féry and S. Ishihara (Eds.): *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press.