HAT PATTERNS AND DOUBLE PEAKS: THE PHONETICS AND PSYCHOLINGUISTICS OF BROAD VERSUS LATE NARROW VERSUS DOUBLE FOCUS INTONATIONS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the Graduate

School of the Ohio State University

By

Stefanie Jannedy, M.A.

* * * * *

The Ohio State University
2002

Dissertation Committee:

Prof. Mary E. Beckman, Adviser

Prof. Keith Johnson

Prof. Shari Speer

Approved by

Adviser
Department of Linguistics

ABSTRACT

This dissertation explores some of the less well−studied aspects of prosodic focus in English. Prosodic focus refers to the ways in which speakers use acoustic means such as fundamental frequency (F0) modulation to make one part of an utterance more prominent than other parts. A fundamental assumption is that the speaker aims to facilitate the hearer's understanding of the message at any particular time in a discourse, and uses intonation to guide the listener's interpretation of the utterance in relationship to the larger discourse context. We assume that the speaker is signaling the discourse purpose of the utterance presented by using differences in the intonational tune. One of the challenges for psycholinguistics, therefore, is to devise tasks that tap the listener's competence in interpreting the intended discourse purpose. That is, a challenge for psycholinguistics is to determine how exactly different prosodic manipulations contribute to the highlighting of a part of an utterance and how this highlighting contributes to the interpretation of the utterances.

While previous experimental work on the contribution of prosodic focus to the interpretation of an utterance has concentrated on matching question-answer pairs (Gussenhoven, 1983; Birch & Clifton, 1995), the experiments devised for this study required listeners to anticipate and make forward interpretations of the context and the moves set up by a speaker. The context in all cases is followed by a stimulus sentence with one of the following intonation contours:

1. H* H* L-L% (hat pattern)

2.	H*		L-L%	(early nuclear accent)
3.	H* L-	H*	L-L%	(double nuclear accents)
4.	H*	L+H*	L-L%	(rising nuclear accent)

In one experiment, listeners were required to select the question that was best addressed by the stimulus sentence having one of these four intonation contours (Question Under Discussion), while in the second experiment, the listener was required to find the most appropriate continuation to the narrative, given the intonation contour.

Results from the QUD paradigm indicate that listeners reliably interpret the early peak contour as an early narrow focus, and that they can reliably differentiate between the phonetically rather similar contours 3 (more double focus responses) and 4 (more late focus responses). These phonetically similar contours differing primarily with regard to the alignment of the intervening low tone, were interpreted to be structurally different, making a phonological analysis of intonation contours imperative. This result points out deficiencies in studies using phonetic descriptions of tonal patterns, such as F0 only, to describe focal prominences.

Results from the continuation paradigm revealed that listeners needed to be pressed for time to perform the task, else they would only show a general preference for broad focus responses for all contours. We take this as an indication that listeners have to respond to a still fresh auditory memory to perform this task since listener's interpretation changed, when the time interval allowed to reply to a stimulus was shortened. As expected, listeners then gave more double focus responses to contour 3 than 4, and more late ratings to contour 4 than 3, showing that they are sensitive to subtle phonetic differences when they are linguistically relevant.

Dedicated

to my Mother

Astrid Jannedy

and my Brother

Benjamin.

ACKNOWLEDGMENTS

Over the years, many people have generously given their time, support, guidance, knowledge and advice to me. First and foremost, I would like to thank Mary Beckman, my advisor for supporting my various research interests in whatever way she possibly could, and for having been a great inspiration. Mary always knew when to push and when to let go. In the early years (when making spectrograms was still the greatest wonder) Mary sparked my interest in phonetics, taught me to listen, to appreciate data - and she set out to educate my palate. She still manages to amaze me with her energy, and her enthusiasm, her willingness to learn and her ability to challenge me to follow. Mary's immensely insightful comments, and the amount of time she's given me to help me finish on time have been invaluable.

A special thank you goes to Keith Johnson who has been a great teacher and resource to me, I have much benefitted from his encouragement, his time and his attitude towards data, and admire his playful and enthusiastic attitude towards phonetics. A big thank you also goes to Shari Speer who repeatedly provided me with needed feedback and interesting insights. I also thank Scott Schwenter for many incredibly thoughtful comments as the graduate school representative during the thesis defense.

My good friends and colleagues have done more for me than they'll ever know: Sun-Ah Jun has educated me on many aspects of phonology, Liz Strand made our office a cozy home and provided much food for thought on various topics, Christian

Uffmann has worked tirelessly to keep me on top of current phonological theory while cooking dinner, and Jennifer Venditti (who watched with me the Magnolia tree bloom for many years), kept me company for many late nights in the lab; Pauline Welby I thank for her extreme generosity from which I have benefitted many times. I am also greatful to Ilse Lehiste and Don Winford for holding me to my promise to finish what I had started.

I am very thankful to the lab rats for their time, help, and encouragement and for making the lab such an extraordinary place to learn, work, play and live: Kevin Cohen, Kim Ainsworth-Darnell, Gayle Ayers Elam, Allison Blodgett, Mariapaola D'Imperio-Piterman, Barbara Gili Fivela, Janice Fon, Svetlana Godjevac, Jamie Green, Rebecca Herman, Craig Hilts, Tsan Huang, Satoko Katagiri, Matt Makashay, Laurie Maynell, Julie McGory, Amanda Miller-Ockhuizen, Ben Munson, Ho-Hsien Pan, Shu-hui Peng, Georgios Tserdanelis, Steve Winters, Peggy Wong, and Kiyoko Yoneyama. Many thanks also to Jim Harmon, Jane Harper, Matt Hyclak, Jignesh Patel, Claudia Stephens, and Sherif Barsoum at OIE for all their help.

For making life in Jones Tower such a fun and happy experience, I sincerely thank Kristin Fishburn, Cemil Ünal, Anne Hoendgen, Tilmann, Wendel and Murat Yildiz.

For many reasons, too countless to mention, I owe much gratitude to a number of teachers, colleagues and friends who, over the years, have extended their friendship to me and have taken the time to explain, to discuss, to challenge, to criticize, to help and support me in whatever way they knew: Anthony Allen, Mary Bradshaw, Allison Blodgett, Julie Boland, Mike Calcagno, Paul Davis, Jan Edwards, Marios Fourakis, Rob Fox, Cecile Fougeron, Rudi Gaudio, Karin Golde, Shelome Gooden, Steve Hartman-Keiser, Beth Hume, Rich Janda, Martin Jansche, Andreas Kathol, No-Ju Kim, LN Loewenbruck, Duncan Markham, Nasiombe Mutonyi, Bernd Möbius, Jen Muller, Dave Odden, Frederick Parkinson, Robert Poletto, Michelle

Ramos-Pellicia, Craige Roberts, Nasiombe Mutonyi, Panos Pappas, Misun Seo, Tom Stewart, Shravan Vasishth, Tracey Weldon, and Jacques Transue & Arnold Zwicky.

I'd also like to express my deepest gratigude to all those who participated in experiments or who let me recruit listeners and speakers in their classes. And I am greatful to Hope Dawson and the OSDL crew for continuing what was begun. I owe much to the Department of Linguistics, for sending me to many conferences, far and near, for supporting me all the way through the program, and for continuously granting funds to conduct experiments.

My friends and colleagues at Lucent Technologies and Bell Labs have been a great resource and support while writing the thesis. I could not have wished for a more supportive work environment, and colleagues who held me to it and expected for this thesis to get done. Thank you, Vic Faubert, John Holmgren, Igor Levin, Amy O'Hara, Joe Olive, Yuriy Pavlov, Leslie von Pischke, Sean Ramprashad, Greg Szeszko, Igor Sominskiy, Michael Tanenblatt, and Evelyne Tzoukerman.

Finishing this thesis would not have been possible without the encouragement of a few who have given much of their time and friendship to me. I consider myself to be extremely lucky for having been able to surround myself with loyal, fun, and smart friends, some of whom took the role of family and cared for me deeply, who reminded me of some of the important things in life, and who chose to share with me some of their thoughts and experiences. Thank you, Jen Hay, Bettina Migge, and Norma Mendoza-Denton.

Finally, I would like to express my sincerest gratitude to my friends and family at home–home. Bettina von Eitzen and Helly Hansen have undertaken every effort to stay in my life, their words of encouragment have echoed in my ears over the years. A thank you is also due to Ulrike Schulte for help with the revisions and for coffee and chocolate.

I'm thankful to Uschi Hesse for all her kindness and care and my mother Astrid Jannedy who insisted that I seek the challenge of a graduate program in the US, and to my brother Ben for his constant encouragement, computer support and for believing that it was possible.

VITA

1966	Born in Hamburg, Germany
1994	M.A. in Linguistics The Ohio State University
2001	Ph.D. in Linguistics The Ohio State University
1999 - present	Lucent Technologies Bell Laboratories, Language Modeling Research and Development, Murray Hill, New Jersey.

PUBLICATIONS

Bod, R., Hay, J. & Jannedy, S (eds.) (2002) *Probabilistic Linguistics.* MIT Press.

Mendoza-Denton, N. Hay, J. & Jannedy, S. (2002) Probabilistic Sociolinguistics: Beyond Variable Rules. In *Probabilistic Linguistics*, Bod, R. Hay, J. & Jannedy, S. (eds). MIT Press.

Jannedy, S. & Tanenblatt, M. (2001) Is it Live? Or is it a Machine? Making TTS sound more Human. *Speech Technology Magazine.*Vol.6−4, pp. 30-32

Jannedy, S. (1999) with Hay, J. & Mendoza-Denton, N.Ȯprah and /ay/: Lexical Frequency, Referee Design and Style. In *Proceedings of the International Congress of Phonetic Sciences* (ICPhS) '99 in San Francisco, August 1-7, 1999.

Jannedy, S. (1997) Development of Narrow Focus Prosody. In *Proceedings of the GALA'97 conference: Language Acquisition, Knowledge Representation & Processing.*

Jannedy, S. & Möbius, B. (1997) Name Pronunciation in German Text-to-Speech Synthesis. In *Proceedings of the Fifth Conference on Applied Natural Language Processing*, Association for Computational Linguisitics, pp. 49–56.

Jannedy, S. (1995) Gestural Phasing as an Explanation for Vowel Devoicing in Turkish. In Jannedy, S. (ed.) *Papers from the Linguistics Laboratory.* Ohio State University Working Papers in Linguistics 45, pp. 56-84, Department of Linguistics, The Ohio State University.

Jannedy, S. (ed.) (1995) *Papers from the Linguistics Laboratory*, Ohio State University Working Papers in Linguistics, Vol. 45.

Jannedy, S. (1995) Fast Speech Reduction in German. In Proceedings of *CONSOLE II*, Eckardt, R. & van Geenhoven, V. (eds.) pp. 68–79. Holland Academic Graphics, The Hague.

Jannedy, S. Poletto, R. & Weldon, T. (1994) *Language Files - Materials for an Introduction to Language and Linguistics.* 6th Edition. Ohio State University Press, Ohio State University.

Jannedy, S. (1994) Prosodic and Segmental Influences on High Vowel Devoicing in Turkish. In *Proceedings of the Fifth Australian Conference on Speech Science and Technology*, Volume II, pp. 674–679. Perth, December, 1994.

FIELDS OF STUDY

Major field: Linguistics

Specialization: Phonetics

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

This dissertation investigates the contribution of four different intonation patterns to a listener's interpretation of the speaker's discourse intent in uttering a sentence within a monologue. Discourse intents for which intonation has been shown to be relevant include such dialogue functions as posing a yes-no question by using a low-rising final contour, and attempting to hold the floor by using a falling-rising intonation (termed a continuation rise) if the speakers is not yet done with his/her turn. Discourse intents also include functions related to the notion of focus of attention, such as the introduction of new discourse entities, or the highlighting of old entities in a new discourse role. When intonation facilitates the recovery of these sorts of attention-related intents, it is useful to use the term "prosodic focus". More specifically, prosodic focus refers to the ways in which speakers use acoustic means such as fundamental frequency (F0) modulation to signal prominence for the purpose of making a word or larger constituent more salient in the discourse context.

In studying prosodic focus, there are two over-arching questions that are more effectively addressed if they are addressed together. One question pertains to the mechanics of how the speaker imparts prominences to some parts of an utterance but not to others, while the other question addresses the listener's interpretation of such prominences - i.e., the function of prosodic focus from the listener's point of view. A fundamental assumption in posing the first question is that the speaker has various

methods at his/her disposal to make some part of an utterance prosodically more prominent than other parts. In English and languages like English, for example, one means of making a particular word more prominent than surrounding words is to align a pitch accent - a prominence lending tonal morpheme - with the syllable in a word that bears primary stress. Most current accounts of prosodic focus in English recognize this mechanism of putting a constituent in prosodic focus, and in one particularly influential account, due to Selkirk (1984, 1995), this is the only mechanism recognized. Other accounts, however, suggest that other aspects of the tune also may play a role in imparting prominence. For example, the accented word that is the last accented material in its phrase is also aligned to another tonal morpheme, the phrase accent, which is simultaneously aligned to the end of the phrase as well. When it is followed immediately by the phrase accent, a pitch accent becomes the "nuclear accent" in its phrase and in the account of Pierrehumbert (1980) and her colleagues (e.g., Beckman & Pierrehumbert, 1986; Beckman & Edwards, 1994), any nuclear accent is more prominent than all earlier, non-nuclear accents. (This is related to Ladd's, 1980, notion of "deaccenting", which says that an accented word can be made prominent if all following material is left unaccented, effectively positioning the nuclear accented word early in its phrase).

An equally fundamental assumption underlying the second question is that speakers use prosody and prosodic focus to facilitate and guide the hearer's understanding and comprehension of the message being conveyed at any particular time in a discourse. Thus, one of the uses of intonation is to guide the listener's interpretation of the utterance in relationship to the larger discourse context. Different intonational structures, then, are used to distinguish one discourse purpose, one extension of the current discourse state, from other possible moves in the mutual building of the discourse structure by the speaker and hearer. This function of intonation makes it

difficult to test claims that two or more intonation patterns differ categorically. This differs markedly from claims about the number of tones in contrast in languages such as Cantonese, which can be tested by seeing whether the tune distinguishes one word from any other word that could have occurred in the same place. Listeners are generally very good at identifying which of two minimally contrasting words they heard. They are generally much less facile at identifying different discourse intentions, unless the differences also trigger a difference in truth conditions. One of the challenges for psycholinguistics, therefore, is to devise tasks that tap the listener's competence in interpreting the intended discourse purpose. In studying the functions of prosodic focus, for example, the psycholinguist must find an experimental design that can be used to determine how exactly different prosodic manipulations contribute to the introduction of new entities or highlighting of old entities in the interpretation of the discourse purpose of an utterance.

1.1 The Experiments

In this thesis, three experiments are described that test naive listeners' interpretations of the following intonation contours, associated with simple SVO sentences such as *Benjamin heard shots*:

(1.1)	H*	H*	L-L%	(hat pattern)
(1.2)	H*		L-L%	(early nuclear accent)
(1.3)	H* L-	H*	L-L%	(double nuclear accents)
(1.4)	H*	L+H*	L-L%	(rising nuclear accent)

In all of these contours, the subject is accented, with a simple H* accent. In the hat pattern in (1.1), the very next tone also is a H* accent, placed on the object. The

subject and predicate are grouped into one intonational phrase, ending with a fall to a sequence of L- phrase accent and L% boundary tone. This makes the object accent the nuclear accent. The accent on the subject is then a pre-nuclear one. In the early nuclear accent pattern in (1.2), the subject and predicate are again grouped together into one phrase, but in this contour the L- phrase accent immediately follows the subject H*, making it the nuclear accent, with the object being deaccented. In the double nuclear accents pattern in (1.3), the subject and predicate are separated into two intermediate-level intonation phrases. That is, each of two H* accents is followed by a L- phrase accent, making both the subject and the object bear nuclear accents. The rising nuclear accent pattern in (1.4) is metrically identical to the hat pattern in (1.1) and differs only in the type of pitch accent on the object. Here it is a rising L+H* accent, and not a simple H* tone.

Given these materials, the following closely related questions can be addressed. First, what is the effect of deaccenting the verb phrase? The early peak pattern can be compared to the other contours, all of which differ from the pattern in (1.1) in having a second accent on the object. Accounts of prosodic focus in English and other similar languages agree that an early accent followed by deaccented material can only be interpreted as narrow focus of attention on the accented word. In our experiments, then, we can use the early peak pattern in (1.1) as a control to evaluate the experimental paradigm. If we get results that suggest that this pattern is no less compatible with a broad focus interpretation than the other three, then the task is not sensitive enough for our purposes.

Second, what is the function of the phrase accent? As noted above, in the phonological model of intonation assumed here (Pierrehumbert, 1980), English has a structural distinction between nuclear accents (which immediately precede the phrase accents) and non-nuclear accents. Comparing the double nuclear accents pattern in

(1.3) to the hat pattern in (1.1) can tell us something about the interpretive significance of this metrical difference. Experimental studies of the interpretation of verb phrase accentuation patterns by Gussenhoven (1983), Birch & Clifton (1995), and Welby (2001) suggest that there is an accompanying difference in interpretation. A pre-nuclear accent on the verb in a VO sequence can be interpreted as non-focal, whereas a nuclear accent cannot. That is, the hat pattern is compatible with a broad focus pattern that makes both the verb and the object new information, but it does not preclude an interpretation of narrow focus on the object, with the verb placed in the background as old information. By contrast, a double nuclear accents pattern, with a sequence of H* L- aligned both to the verb and to the object, seems to be compatible only with an interpretation in which both the object and the verb are in focus. We want to know whether there are comparable differences in the interpretation of pre-nuclear versus nuclear accents on subjects. One possible prediction is that the hat pattern in (1.1) will be compatible both with narrow focus on the VP and with broad focus on the sentence as a whole, but that the double nuclear pattern in (1.3) will be compatible only with an interpretation that puts the subject in focus, either as part of a broad focus pattern or as its own separate pocket of new information in a double focus construction.

Third, if we find the predicted difference between the double nuclear pattern in (1.3) and the hat pattern in (1.2), is this difference in interpretation due to the structural difference between these utterance types (two separate phrases each with its own nuclear accent versus one phrase with a pre-nuclear accent followed by a nuclear accent), or it is simply an artifact of the fall-rise shape of the double nuclear contour?

Gussenhoven (1983) noted that the predicted differences in interpretation of accented versus unaccented verbs in phrases containing a verb plus a non-object complement "depend on the extent to which the presence or absence of the pre-final accent can be perceived" (p. 156). Beckman (1996) made a similar suggestion regarding the interpretation of hat patterns with respect to subject focus. Specifically, she suggested that the rise to the pre-nuclear H* is less salient than the fall from the nuclear H* to the following L-, and that the location of pre-nuclear accents therefore might be difficult to parse, unless there is a following fall, as in a H* L- H*, or H* L+H* sequence.

Therefore, if we find the predicted difference between the hat pattern and the double nuclear pattern, we cannot conclude immediately that this is a difference in meaning between the two structures. It may be just an instance of phonetic ambiguity for the hat pattern. Could it be simply that the accent on the subject is easier to hear when there is a fall separating it from the accent on the object? To answer this question, we can compare the responses to the double nuclear accent pattern (H* L- H* L- L%) and the rising nuclear accent pattern (H* L+H* L-L%) . If responses to these two differ, with the double nuclear pattern receiving more double focus interpretations, then we can preclude the alternative phonetic explanation of the difference between the double nuclear pattern and the hat pattern.

Finally, are all nuclear accents alike? English has a distinction between accent types that explicitly evoke a scale or partially ordered set (L+H*, L*+H) and accents that are less restricted (H*) (see Pierrehumbert & Hirschberg, 1990; Ward & Hirschberg, 1985). The nuclear accent in pattern (1.4) is of the former type, whereas the nuclear accent in the hat pattern in (1.1) is not. Intuitively, it seems that an accent that evokes a partially ordered set is more likely to be interpreted as late narrow

focus. That is, it seems easier to recover the intended partial ordering, if the ordering is an ordering over a homogeneous set of elements which are all simple semantic types, such as discourse entities that are all potential patients of some named action, or one-place properties that all might plausibly be predicated of some named agent. Placing a L+H* accent on a particular word in a sentence thus seems more likely to be interpreted as picking out the word itself or some only slightly larger constituent, something small enough to denote a simple semantic type. For example, if *Benjamin heard shots* is uttered in the context of a story of a walk with friends in the hills of southeast Ohio, and it is produced with the rising nuclear accent pattern in (1.4), we can imagine a set of discourse elements such as {{nothing}, {shots}, {birdsong}, {other hikers' voices}, {shots and birdsong}, {birdsong and other hikers' voices}, {shots, birdsong, and other hikers' voices}, ... } – the set of (sets of) things that might be heard on a walk in the bush. We can also easily imagine a partial ordering, from the smallest number of things (the null set) to the largest (everything that one might hear). It is more difficult to imagine what the scalable set is if the L+H* were interpreted as focusing on the complex proposition denoted by the sentence as a whole. Thus, L+H* seems more compatible with narrow focus interpretations, and we might predict that, by comparison to H* H* L-L% and H* L- H* L- L%, the H* L+H* L-L% pattern should receive more late narrow focus responses. The prosodic focus is in a sense a more localized prominence.

To jump ahead to our results, then, all four of these predictions were borne out to some extent in our data. Thus, a major contribution of this dissertation was to partially answer the first question about prosodic focus posed above. In addition to the expected difference between the early nuclear accent pattern and the other three types with object not deaccented, we also found more subtle prosodic focus effects. In particular, we found that intonational phrasing contributes to prosodic focus by

making the metrical difference between nuclear and pre-nuclear pitch accents, and we found that this was not an artifact of the greater phonetic salience of the subject accent when it was separated from the object accent by a fall-rise. We also found that L+H* marked a more local prosodic focus than H*, other things being equal. This result is in keeping with Pierrehumbert & Hirschberg's (1990, p. 286) observation that a "common use of L+H* ... is to mark a correction or contrast". It is also in keeping with Ayer's (1996) result showing that listeners are quicker to respond to words with L+H* accents in phoneme monitoring tasks.

A second contribution of this dissertation is a methodological one. Previous experimental work on the contribution of prosody to the interpretation of focus has concentrated on matching question-answer pairs in a dialogue (Gussenhoven, 1983; Birch & Clifton, 1995; and Welby, 2001, for English; Jin, 1996 for Mandarin Chinese; D'Imperio-Piterman, 2000 for Neapolitan Italian). The (usually unstated) assumption behind this method is that we can generalize from the relationship between an answer by one speaker and an explicit question from the other speaker in a dialogue, to the relationship between a statement by one speaker and (his/her understanding) of an implicit question in the other speaker's immediately preceding turn. However, we know that discourse is structured for the listener not just in dialogue exchanges, but also in monologue narratives. Recent work on discourse structure suggests that monologues also can be modeled as a sequence of question-answer pairs (Roberts, 1996; Büring, 2000, 2001), with the difference from dialogues that the questions are almost always implicit and are supplied by the same speaker who answers them. Two types of tasks were used in the three experiments described in this dissertation to extend the previously deployed experimental methods to such monologue narratives.

The first task, which was used in the first experiment, adapts the previous methodology of matching answers to questions by presenting a context monologue ending in

the stimulus sentence to the listener, and then requiring the listener to select the question (from among four alternatives) that was best addressed by the target utterance. The target sentence is a simple SOV sequence produced with one of the four intonation patterns listed above in (1.1)-(1.4), and the list of questions are all WH-questions which focus on the sentence as a whole (*What happened?*) or on the subject (*Who heard shots?*), the object (*What did Benjamin hear?*), or both (*Who heard what?*). In essence, the listener matches the focus pattern of the last sentence uttered before the narrative is interrupted with a possible "question under discussion" at that point in the discourse. Since the prior context is identical, and the words in the target sentence are the same for the four utterances in any given item set, any differences in response patterns across the four utterances must be due to the intonation contour and the question that is implicitly invoked by the speaker in uttering the sentence with that prosodic focus pattern.

This paradigm allows us to set a baseline of responses for the second paradigm, which was tested and refined in the second and third experiments. Here, we simulate the conversational environment more closely, by tapping the listener's ability to predict what will come next in the narrative, given the informational structure that the speaker has set up. That is, in listening to a coherent narrative, the competent native listener should not only be able to recover and accommodate to the sequence of "questions under discussion" that the speaker has implicitly set up, but also to anticipate the next question to some extent. The task then is one that requires a listener to anticipate or make forward interpretations as a speaker is setting up the context and the moves (implicit questions). After hearing the interrupted narrative, the listener is then required to select the best/most appropriate continuation to the "story" just heard.

In both of these paradigms, the information shared by the speaker and the listener (mutual belief space or common ground) is provided solely by the speaker. In all experiments, listeners were asked not only to pay attention to what was said but also to how the target sentence was rendered. Since the stimulus utterances varied in terms of the accent location, accent type, and phonological phrasing, this dissertation addresses the relative contributions of these variations to the interpretation of simple subject-verb-object (SVO) utterances in American English.

1.2 Organization

The organization of the dissertation is as follows: Chapter Two gives an overview of the experimental literature addressing how prosodic focus is interpreted, while Chapter Three gives a detailed phonetic account of the stimuli used in this study. Chapter Four introduces the concept of Focus as used in the semantics/pragmatics literature, and the relevance of congruence between question and assertions. Chapter Five describes the methods and results of the experiment in which listeners were required to find the local question under discussion, while Chapter Six describes the methods and results of the novel continuation paradigm. Chapter Seven provides some concluding remarks.

CHAPTER 2

BROAD AND NARROW FOCUS

2.1 Accenting & De-Accenting

This dissertation explores two of the less well-studied aspects of prosodic focus in English. One has to do with the contrasting choices of intonational phrasing and the other with the choice of pitch accent type. To place these phenomena in context, we will begin by reviewing the better-studied aspects of prosodic focus, having to do with the location of the most prominent accents in an utterance, and introduce the concepts and some technical terms that we will be referring to throughout this dissertation. We will illustrate these concepts using the contrasting utterances of the sentence *Benjamin served lemonade.* shown in Figures 2.1 and 2.2.

It has long been noted that when a sentence such as *Benjamin served lemonade.* is produced out of the blue, or in response to a very general question such as *What happened?* or *Tell me about the block party.* there often is a prominent fall in pitch from the stressed syllable of the object noun. This is illustrated by the fundamental frequency (F0) contour for such an utterance of this sentence in Figure 2.1. However, if the same prominent fall in pitch occurs earlier, on the subject, and the verb and object noun are produced with low flat pitch, as in Figure 2.2, then the utterance suggests a more particular question such as *Who served lemonade?.* That is, an

utterance with this contour an F0 fall early is more felicitous in response to a question that presupposes that someone served lemonade.

The standard understanding of these facts is that the different locations of the prominent fall in the two utterances is due to a difference in the "accentuation" . By accentuation we mean the alignment of 'pitch accents' - prominence lending tonal morphemes - to one or more stressed syllables in the sentence. A word containing a syllable that is associated to a pitch accent is then accented. Thus in Figure 2.1, the object noun *lemonade* is accented (the stressed syllable in the word bears a pitch accent), whereas in Figure 2.2, the object is "de-accented".

Occasionally throughout this thesis, we will indicate such accentuation by capitalization of the entire accented word, as in (2.1) and (2.2), even though pitch accents are aligned with the lexically stressed syllable of the word they fall on. This convention of capitalization is used to refer to the link between accentuation and prosodic focus without committing to any particular model of intonational phonology. This is useful to do in cases such as the utterances in Fig. (2.1) vesus Fig. (2.2) because different researchers disagree about how to analyze the tune. For example, Gussenhoven (1983) and Selkirk (1995) analyze the fall as a single event, a H*L pitch accent, whereas Pierrehumbert (1980) and Selkirk (1984) analyze it as a sequence of events: a H* pitch accent, associated with the stressed syllable and a following L phrase accent that has its primary association to the phrase end. Despite such disagreement, however, all researchers agree that this intonation contour makes *lemonade* accented in Fig. 2.1 as contrasted with *Benjamin* accented but following material deaccented in Fig. 2.2.

(2.1) Benjamin served LEMONADE.

(2.2) BENJAMIN served lemonade.

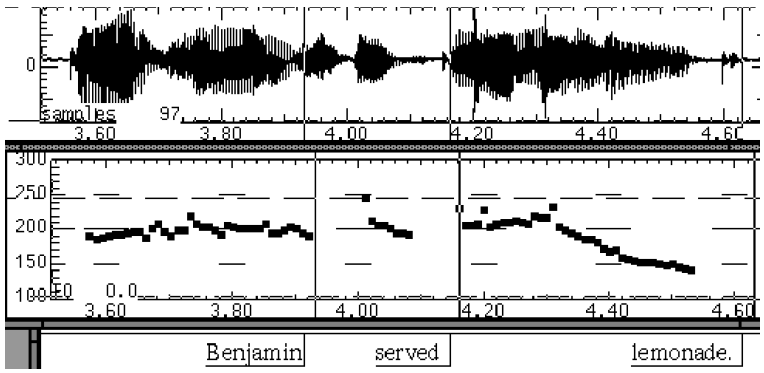

Figure 2.1: Accented Object in *Benjamin served LEMONADE.*

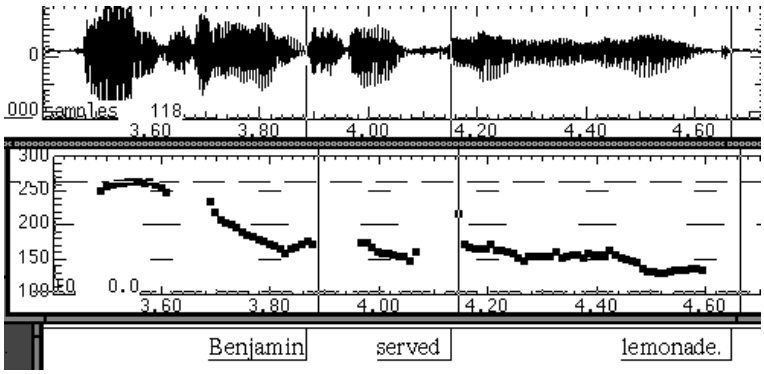

Figure 2.2: Object deaccented in *BENJAMIN served lemonade.*

In the example in Figure 2.2, where the prominent fall in F0 is on the subject *Benjamin* and all following material is deaccented, the link between accentuation and prosodic focus is especially clear: there is only one pitch accent and that pitch accent signals focus only on the word that is accented. If focus is restricted in this way, we refer to it as "narrow focus". The fall in the F0 indicates that the scope of the focus is not to go beyond the entity denoted by the word *Benjamin*. Thus, the prosodic structure partitions this sentence into information that is presupposed, that is known in the context, and information that is the focus of the utterance. A Wh-question such as *Who served lemonade?* already presupposes that there was lemonade, and that it was served.

By contrast, the case illustrated in Figure 2.1, where the last pitch accent falls on the first syllable of the word *lemonade*, is more difficult to describe and to understand because the status of the material before the fall in the F0 is unclear. In the model I am assuming to describe intonation contours (Pierrehumbert, 1980), there is also an accent on the subject *Benjamin*. This early accent however, is a structurally different type of accent since it is not the last accent within the domain. Following Pierrehumbert (1980), and many others, we will refer to the last accent as the "nucelar accent" and he earlier accent on *Benjamin* as a "prenuclear accent". Beckman (1996) suggests that when linguists such as Chomsky (1972), Selkirk (1984), and Rochemont & Culicover (1990) refer to "the accent" in an utterance such as Fig. 2.1, they are referring specifically to nuclear accents and ignoring any prenuclear accents. Thus, in (2.2) we have indicated the nuclear accent with capitalization but left the prenuclear accent unmarked.

Experimental evidence reported in Gussenhoven (1983), Birch & Clifton (1995), and Welby (2001) indicates that the scope of focus can be quite unclear when there is a nuclear accent late in the utterance. The focus can be just the last word (late

narrow focus), the VP (VP-broad focus) or the entire sentence (broad focus). That is, there is some degree of ambiguity with regard to what the focus is in this case. Due to the phenomenon of *focus projection*, described by Selkirk (1984, 1995) and others, a late accent on the object can project focus to the entire sentence, and it has even been claimed e.g. by Chomsky, (1972), Ladd (1980) and Selkirk (1984, 1995) that a late narrow focus is indistinguishable phonologically from a broad focus, that both are marked by a single late accent. (Beckman, 1996, on the other hand, suggests that this "lack of distinctiveness" is due only to the lesser phonetic salience of the pre-nuclear H* in the hat pattern, and that it is quite implausible to suppose that a sentence with a full referring expression as it subject would not also have a pre-nuclear accent on the subject, a conjecture that is congruent with Nakatani's (1997) corpus study of the distribution of accents and pronouns in subject and object position.)

Therefore, our example sentence with a late nuclear accent on the object, can satisfy any of the three questions in (2.3)-(2.5), wherease the example with early nuclear accent is restricted to the interpretation in (2.6). We use the bracketing notation and the 'F' ([...]F) to denote the focus domain in the examples below.

(2.3) late narrow focus

What did Benjamin serve?

Benjamin served [LEMONADE.]F

(2.4) VP-focus

What was Benjamin doing?

Benjamin [served LEMONADE.]F

(2.5) broad focus

What went on at the Block Party?

[Benjamin served LEMONADE.]F

(2.6) early narrow focus

Who served lemonade?

[BENJAMIN]F served lemonade.

In (2.3), the focus domain corresponds only to the object NP *lemonade* and we call this type of focus domain 'narrow focus'. In (2.4), the domain of the focus is less narrow; it spans the verb phrase (VP) *served lemonade.* Finally in (2.5), the domain of the focus is the entire sentence *Benjamin served lemonade.* When the entire sentence is in focus, we describe the focus domain as 'broad focus'. The terms *broad focus* and *narrow focus* were introduced by Ladd (1980), who also suggested that the distinction between narrow and broad is a continuum and not a mere dichotomy. That is, the VP focus in (2.4) is broad elative to the object focus in (2.3) but narrow relative to the whole sentence focus in (2.5).

The contrast between the accentuations in (2.1) and (2.2) suggests that, we can use Wh−questions such as the one in (2.3)-(2.6), to probe for focus. The focus domain of a sentence is the part of an utterance that answers a Wh−question. Once we know what Wh−question is being answered, we can identify the focus. Identifying the focus domain is easy if the questions are posed explicitly, as in our example above. In most instances, however, such questions are not posed explicitly, but are implicit: communicative interactions between speakers involves the calculation of implicitly posed questions. Because prosodic focus involves the alignment of prominence lending events in the tune to focused elements in the sentence, another, less direct probe for the focus domain is judgements of prominence or stress.

The prominent early fall in the F0 in Figure (2.2) indicates the focused information in this context, and the presupposed background is produced with lesser prominence. This idea is well supported in the experimental literature: Listeners are very good at recovering the early narrow focus question that is felicitous for the pattern with an early accent peak followed by substantial de-accented material (Gussenhoven, 1983; Birch & Clifton, 1995) and they reliably judge a word that has early narrow focus to be more promient than a following word (Gussenhoven, 1983; Krahmer & Swerts, 2001). Prominence judgements are less robust with regard to a late nuclear accent. A question to ask, then is how this fact might be related to the purported ambiguity of focus domains. Are prenuclear accents different from nuclear accents? We will now turn to the experimental literature addressing this issue of nuclear and prenuclear accents (and how their processing has been studied).

2.2 Listening to Early Nuclear Accents

Gussenhoven (1983) conducted a context retrievability task in which he evaluated listeners focus interpretations. He presented listeners with a series of question-answer or context-reply combinations in which the answer/reply either had been elicited in response to that question or as a reply in another context (Question 1 matched with Reply 2 or Question 2 matched with Reply 1). The types of sentences he tested includes the responses shown in (2.7). Note that in R2, the last accent is early in the utterance.

(2.7) Structure C Stimuli:

Dialogue One ([+focus] on object):

Q1: Anything in particular you want me to say to her?

R1: Tell her I'm planning another COUP.

Dialogue Two ([-focus] on object):

Q2: She seemed sort of sorry there'd never be another coup.

R2: Tell her I'm PLANNING another coup.

Results indicated that listeners were able to determine whether the context question and the reply came from the same or from different dialogues. The question in Dialogue One sets up a broad focus context while the context in Dialogue 2 sets up an early narrow focus on the verb *planning* since the object *coup* already is in the context. Thus, listeners were able to infer the early narrow focus context (Context 2) from the early nuclear accent placement in Reply 2.

Birch & Clifton (1995) obtained appropriateness ratings for three types of utterances, in the context of two types of questions. The questions either did or did not explicitly refer to the NP which would be argument of the verb in the response. We assume 1 to ask for a broad focus since it invokes a more general question such as *Remind me of the evidence that she is smart. For example, what does she do?*, querying the predicate as a whole, while Q2 presupposes the NP and therefore queries the value of the V. However, a felicitous reply to Q2 could also be *Yes, she is an ACCOUNTANT.* with the NP in focus. Thus, it is not just the presence of the NP in Q2 that results in the query of the verb.

(2.8)

Questions:

Q1) Isn't Kerry pretty smart?

Q2) Isn't Kerry good at math?

Answers:

R1) Yes, she TEACHES MATH.

R2) Yes, she teaches MATH.

R3) Yes, she TEACHES math.

The reply sentences either had accents on the verb and object, only on the object, or only on the verb. The appropriateness rating of Q2 matched with R3 was highest, and the rating for Q1 matched with R3 was lowest, indicating that listeners were able to recover the more appropriate early narrow focus context for the reply sentence that contained a single early accent. This result fits well with Gussenhoven's findings for his stimulus sentences.

Birch & Clifton (1995) described the prosodic focus in their stimuli by having a trained phonetician transcribe the accent patterns. However, they only tested the focus interpretation by naive subjects, and did not examine the perceived prominence directly. By contrast, Gussenhoven did both. He instructed listeners to rate the level of stress on the object in sentences such as Answer 1 and Reply 2 in (2.7). He found a robust difference in the perceived stress between these two conditions, the biggest effect was on the perceived stress of the object. It got substantially lower stress ratings when the preceding verb was in narrow focus in the structure C type sentences, indicating that listeners can perceive a difference between the two conditions.

The , the biggest effect was on the perceived stress of the object was on average 1.2 points higher (on a five-point scale) in the Answer 1 type utterances than in the Reply 2 types utterances, and the degree of stress. It got substantially lower stress ratings when the preceding verb was in narrow focus in the structure C type sentences, indicating that listeners can perceive a difference between the two conditions.

Krahmer & Swerts (2001) got similar results in their study of elicited spontaneous phrases. They gathered ratings of "prominence" for adjective-noun phrases (such as *red square* versus *BLUE square*). These phrases were elicited from several pairs of speakers within a dialogue game that prompted four different focus patterns: (1) broad focus on the phrase as a whole at the beginning of the game, (2) narrow focus on the adjective when a move involved a piece with a different color from the piece in the preceding move, (3) narrow focus on the noun when the piece in the current move differed in shape but not color from the previous move, and (4) double narrow focus on both the adjective and noun when the piece involved both a new color and a new shape. They also had two trained phoneticians judge the distribution of accents within the phrase.

In the rating task (a pairwise comparison of adjective-noun pairs with contrasting focus patterns) they found that the noun was reliably judged to be substantially less prominent (in comparison to other focus patterns) when the context exlicted early narrow focus on the adjective. This was also the context where the two phoneticians reliably judged the adjective to be accented, and the noun to be deaccented.

All of these results together confirm the introspective descriptions of "sentence stress" or "the accent" in the prior literature, such as Chomsky (1972) and Selkirk (1984). That is, when an utterance is transcribed as having an early accent and following deaccented material, listeners reliably interpret the utterance as appropriate for a context eliciting early narrow focus and they reliably perceive the deaccented material as extremly non-prominent.

2.3 Listening to Late Nuclear Accents

In contrast to the evidence we have about the interpretation of early nuclear accents, the facts and our understanding of late nuclear accents are more controversial. In addition to the interpretation of late narrow focus, a late nuclear accent also can be associated with a broad focus interpretation. A third possibility that is often overlooked is that of double narrow focus. The same literature that shows robust and consistent responses to sentences with early nuclear accent and following material deaccented shows less robust and seemingly contradictory results for sentences with late nuclear accents.

For example, Gussenhoven (1983) tested two other types of sentences in his Context Retrievability task. These are shown in (2.9) and (2.10). He tested sentences with an early nuclear accent in (2.7) in large part as a control for interpreting responses to sentences with late nuclear accent and no accent on the verb in the Dialogue Two sentences in (2.9) and (2.10). In the R2 conditions (or [-focus] conditions, as Gussenhoven refers to them), there was only a late nuclear accent on the object. In the R1 instances (or [+focus], in Gussenhoven's terms), there was a prenuclear accent on the verb, and a nuclear accent on the object or complement of the sentence.

(2.9) Structure A Stimuli

Dialogue One ([+focus] on verb):

Q1. What is the nature of your business?

R1. We repair RADIOS.

Dialogue Two ([-focus] on verb):

Q2. What is it you repair?

R2. We repair RADIOS.

(2.10) Structure B Stimuli

Dialogue One ([+focus] on verb):

Q1. Please tell me what happened that night.

R1. I remember NOTHING.

Dialogue Two ([-focus] on verb):

Q2. What do you remember from the last lesson?

R2. I remember NOTHING.

He also varied the number of syllables between the stressed syllable in the verb and the stressed syllable in the direct object or complement to see if there is an interaction with the length of the utterance. His hypothesis was that with an increasing number of syllables between the prenuclear and the nuclear accent, the Context Retrievability Score (CRS) should improve since listeners would be more likely to perceive the presence or absence of a prominence-lending rise on the verb.

More intervening syllables would result in an F0 shape like the one shown in the stylized Figure 2.3, while fewer or no interevening syllables would resemble an F0 pattern like the one in Figure 2.4, where it is not clear whether the rise is located on the verb or is due to the accent on the object.

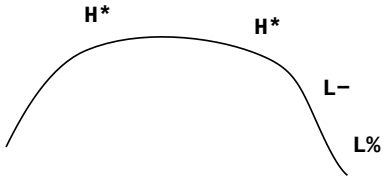

Figure 2.3: Schematized contour for a prenuclear accent and a late accent as in Gussenhoven's Reply 2.

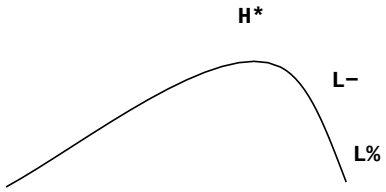

Figure 2.4: Schematized contour for a late accent on the object as in Gussenhoven's Reply 1.

Beckman (1996) suggests that prenuclear accents in a flat hat pattern are less salient and hence that it would be difficult for listeners to tell if material before the final fall in the F0 at the end of the sentence is accented or not. (This is the background for her further suggestion, described above, that when linguists such as Selkirk, 1984, refer to "the accent" in an utterance, they are thinking of nuclear accents, and not pre-nuclear ones, Selkirk's own claim to the contrary.) Gussenhoven describes the [-focus] utterances (that only have an accent on the object) as 'not display[ing] a clear step-up in pitch on the verb' while the [+focus] utterances did

display such a step up in pitch. Thus, we interpret the pattern in terms of the tunes transcribed above the schematic F0 contours in Figures 2.3 and 2.4.. Since these two tunes differ phonetically primarily in the location of the rise in pitch to the first H* accent, and that first H* accent can be either a pre-nuclear accent on the verb or the nuclear accent on the object, there is a potential for phonetic ambiguity. The closer together the two potential loci for the first accent are, the harder it is to parse where the accent is.

The trial with a pitch step up on the verb would strongly resemble the trial materials with no pitch step up on the verb and little intervening segmental material before the nuclear accent on the object would be difficult to distinguish from the trial items with no pitch step up on the verb.

The Structure A sentences as in (2.9) consisted of a Pronoun Subject + Verb + direct Object (*We repair radios.*) which Gussenhoven refers to as a "merging predicate". The Structure B sentences, as in (2.10) consisted of a Pronoun Subject + Verb + Complement (*I remember nothing.*), which Gussenhoven describes as being a "non-merging predicate". In terms of the overall F0 shape of these two types of structures, however, they are similar. The [+focus] utterances with accents on both the verb and object should, with additional segmental material intervening, resemble the hat pattern shown in Figure (2.3), while the same types of utterances with much less or no intervening segmental material should be harder to distinguish from the stylized contour shown in Figure (2.4). Utterances with accent on the object only, should also resemble the stylized F0 contour shown in Figure (2.4). Thus, there is phonetic ambiguity between the utterances with a prenuclear and a nuclear accent when there is no or very little intervening segmental material, and the utterances where there is only one late nuclear accent to begin with. Gussenhoven tested whether

both of the Structure A and Structure B utterances are compatible with a late narrow focus and a broad focus interpretation.

Gussenhoven's results indicate that listeners performed no better than chance for the Structure A type sentences, when judging whether the questions and answers were correctly matched, regardless of the number of intervening syllables. Gussenhoven interprets this result as supporting his hypothesis that the verb and direct object 'merge' into a single focus domain, which is totally in focus whether there is a single accent on the object or two accents on the verb and the object.

On the other hand, since the number of intervening syllables was never more than three, this result might still be in keeping with Beckman's (1996) suggestion that there is a chronic phonetic ambiguity between flat hat and single late nuclear accents, and that this is the source of the long-noted ambiguity of focus interpretation for utterances with late nuclear accent. That is, on the basis of the Structure A sentences, we do not know whether it is just the difficulty of the phonological parse that is operative in the focus ambiguity, or whether there are in fact two possible focus interpretations for utterances with the intonation contours transcribed in Figures 2.3 and 2.4.

Gussenhoven's results for the Structure B sentences, however, call Beckman's (1996) suggestion that material before the final fall of a nuclear accent is less salient, into question. Listeners clearly can hear a difference between a structure with a prenuclear and a nuclear accent versus one with just a nuclear accent if there is enough room for the step up in pitch to be distinguished, and they respond differently, if the material following the verb is not a direct object but a PP or another comparable non-argument complement: in the Structure B sentences, R1 also contained a prenuclear accent on the verb and a nuclear accent on the object, while R2 again only contained a nuclear accent on the object NP. For these sentences, Gussenhoven does find an

interaction between the context retrievability score and the number of intervening syllables: With increasing number of syllables between the two accents, the context retrievability score improves.

If the improvement was only due the amount of intervening segmental material, we should have also found this effect for the Structure A sentences. Instead responses were at chance regardless of the number of intervening syllables. Gussenhoven concludes that the difference between having a direct object in the structure A type sentences ("merging predicate") and prepositional phrases or other non-argument complements in the Structure B sentences ("non-merging predicates") accounts for these findings. That is, Beckman's suggestion does not hold because the complement structure of the sentence also influence the processing of the tonal structure.

In the 'degree of stress judgement task' the verb itself was rated for Structure A and Structure B type sentences. Gussenhoven found differences between the answers elicted in response to broad VP focus questions and those elicited in response to questions putting narrow focus on the object. The verb was perceived as less stressed in the latter case, in keeping with Gussenhoven's claim that he took care to make sure that there was a step up to a prenuclear accent in both structures in the broad focus cases but not in the narrow focus on object cases. (The stress judgments tended to be higher in both cases for the Structure B types sentences, but the difference in mean degree of stress between the [+focus] and [-focus] conditions for the two syntactic structure types was the same.)

Gussenhoven also makes the important point that the perceived degree of stress did not correlate with the ability to recover the context question in the structure A cases, supporting his claim that the accent on the verb here is not focal, and that focus interpretation is not in a simple one-to-one mapping to accent placement. Gussenhoven's claim gets further support from Krahmer & Swerts's (2001) finding

in their ratings of "prominence" task, that the adjective was not substantially and reliably less prominent when there was narrow focus on the following noun. That is, judgments of high prominence on the adjective (which corresponded strongly to the phonetician's transcriptions of accents) were not limited to the utterances elicited in contexts that put focus on the color as well as on the shape of the piece.

Birch & Clifton's (1995) results, however, seem to pose a difficulty for Gussenhoven's (1983) interpretation of his results for the Structure A type sentences. That is, even though they were looking at the same type of *Pronoun Subject + Verb + Object* sentences, and comparing tunes that had low pitch on the verb with tunes that had a clear step up onto the verb (albeit with a following dip), listeners preferred the forms with a high pitch accent on the verb (R1, *Yes, she TEACHES MATH.*) in response to broad focus questions such as *Isn't Kerry pretty smart?*). That is, where Gussenhoven found that listeners could not reliably tell if an utterance with accents on both the verb and the object, and an utterance with an accent just on the object, were originally elicited in the context of a question asking for VP focus or for Object NP focus, Birch & Clifton found that listeners preferred the reply with an accent on the verb as well as an accent the object in response to such a broad focus question.

This apparent discrepancy in results between Gussenhoven's (1983) Structure A sentences and Birch & Clifton's (1995) results can be resolved by understanding the exact prosodic structure of the stimulus sentence for these two experiments. From Birch & Clifton's sample F0 contours and their description of the stimulus however, we do know that between the two accents, there was an intermediate phrase boundary, thus, rendering two intermediate phrases with two nuclear accents. We can interpret the difference to Gussenhoven's results, then, as suggesting that while a pre-nuclear accent need not mark focus, a nuclear accent does bring the accented constituent into focus.

Further support for such an interpretation of the discrepancy in the results between Gussenhoven's and Birch & Clifton's experiments is provided by two other studies: The first is a phonetic study rather than a psycholinguistic one. Eady et al. (1986) elicited sentences with the structure SVO-PP (such as *Don shot the puck to Kent.*) in response to four different Wh−questions (such as those in (2.11)) from a group of speakers, and measured stressed syllable durations and peak F0 values. Note that the fourth Wh−question contains two Wh−words, indicating that the best matching reply will contain two independent narrow focus constituents (*[DON]F shot the puck to [KENT]F.*

(2.11) Context Questions in Eady et al. (1986):

1. What happened?
2. Who shot the puck to Kent?
3. Who did Don shoot the puck to?
4. Who shot the puck to whom?

They found clear phonetic differences between broad focus and late narrow focus utterances (as well as between the broad focus and the double focus utterances on the subject and PP). Although the aim of the study was to look at phonetic differences across different focus structures, they also report a listener's interpretation of these elicited contours: One of the experimenters listened to all elicited sentences (initial, final, dual, neutral focus) in a blind test. The purpose was to identify those utterances for further study for which the listener could recover the Wh-question to which it was a response. They report that the listener was able to reliably match the utterances with the original question in 96% of all cases, even though the contexts included as many late narrow focus questions (e.g., *Who did Don shoot the puck to?* and double

focus questions (*Who shot the puck to whom?*) questions as early narrow focus (*Who shot the puck to Kent?*) and broad focus (*What happened?*).

Measurements of the peak F0 for the replies elicited as a response to the double focus question (*Who shot the puck to whom?*) indicated a dip in the F0 between the subject and the PP, while the responses to the broad focus questions did not. Moreover, Eady et al. observed that the F0 rise onto the subject peaked late in the stressed vowel in the broad focus case, in keeping with the alignment of an initial pre−nuclear H*, whereas F0 on the subject peaked early in the stressed vowel and then fell in the double focus case, in keeping with the alignment for a following L- phrase accent.

From the description of Birch & Clifton's results, we also know that the preferred broad focus response was the contour with a fall from H* to a L- phrase accent on the verb as well as one the object. And of course, this contour contained a dip in the F0 between the two high tone targets. That is, it was the contour that put a nuclear accent early as well as late in the sentence that was interpreted as clearly having focus both early and late.

When the late accent is on a PP complement, as in Eady et al. (1986), the two nuclear accents yields double focus, whereas when the late accent is on an object, as in Birch & Clifton (1995), the two nuclear accents merely precludes a single late narrow focus, but the necessarily focal nature of the early nuclear accent is the same. The accent on the verb in Gussenhoven's (1983) study, by contrast, was merely a "step up" to a pre-nuclear H*. When the following noun was an object, then, this merely pre-nuclear accent was compatible with focus on the verb phrase as a whole or with focus just on the object.

Support for this interpretation of the necessarily focal nature of the double nuclear accents in a verb-object sequence comes from a study by Welby (2001). As in the two

earlier studies, Welby presented listeners with short question-answer dialogues, such as the ones shown in (2.12), with 'answer' varying intonation pattern. The task was like the one used in Birch & Clifton's (1995) first experiment. She asked listeners to indicate the appropriateness of the match between the question and the answer on a scale from 1 through 5.

(2.12)

Q1: VP-focus Q: How do you keep up with the news?

A1: I READ the DISpatch.

Q2: Object-focus Q: What newspaper do you read?

A2: I read the DISpatch.

As the sample dialogues in (2.12) indicate, Welby tested the same simple *Subject Pronoun + Verb + Object* sentences as in Gussenhoven's Structure A type sentences and all of Birch & Clifton's stimuli. More crucially for us, she included all four of the intonation contours that were tested in either of the earlier studies. These are listed in (2.13). That is, she tested the single accent contours that both earlier studies had used. These were the contour with the early nuclear accent on the verb and the object de-accented (Gussenhoven's [-focus] condition in the Structure C type controls and Birch & Clifton's *She TEACHES math.* type sentences) and the contour with the single (nuclear) accent on the object (Gussenhoven's [-focus] condition in the Structure A type sentences and Birch & Clifton's *She teaches MATH.* type sentences). And she also tested *both* types of double accent contours: the contour with a pre-nuclear "step up" in pitch on the verb (Gussenhoven's [+focus] condition in the Structure A

type sentences) and also the contour with two nuclear accents separated by a fall to a L- phrase accent (Birch & Clifton's *She TEACHES MATH.* type sentences.)

(2.13)

	H*	L-L%	late peak contour
H*		L-L%	early peak contour
H*	H*	L-L%	hat pattern contour
H* L-	H*	L-L%	double nuclear accent contour

Her results indicate that the F0 pattern with a pitch step up on the verb and no dip before the object (Gussenhoven's [+focus] contour) is equally appropriate to the contour containing a single late accent on the object (Gussenhoven's [-focus] contour and Birch & Clifton's *She teaches MATH* contour) in responses to both the VP-focus questions and the Object-focus questions. However, the contour with two nuclear accents (Birch & Clifton's *She TEACHES MATH* type) got lower appropriateness ratings in response to the Object-focus questions. In other words, we might infer from the results reported in Gussenhoven (1983), Birch & Clifton (1995), and Welby (2000) that narrow focus on the object is not precluded by placing an accent on the verb, so long as that accent is not a nuclear accent.

Before we make that inference, however, we must first eliminate an alternative account of the difference between Gussenhoven's [+focus] stimuli and Birch & Clifton's *She TEACHES MATH* stimuli. We have described the difference between these two contours as a stuctural one, a metrical contrast between a merely pre-nuclear accent on the verb in Gussenhoven's stimuli and a (more prominent, more stressed) nuclear accent in Birch & Clifton's stimuli. However, the two contours also differ phonetically. In Gussenhoven's [+focus] stimuli, there is a step up in pitch onto the verb, but no fall-rise before the object to make the first accent salient even when there is

little intervening segmental material. In Birch & Clifton's two-accent contours, by contrast, there is a fall-rise intervening. Perhaps it is this fall-rise and not the attendant structural difference that makes the first accent necessarily focal in the double peak contour.

2.4 Summary

To understand the apparent discrepancy between Gussenhoven's results and those of Birch & Clifton (1995), we will need to describe other aspects of the intonation besides the location of the accents and the gross shape of the F0 contour. In this study, we will will be comparing 2 contous that contain dips in the F0, and we are testing the assumption that the utterance that contains a low phrase tone rather than just a dip in the F0 between a fully specified subject and an object, lends greater phonetic salience to the accent preceeding the dip in the F0. That is, we are going to show that the difference is not simply one of greater or lesser phonetic salience, but one of the status the accent, if it is nuclear and hence focal, or if it is prenuclear and hence does not have to be focal but can be rhythmic.

CHAPTER 3

THE PHONETICS OF FOCUS

3.1 Introduction

In this chapter we will describe the phonological (including metrical) structure of the materials used in the three experiments to be described in Chatpers 5 and 6. We will then examine the phonetic properties of the stimulus sentences used in these experiments in order to motivate why the particular phonological description of the intonation patterns that we are providing is needed to understand their focus interpretations.

That is, we are going to motivate a claim that not all sentences that are written with only the last word capitalized (such as the *Benjamin served LEMONADE.* depicted in Figure 2.1) necessarily have the same metrical structure, in preparation for testing our hypotheses about differences in preferred focal interpretation that go along with the claimed differences in metrical structure. Further, we are going to suggest that measures of duration or the F0 top line are not sufficient to describe prosodic focus patterns; rather, prosodic focus patterns need to be interpreted in terms of the tune, and the patterns of accentuation, as we did above in Chapter 2, when we interpreted Eady et al.'s (1986) data in terms of intonation contours similar to the ones transcribed for the stimuli in Birch & Clifton (1995) and Welby (2001). This is important in order to relate our results not just to the VP−focus versus

object−focus interpretations tested in these two psycholinguistic experiments, but also to the subject−focus and broad (whole−sentence) focus productions elicited and measured in Eady et al. (1986).

Eady et al. (1986) elicited utterances that varied with regard to their focal properties by posing various Wh−questions. They then measured stressed syllable durations and peak fundamental frequency values in the target words of these elicited utterances to provide an acoustic analysis of different focal structures. If we are able to elicit specific patterns of prominences with questions that query specific values of a proposition, then the reverse may be true also, that we should be able to recover a question based on the prominence structure of an utterance.

To show principled phonetic and hence phonological (structural) differences between the contours used as test materials in the experiments, we measured midpoint F0 values and durations of the stressed and accented vowels in subject, verb and object position for sentences such as *Benjamin served lemonade.*. In this context, it is relevant for us to show that a description of the F0 top line alone or a tonal pattern without specification of the anchor points for the tones will obscure the metrical structure that corresponds to these tonal patterns, and will prevent us from making distinctions that are necessary to understand how prosodic focus mediates the interpretation of an utterance.

3.2 Materials

The materials consisted of twenty short monologue narratives, each leading up to a particular stimulus sentence, just before the 'story' was interrupted. In Experiment 1 (described in Chapter 6), the listener then identified the discourse purpose of the stimulus in terms of the implicit question being answered by the stimulus sentence

(the "Question under Discussion"; see Chapter 5), whereas in Experiments 2 and 3 (Chapter 6), the listener chose the most appropriate continuation. However, the stimulus utterances were identical between the two types of experiments. The stimulus sentence for each story consisted of the same segmental material, with the intonation contour varying across the four patterns listed in (3.1) and illustrated in Figures 3.1 to 3.4:. (The sentence used in the figures, *Jeremy will marry Marrianna*, is not part of our corpus of twenty stimulus sentences, but it was chosen to illustrate the contours, since it consists primarily of sonorants, for which the pitch contour is more easily tracked and explained.) (3.1)

1.	H*	H*	L-L%	(hat pattern)
2.	H*		L-L%	(early nuclear accent)
3.	H* L-	H*	L-L%	(double nuclear accents)
4.	H*	L+H*	L-L%	(rising nuclear accent)

The first contour is called a "hat pattern". Described in terms of the theory of intonational phonology that we are assuming here, this contour has two high tone targets, one on the stressed syllable in the subject and one on stressed syllable in the object, between which there is a simple interpolation, resulting in an F0 plateau between the very high F0 on the first accent and the prominent fall in pitch at the nuclear accent. Also, this contour contains only a single nuclear accent on the object since it consist of only a single intermediate phrase, making the accent on the subject a pre-nuclear one.

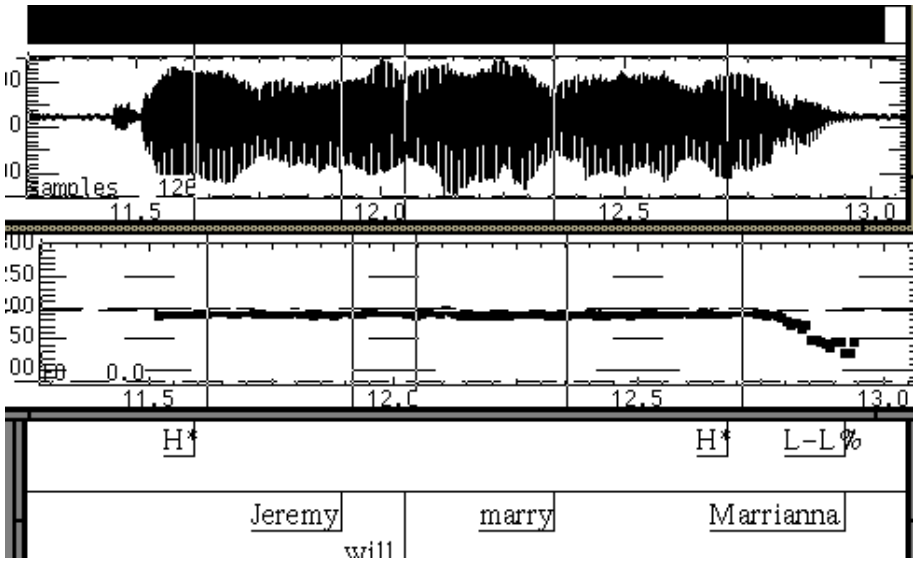

Figure 3.1: Hat pattern (H* H* L-L%).

The second contour (which we will call the "early nuclear accent" pattern) also consists of a single intermediate phrase. However, it has only one accent, rather than the two of the hat pattern. That is, Itit has an early high tone target on the subject and then the F0 falls steeply to the low phrase accent and stays low and level untilit breaks into creaky voice for the low boundary tone at the end.

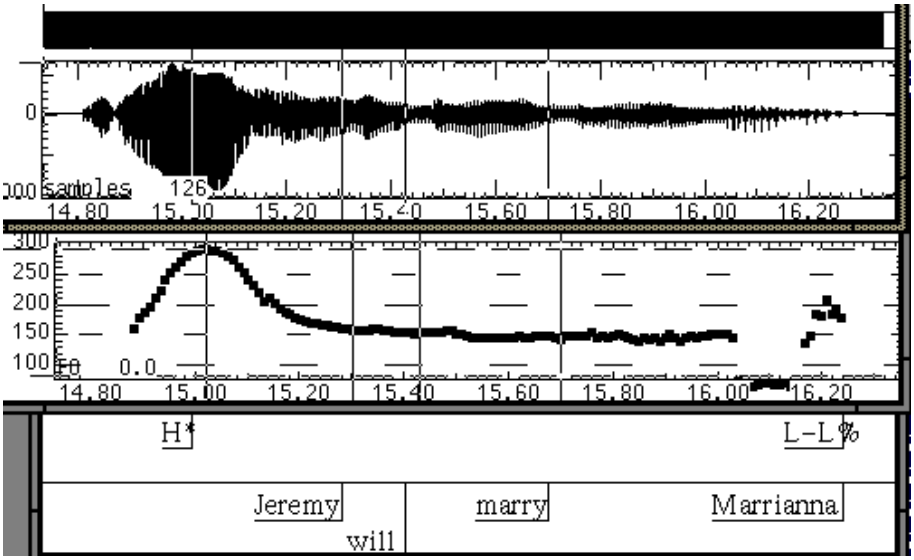

Figure 3.2: Early nuclear accent pattern (H* L-L%).

The third contour (the "double nuclear accents" contour) has high tone accents on both the subject and the object, as in the hat pattern. However, these two high tone targets are separated by a low phrase tone. Each of the high tone accents occur in its own intermediate phrase, and since each is the only accent in its intermediate phrase, they are both nuclear accents.

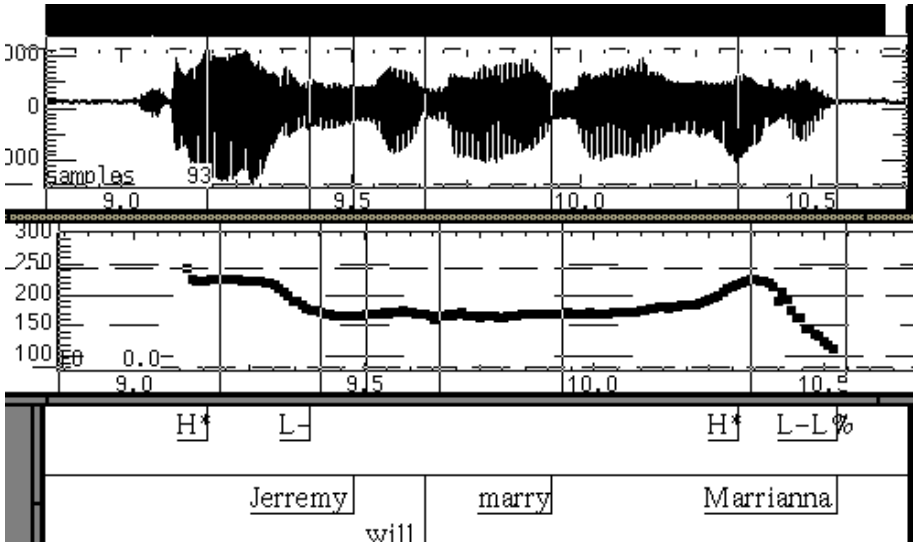

Figure 3.3: Double nuclear accent pattern (H* L- H* L-L%).

The fourth contour (with a late "rising nuclear accent") is metrically identical to the hat pattern, but phonetically more like the double nuclear accents pattern. That is, metrically it consists of a single intermediate phrase, with a simple high tone accent on the subject, and a bitonal L+H* nuclear accent on the object. The leading low tone of the rising nuclear accent makes for a fall−rise rather than a plateau between the two accents, but the contour differs from the double nuclear accent contour in terms of affiliation of the low tone. In the double nuclear accent contour, the trough in between the two accents is an independent tone affiliated with the intervening phrasal boundary. In the rising nuclear accent contour, by contrast, the fall−rise is due to the leading low tone of with the second accent.

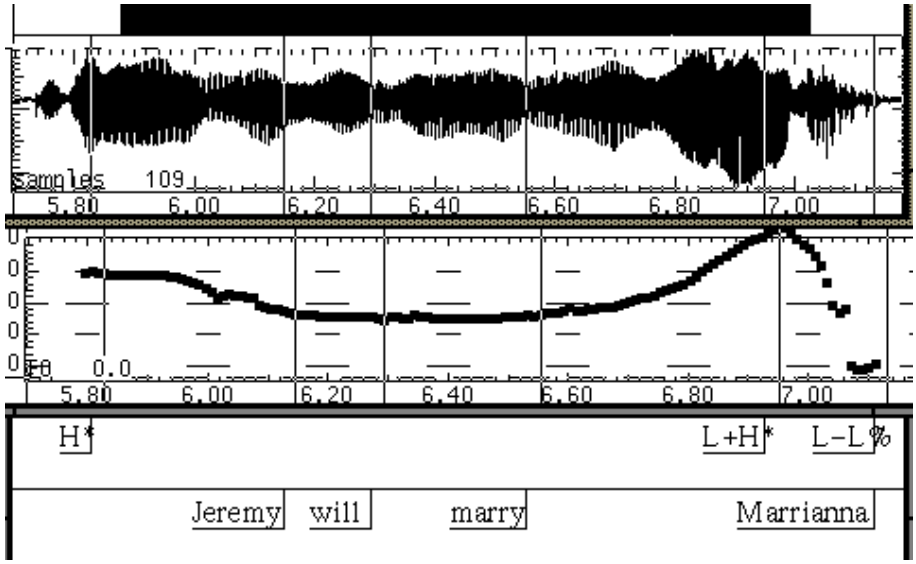

Figure 3.4: Late Peak contour (H* L+H* L-L%).

This phonological difference in the affiliation of the low tone results in a more subtle difference in the F0 contour than can be captured in the description of a fall-rise before the (final) nuclear fall on the object. This phonetic difference involves the timing of the minimum and (to a less extent) the target value at the minimum. In contour 3, the F0 steeply falls down from the first H* to the bottom of the local ptich range at the low phrase tone and then gradually slopes up again to the second H* accent. In contour 4 however, the low tone is part of the bitonal pitch accent, and the pitch falls gradually from the first high tone accent to the onset of the syllable that the second accent is associated with, and then sharply rises to its high tone target.

Another phonetic difference that is even less obvious in these figures involves the segmental timing. Since the low phrase accent in the double nuclear accents pattern marks a kind of prosodic phrase boundary (the intermediate phrase boundary) not

present in the rising nuclear accent contour, we might predict there also to be segmental correlates of the boundary, such as phrase-final lengthening. In addition, since the preceding accent is a nuclear accent in this case, it might show other hallmarks of the more prominent stress, such as a higher F0 peak for the accent, or longer duration for the accented syllable. These segmental effects are difficult to assess in the mostly sonorant segments of *Jeremy* in Figures 3.1 to 3.4, but we can measure segment durations more easily in many of the proper nouns that we used as subjects in our test stimuli.

3.3 Acoustic Analysis

In the next section, we will review the theory of English intonational phonology that we were assuming in our review of the earlier experimental literature in the previous chapter and in the descriptions of our own materials in the previous section of this chapter. Before we review the phonological account, however, we want to describe in more quantitative detail the phonetics of our stimuli. This quantitative description is important for two reasons. First, we want to show that each of our stimulus sentences is an example of the intonation pattern that we claim it has. Second, we would like to relate our stimuli to the patterns in previous experiments where the description of the stimuli was not couched in the same phonological framework. Since the most detailed description is in Eady et al.'s (1986) production study, we will present some of our acoustic data in the same format as in Eady et al's figures, to bolster our interpretation of the phonological structures that they elicited in equating Eady et al.'s double focus utterances to Birch & Clifton's (1995) two-accent stimui (and, by extension, to Welby's (2001) double nuclear accent pattern).

For the purpose of describing the phonetic properties of the stimuli sentences used

in this thesis, we measured the duration and the peak F0 in the lexically stressed vowel in the subject, verb and object position for all sentences. In addition, since all names occurring in subject position were tri-syllabic, we measured the midpoint F0 value of the other two vowels in the subject to get a sense of the slope of the fall (or lack thereof) across that part of the contour.

3.3.1 Duration Measurements

The duration of each target vowel was measured with the help of spectrogram and waveform displays. In most instances the onset and offset of the vowels were easily determined, for example from the release of the consonant. In other cases, we relied on standard acoustic criteria such as for a lowering in the formant values of F2 and F3 to determine offset of the vowel after [r].

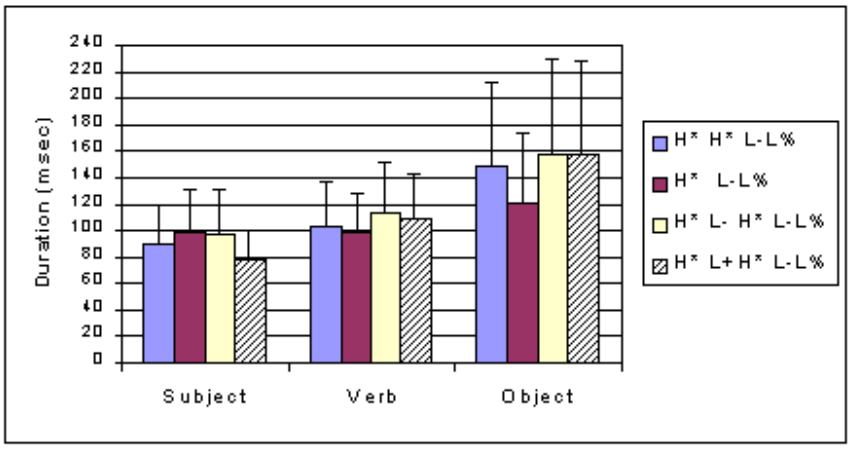

Figure 3.5: Plot of the mean duration of the lexically stressed vowel in subject, verb and object position for all stimulus sentences.

The graph in Figure 3.5 shows the mean duration data for the three target vowels in the four contrasting intonation contours. Three one-way ANOVAs on the durations of the stressed vowels in each of the three positions did not reveal any significant effect of intonation contour on the durations, probably because there were only 20 tokens per type distributed across a wide range of syllable structures and segment types. While the differences are not significant in any of the analyses of variance, the graph does show two trends that can be compared to significant differences in the durations of the stressed vowels in the subject and PP in Eady et al. (1986).

First, in the earlier study, the target vowel in the PP was shorter in the utterances with early narrow focus on the subject, as compared to each of the three other focus patterns. This is just what we would expect if the material after the focused subject were deaccented. In our data, too, the deacccented object in the H* L- L% early nuclear accent pattern was the shortest.

Second, in the earlier study, the stressed vowels in "focused" words were longer. In particular, the stressed vowel in the subject was longer in the utterances with early narrow focus on the subject or double focus on the subject and PP, as compared to the utterances with broad focus on the sentence as a whole or with late narrow focus on the PP. In chapter 2, we suggested that the double focus utterances in Eady et al. were produced with the same double nuclear accents pattern as in Birch & Clifton's (1995) *She TEACHES MATH*, which is the contour illustrated in Figure 3.2 above. This suggestion is supported by the duration data in our stimuli, which show longer stressed vowels for subjects in the patterns with early nuclear accent and with the double nuclear accents as compared to the other two intonation contour types.

In our data, the difference between the double nuclear accents and rising nuclear accent patterns is especially noteworthy, since this is the phonetically similar pair that we have claimed to differ structurally, with a metrically stronger nuclear accent

on the subject in the first case, but a mere pre-nuclear accent on the subject in the second case. A paired samples t-test showed this difference between to be significant ($t[38]=2.13$, $p<0.03$).

3.3.2 Fundamental Frequency Measurements

As a second phonetic descriptor, Eady et al. measured the F0 peak of the vowels in the words in subject, direct object, and PP (functioning as indirect object) position. The F0 on the subject was slightly higher in the double focus contour than in the broad focus and late narrow focus conditions, but this difference was not reliable across subjects and items, and it contradicted the results of two earlier studies with even longer sentences than these SVO-PP sentences. Therefore, Eady et al. could not safely conclude that focus results in an increased F0 peak value. The authors did, however, notice another, more reliable difference in the shape of the F0 on the target vowel. The F0 peak in the subject consistently occurred earlier in the stressed vowel, to be followed by a subsequent fall, in the utterances with early narrow focus on the subject and with double focus on the subject and the PP.

For our set of stimuli sentences, we measured the peak F0 at the approximate midpoint of the vowel in the subject, verb, and object, and logged the F0 value at those time points. The mean F0 peak values and standard deviation are given in Table 3.1 and the means repeated in Figure 3.6, in order to better show the "topline" trends.

F0 in Hz	Subject	Verb	Object
H* H* L-L%	258.43 (29.63)	262.23 (26.39)	267.84 (28.02)
H* L-L%	310.96 (35.62)	111.10 (47.27)	108.32 (40.64)
H* L- H* L-L%	283.43 (34.02)	197.42 (22.12)	256.74 (27.19)
H* L+H* L-L%	259.74 (27.03)	194.83 (18.89)	277.00 (32.42)

Table 3.1: Mean F0 peak (Hz) and standard deviation for all lexically stressed V in Subject, Verb and Object position for all sentences.

A 4 (contours) x 3 (Vowel positions) two factor Anova with repeated measures indicates a significant difference in the means of the F0 for the word positions 'subject' vs. 'verb' vs. 'object' ($F(2,38)=171.26$, $p<0.0001$). There is also a significant interaction between the intonation contour and the position of the word ($F(6,114)=131.27$, $p<0.0001$), showing that the F0 in the 3 measurements differs as a function of the particular intonation contours they occur in.

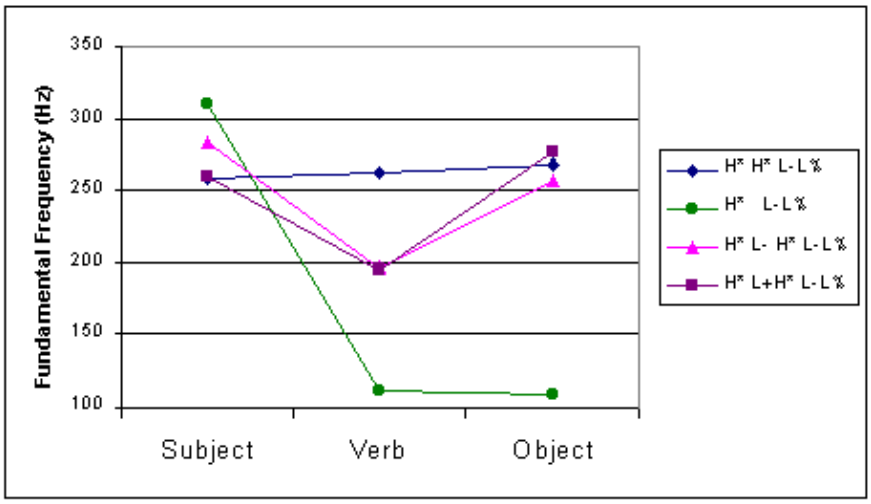

Figure 3.6: Plot of the mean F0 (fundamental frequency) in the lexically stressed vowel in subject, verb and object position for all stimulus sentences.

Two aspects of the F0 patterns in Figure 3.6 replicate the patterns of (what we claim to be) analogous intonation contours in Eady et al. (1986). First, the F0 peaks on the subject in our early nuclear accent and double nuclear accents contours are higher than the peak on the subject in the hat pattern. This is like the difference in subject peak values for the early narrow focus and double focus utterances as compared to the broad focus utterances in Eady et al. (1986). Second, there is a substantial fall in F0 from the first measurement point to the second measurement point in the double nuclear accents contour, just as there is in the early nuclear accent contour, but the minimum reached at the verb is not as low in the double nuclear accents contour. This is like the difference between the fall in F0 from the peak on the subject to the peak on the direct object in the double focus versus early narrow focus utterances in Eady et al.'s study.

If our interpretation of the intonation patterns in Eady et al. is correct, we might ascribe the difference in peak value on the subject to the greater metrical prominence of a nuclear accent relative to a prenuclear accent in the early narrow focus and double focus utterances, and we can ascribe the fall in F0 from the subject to the subsequent measurement point in the double focus cases in part to the contrast between a H* target on the subject and a subsequent L- phrase accent that governs the pitch up to the end of the intermediate phrase, which should occur somewhere between the subject and the last noun in the sentence. The fact that the fall is not as great in the double nuclear accents contour as in the early nuclear accent contour in Figure 3.6 can be explained by the placement of the L- phrase accent. Since the intermediate phrase boundary typically divides the nuclear accented subject from verb phrase as a whole, the verb in our stimuli is somewhere in the rising transition from the L- to the subsequent H* on the object. The analogous pattern in the F0 peak values in Eady et al.'s figure might be ascribed similarly to the placement of the L- phrase accent. If the intermediate phrase boundary that makes the first accent a nuclear one in their double focus utterances is similarly placed immediately after the focused subject, then the object peak would be even later in the rising transition from the L- tone to the subsequent high target on the accented noun in the PP.

A third aspect of the F0 patterns in Figure 3.5, on the other hand, does not replicate the patterns in Eady et al. (1986). In our hat pattern utterances, the mean peak values are nearly the same across the subject, verb, and object positions, suggesting a plateau very much like the one described for the hat pattern contour in Figure 3.1. In Eady et al.'s analogous figure, by contrast, the broad focus utterances show a decline in F0 across the subject, direct object, and head noun of the prepositional phrase. One possible explanation of this difference is that the broad focus utterances in Eady et al. (1986) were not all "plain" hat patterns, but intermingled plain hat patterns

and downstepping hat patterns (see below). Whatever the explanation, however, we cannot plausibly ascribe the fall in peak F0 from subject to object in their broad focus utterances to a transition from a H* on the subject to an immediately following L- phrase accent, given the more local pattern of late peak placement within the stressed vowel in the broad focus utterances relative to the early peak placement in the double focus and early narrow focus utterances. This fact suggests that the "topline" pattern shown in figures such as our Figure 3.6 is not a completely informative measure. It does not reliably distinguish the fall from a nuclear H* accent to a L- phrase accent from other sorts of decline. This is particularly a problem for us, if we want to capture the more subtle differences suggested above for the contrast between the double nuclear accents contour and the rising nuclear accent contour.

The H* L+H* L-L% rising nuclear accent pattern bears much phonetic similarity to the H* L- H* L-L% double nuclear accent pattern, both have two distinct F0 peaks with an intervening low tone creating a fall in the F0 before the rise to the second peak. However, they are structurally different. The double nuclear accen contour contains two intermediate phrases whereas the rising nuclear accent contour contains a single intermediate phrase andhence a single nuclear accent. The rising nuclear accent contour is structurally more similar to the hat pattern *H* H* L-L%), despite their phonetic differences. Both of these contours consist of a single phrase, with a prenuclear accent followed by a late nuclear accent, although the rising nuclear accent contour has a dip in F0 before the second accent while the hat pattern does not.

The phonetic similarities between them are more subtle than the differences. In Figure 3.5, both show a lower F0 peak on the first accent relative to the high F0 peak of the (first) nuclear accent in the double nuclear accents and early nuclear accent contours, and in Figure 3.4, both have shorter mean durations for the stressed syllable of the subject relative to the metrically stronger stressed syllables in the two patterns

where the subject bears a nuclear accent. How reliable are these small differences in mean F0 peak value on the subject and stressed syllable duration that class the rising nuclear accent pattern with the hat pattern rather than with the double nuclear accents pattern?

We would like to demonstrate that the double nuclear accent and the rising nuclear accent contours are phonetically more different than this suggests, to show that our two types of stimulus sentences with two peaks seperated by a dip in F0 could be reliably perceived to be of different structure. We can do so by looking more closely at the fall-rise shape separating the two peaks on the subject and the object.

In both of these contours, there is a low tone between the two accent targets, yet our expectation is that the anchor point for the low tone is earlier when the low tone is a phrase accent than when it is a leading tone for the second pitch accent. Since we expect listeners to parse these two contours differently, we need to show that these contours actually differ in a reliable way, across all items.

For this purpose, we made additional measurements of the mid-point F0 values in the second and third syllables of the subject. The prediction is that the F0 fall should be steeper in the contour containing the low phrase accent than in the contour where the dip is due to a low leading tone for the second pitch accent. It also turns out that (for our speaker, at least) the low phrase accent is a lower-pitched target than the low leading tone of the rising nuclear accent.

That is, for all sets of items we did find a difference in the F0 on the third syllable of the word in subject position (see Appendix B for comparative graphs of the hat pattern, the double nuclear accent contour and the late rising nuclear accent contours for each item). The aggregate graph is shown in Figure 3.7.

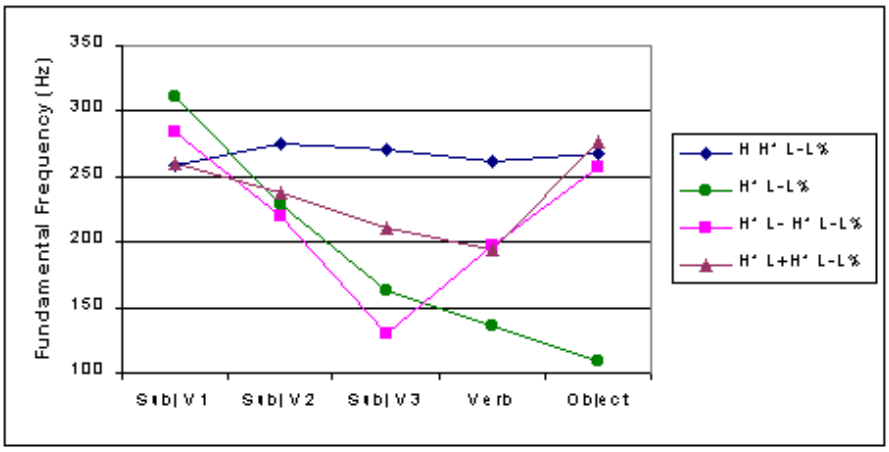

Figure 3.7: Summary graph of mean F0 for five points across the four intonation contours.

To compare the mean F0 of the last vowel in the subject in all four contours, we performed a One Way Anova. The results indicate a significant difference among these data points ($F(3,76)=90.33$, $p<0.0001$). A Tukey HSD test for posthoc comparisons indicates that all pairwise comparisons of the mean F0 are significant at the $p<0.01$ level.

Note that the fall of the F0 in the double nuclear accent contour is not only to a lower value at the end of the subject, but it also reaches the minimum during the third vowel of the subject (Sub V3) whereas the minimum F0 of the late peak contour is not reached until the verb. The difference in timing of the minimum (later in the rising nuclear accent pattern) and the F0 value at the minimum (higher in the rising nuclear accent pattern) is consistent across all twenty stimulus sets, as can be seen by looking at the individual items graphs in Appendix B.

Unlike in Figure 3.7, the x-axis in each of these individual items graphs plots the actual vowel durations rather than the ordinal position of the target syllable. A comparison of these x-axis values shows that the third vowel typically is longer in the double nuclear accents pattern, as well as lower in F0, suggesting a phrase-final lengthening that is in keeping with the posited structural difference. Also in several instances of the double nuclear accent contour, we can observe glottalization on the spectrogram and waveform, which is an indication of non-periodic vocal fold vibration, which may also serve as a perceptual cue to listeners that the F0 is extremely low in these instances. For our speaker, glottalization was also a frequent cue to the L% tone at the end of the utterance, as illustrated above in Figure 3.2. The presence of glottalization at the end of the subject in the double nuclear accents pattern is thus another cue to the presence of the medial phrase boundary.

Thus, while both the double nuclear accents pattern and the rising nuclear accent pattern consistently show the fall rise pattern that distinguishes the two from the hat pattern, they also show reliable differences in the timing and the extent of the fall. These differences are apparent not just in the aggregate graph in Figure 3.7 but also in all of the graphs for the individual item sets in Appendix B.

The aggregate graph in Figure 3.7 also indicates that on average, the F0 in the hat pattern does not vary much, but stays flat between the initial and the last accent. This pattern, however, was less consistent across the individual item graphs in Appendix B. Sometimes there was a perfectly flat plateau that matched the aggregate plot (as in items 1 and 5). Sometimes there was a small rise onto a higher peak on the verb, in keeping with its greater metrical strength (as in items 11 and 17). And sometimes there was a more or less pronouncced declination across the utterance (as in items 10 and 15). In some of the last cases, two transcribers, trained in the account of intonational phonology that we are assuming here, transcribed a different accent

type on the verb, a "downstepped" !H accent, which Pierrehumbert & Hirschberg (1990) describe as differing from the plain H in invoking a more explicit inference chain between the current utterance and propositions already in the mutual belief space. This different choice of accent type, however, does not change the metrical structure of pre-nuclear peak accent followed by a nuclear accent that is consistent with broad focus on the utterance as a whole.

The "topline" plot of the utterances with broad focus in Eady et al. (1986) also shows this gradual slope from the initial through the second and the third peak F0 measurement point. Thus, we believe that the contour described as 'neutral focus' in Eady et al. and the one we describe as 'hat pattern' are basically the same type of intonation contour.

Thus, based on the duration and F0 measurements, we find great similarities between the elicited productions in Eady et al. (1986) and our stimulus sentences. It is important to note that the authors were able to elicit these prosodic patterns on longer sentences such as *Don shot the puck to Kent.* as replies to the four types of questions listed below in (3.2):

(3.2)

1. What happened?
2. Who shot the puck to Kent?
3. Who did Don shoot the puck to?
4. Who shot the puck to whom?

Thus, the hat pattern that Gussenhoven (1983) found to be felicitous for both broad VP-focus and late narrow focus when the prenuclear H* was on the verb should be felicitous for even broader whole-sentence focus (as in response to question 1 above) when the prenuclear H* is on the subject. And the double nuclear accents pattern

that Birch & Clifton (1995) and Welby (2001) found to be more felicitous for focus on the verb as well as on on the object when the first of the two nuclear accents was on the verb should be felicitous for an explicit focus on the subject (as in response to question 4 above) when the first nuclear accent is on the subject.

Our listeners should be able to recover the (implicit) question when prompted with one of these four intonation contours on a longer constituent than the verb phrase. This is one hypothesis we will test in this thesis.

3.4 Intonational Phonology

Since we find great similarities in the prosodic structures of three of the four types of utterances from Eady et al. (1986) and our stimuli, we attributed their phonetic properties to structural specifications within a phonological model that allows us to interpret F0 patterns in terms of tone targets in different intonational contours, where the intervening low target in the two double peak cases is aligned differently to the text. In this section, we will describe that phonological model in more detail.

For the purpose of describing intonation contours, we will assume the "Autosegmental/Metrical" account developed by Pierrehumbert and her colleagues (see, e.g., Pierrehumbert, 1980; Liberman & Pierrehumbert, 1984; Beckman & Pierrehumbert, 1986; Beckman & Ayers Elam, 1994/1997). In this model of intonation for English, intonation contours can be described by an intonational grammar in which the basic elements are intonational morphemes: This model assumes tonal elements such as pitch accents, phrase tones and boundary tones. While accents mark the heads of the phrases, phrase tones and boundary tones mark the edges of the phrases.

Pitch accents are local prominences which are aligned with the lexically stressed

syllable [1] of a (pitch accented) word, and in this model the tone target associated with the pitch accented syllable is starred (*). Phrasing is the grouping of words into prosodic constituents and is indicated by tonal events on the right edge of a phrasal domain. There are two levels of phrases in an utterance or sentence, the larger phrase is called intonational phrase (IP) and the smaller one is called intermediate phrase (ip). The right edge of the intermediate phrase (ip) is marked with a phrase tone (marked by -), and the right edge of an intonational phrase (IP) is marked with a boundary tone (marked by %). In Pierrehumbert's system, and the ToBI transcription system of English intonation that is based on it, each intonation phrase (IP) needs to contain at least one intermediate phrase (ip) which in turn must contain at least one accent. The last, or the only accent within an intermediate phrase is (structurally) defined to be the nuclear accent.

One way to represent the phrasal structure is by bracketing the prosodic constituents, as in (3.3):

```
(3.3)     H*    L-            L+H*      H*   L-  L\%
          |                     |        |
     [ [ Phebe ]ip      [wrote another song]ip ]IP.
```

The structure in (3.3) consists of a single Intonation Phrase (IP) which in turn consists of two intermediate phrases (ip). The first intermediate phrase contains a single accent, which therefore is also the nuclear accent of this ip. The second intermediate phrase contains two pitch accents, one associated with the lexically

[1] Bolinger (1986) points out that for contrastive purposes, the pitch accent can also be aligned with a syllable that is not lexically stressed (The whiskey wasn't EXported from Ireland, it was DEported.).

stressed syllable of the word 'another', and the second one with the word 'song'. The last accent, associated with the word 'song' is the nuclear accent of this phrase, while all preceding accents (in this case there is just one) are prenuclear accents.

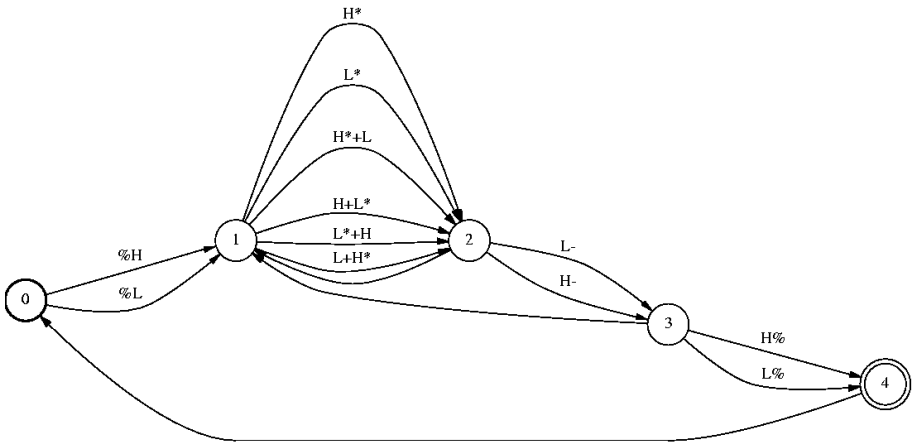

Figure 3.8: Finite State Automata representation of the recursive tonal grammar of English, illustrating the possible combinatorics of accents with intermediate phrase tones and boundary tones.

Figure 3.8 is a Finite State Automata representation of the tonal grammar of English. Going through it f rom left to right, an optional initial boundary tone can be followed by any number of accent tones targets, which have to be followed by one of the two intermediate phrase tones. If there is more than one intermediate phrase in an intonation phrase (IP), there is an arc from state 3 back to state 1, where any accent type and any number thereof can occur in the ip. Eventually, the last

intermediate phrase in an intonation phrase will be delimited with a boundary tone to it's right.

No spoken utterance in English exists without intonation. The tune, composed of accents and phrase tones, serves to perceptually highlight words and to structure information: Accents, associated with syllables within words can highlight discourse relevant entities (e.g. information newly introduced to the discourse) and phrase tones, marking the right edge of a phrase can indicate discourse relevant structure (e.g. the continuation of a turn). A speaker has two choices to make when aligning the tune to the text. The first choice is syntagmatic in nature and determines the location of the accent or accents given a particular text. The second choice is paradigmatic in nature and determines the type of accent in a given location. That is, a speaker has choices to make in how to align the tune to the text, there is no single one correct versus incorrect mapping. The alignment of tune, in particular the pitch accent, to text can indicate prosodically marked focus and thus can yield categorical differences in the interpretation of an utterance .

3.4.1 What Justifies a Phonological Layer?

In their study, Eady et al. (1986) relate their data on F0 excursions and the like directly to the concept of focus, without this intervening layer of phonological interpretation. We showed in the previous section that adopting the same strategy for our stimuli would make it difficult to describe systematic differences between two contour types that we predict will not be felicitous in the same contexts. Specifically, we predict that the double nuclear accents pattern will be more felicitous for double focus on subject and object whereas the rising nuclear accent pattern will be more felicitous with late narrow focus. Particularly for cases where the association of the

tone to text material is crucial to the disambiguation of focus structures, such as the difference in the alignment of the low tone in a tune such as H* L+H* L-L% where the fall of to the low tone is more gradual versus the double nuclear accent contour H* L- H* L-L% where the fall is steeper and where the low phrase tone is anchored at the right edge of the intermediate phrase boundary, a phonological layer allows for unambiguous descriptions of the F0 contour.

Assuming the phonological layer also allowed us to relate Eady et al.'s materials to findings in the psycholinguistics literature on focus patterns in verb phrases (Gussenhoven, 1983, Birch & Clifton, 1995, and Welby, 2001), leading to clear predictions about how those findings might extend to focus patterns in larger constituents such as the whole SVO sentences of our materials.

Further, assuming this particular model of the phonological layer allowed us to interpret the descriptions of the [+focus] conditions in Gussenhoven (1983) in the same terms as the materials in Birch & Clifton (1995). This allowed us to suggest a difference in the intonational phrasing between the sentences tested in Gussenhoven (1983) and those tested in Birch & Clifton (1995), which in turn helped us understand and possibly resolve the discrepancy between their results.

A more general justification for wanting a phonological layer rather than going directly from an acoustic measure to interpretation comes from other experiments, such as a study conducted by Pierrehumbert & Steele (1987) where they investigated timing properties of the 'rise-fall-rise' contours (L+H* L-H% and L*+H L-H%), which have also received attention in Ward & Hirschberg's (1985) research and in Pierrehumbert & Hirschberg's (1990) work on the inherent meaning of pitch accents. According to Pierrehumbert & Steele's research, the two related pitch accents L*+H and L+H* when associated with the syllable "-mil-" as in the utterance *Only a millionaire!*, are either interpreted as *incredulity* or as *assertive*. Thus, seemingly subtle differences

in the alignment of the tonal target with the syllable change the overall meaning of the contribution to the discourse. Ward & Hirschberg (1985) suggest that the intonation contour L*+H L-H% could also cue a speaker's *uncertainty*. Pierrehumbert & Hirschberg (1990) suggest that both of these pitch accent types are used to convey the salience of some scale (see Chapter 4)which links "the accented item to other items salient in [the] H[earer]'s mutual beliefs".

One last point to make is that not all local pitch prominences must indicate focal prominence. Some pitch accents, particularly prenuclear ones occurring early in an utterance, might serve the purpose of having an early rhythmic peak in the utterance before a stretch of unaccented material. Thus, pitch accents can also only have rhythmic functions to satisfy rhythmic constraints. Gussenhoven (1984) suggested something like this when he described the "step up" in pitch onto the verb in the [+focus] condition as not being of the same focal accent status as the peak on the following object. Couching his idea in terms of the particular model that we assume allows us to interpret it as a difference between prenuclear and nuclear accent. That is, another hypothesis that we will test is that the two metrically similar patterns shown in Figures 3.1 and 3.4 will both be compatible with late narrow focus in a way that the metrically contrasting pattern in Figure 3.3 is not.

As this brief overview indicates, fundamental frequency excursions, when interpreted within a phonological model, are interpreted as pitch accent types. These pitch accents serve to highlight information for the purpose of attracting attention to particular lexical items in larger strings of such words, but also indicate which words do not receive such attention (Bolinger, 1958). Pierrehumbert & Hirschberg (1990) further suggest that different pitch accent types can have different discourse functions: while both H* and L+H* accents add information to the shared knowledge between the hearer and the speaker, the L+H* accent - they suggest - has the added

function of restricting or limiting the scale of values induced. This will be explained further in Chapter 4.

CHAPTER 4

THE PSYCHOLINGUISTICS OF FOCUS

4.1 How Prosodic Focus is Studied

In this section we will review some of the methods used in earlier studies to address the processing and interpretation of prosodic focus. The three dominant paradigm types are *prominence ratings*, *question/context selection in a dialogue*, and *appropriateness ratings*. Since the relevance of results using each of these paradigms has already been pointed out previously, we will only briefly review the methods and point out their limitations.

One general issue with all of the experimental literature to date is that the meaning and semantics of focus is not very well defined. Since the adequacy of any behavioral measure of focus interpretation depends on what "focus" means (and how different responses correspond to different focus meanings), we will attempt to define what we mean by *focus* to show how measures of the congruence between questions and answers, mediated by prosodic focus, is relevant to the interpretation of discourses outside the laboratory including larger narratives that are not simple sequences of questions and answers. We will motivate the relevance of the question selection paradigm to the larger discourse structure via Büring's (2000) notion of discourse trees. This chapter closes with a motivation of the experimental paradigms used in this study.

4.1.1 Rating of Prominence or Degree of Stress

In several studies, primarily on Dutch (Rietveld & Gussenhoven, 1985; Gussenhoven & Rietveld, 1988; Terken, 1991, 1994; Krahmer & Swerts, 2001) but also for English (Ladd & Morton, 1997) listeners were asked to rate the preceived prominence of synthetic or natural fundamental frequency peaks in relation to a second peak. Krahmer & Swerts gathered ratings of "prominence" for either the first or the second word in adjective-noun phrases collected in different contexts (such as *red square* elicited in a context that contrasted a red square to a blue square versus a context that contrasted a red square to a red circle) and found that the noun was substantially and reliably rated to be least prominent in the phrases that were elicited in contexts that put narrow focus on the adjective.

We have interpreted this result in terms of the notion of 'deaccenting' introduced in Chapter 2. When the adjective and the noun were rated about equally prominent, we cannot be sure what the focus interpretation is. It could be a double narrow focus interpretation as well as a broad focus interpretation that would allow for both words to be equally prominent. Further, in the instances where the noun was judged to be most prominent, we cannot reliably gauge the focus interpretation either, we need to be unsure whether listeners interpreted the utterance in terms of a broad focus or late narrow focus. Thus, prominence ratings by themselves do not allow for a differentiation of prominence for the purpose of marking focus, or for the purpose of just saying a word more emphatically due to other discourse constraints.

We have interpreted this result in terms of the notion of "deaccenting" introduced in Chapter 2. When there is early narrow focus on the adjective, it bears the nuclear accent, and the noun is deaccented. Here the relationship between prominence and focus is clear. However, as the focus literature has long noted, the relationship

between prominence and focus interpretation is much less clear for any other value of relative prominence on the last word in a phrase. Thus, prominence ratings by themselves are not a very informative measure of how prosodic focus is interpreted in terms of the speaker's intention to focus the listener's attention on one or another discourse element.

This inadequacy comes out especially clearly in the study by Gussenhoven (1983) who also used ratings of "degree of stress" for the three types of sentences introduced in Chapter 2. In the structure C type dialogue pairs, where sentences such as *Tell her I'm planing a coup.* were elicited in the context of either *She seemed sort of sorry there'd never be another coup* or *Anything in particular you want me to say to her?*, the object noun in the response received substantially lower stress ratings when the sentence was elicited in the context that put the preceding verb in narrow focus than when the sentence was elicited in the broad focus context. Again, we can interpret this result in terms of the lesser prominence (or "stress") of a deaccented word relative to a nuclear accented word. In the structure A and B sentences, also, where the verbs were rated, he found differences between the ratings elicited in response to broad VP focus questions (*What is the nature of your business?*, and *Please tell us what happened that night.*), as compared to those elicited in response to questions putting narrow focus on the final word (*What is it you repair?* and *What do you remember from your last lesson?*).

We can interpret this result as indicating that a word with a prenuclear accent ("step−up in pitch") is more prominent that one without an accent ("clearly lack a step−up in pitch"). However, because Gussenhoven elicited another kind of judgment more pertinent to the focal interpretation, he could directly correlate the stress ratings to the focus interpretation. He observes that the perceived degree of stress on the verb in the Structure A type sentences was not a good predictor of the ability to

recover the context question. This result is in keeping with his point that the accent on the verb in the Structure A sentences is not becessarily a focal prominence, and that not every accent denotes focus. This also suggests that Selkirk's focus projection theory (Selkirk, 1984, 1995) where every accent is a focus, and where there is a one-to-one mapping between accent placement and focus (except in cases where the accent on an argument "projects" a focus onto the preceding unaccented verb or preposition to make a broader constituent-wide focus), cannot account for Gussenhoven's results. We assume that in these cases, listeners give a rhythmic interpretation to the prenuclear accent. That is, some early accents might be produced for purely metrical reasons, such as to provide an early target that can define the phrasal pitch range or to mark the beginning of words for lexical access, and in that case, the listener need not interpret them as necessarily marking a focused constituent. The possibility that not all accents are focal makes phonological judgments such as prominence ratings less than satisfactory measures of focus.

4.1.2 Appropriateness Ratings

A second type of task devised to understand how listeners interpret accent patterns gauges the appropriateness of an intonation pattern of a target sentence with respect to a context sentence. Birch & Clifton (1995) obtained appropriateness ratings on a scale from 1 through 5 for three types of utterances, very much in structure like Gussenhoven's Type A and Structure B sentences, in the context of two types of questions. Their results indicate that listeners preferred a reply such as *Yes, she TEACHES MATH.* in response to the broad focus question *Isn't Kerry pretty smart?*. Results also indicated that listeners were able to recover the more appropriate early narrow focus context for the reply sentence that contained a single early accent. In a

subsequent reaction time study, they obtained reaction times for simple "yes, makes sense" and "no, does not make sense" responses. The reaction times for 'Yes' were fasted for the broad focus question paired with R1 (Yes, she TEACHES MATH.), and considerably slower when paired with R3 (Yes, she TEACHES math.). The "no" responses were not reported.

A Yes-No decision task as to the Appropriateness of a stimulus with a specific intonation pattern to a question was performed by Ayers (1996). The time interval between presenting the response and returning a "yes, appropriate" and "no, not appropriate" response were measured. Just as Birch & Clifton, Ayers found that the response time to inappropriate matches was longer than to appropriate matches, indicating an added processing load for the non-matching instances.

Welby (2001), also presented listeners with short question-answer dialogues in which the 'answer' varied with regard to the intonation pattern. She asked listeners to indicate the appropriateness of the match between the question and the answer on a scale from 1 through 5. Her results indicate that in VP-focus questions such as *How do you keep up with the news?* versus Object-focus questions such as *What newspaper do you read?*, the hat pattern contour with a pitch step up on the verb is equally appropriate to the contour containing a single late accent on the object.

This task has the advantage that it gets at the discourse function of prosodic focus more directly. That is, the implicit claim is that the "meaning" of focus has to do with the interpretation of an answer in response to a question. When presented with an utterance with a given prosodic focus pattern, the listener should be able to recover the (implicit or explicit) context. This implicit claim is brought out even more clearly in a related task, that of matching an utterance with its context.

4.1.3 Choice of (Written) Context Question in a Dialogue

To study the interpretation of accent patterns, Gussenhoven (1983) conducted a context retrievability task in which he asked listeners to match question-answer pairs. He presented each answer in combination with a question and asked listeners if this mini-dialogue came from the same interaction or from different ones. Due to the design of this task, each subject got to hear each context question and they also listened to each answer twice, once in combination with the right context question and once in combination with the wrong context question (Q1-A1; Q1-A2; Q2-A2; Q2-A1). They also saw both choices of context question at all times on their response sheets.

Results indicated that listeners were unable to reliably determine the context for the Structure A sentences (SVO with narrow focus on the object contrasted to broad VP focus) while for the other two structures the retrievability scores were substantially better. The Structure C sentences were syntactically identical to the Structure A sentences, but differed in the prosodic structures compared, contrasting a pattern with nuclear accent on the object to a pattern with nuclear accent on the preceding verb and the object deaccented. The difference in context retrievability here thus merely confirms what we already know about the clarity of focus interpretation when early narrow focus results in deaccenting of post-focal materials, and Gussenhoven included the Structure C stimuli primarily as a control, to gauge the range of scores to expect in clear cases. Structure B sentences, on the other hand, were prosodically similar to the Structure A type sentences, contrasting narrow focus on the final word with broad focus on the VP as a whole, and differed only in the semantic/pragmatic relationship between the verb and the following word. The difference in context retrievability ratings between the Structure A and Structure B type stimuli, then, support the idea

that this task is a reliable measure of focus interpretation per se, and not just of the phonetic salience of the potentially focus-marking prosodic prominences.

In contrast to Gussenhoven who presented listeners with Question/Answer pairs, D'Imperio (1998) and Jin (1996) presented listeners with a statement and asked them to select the one WH-question from a set of written questions that identified what was being questioned. In this way, the subjects heard each response only once.

Since the context-choice task was the same for both studies, we will describe it for the Neapolitan Italian experiment. D'Imperio (1997) had 2 speakers produce the SV and SVO stimulus sentences she tested. She collected the same syntactic structures for both questions and statements. In Neapolitan Italian, yes/no questions and statements are differentiated by accent type only (D'Imperio, 2000). She collected utterances such as the ones in (4.1):

(4.1)

1.	Mario esce.	'Mario goes out.'
2.	Mario esce?	'Is Mario going out?'
3.	Mario ama Giovanni.	'Maria loves John.'
4.	Mario ama Giovanni?	'Does Maria love John?'

In her first experiment, she listed all sentences (without punctiation) on a response sheet and asked listeners to indicate which word appeared to be the "most important" in the sentence that was played back to them. She intended this task to yield judgments like Krahmer & Swerts's ratings of "prominence" or Gussenhoven's ratings of "degree of stress" except that "importance" was judged relative to other words in the same sentence rather than relative to an absolute paradigmatic scale. This task yielded similar results to those of Krahmer & Swerts: the judgments were most consistent for utterances where an early nuclear accent marked narrow focus.

In the focus interpretation task, then, she presented written lists of preceding contexts of the sort listed in (4.2):

(4.2)

1. Dimmi qualcosa di quella coppia.
   'Tell me something about that couple.'

2. Chi ama Maria?
   'Who does Maria love?'

3. Chi ama Giovanni?
   'Who loves Giovanni?'

4. Che pensa Giovanni di Maria?
   'What does Giovanni think about Maria?'

In the blocks where the listeners heard declarative statements, they were instructed to "match the statement they heard to one of several questions setting up a specific context.". In the blocks where listeners heard yes-no question, they were similarly instructed to choose the Wh−question whose meaning most resembled that of the stimulus question, or that was felicitous as a subsequent question to the question they heard. For example, if a question such as *Who loves Maria?* is posed, then a subsequent question such as *Does Giovanni love Maria?* is felicitous.

We can identify the same task in Eady et al.'s study. That is, while Eady et al. (1986) did not rigorously study the interpretation of their four types of elicited sentences, one of the experimenters listened to the entire set (initial, final, dual, neutral focus) in a blind test. The purpose was to identify those utterances for

further study, for which they could recover the Wh−question it was a response to. They report that the listener was able to match the utterances with the question in 96% of all cases.

Matching answers to the question that might have elicited them using either Gussenhoven's same-different dialogue judgment task or D'Imperio's choice of Wh−question task allows us to gauge the focus interpretation of a sentence, if focus is defined as congruence between a question and answer in a sequence of dialogue turns. As we will show later in this chapter, this definition is related to one of the formal definitions of focus that has been proposed in the literature on the semantics and pragmatics of focus. This is a promising task to use. However, its major drawback is that in uses to date, the dialogue task has not been extended specifically to other types of discourse. Even if we were to define focus in terms of the semantics of questions, however, focus is not limited to such explicit question/answer pairs. Rather, the semantics of focus needs to be specified more generally, and related more clearly to the semantics of questions in general, so that we can adapt these task paradigms and extend them to other contexts so as to address the issue of focus patterns in monologue narratives also, where questions are only implicit and do not necessarily immediately precede their answers. We turn to this issue of the semantics of focus in the next section.

4.2 The Denotation of Focus

The notion of "focus" has been approached from different areas within linguistics, among them phonology (Bolinger, 1958; 1986, 1989; Pierrehumbert & Beckman, 1988; Ladd, 1980, 1992), syntax (Jackendoff, 1972; Selkirk, 1984, 1995), semantics (Jackendoff, 1972; Rooth, 1985, 1992; Gussenhoven, 1984), formal pragmatics (Roberts, 1996; Büring, 2000, 2001), and discourse/topic structure (Grosz & Sidner,1986). To

reveal a potential strategy in recovering the discourse intent of an utterance within a monologue, we will have to understand what "focus" is. We will review how the semantics and formal pragmatics literature addresses the issue of "focus" in questions and statements, and how prosody (including the choice of accent location, accent type and phrasing) is understood to mediate between questions and assertions. In this review, we will draw on research by Rooth (1985, 1992), Roberts (1996) and Büring (2000).

The dominant Generative Linguistics theories assume a primitive feature of "focus", and most of these theories, following Jackendoff (1972) assume that focus is a property of (surface) syntactic constituents, possibly extending the scope of focus beyond the word receiving prosodic prominence via a process termed "focus projection" (Selkirk, 1984, 1995). Jackendoff (1972) and Rooth (1985) both interpret focus to be a semantic property which manifests itself in syntactic structures which are dominated by the feature "F" (for focus). The feature "F", on the other hand, is not a syntactic feature, but a semantic one. It contributes to the interpretation of a sentence in context, and that interpretation is related to the semantics of questions.

4.2.1 The Semantics of Assertions

In defining the semantics of focus, it is useful to compare "focus semantics" to "ordinary semantics" for the same sentence. The focus denotation of a sentence is a semantic object. It contrasts with the ordinary semantic value as follows: In a sentence such as *Benjamin heard shots.*, the ordinary semantic value is a single proposition which is constructed from the ordinary semantic values of its parts which are *Benjamin*, *heard*, and *shots.*

The ordinary semantic value of the noun *Benjamin* would be an element from a

domain of *individuals*, while the ordinary semantic value of the noun *shots* would be an entity from a domain of *things that can be heard*. The ordinary semantic value of a verb in an SVO utterance is a function (Rooth, 1985) or a relation (Roberts, 1995) over two entities. Thus, the ordinary semantic value of the sentence as a whole is the proposition:

heard(x, y), with x = Benjamin and y = shots

This ordinary semantic value of the sentence is identical, no matter what the intonation pattern is. By contrast, the focus semantic value differs, depending on the intonation.

In this chapter we will adopt a usage of the semantics and formal pragmatics literature: double bars are used to indicate the focus semantic value or the focus denotation of an utterance:

(4.3)

H* L-L\%

|| [Benjamin]F heard shots. ||

The phonology, via a particular prosodic pattern on an utterance, marks prosodic focus. Thus, if there was an early nuclear accent (H* L-L%) on the subject of this sentence, a rather likely interpretation would be an early narrow focus on the subject NP.

If a H* accent is located late in the utterance, on the object for example, the status of the focus is less clear. It could be interpreted as an Object focus only, or via focus projection, it could be VP focus or broad sentence focus.

(4.4)

H* L-L\%

|| Benjamin heard [shots.]F||

(4.5)

H* L-L\%

|| Benjamin [heard shots.]F||

(4.6)

H* L-L\%

|| [Benjamin heard shots.]F||

In contrast to the ordinary semantic value of our example sentence which is always this single proposition, the focus semantic value of this sentence is a set of propositions. The set of propositions is obtained by substituting the element in focus (marked with the feature F) with other variables of the same semantic type in the discourse domain (D). The focus semantic value is also termed the *focus alternative set*. Thus we can fill out the focus semantic value (4.4) with a set of propositions, such as the one shown in (4.7).

(4.7)

|| Benjamin heard [shots.]F|| = {'Benjamin heard birds', 'Benjamin heard Maya sneeze', 'Benjamin heard John and Maya laugh', 'Benjamin heard nothing', 'Benjamin heard Maya call John names',...}

(4.8) is a formalization of this set:

(4.8)

——Benjamin heard [shots]F—— = Benjamin heard u: u ϵ D

If the prosodic focus pattern is different however, and there was only a single early H* accent on the subject NP *Benjamin*, the focus semantic value of this utterance also would be different, as shown in (4.9).

(4.9)

||[Benjamin]F heard shots.|| = {'Maya heard shots', 'John heard shots', 'I heard shots', 'John and Maya heard shots', 'Maya and I heard shots', 'noone heard shots'....}

If a prosodic pattern such as the hat pattern is interpreted to yield broad focus over the entire utterance, the focus semantic value of this utterance would different again and much less specific since here the entire proposition can be substituted with other variables of the same domain (D), as shown in (4.10).

(4.10)

||[Benjamin heard shots.]F|| = {'Anna watched TV', 'Nobody liked Raymond', 'Ross got married to Rachel', 'Everybody fell asleep', ...}

It is important to note from these examples, that the focus alternative set differs for different focus patterns, which can be indicated by different prosodic patterns.

4.2.2 The Semantics of Questions

In contrast to an assertion, the ordinary semantic value or the denotation of a question such as *What did Benjamin hear?* is not just a single proposition, but a set of propositions, , much like the focus denotation of an assertion. Rooth (1992), notes this, citing Hamblin (1973) for the semantics of questions:

> "A particularly direct formulation of a constraint relating questions to focus in answers can be given in terms of the semantics for questions of Hamblin (1973). Hamblin constructs a theory in which a question determines a set of potential answers by actually identifying the semantic value of a question with a set of potential answers, including both true and false answers." (p85)

In other words, the denotation of a question is a set of alternative propositions, much like the focus alternative set of a statement or assertion. The Question alternative set (Q-alt) is formalized as follows (following Roberts (1996) who cites von Stechow (1989):

Q-alt(What did Benjamin hear?) = Benjamin heard u: ϵ D

Note that the Q-alternative set in (4.11) for the question *What did Benjamin hear?* and the focus alternative set for the statement *Benjamin heard [shots]F.* in (4.8) are identical (Q$-$alt = F$-$alt).

This is what we mean, formally, when cases we speak of congruence between questions and assertions. And (we assume) this congruence relationship between assertions and questions is what is being recovered in experimental paradigms in which listeners are required to match questions to assertions, or assertions to questions. Thus, in questions such as *What happened?*, or *What is the way things are?* where the entire proposition is queried, the Q-alt contains the same set of alternative as the focus semantic value or focus alternative set of a broad focus statement. The issue then becomes, how can we extend this notion of congruence from single question-answer sequences in the two-turn dialogues of Gussenhoven (1983) or Birch & Clifton (1995) to more complex discourses such as a long monologue narratives.

4.2.3 Information Structure

Roberts (1996) suggests the flow and the structure of information between two interlocutors to be a co-constructed game of questions and assertions for which the ultimate goal is to answer the *big question* which is *What is the way things are?*. This main goal of answering the "super question" can be addressed by answering intermediate or subordinate questions, satisfying intermediate goals. Questions are defined as set-up moves in this game, while the pay-off moves are assertions, or answers to questions. The context in which these questions and assertion are interpreted is the *common ground*, which is a set of propositions which are believed (or temporarily accepted) to be true by the interlocutors.

Coherence in a discourse is achieved by having congruence between the questions and the answer moves. And congruence between a question and and answer is attained when the focus alternative set of an assertion matches the Q-alt set (which is the ordinary denotation of the question).

To illustrate, consider the two-turn dialogue in (4.12), and the associated semantic values in (4.13).

(4.12)

Speaker A: Who heard shots?

H* L- L%

Speaker B: Benjamin heard shots.

(4.13)

Q-alt:(Who heard shots?) = {'No one heard shots', ''Benjamin heard shots', 'Maya heard shots', 'John heard shots', 'Susi heard shots', 'Ben and

Maya heard shots', 'Maya and John heard
shots', 'the entire group heard shots',
'Everybody but Susi heard shots', ...}

F-alt:||[Benjamin]F heard shots.|| = {{'No one heard shots',
'Benjamin heard shots', 'Maya heard
shots', 'John heard shots', 'Susi
heard shots', 'Ben and Maya heard
shots', 'Maya and John heard shots',
'the entire group heard shots',
'Everybody but Susi heard shots', ...}

Here, the set of alternatives for the question and the assertion are identical, or at least intersecting, and hence, there is congruence between the question and the reply. We have already noted that the set of focus alternatives differs for different prosodic focus patterns. Thus, prosodic focus mediates between question alternatives and focus alternatives. If the question and answers are adjacent turns in a dialogue, as in (4.12) and many of the stimuli in Gussenhoven (1983) or Birch & Clifton (1995), this relationship is obvious and explicit. However, as Roberts (1996) suggests, because skillful interlocutors assume congruence, the same relationship also holds between the F-alt of an assertion and the Q-alt of an implicit question that is only invoked by the prosodic focus of an assertion, as in the discourse in (4.14). Here Speaker B does not simply answer the question, but expands with a statement that has the F-alt in (4.15), invoking an implicit question which has an appropriately congruent Q-alt set, such as that in (4.16).

(4.14)

Speaker A: Isn't Kerry good at math?

H* L- L%

Speaker B: Yes, she teaches math

(4.15)

||she [teaches]F math|| = \{'She dabbles at math', 'She reads math books every night', 'She failed math in grade school', ...}

(4.16)

What is Kelly's relationship to math (that makes you agree that she's pretty good at it)?

A second thing to note is that, while the ultimate goal is to satisfy the super question, a felicitous reply can also only address a sub−part of the question.Thus, a question such as *What frightening thing happened to you?* can have subquestions such as *Did one of you hear something unusual?* and *Who heard shots?*.

Similarly, a very broad question such as *What happened?* can have subquestions such as *What was the setting?*, or *What frightening thing happened to you?*. We can understand the prosodic focus in monologue narratives in terms of the Q-alt sets of (usually implicit) questions if we invoke such a hierarchy of broader questions and their nested smaller subquestions. That is, the prosodic focus of any particular statement in the narrative invokes the local "question under discussion" at that point in the unfolding discourse hierarchy of larger and smaller questions that is the purpose of the narrative. Büring (2000) provides a formal account of this hierarchy that is particularly useful for understanding how the speaker of a monologue narrative sets up expectations about the next questions to be discussed by the narrator. Before

describing this account, however, let us first apply the notion of the implicit question under discussion to understanding how we might extend the question-answer matching paradigm developed for two turn dialogues to the somewhat longer monologue stories that are the context for our stimuli in the first experiment.

4.2.4 Identifying the Question Under Discussion

We have already shown above in (4.12) and (4.13) how congruence between the F-alt and the Q-alt is established in the case of an SVO statement with narrow focus on the subject uttered in response to an explicit question with a WH word for its subject. This understanding of the discourse function of focus on a statement can be extended to a monologue narrative such as the one in (4.17) if we assume that the congruence is between the F-alt set on the explicit assertion and the Q-alt set of an only implicit question, the question under discussion that is invoked by that prosodic focus in that context, as shown in (4.18).

(4.18)

Invoked QUDs	Narrative
What happened?	
What was the setting?	
How was the weather?	It was a cold day, and
Where were you?	we were walking around Hocking Hills.
What warnings were there that something might happen?	
What sounds of others did	

did you hear?	It was very quiet out there, and
Who did you see?	there was nobody else there but us,
	and it looked completely deserted.
What frightening thing	
happened to you?	
Did you hear shots or something?	
Who heard shots?	BENJAMIN heard shots.
1. Who heard shots?	

Our first experiment exploits this relationship between the focus semantics indicated by prosodic focus of the stimulus sentence and some implicit question under discussion (QUD) that is invoked by that F-alt set. This is what it means to match the stimulus sentence that is the last assertion in an interrupted narrative to a Wh−question. We predict, then, that the early nuclear accent pattern invokes the last QUD in sequence shown to the left in (4.18).

What about the other intonation patterns? If the hat pattern is interpreted as broad focus, it should invoke only the larger question *What happened?* and not the embedded sub−question *Who heard shots?*. However, there are other possibilities. We have suggested above that a late nuclear accent is not only compatible with broad focus on the sentence as a whole, but also with with Object NP focus, thus, effectively answering a different embedded sub−question *What did Benjamin hear?*. Of course, it is also compatible with VP−focus, answering a third alternative embedded sub−question such as *What is the frightening thing that happened to Benjamin that day?*. In the choices we presented to our participants, however, we listed only narrow focus on the object and the broadest of all possibilities, with focus on the

entire sentence. The focus semantics for these two interpretations are given in (4.19) and (4.20).

(4.19)

QUD: What happened?
F-Alt:||[Benjamin heard shots]F|| = {‘Suddenly Prof. Smith walked into view with Angela on his arm’, ’John slipped in a stream and broke his ankle’, ‘Maya found the bracelet that she lost the last time we came here’, ...}

(4.20)

QUDs: What happened?
What did Benjamin hear?
F-Alt:||Benjamin heard [shots]F|| = {‘Benjamin heard shots’, ‘Benjamin heard a woman screaming’, ‘Benjamin heard someone breaking into our car’, ...}

\end(verbatim)

The same sentence with the double nuclear accents pattern can be a felicitous reply to two types of questions, one asking for broad focus

```
({\em What  happened?}  if  it is interpreted  as a  rather deliberate
speaking style,  or a  double question asking  for two  separate foci,
such as  {\em Who heard what?}. The  F-alt for the latter  is shown in
(4.21).

\begin{verbatim}
(4.21)

QUDs: What happened?
      Who heard what?
F-Alt:||[Benjamin]F heard [shots]F|| = {‘Benjamin heard shots’,
                                       ‘Maya heard screams’,
                                       ‘John heard laughter’,
                                       ‘Susi heard nothing’, ...}
```

The above predictions about the invoked QUD follow from our discussion congruence and our interpretation of the results of the earlier experimental literature reviewed in Chapter 2. That is, we have attempted to explain how the semantics and formal pragmatics literature addresses the issue of 'focus' in questions and statements, and how prosody (including the choice of accent location and phrasing) mediates between questions and assertions. A last issue to address is the choice of accent type and how this choice can be relevant to the interpretation of focus.

We have repeatedly referred to the discrepancy in results between Gussenhoven (1983) and Birch & Clifton (1995) and indicated that factors beside the location of the accent may be relevant for the understanding of those results. From descriptions that Birch & Clifton provide, we know that their nuclear accent types were L+H*

accents, whereas out best guess regarding the intonation contours in Gussenhoven's experiment is that the stimuli involved something like our hat pattern, with a different nuclear accent type. The question of the type of accent becomes relevant in light of some research discussed below.

We assume that both the H* H* L-L% as well as the H* L+H* L-L% are structurally similar, they both consist of a single intermediate phrase only, and both have a single nuclear accent. We also assume that both these contours are felicitous in the context of the broad focus and late narrow focus questions stated above. However, Pierrehumbert & Hirschberg (1990) suggest that the L+H* accent type has a different meaning from the H* accent type in that it not only adds information to the mutual belief space (previously referred to as *common ground*) but also invokes a scale with the focus as an instantiation (item) of that scale:

> 'L+H[*] accents are employed by [the] S[peaker] to convey the salience of some *scale* (defined here following Ward & Hirschberg 1985 as a partial ordering) linking the accented item to other items salient in H[earer]'s mutual beliefs.' [p.294]

A scale is more easily understood in terms of an inherently ordered set such as *grades in school*, or *temperature*, but it could also be more abstract and be constructed over sets such as *things I really like to eat*, or *more or less bad things that could happen at a reception*, etc. All that is really necessary is to invoke a set of discourse entities and an obvious partial ordering over (subsets) of those entities. It is often the case that the L+H* accent is used for contrastive or corrective purposes, but it is not necessarily so. If there was a late nuclear accent of the type L+H* on the object NP *shots* in *Benjamin heard shots.*, then the invoked scale could be over the set *more or less dangerous sounds*, or *more or less loud sounds*, whatever is a salient ordering in

this context. However, as should be obvious, the ordering (the scale) will be more salient if the set is something homogenous and well-defined such as *more or less loud sounds*, rather than something like *everything that could have happened on a walk in the Hocking Hills*. Thus, a late narrow focus interpretation is particularly plausible for this contour.

We have argued how the congruence between questions and answers, mediated by prosodic focus, is relevant to the interpretation of discourses. In the following section, we will motivate the relevance of the question selection paradigm to the larger discourse structure via Büring's (2000) notion of discourse trees. That is, the prosodic focus of any particular statement in the narrative invokes the local "question under discussion" at that point in the unfolding discourse hierarchy of larger and smaller questions that is the purpose of the narrative. Büring (2000) provides a formal account of this hierarchy that is particularly useful for understanding how the speaker of a monologue narrative sets up expectations about the next questions to be discussed by the narrator.

4.3 Discourse Trees

Büring (2000) provides us with a hierarchical model or visualization of the information structure of a discourse. In this dynamic tree model, non-terminal nodes are question moves, and terminal nodes are answer moves. When applying Büring's Discourse Tree model to a monologue narrative such as the stories presented, the only explicitly expressed moves typically are the answer moves. Thus, all question moves are implicit. An answer move invokes an implicit, immediately dominating node of the question tree, a local QUD that is congruent with the assertion. In order to construct such a tree, the listener first has to interpret the focus structures of the explicit answer

moves and recover the implicit QUDs that immediately dominate the answer moves. Moreover, since all (or most) of the QUDs are implicit, the competent listener might also need to recover other QUDs that do not immediately dominate any answer move, but are the mothers of one or more of the subquestions that immediately dominate, to link the current QUD into the QUD tree. That is, the listener might need to project a path upward to another larger question, in such a way that the current QUD can be parsed as a sister to the QUD dominating one or more other answer moves that together can constitute a strategy for answering the larger question. The other answer moves that are parsed as sisters might be ones that have already been uttered, as in the hierarchy of questions that links *we were walking around Hocking Hills* as a sister to *It was a cold day* in the tree dominated by the larger QUD *What was the setting* in the QUD hierarcy in example (4.18) above. Or they might be ones that the listener is assumed to be able reconstruct from the focus pattern in combination with Gricean principles such as the Maxim of Quantity. For example, when the listener hears *BENJAMIN heard shots* in the context of the narrative in (4.18), he or she must accommodate to the early narrow focus by assuming some implicit answer move such as *We were frightened by strange sounds such as gunshots.* Some such implicit move is invoked by the presupposition *Someone heard shots* that is shared by all of the propositions in the F-alt set for the sentence *BENJAMIN heard shots* uttered in that context, and the assumption that the discourse hierarchy will be coherent then links the implicit answer move together with explicit one into the hierarchy. In the second experiment, we are testing to see if and how listeners can reconstruct the Discourse Tree in this way. We do so by asking them to choose among alternative continuations, such as the sentences in (4.22) for the narrative in (4.17), to probe what kind of tree structure they have constructed at the point where the narrative is interrupted. That is, we provided listeners with an interrupted narrative that was

held constant, ending in a stimulus sentence with one of four different intonation contours, and see whether the choice of continuation is influenced in any way by the intonation of the last sentence before the interruption. The different continuations are designed to be compatible with three different Discourse Trees which might be constructed to accommodate to the different local QUD that is invoked as dominating broad focus versus late narrow focus versus double focus interpretations of the this last sentence.

(4.22)

1. And we ran back to the car as fast as we could. (broad focus)
2. And then he heard screams. (late narrow focus)
3. And John heard screams. (double focus)

We illustrate with the Discourse Tree for the narrative in (4.18), which we have already shown in outline format in (4.18). Figure 4.1 shows this same QUD structure as a Discourse Tree. A Discourse Tree always begins with the broadest of such question, querying the largest proposition, such as the entire sentence *What happened?*. In the beginning of the discourse, various aspects of the proposition are queried. Thus, more specific follow up questions to *What happened?* are, for example, *What was the setting?* and *What frightening thing happened there to you?*. Felicitous subquestions to *What was the setting?* might be *How was the weather?* or *Where exactly were you?*. Another subquestion to *What was the setting?* could be *What warnings were there that something might happen?*, which subsequently has sub-sub-questions *What sounds of others did you hear?* and *Who did you see?*. All of these questions are addressed by part of the narrative monologue which is the context to the stimulus sentence. The listeners' task then is to incorporate the stimulus sentence into the

Discourse Tree under the *What frightening thing happened to you?* node, and project a node for the continuation.

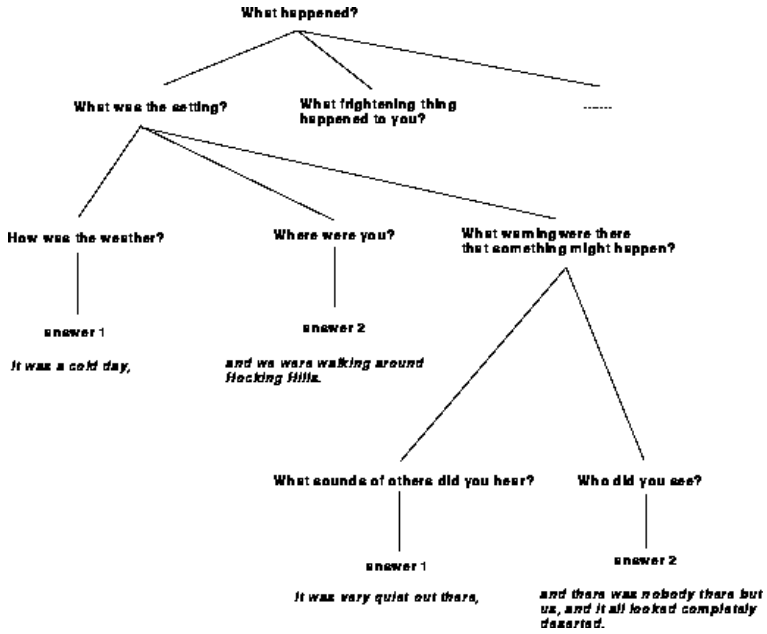

Figure 4.1: Discourse Tree for the Beginning Part of the Narrative.

We predict that the focus interpretation of the stimulus sentence will affect how it is incorporated into the Discourse Tree and, thus, what kinds of projections are available for the continuation. This prediction is based on two assumptions. First, we assume that an answer move must be congruent with the QUD that immediately dominates it. Second, we assume that the QUD immediately dominating the stimulus sentence must be part of a coherent strategy for answering the last open question, which in this case is *What frightening thing happened to you there?* If the stimulus sentence *Benjamin heard shots.* is produced with a prosody indicating broad focus,

this in itself is a felicitous reply to that larger question. That is, the QUD invoked by the stimulus sentence is the larger question, and the projected QUD path that dominates the continuation, then, can be a similarly broad focus question such as *How did you react?* This Discourse Tree is depicted in Figure 4.2.

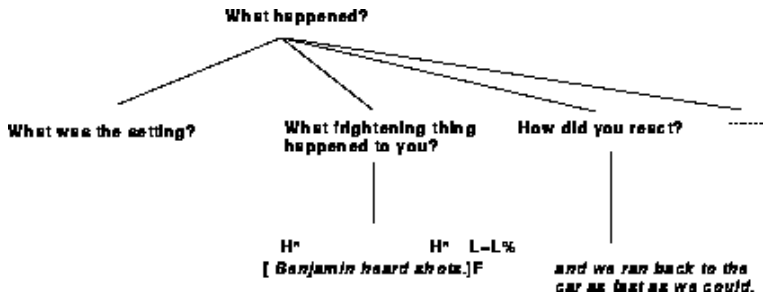

Figure 4.2: Discourse Tree Evoked by Broad Focus Interpretation.

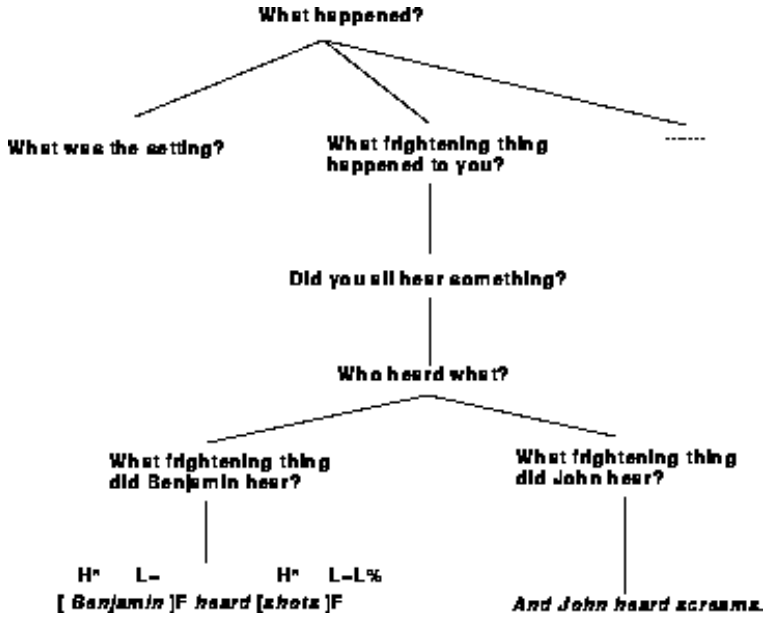

Figure 4.3: Discourse Tree Evoked by double Focus Interpretation.

If the intonation contour on the stimulus sentence invokes a double narrow focus interpretation, on the other hand, the question that is implicitly evoked cannot be equated with the broad focus QUD that is the sister to *What was the setting?* in the discourse tree in Figure 4.1. The listener must infer a more complicated intervening structure involving a double narrow focus question such as *Who heard what?*. Büring (2000) suggests that such a double-barrelled QUD does not directly dominate the double focus answer move, but instead is the mother node for a strategy that makes one or the other focused element be interpreted as a "Contrastive Topic" (CT). For example, *BENJAMIN heard SHOTS* could be interpreted as part of a strategy where the thing heard is the CT, so that *Who heard what?* dominates a series of QUDs such as *What about shots? Who heard that?*, *What about screams? Who heard that?*, and *What about running footsteps? Who heard that?* Alternatively, the strategy might

be to expand on the *What?* part of the larger question, as in Figure 4.3. In either case, a double focus stimulus sentence would project intervening structure involving a double focus question, which would automatically provide a path for incorporating another double focus sentence as the continuation.

Of course, a double focus interpretation is not the only interpretation that requires intervening structure to incorporate the QUD that immediately dominates the stimulus sentence. We have already described above in (4.18) the kind of intervening structure that might accommodate an answer move with an early narrow focus. Figure 4.4 similarly shows how the QUD for a stimulus sentence with a late narrow focus interpretation might be incorporated into the Discourse Tree that is set up by the background narrative.

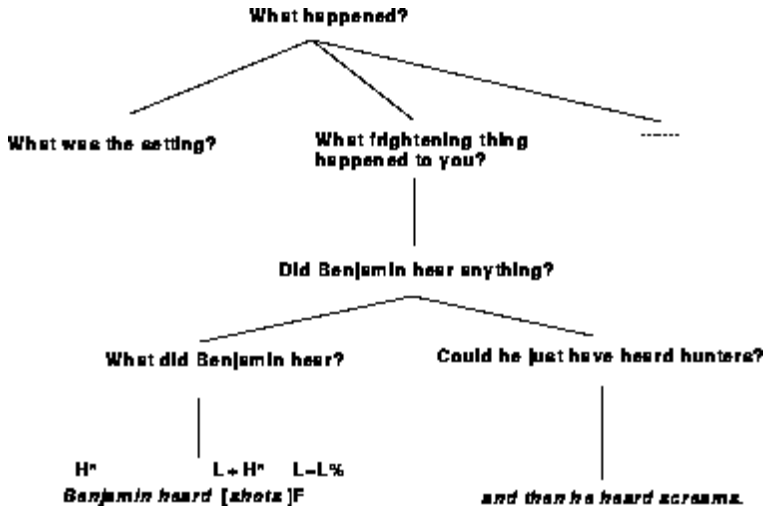

Figure 4.4: Discourse Tree Evoked by Late Narrow Focus Interpretation.

The Discourse Tree model, then, is the assumed background for the two types of experiments that we will describe in Chapters 5 and 6.

In the first experiment described in the next chapter, we will be testing whether or not listeners can reliably recover the local or immediate Question under Discussion (QUD), the question that can be the set up move for this particular response. In this experiment, we are testing whether or not listeners will reliably match one of the four intonation patterns pattern with a specific Wh−question.

In the second and third experiments, described in Chapter Six, we will test whether or not listeners are able to uncover the QUD strategy at a particular time in the discourse, as cued by the intonation contour on the stimulus sentence.

CHAPTER 5
EXPERIMENT 1: RECOVERING THE LOCAL QUESTION UNDER DISCUSSION

5.1 Introduction

The theoretical work on focus (Jackendoff, 1972; Rooth, 1985; Roberts, 1996; Büring, 2000, 2001) suggests that congruence between questions (Q(uestion)−alternative sets) and assertions (F(focus)−alternative sets) is required to achieve coherence in a discourse. Congruence between a question and answer is attained when the focus alternative set (or focus semantic value, the focus denotation of an answer, the set of alternative propositions) of an assertion matches the Q-alt set, which is the (ordinary) denotation of the question, that is, a set of alternative propositions. (See Chapter 4 for a more detailed discussion of these concepts.)

In simpler terms, the idea is that the abstraction over the Wh-element of a question will allow us to identify the focus of a sentence that is an appropriate answer to the question. If this is so, then the opposite should be true as well: given a statement with a specific prosodic focus structure, a listener should be able to identify a congruent question. When listening to a coherent sequence of statements, then, the listener should be able to identify the sequence of questions that the speaker is implicitly posing. That is, parsing the intended discourse structure of a monologue narrative,

can be described as a process of identifying the sequence of larger and smaller sub-questions that the narrator is answering on the way to answering the largest question of all, namely, *What is the way things are? Tell us the story that is the story of your life right now.*

The first experiment described below required subjects to select the question they felt was best addressed by the last sentence to be heard in an interrupted monologue narrative, which was the stimulus sentence, carrying one of four different intonation patterns. With this experiment, we wanted to make sure that listeners can recover a congruent question when it is presented not as an explicit preceding question move in a dialogue, but as an implicit question (QUD) at some point in a monologue narrative.

Different groups of listeners hear the same sentence bearing different intonation patterns. If that variation in the prosodic structure triggers variation in the listeners' focus interpretation of the stimulus sentence then the different utterances should be identified with different questions at that point in the narrative. A different focus structure implicitly invokes a different QUD that is being addressed. In this way, we can probe the focus interpretation. That is, we are making the assumption that while explicit questions can elicit certain focus structures in a reply, a certain focus structure can also evoke the implicit QUD that was addressed by the sentence. Since this paradigm asks listeners to identify the implicit question under discussion at the interruption point, we will call it 'the "QUD paradigm."

(As an extension of the QUD paradigm, the "continuation paradigm" probes the listeners' focus interpretations of an intonated stimulus sentence by asking them to find the most appropriate/best continuation to the context/story they were just presented with, paying particular attention to how the stimulus sentence was said. The experiments using this second paradigm will be described in Chapter Six.)

5.2 Methods

5.2.1 Materials and General Design

For a phonetic and phonological description of the stimulus utterances, see Chapter 3. These stimulus utterances were embedded in a story or narrative. For an example, see Chapter 4. Since there were 20 different stimulus sentences (i.e., 20 sets of utterances with the four different contours produced on the same text), 20 stories were constructed. The story as a whole consisted of two parts: a background sentence spoken by a female speaker to establish the setting of the story, and then the narrative proper, including the stimulus sentence, spoken by a different female speaker. This second speaker is very well trained in producing intonation contours on demand. The stimulus sentence could have any one of the 4 intonation contours described in Section 3.2, which are listed again in (5.1).

(5.1) The four target intonation contours:

1.	H*	H*	L-L%	(hat pattern)
2.	H*		L-L%	(early nuclear accent)
3.	H* L-	H*	L-L%	(double nuclear accents)
4.	H*	L+H*	L-L%	(rising nuclear accent)

The recordings for the background setting and the narrative proper were made separately in a sound attenuated room at 44KHz.

The setting and the corresponding narrative recording were concatenated to make a single coherent "story". The setting & narrative were auditorily presented and the stimulus sentence was repeated once. A sample story (with stimulus sentence in italics) is given in (5.2).

(5.2) Setting: Susi was telling a story of what happened on a walk with her friends Benjamin, Maya and John last February. Monologue Narrative: It was a cold day, and we were walking around Hocking Hills. It was very quiet out there, and there was nobody else there but us, and it looked completely deserted. *Benjamin heard shots.*

After the stimulus sentence was presented, the narrative was interrupted. In this experiment, as the narrative is interrupted, a set of 4 questions is presented, as in (5.3) for the story in (5.2). The questions were presented visually.

(5.2) QUD Paradigm: Experiment 1 (visual presentation on paper):

1. What happened?	broad focus
2. What did Ben hear?	late narrow focus
3. Who heard what?	double focus
4. Who heard shots?	early narrow focus

Each listener heard the narrative and only one version of the stimulus contour. (The complete set of narratives is listed in Appendix A.) Thus, there were 4 different groups of subjects to avoid playing the same background with different stimulus sentences to the same subject.

5.3 Predictions

Based on the results of earlier experiments involving shorter two-part dialogues and on some (as yet untested) suggestions in the literature on intonational meaning, we can make the following predictions about the relative interpretive preferences listeners will have for the four intonation contours. **Hat pattern: H* H* L-L%**
Contour 1 should be the most ambiguous: it can be interpreted as a broad focus

utterance (identified with the QUD *What happened?*), as a late narrow focus utterance (em What did Ben hear?), or as a double focus utterance with focus on the subject NP and an independent focus on the object NP (*Who heard what?*). In the first and second interpretations, the initial H* accent would be interpreted only as a rhythmic prominence, and not as a focal prominence. (This is in keeping with Gussenhoven's (1983) and Welby's (2001) results, but it would not be predicted by Selkirk's (1984; 1995) theory of focus projection, in which every accent is a focus.)

Early Nuclear Accent Contour: H* L-L%

Since there is only a single early pitch accent in Contour 2, this early accent is the nuclear accent and should be unambiguously identified as an early narrow focus accent (identified with *Who heard shots?*).

Double Nuclear Accent Contour: H* L- H* L-L%

Contour 3 should receive more double focus interpretations than Contour 1, because of its two nuclear accents. If listeners interpret the phrasing as simply a more deliberate speaking style however, then a broad focus interpretation is plausible, also.

Late Rise Nuclear Accent: H* L+H* L-L%

Given the meaning of the L+H* accent, Contour 4 should receive more late narrow focus interpretations. That is, due to the nuclear pitch accent type (L+H*), which invokes a scalar implicature (a commitment to some choice from a partially ordered set), which we predict the contour to be easier to interpret if the scope of the focus of this accent is limited to the word it is associated with. Although this contour is structurally similar to the hat pattern (only a single intermediate phrase containing a prenuclear and a nuclear accent), we expect a different response pattern, with relativce more listeners choosing the late narrow focus question *What did Ben hear?* over the other two focus interpretations that we predict are possible with this metrical structure (see description of hat pattern).

5.3.1 Subjects

Four groups of Ohio State University Undergraduate students participated in this study, which was conducted as an in-class experiment. That is, four instructors volunteered their classes for this experiment. The students could participate in the experiment in exchange for extra credit, but each student could individually choose whether or not he or she wanted to participate. Anyone who chose to participate signed a consent form and filled out a language background questionnaire. Although responses were collected from anyone in the class who chose to participate, only data from native speakers of US-English from the MidWest were kept. After culling non-native speakers, there were 92 participants in all, distributed across the groups as follows: 30 listeners in the first group, 21 listeners in each of the second and third groups, and 20 listeners the fourth group.

5.3.2 Presentation of Stimuli

Since there were four groups of subjects, four stimulus tapes were created. All 20 stories (with the appropriate versions of the stimulus sentences) were recorded onto each DAT tape. After the end of each interrupted narrative, the stimulus sentence was repeated once. Then there was 24 seconds of silence, a warning beep, and another second of silence during which time the listeners could choose the appropriate QUD for the stimulus sentence. The listeners were given answer sheets with the QUDs on which to mark their choices. This meant that the set of responses was available and known to the participants as they listened to the stimulus sentence. For item sets 10 and 11, there was an error on the answer sheet. For these two items, the late narrow focus QUD was written twice, and there was no double focus QUD. Responses for these items, therefore, had to be discarded.

This experiment was conducted as an in-class experiment. All 18 stories (with the appropriate versions of the stimulus sentences) were recorded onto DAT tape. After the end of the interrupted narrative, the stimulus sentence was repeated once. The time interval between the presentation of the stimulus sentence and the onset of the next trial was 25 seconds, and listeners had 24 seconds before a beep alerted them to pay attention to a new trial. The beep was followed by 1 second of silence. Since all listeners were given answer sheets to mark their responses, the questions were known to the participants as they listened to the stimulus sentence.

5.3.3 Procedure

The instructions given to the listeners are shown in (5.4). Note that the listeners were specifically alerted to listen not only to what was said, but also to how it was said. (5.4)

> "Please listen carefully. You will hear a recording of the beginning of a story. The story will be interrupted before the speaker finishes. The last sentence before the interruption will the be replayed. There are four possible questions which you should read after you hear the repeated sentence. Pick the question which is best answered by the repeated sentence. Make sure to listen not only to what is said but also to HOW (particularly the last sentence) it is said."

These instructions were printed on the first page of a response booklet. Each subsequent page of the booklet contained the questions for any particular story, presented in a constant order, the order shown in (5.3) above for the story in (5.2). (The participants saw only the questions, of course, and not the identifying tags of "broad focus", "late narrow focus", and so on.)

5.4 Results

Contours	H* H* L-L%	H* L-L%	H* L- H* L-L%	H* L+H* L-L%	Total
Grp.1: 30	150 (5x30)	150 (5x30)	120 (4x30)	120 (4x30)	560
Grp.2: 21	84 (4x21)	105 (5x21)	105 (5x21)	84 (4x21)	378
Grp.3: 21	84 (4x21)	84 (4x21)	105 (5x21)	105 (5x21)	378
Grp.4: 20	100 (5x20)	80 (4x20)	80 (4x20)	100 (5x20)	340
Total: 92	418	419	410	409	1656

Table 5.1: Summary of responses collected for the four groups of listeners.

Due to the unequal number of listeners in each group, there was an unequal number of items in each condition. These numbers are listed in Table 5.1. Table 5.2 shows the number of times a particular question was chosen for any particular intonation contour type, totaled across the items, and Figure 5.1 summarizes these numbers as percentages. (See Appendix C for an item−by−item listing of the percent broad focus, late narrow focus, double focus, and early narrow focus questions chosen for each contour type.)

A 4x4 Chi-Square analysis yielded significant results ($\chi^2 = 983.3$, df $= 9$, $p<0.0001$), indicating that listeners responded differently to the four intonation contours. Thus, the H0 which is that listeners interpret all four intonation contours the same, must be rejected.

Contours	H* H* L-L%	H* L-L%	H* L- H* L-L%	H* L+H* L-L%
broad	113 (27.03%)	6 (1.43%)	116 (28.30%)	79 (19.32%)
late	208 (49.76%)	25 (5.97%)	120 (29.27%)	232 (56.72%)
double	68 (16.27%)	13 (3.10%)	101 (24.63%)	63 (15.40%)
early	29 (6.94%)	375 (89.50%)	73 (17.80%)	35 (8.56%)

Table 5.2: Raw numbers and percentages of elicited broad, late, double and early continuation responses for all four intonation contour types.

As predicted, the early nuclear accent contour is a rather salient phonetic pattern, as can be seen by the second group of bars in Figure 5.1. The late narrow focus question (e.g., *What did Ben hear?*) is easily singled out as the most likely QUD for the early nuclear accent stimulus sentence.

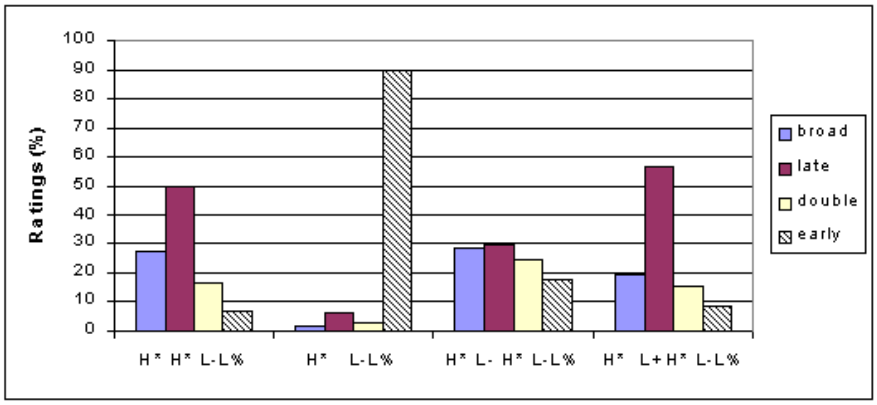

Figure 5.1: Percent responses of broad (b), late (l), double (d), and early (e) ratings given in response to the stimulus sentence having one of the four listed intonation contours.

While none of the other contour types elicits such a clear preference of one QUD over all other, the distribution of the QUD choices indicating a broad focus, late narrow focus, double focus, or early narrow focus interpretation of the intonation contour on the stimulus sentence does vary across the other three contours. The hat pattern and the rising nuclear accent contour receive more late focus interpretations, while the double nuclear accents contour shows the most even distribution of all types of replies. In the rest of this chapter, we give a more detailed comparison of the proportion of choices of different QUDs for the different contours, to test our predictions above.

H* H* L-L% versus H* L- H* L-L%

We predicted that the double nuclear accents pattern would differ from the hat pattern in eliciting more double focus interpretations. A 2x4 χ^2 comparison of all responses to the hat pattern and the double nuclear accents contour indicates a difference in the overall distribution of QUD choices (χ^2 = 49, df=3, p<0.0001).

A one-tailed Wilcoxon-Signed Rank comparison of the proportion of choices of the double focus QUD for these two contours indicates that listeners chose the double focus QUD as appropriate question significantly more often for contour 3 than for contour 1 (W = -1201, N = 76, P< 0.001 in the subject analysis; W = -103, N = 18, p< 0.05 in the item analysis).

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

A 2x4 χ^2 analysis of all responses to contours 1 and 4 (H* H* L-L% and H* L+H* L-L%) shows difference in the distribution of QUD choices for these two contours (χ^2 =7.99, df=3, p<0.05).

Our prediction was that, other things being equal, the rising nuclear accent contour would be more easily interpreted as a late narrow focus than as a broad whole-sentence focus. When comparing across these two metrically identical contours, then, this predicts that the late narrow focus QUD would be judged as appropriate more often for the rising nuclear accent contour than for the hat pattern, and the broad focus QUD should be judged as appropriate less often for the rising nuclear accent. Wilcoxon Sign Rank tests showed these predictions to be borne out statistically in the subject analyses (W = 715, N = 78, p <0.05 for the late narrow focus QUD; W = 737, N = 73, P<0.05 for the broad focus QUD), but not in the item analyses (W = 55, N = 16, p = 0.0793 for the late narrow focus QUD; W = -69, N = 18, p = 0.068 for the broad focus QUD).

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

A 2x4 χ^2 analysis for all responses to the double nuclear accents contour (H* L- H* L-L%) versus the rising nuclear accent contour showed that the distribution of QUD choices to these two contours differed significantly ($\chi^2 = 64.83$, df = 3, $p < 0.0001$).

A one-tailed paired Wilcoxon Sign-Rank test confirms that the late narrow focus QUD was chosen significantly more often for the H* L+H* L-L% rising nuclear accent contour than for the H* L- H* L-L% double nuclear accents contour (by subject analysis: W= -2565, n = 83, p< 0.0001; by item analysis: W = -171, n = 18, p< 0.0001). This result is important for two reasons. First, it bears out our prediction that the double nuclear accents contour would favor a double focus interpretation. Second, it shows that listeners were able to reliably differentiate the low phrase tone from the leading low tone of the bitonal pitch accent despite the rather short stretch of intervening segmental material that might allows for a misparsing of the alignment of the low tone.

5.5 Discussion

Listeners' capability of differentiating between the double nuclear accent contour (H* L- H* L-L%) and the late peak accent contour (H* L+H* L-L%), particularly with regard to the double focus QUD choices makes a strong case for a phonological interpretation of tonal structures. These phonetically rather similar contours that differ only with regard to the alignment of the intervening low tone. Yet they were interpreted to be structurally different, making a phonological analysis of intonation contours imperative. This result in particular points at deficiencies in studies using phonetic descriptions of tonal patterns such as F0 only, to describe focal prominences (Eady et al. 1986; Krahmer & Swerts, 2001).

CHAPTER 6
EXPERIMENTS 2 AND 3: RECOVERING THE LARGER QUD STRATEGY

6.1 Introduction

An underlying assumption in both experimental tasks used in this dissertation is that speakers use intonation to guide listeners' interpretation of the intended meaning and discourse purpose of an utterance. In the model of discourse structure that we have been using - namely, Roberts's (1996) theory of Information Structure as formalized in Büring's (2000) Discourse Tree model - the discourse purpose of an utterance that is an assertion can be equated with the (explicit or invoked) question that dominates it in the QUD. In the previous chapter, we showed that listeners seem to be able to recover the local QUD that dominates the last assertion move before a monologue narrative is interrupted. That is, we adapted the question-matching paradigm of D'Imperio (2000) and others, by equating the question that is matched to the stimulus with an implicit move in the Discourse Tree. This task probes the focus interpretation assigned to the utterance by tapping the conditions on congruence between the Q-alt set of the local QUD and the F-alt set of the answer move that it dominates.

In this chapter, we will describe the results of piloting and using a completely novel paradigm to probe for listeners' focus interpretations. In this novel paradigm, listeners had to select the most appropriate continuation to a story that is interrupted

after the stimulus sentence, based not only on what was said, but also on how it was said. The task requires the listeners to uncover the place of the discourse purpose of an utterance within the larger QUD hierarchy that is being set up by the monologue narrative context. That is, we assume that the local QUD that listeners recovered in the first experimental task is only one of many QUDs forming a larger Discourse Tree. The implicit question move that is invoked by the stimulus sentence in the monologue narrative must be congruent with the answer move itself, but it also must satisfy constraints on coherence across a larger subset of the nodes in the Discourse Tree. Thus the prior context of the narrative not only provides the domain of discourse entities that are included the F-alt set of the stimulus sentence. It also provides the basis for recovering the larger hierarchy of QUDs in which the local QUD is situated. Listeners will have to form an interpretation of the stimulus sentence in order to achieve congruence between the implicitly posed question move and the stimulus sentence, and they also will have to connect that dominance path into a longer path up to the root of the tree, which is the super-question *What is the way things are?* for the story as a whole. The focus interpretation of the stimulus sentence, then, lets the listeners extract the discourse purpose of the utterance by constructing a Discourse Tree, as described in Chapter 4. If this larger Discourse Tree is constraining enough, it should also let the listener anticipate the speaker's next discourse move to some extent, because that move also must fit into the Discourse Tree that is being constructed to embed the QUD invoked by the stimulus sentence in the context of the prior narrative.

We conducted two experiments, so that the first could serve as a pilot experiment to work out problems with the task.

6.2 Experiment 2: Methods

6.2.1 Materials

For this pilot study, the same twenty stories as in the QUD experiment were used. For a full description of the intonation contours, see Chapter 3. The stimulus sentence, carrying one of the four tested intonation tunes, was followed by a visual presentation of three possible continuations to the story, such as the continuations in (6.1), for the *Benjamin heard shots* story shown in (5.2) in the previous chapter. (See Appendix B for the continuations for all of the stories.)

(6.1) Continuation Paradigm (visual presentation):

1. And we ran back to the car as fast as we could.	(broad)
2. And then he heard screams.	(late)
3. And John heard screams.	(double)

These continuations are intended to be plausible in the context of different discourse trees, such as those shown earlier in Figures 4.2 through 4.4. Continuation 1, for example, simply proceeds with an answer to another fairly general question that might be a sister to a broad-focus question, as in Figure 4.2. To construct this Discourse Tree, listeners can assume that the local QUD for the stimulus sentence was a fairly general broad-focus query immediately dominated by the root QUD. The incorporate the continuation, they only need to add another broad-focus QUD that also is immediately dominated by the root QUD of the story as a sister to the stimulus sentence.

Continuation 2 is not as plausible as the sister move that answers the next broad-focus QUD in this simple Discourse Tree. However, it would be quite compatible with a late narrow focus utterance which requires listeners to construct a more complex

intervening structure between the root QUD and the local QUD of the stimulus in order to accommodate to a presupposition that Benjamin heard something - i.e., to accommodate to the presupposition of a late narrow focus. This is the Discourse Structure shown in Figure 4.3.

Continuation 3 suggests a comparable complicated intervening structure, to accommodate to the presuppositions of a double focus QUD, as shown in Figure 4.2.

6.2.2 Subjects

54 Ohio State University Undergraduate students with only limited knowledge of linguistics, naive to the purpose of the study participated in this study. All students were native speakers of American English, which was the only language spoken in their homes during childhood. None of them reported hearing problems and all of them were native speakers of US-English as spoken in the Midwest. Each participant was assigned to one of four groups so that no participant would hear more than one version of any stimulus sentence. Groups 1 and 2 had 14 subjects each, while groups 3 and 4 had 13 subjects each.

6.2.3 Procedure

The stories and the continuations were presented individually to the subjects using a Visual Basic program. The stories were downsampled from 44KHz to 8KHz because the sound card in the PC running the Visual Basic Program could only play 8KHz files.

On each trial, he participant was presented with the background setting and story ending in one version of the stimuli contour, and then visually presented with the three continuation sentences. Note that there was no appropriate continuation for

the early narrow focus contour available for listeners to select, which in the previous experiment, listeners overwhelmingly interpreted as late narrow focus on the object.

Participants were tested individually. They were seated inside a sound attenuated room in front of a computer screen. Via a Visual Basic program, they were auditorily presented with the generic part of the monologue narrative and one of the possible four stimulus utterances (i.e., the sentence with one of the four tonal patterns), repeated twice. As the stimulus sentence was played the first time, three possible continuations to the narrative were presented visually on the computer screen. The stimulus sentence was then repeated once more so that the participant would hear the stimulus sentence twice in total. The participant's task was then to select from the three possible options, the one continuation that they felt best fit the narrative. There was no time pressure on the listeners to make their selection; they went at their own pace from trial to trial, taking as much time as they chose to make a selection from the three continuations before they initiated the presentation of the next mini-dialogue.

6.3 Experiment 2: Results

Results are shown in Table 6.1 and Figure 6.1. Note that data in Table 6.1 shows raw response counts and percentages, while the graph in 6.1 only shows percentages.

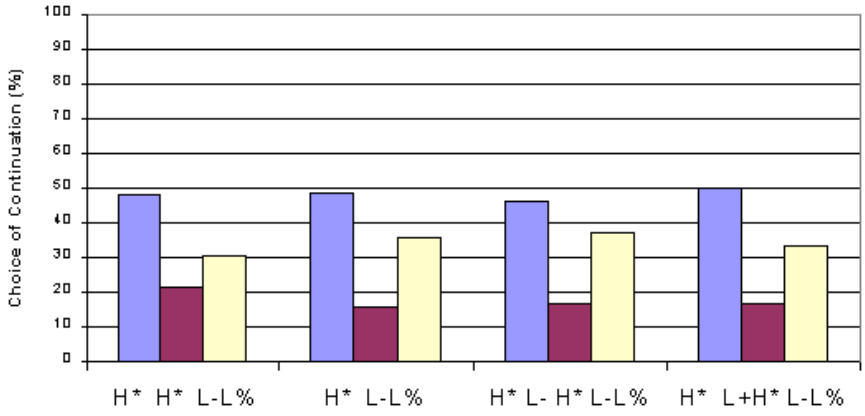

Figure 6.1: Percentage of broad focus, late narrow focus, and double focus continuations chosen in response to the stimulus sentence having one of the four listed intonation contours.

As the graph in 6.1 indicates, listeners did not seem to be influenced by the intonation pattern in the stimulus sentence when choosing the continuation. They appeared to show a slight preference for the simple Discourse Tree that is compatible with the broad focus interpretation when selecting a continuation to the story and a slight dispreference for the more complex Discourse Tree that would accommodate a narrow focus interpretation, and this distribution of broad, late and double responses is very similar across the four contours. (The consistently low percentage of late narrow focus continuations is in marked contrast to the results of Experiment 1, where the late narrow focus QUD was chosen more often than any other QUD for the three contours with accents on both the subject and the object. This suggests that the late narrow continuations were not as well-designed as the broad focus or double focus continuations.)

Contours	H* H* L-L%	H* L-L%	H* L- H* L-L%	H* L+H* L-L%
broad	130 (48.15%)	131 (48.52%)	125 (46.30%)	135 (50.00%)
late	58 (21.48%)	43 (15.92%)	45 (16.67%)	45 (16.67%)
double	82 (30.37%)	96 (35.56%)	100 (37.04%)	90 (33.33%)

Table 6.1: Raw numbers and percentages of broad focus, late narrow focus, and double focus continuation responses elicited after each of the four intonation contour types.

In short, while listeners were able to perform the task of choosing a continuation, and even showed a consistent preference for the continuation that required the simplest Discourse Tree, there was no effect of the intonation contour. As a probe for the preferred focus interpretation of the prosodic structures being compared, it yielded negative results.

We therefore decided to run the paradigm a second time with some small modifications. In particular, instead of waiting until the story was interrupted to present the continuations, we gave participants the three continuations to read while they were listening to the story, and we changed the choice of continuation from a self-paced task to an experimenter-paced task with a short inter-stimulus interval, so that we would be getting the listeners' immediate responses to the auditory stimuli without waiting for them to read the set of continuations. Also, we elicited a confidence rating as a second response measure, to see if there were differences in how confident listeners were in their choice of the same continuation for different intonation contours.

6.4 Experiment 3: Methods

6.4.1 Subjects

In total, 64 Ohio State University Undergraduate students with only limited knowledge of linguistics participated in this study. All students were native speakers of American English (as spoken in the Midwest), which was the only language spoken in their homes during childhood. None of them reported hearing problems. Again, subjects were assigned to one of four equal-sized groups, so that none would hear more than one version of any stimulus sentence.

6.4.2 Stimuli

The same background and stimuli sentences as in the pilot study (see Appendix B for details) were used in this experiment. They were recorded onto DAT tape at 22KHz.

6.4.3 Procedure

As in Experiment 2, each experiment participant was individually seated inside a sound attenuated room and was given headphones to wear. In contrast to the first continuation experiment, where listeners interacted with a computer screen, in this set up listeners the three continuation sentences on a paper response sheet, so that they could be reading the continuations as they were listening to the background narrative and stimulus sentence played to them off a DAT tape. The interstimulus interval was 25 seconds. That is, after hearing the stimulus sentence, the listeners had 24 seconds to respond before a beep alerted them to pay attention to the next trial. The beep was followed by one second of silence before the next trial started.

The instructions in (6.2) were given to the listeners. Note that the listeners was

specifically alerted to listen not only to what was said but also to how it was said. (6.2)

> Please listen carefully. You will hear a recording of the beginning of a story. The story will be interrupted before the speaker finishes. The last sentence before the interruption will then be replayed. There are three possible continuations which you should read after you hear the repeated sentence. Pick the continuation that fits the story best.
>
> Make sure to listen not only to what is said but also to HOW (particularly the last sentence) it is said.
>
> Once you have made your choice, indicate on the scale below with a line, how good a choice the continuation was.
>
> 1. And we ran back to the car as fast as we could.
> 2. And then he heard screams.
> 3. And John heard screams.

```
very good                                        very bad
     |----------|----------|----------|----------|
     1          2          3          4          5
```

Please indicate how well you think your choice of continuation fits the story.

6.5 Experiment 3: Results

6.5.1 Distribution of Continuation Choices

Table 6.2 shows the raw numbers and percentages of the different continuation types elicited in response to each of the intonation patterns in the continuation task, and Figure 6.2 summarizes the percentages as a barchart. (See Appendix D for a by-item listing of the percentage of broad, late and double rating results for this continuation choice experiment.)

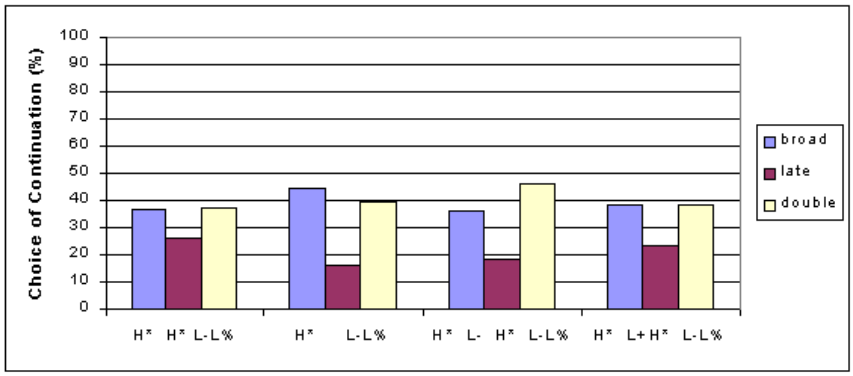

Figure 6.2: Percentage of broad (b), late (l), and double (d) ratings given in response to the stimulus sentence having one of the four listed intonation contours.

Table 6.3 shows the same data organized in a slightly different way, as a "deviation score". That is, we calculated the mean response (i.e., the average number of broad, late or double response would be across all contours) and subtracted these mean from the corresponding percentages for that continuation in response to a particular contour type.

	H* H* L-L%	H* L-L%	H* L- H* L-L%	H* L+H* L-L%	mean %
broad	117 (36.56%)	143 (44.69%)	115 (35.94%)	123 (38.44%)	38.9
late	83 (25.94%)	51 (15.94%)	58 (18.13%)	75 (23.44%)	20.8
double	120 (37.50%)	126 (39.38%)	147 (45.94%)	122 (38.13%)	40.2

Table 6.2: Raw numbers and percentages of elicited broad, late and double continuation responses for all four intonation contour types.

The data show the same dispreference for the continuation that accommodated a late narrow focus interpretation as in Experiment 2, but now there was not the preference for the continuation that assumed the simplest Discourse Tree of the broad focus interpretation. Also, the distribution of continuation choices now differed across the four contour types, with the double nuclear accents contour eliciting the fewest broad focus continuations and a greater number of double focus continations. A 3x4 χ^2 analysis of all contours and all responses leads us to reject the Null-Hypothesis that listeners reacted in the same way to all contours. That is, this test revealed a significant difference in the continuations chosen in response to the different contours ($\chi^2 = 17.38$, df $= 6$, p< 0.01). While this effect certainly is not as strong as the effect of contour type in the QUD paradigm, it is significant, showing that listeners were influenced by the prosody not just in recovering the local QUD for the stimulus, but also in constructing a plausible larger Discourse Tree in which that local QUD could be embedded.

The data in Table 6.3 are the deviations from the mean score for each contour for each response. This score tells us which by how much and in which direction the

listeners' responses on average deviated from the mean for any particular response type. The largest deviations from the mean score for each contour (in each column) are highlighted by the italics.

Deviation Score	H* H* L-L%	H* L-L%	H* L- H* L-L%	H* L+H* L-L%
broad	-2	*5.8*	-3	-0.5
late	*5.1*	-4.9	-2.7	*2.6*
double	-2.7	-0.8	*+5.7*	-2.1

Table 6.3: Deviation Score (% − mean %) for broad, late and double responses.

The largest deviation score for the hat pattern (H* H* L-L%) is 5.1 for the late continuation responses, indicating that listeners gave more than the average number of late focus responses to this contour. This is in line with the predictions we made that a hat pattern could receive late narrow focus interpretations, and that the prenuclear accent is not focal but serves rhythmic purposes. The deviation score of 5.7 for the double nuclear accents contour indicates that listeners on average gave more than the average number of double focus continuation responses to this contour. This result also is as predicted: listeners prefer a double focus continuation when prompted with a double nuclear accents contour. For the rising nuclear accent contour , we find the largest deviation from the mean to be 2.6 for the late narrow focus category. This indicates that listeners did show some preference for the late continuation when prompted with the rising nuclear accent contour, even though this was still the minority response even for this contour.

As in Chapter 5, we give a more detailed statistical description of how well the continuation choices conformed to our expectations from the predicted difference in focus interpretation.

H* H* L-L% versus H* L- H* L-L%

A 2x3 χ^2 analysis of all responses to the hat pattern (H* H* L-L%) and the double nuclear accents pattern (H* L- H* L-L%) revealed a significant effect of intonation contour on the distribution of continuation choices ($\chi^2 = 7.18$, df $= 2$, p$>$.05), indicating that listeners showed a different pattern of response to these two contours. A one-tailed Wilcoxon Signed-Ranks test of the percentage of double focus continuations showed a significant effect of contour for the both the subject analysis (W = 359, N = 51, p $<$ 0.05) and the item analysis (W = -100, N = 17, p $<$ 0.01,), as predicted in Chapter 4.

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

A 3x2 χ^2 analysis on all responses (broad, late, double) to these two contours yielded non-significant results ($\chi^2 = 0.57$, df $= 2$, p$>$ 0.05), despite the difference in deviation scores in Table 6.3. (This adds further support to our suspicion that the late narrow focus continuations were not as well-designed as the other continuations.)

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

A 3x2 χ^2 analysis on all responses (broad, late, double) to these two contours yielded non-significant results ($\chi^2 = 4.77$, df $= 2$, p$>$ 0.05. However, Wilcoxon tests did show significantly more late narrow focus continuations in response to the rising nuclear accent contour, albeit albeit only in the item analysis (item analysis W = 49, N=15, p (1-tailed)$<$ 0.05,). This result is in keeping with our prediction, and differs from the negative results of Experiment 2.

6.5.2 Confidence Ratings

In Experiment 2, we obtained a rating of the listener's confidence in his or her judgement of the likelihood of the chosen continuation. Listeners were asked to indicate on a scale from 1 to 5 how confident they were that the continuation they selected was a good fit with the preceding context. A low value indicated good confidence while a high value indicated a lack of confidence. A value of 3 indicated neither strong discomfort nor strong preference. Since each item was presented 16 times over the course of this experiment, we obtained 16 confidence ratings for each item. For each contour, we added all confidence ratings for the broad, late and double responses, and divided them by the total sum of confidence ratings to obtain a single mean confidence rating per contour for each response type. The higher the mean, the less confident listeners were on average.

Figure 6.3 shows these mean confidence ratings plotted as a function of the percentage of times that continuation was chosen. On the scatterplot we have reversed the direction of the y-axis, so that datapoints indicating less confidence on average (higher y values) lie lower in the graph and datapoints indicating more confidence on average (lower y values) lie high in the graph. Thus, a datapoint in the lower left quadrant of the graph means that few participants chose that continuation and they were not very confident when they did make that choice, whereas a datapoint in the upper right quadrant means that many participants chose that continuation and they tended to be confident in their selection.

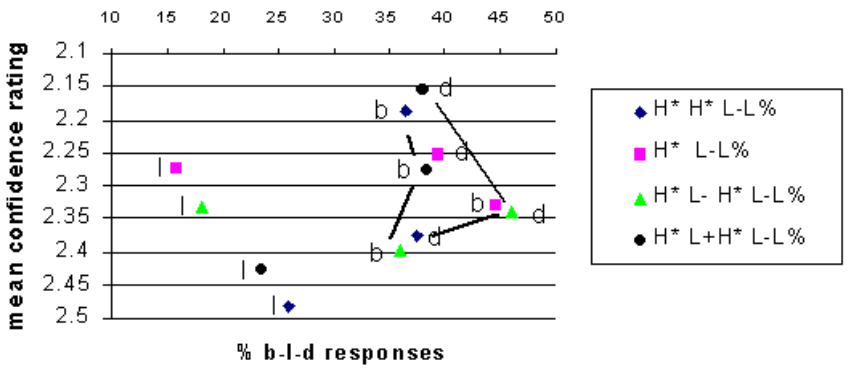

Figure 6.3: Confidence Ratings as a function of the broad, late and double continuation responses to the stimuli sentences.

A first general pattern to note in this graph is that the late narrow focus continuations did not receive very good mean confidence ratings for any of the intonation contours, and got lower confidence ratings than either of the other two continuations in response to the hat pattern and the rising nuclear accent pattern - the two intonation patterns that elicited the greatest number of late narrow focus responses. Thus, even when participants chose this continuation type, they were not confident in their choice, suggesting once again that there was something infelicitous about the continuation sentences here.

A second pattern to note is the relationship between the confidence rating and the preferences for the other two continuation types for the three contours with accents on the object as well as on the subject. For the double nuclear accents contour, the confidence ratings and the preferences go hand in hand. Participants chose the double focus continuation more often than the broad focus continuation, and they tended to be more confident when they did so. For the hat pattern and the rising nuclear accent

pattern, however, confidence ratings give us information that is independent of the preferences. These two metrically similar contours both received an equal number of broad focus and double focus continuations, but confidence ratings differed. For the hat pattern, participants were more confident when they chose the broad focus continuation, whereas for the rising nuclear accent, they were more confident of the double focus continuation. The higher confidence in the double focus interpretations for the rising nuclear accent contour suggests another way of invoking the strategy in Figure 4.3. That is, where the first nuclear accent in the double nuclear accents contour is a "Contrastive Topic" (see Büring, 2000), and quite compatible with this QUD hierarchy, a late nuclear focus interpretation might invoke the almost identical Discourse Tree shown in Figure 6.4. That is, the strategy of answering *Who heard what?* by asking a series of sub−questions about different subjects could be instantiated with sub−questions where the subjects are taken as merely given by the context and not marked as explicitly contrasting with each other.

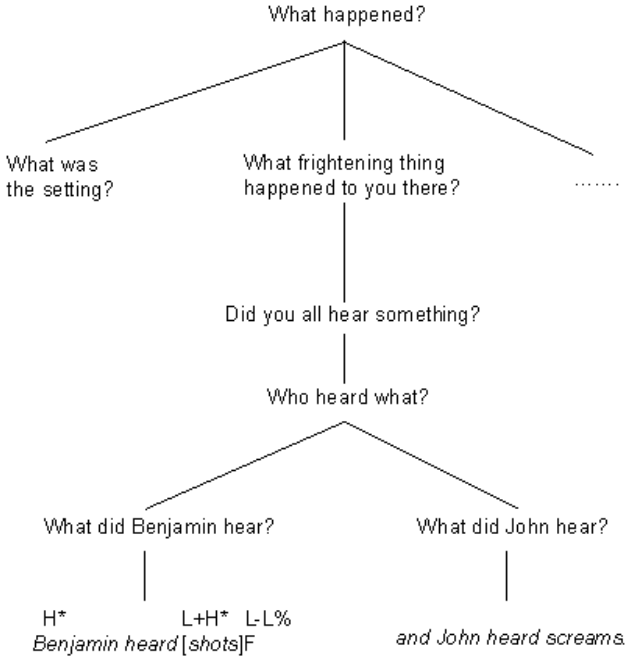

Figure 6.4: Alternative Discourse Tree Evoked by Late Narrow Focus Interpretation.

6.6 Summary & Discussion

The negative results obtained in the pilot study suggest that the first task was not sufficient to test the interpretation of intonation contours in monologue narratives. By changing the mode of presentation and restricting the time listeners had to make a selection from the three continuations to the monologue, we found a different response pattern from the one obtained in the pilot study. This methodological insight suggests that listeners have to respond to a still fresh auditory memory to perform this task. The confidence ratings in the experiment with enforced short inter-stimulus intervals also were informative, suggesting that it would be fruitful to gather multiple responses

of other types as well, such as reaction times. The unexpected results for the late narrow focus continuations, on the other hand, suggest that any further applications of this paradigm be preceded by norming studies to find plausible continuations that can better differentiate the strategies invoked by late narrow focus from the Discourse Trees invoked by double focus interpretations.

CHAPTER 7

SUMMARY AND CONCLUSION

7.1 Summary of Results

In this dissertation we have presented two different experimental paradigms that were designed to assess the contribution of four intonation structures to listeners' interpretation of discourse intents. The two paradigms that we used were the QUD paradigm and the Continuation paradigm. Both paradigms involved listening to an interrupted narrative, and making an interpretive judgment about the last sentence before the interruption. In the QUD paradigm, listeners chose one of four questions that was most appropriately addressed by this stimulus sentence. In the Continuation paradigm, listeners chose one of three sentence that the most appropriate continuation to follow the stimulus sentence. The stimulus sentence could have any one of four different intonation contours. These tasks were devised to tap listeners' competence in interpreting the focus structure of an utterance utilizing tasks that more closely resemble actual processing of discourse intents. Because the judgments involved the kinds of discourse intents that we assume focus is related to, they were an ecologically more natural way to probe for the focus interpretation that the listeners assigned to the varying intonation patterns that they heard. We will begin our summary, therefore, by reviewing what the experiments tell us about the focus interpretations of the four intonation contours.

7.1.1 H* L-L%

Our results from the QUD experiment (Chapter 5) concur with all of the earlier literature discussed in Chapters 2-4. Intonation contours with early nuclear accents and long unaccented postnuclear tails are readily interpreted as early narrow focus.

7.1.2 H* H* L-L% versus H* L- H* L-L%

One of the questions we set out to address is whether or not listeners interpreted the hat pattern and the double nuclear accent pattern differently. The results from the continuation paradigm as well as the QUD paradigm indicate that listeners have a strong tendency to interpret two nuclear accents as double focus, while they interpret the hat pattern as broad focus or as late narrow focus. We take this as supportive evidence that nuclear accent placement determines the focal structure of an utterance.

7.1.3 H* L+H* L-L% versus H* L- H* L-L%

Further support for this conclusion comes from comparing the double nuclear accents pattern to the rising nuclear accent pattern. In the QUD task, the rising nuclear accent pattern got more late narrow focus responses than any other pattern, and in both tasks, it did not get as many double focus responses as the double nuclear accent pattern. These two contours are phonetically very similary, but at the phonological level we can clearly distinguish between them. The H* L- H* L-L% pattern has two nuclear accents, whereas the H* L+H* L-L% is metrically identical to the hat pattern, with the first peak being a prenuclear. The fact that listeners interpret these two contours differently thus indicates that a phonological interpretation of intonational tunes is not only convenient but also necessary. A gross phonetic description of the double peaks separated by an intervening trough does not predict that the rising

nuclear accent contour will be interpreted more similarly to the metrically identical hat pattern instead of to the phonetically similar double nuclear accents pattern. The phonological description of one versus two nuclear accents, on the other hand, predicts the focus interpretation, and allows us to attend to the more subtle phonetic differences that fall out from the different phonological interpretation of the low tone intervening between the accent peaks.

7.1.4 H* H* L-L% versus H* L+H* L-L%

Comparing the hat pattern contour and the rising nuclear accent contour, on the other hand, brings out a consequence of the meaning of the L+H* accent for the focus pattern, and highlights the importance of differentiating between different types of peak accents. In the QUD paradigm, listeners favored late narrow focus responses for both of these contours, but the preference for late narrow focus over broad focus was stronger for the rising nuclear accent pattern. In the continuation paradigm, listeners chose a broad focus continuation equally often for the hat pattern and for the rising nuclear accent pattern, but they were much more confident in this choice for the hat pattern.

7.2 Methodological Issues

Another over-arching result that we can cite here is that the two tasks did work, to the extent that responses were influenced by the intonation contour of the stimulus sentence, in ways that could be interpreted more or less in terms of the predicted focus structures. Nonetheless, there are two methodological issues that need to be addressed. One is the contribution of the task. The distribution of replies in the QUD paradigm is sometimes different from the distribution of replies in either of

the continuation paradigm experiments. The second issue is the contribution of the mode and time course of presentation in the continuation paradigm experiments. The different results for the two continuation paradigm experiments already addresses the second issue. It seems that in order for there to be an effect of the intonation contour on the continuation choice, the response must be elicited soon after the narrative is interrupted.

Turning to the first issue, the biggest effect of task involved the late narrow focus responses. In the QUD task, the late narrow focus Wh−question was the most frequent choice for the three intonation contours other than the early nuclear accent contour, and it was the majority choice for the rising nuclear accent contour. In the continuation paradigm, by contrast, listeners appeared to strongly disfavor late narrow focus continuations for all four intonation contours. This suggests that there was some problem with our choice of sentences to exemplify a late narrow focus continuation. In designing our late narrow focus continuations, we assumed that a Discourse Tree where two sister QUDs contrasting objects in the stimulus sentence and the continuation would be a good narrow focus continuation. Our results, however, suggest that this assumption was a bit too simple. Figure 4.5 suggests an alternative Discourse Tree that would accommodate a late narrow focus interpretation for the stimulus sentence. A conclusion to draw from this result, then, is in future use of the Continuation Paradigm, the materials need to be normed, perhaps in tests where subjects are asked to construct Discourse Trees themselves, using a DT tagging tool.

7.3 Conclusion

In this dissertation, we have presented evidence in favor of a phonological interpretation of the intonation contour. Listeners do not interpret F0 values and vowel durations directly as cues to the focus pattern. Rather, they interpret the phonetic patterns in terms of a small inventory of phonologically contrasting tonal categories and the contrasting metrical structures with which they are associated. Differences in metrical structure are related to information structure, but again there is not a direct mapping between phonological categories and discourse structures. Not every accent is focal and clearly there is a difference between nuclear and prenuclear accents.

We have also pointed out several of the pitfalls when studying the interpretation of intonation contours in context. Of course, much work remains to be done to address this issue. We have also made the case for the need for methodologically evaluating experimental paradigms and their validity in testing claims, as we have seen that some results may be task dependent.

APPENDIX A

LIST OF MATERIALS

A.1 Target Sentence 1: Benjamin heard shots

Setting:

Susi was telling a story of what happened on a walk with her friends Ben and Maya and John last February.

Narrative Monologue:

"It was a cold day, and we were walking around Hocking Hills. It was very quiet out there, and there was nobody else there but us and it looked completely deserted. Benjamin heard shots...

A.1.1 Questions for QUD Experiment:

1. What happened? (broad)
2. What did Benjamin hear? (late)
3. Who heard what? (double)
4. Who heard shots? (early)

A.1.2 Continuations for Pilot Study & Experiment 2:

1. And we ran back to the car as fast as we could. (broad)
2. And then he heard screams. (late)
3. And John heard screams. (double)

A.2 Target Sentence 2: Jonathan wore shorts.

Setting:

John's mother was telling her girlfriends about her son's wedding.

Narrative Monologue:

When John and Mary got married, they had a rather non-traditional ceremony. Johnathan wore shorts.

A.2.1 Questions for QUD Experiment:

1. What was going on? (broad)
2. What did Jonathan wear? (late)
3. Who wore what? (double)
4. Who wore shorts? (early)

A.2.2 Continuations for Pilot Study & Experiment 2:

1. And they hired a Heavy Metal band to play at the party. (broad)
2. He would have worn his dark suit, but he decided to go informal. (late)
3. And Mary wore a bikini. (double)

A.3 Target Sentence 3: Jessica took pictures.

Setting:

The Chicago banker tells her friends about her last family vacation with her mother, father, and her sister Jessica to Ohio:

Narrative Monologue:

We were driving around Hocking Hills, and there we saw the most beautiful 10-foot stalactites. Jessica took pictures.

A.3.1 Questions for Experiment 1:

1. What happened? (broad)
2. What did Jessica take? (late)
3. Who took what? (double)
4. Who took pictures? (early)

A.3.2 Continuations for Pilot Study & Experiment 2:

1. We walked around there most of the afternoon. (broad)
2. And she took slides. (late)
3. And Mom took slides. (double)

A.4 Target Sentence 4: Catherine bought jello.

Setting:

Lydia and her roommates went shopping at the grocery store. It was a hot day, and they were outside and played ball all day. She tells her Dad on the phone what they

did that day:

Narrative Monologue:

First we were hanging out on the Oval 'cause the weather is so nice, and then around five, we decided to go shopping. Catherine bought jello.

A.4.1 Questions for Experiment 1:

1. What happened? (broad)
2. What did Catherine buy? (late)
3. Who bought what? (double)
4. Who bought jello? (early)

A.4.2 Continuations for Pilot Study & Experiment 2:

1. But we had a hard time deciding on the flavor. (broad)
2. And she bought bagels. (late)
3. And Joe bought bagels. (double)

A.5 Target Sentence 5: Jennifer bought shoes.

Setting:

Martha and some of her friends went shopping at the City Center Mall last week and there they observed a robbery. Martha was telling the police.

Narrative Monologue:

First we wanted to go swimming because we didn't feel like shopping, but then we went to the mall to hang out there for a bit. Jennifer bought shoes.

A.5.1 Questions for Experiment 1:

1. What happened? (broad)
2. What did she buy? (late)
3. Who bought what? (double)
4. Who bought shoes? (early)

A.5.2 Continuations for Pilot Study & Experiment 2:

1. We hung out in that store for quite a while. (broad)
2. And then she bought socks. (late)
3. And I bought sandals. (double)

A.6 Target Sentence 6: Christopher made hummus.

Setting:

Last Sunday, a group of frat boys moved into their new apartment on Neil Avenue. Jerry's mother tells how they spent their moving day.

Narative Monologue:

They had so much to do, it was unbelievable. First they went shopping to get some supplies to disinfect the place. And then they started cleaning like maniacs until about dinner time when they got really hungry. Christopher made hummus.

A.6.1 Questions for Experiment 1:

1. What happened? (broad)
2. What did Christopher make? (late)

3. Who made what? (double)

4. Who made hummus? (early)

A.6.2 Continuations for Experiment 2:

1. But they were still hungry. (broad)

2. And then he made pizza. (late)

3. And Joe made pizza. (double)

A.7 Target Sentence 7: Benjamin served lemonde.

Setting:

Laurie tells about the beautiful day that the departmental BBQ-potluck took place.

Narrative Monologue:

"We had a great time and things were going great. The weather was perfect, and the lawn was freshly cut. Benjamin served lemonade.

A.7.1 Questions for Experiment 1:

1. What happened? (broad)

2. What did Benjamin serve? (late)

3. Who made what? (double)

4. Who made lemonade? (early)

A.7.2 Continuations for Pilot Study & Experiment 2:

1. And most people waited in line for the burgers. (broad)
2. And he served beer. (late)
3. And Bob served beer. (double)

A.8 Target Sentence 8: Grandmother made dinner.

Setting:

The famous author gave an interview and was asked about her family. She said:

Narrative Monologue:

In many ways we were a rather traditional family, a bit like 'Father Knows Best'. Grand-Mother made dinner.

A.8.1 Questions for Experiment 1:

1. What happened? (broad)
2. What did Grand-Mother make? (late)
3. Who made what? (double)
4. Who made dinner? (late)

A.8.2 Continuations for Pilot Study & Experiment 2:

1. Most days were the same, but sometime we would go out for picnics at the farm. (broad)
2. And she made lunch, too. (late)
3. And Dad made breakfast. (double)

A.9 Target Sentence 9: Millerman insulted Reynolds.

Setting:

During a recent local election campaign, two of the leading candidates for mayor, Millerman and Reynolds, were debating school funding issues.

Narrative Monologue:

The debate turned into a heated discussion and the news reports talked about a rather specific incident the next day: Millerman insulted Reynolds. Millerman insulted Reynolds.

A.9.1 Questions for Experiment 1:

1. What happened? (broad)
2. Who did Millerman insult? (late)
3. Who insulted whom? (double)
4. Who insulted Reynolds? (early)

A.9.2 Continuations for Pilot Study & Experiment 2:

1. And for that he was cited. (broad)
2. And then he insulted the board of Trustees. (late)
3. And then his wife insulted Reynold's wife. (double)

A.10 Target Sentence 10: Christopher ordered nachos.

Setting:

An article in a gourmet magazine recently started a new series in which college students were asked to rate local greasy spoon places in Columbus. For the June edition,

a group of freshmen tested 'Greasy Spoon Inc.', a new hang-out on High Street. Here is Mary's report:

Narrative Monologue:
We went there on a Monday. We were in a great mood and really hungry. Luckily, the waiter attended to us right away. Christopher ordered nachos.

A.10.1 Questions for Experiment 1:

1. What happened? (broad)
2. What did Christopher order? (late)
3. Who ordered what? (double)
4. Who ordered nachos? (early)

A.10.2 Continuations for Pilot Study & Experiment 2:

1. And then we discussed what entrees to order. (broad)
2. And he ordered fish. (late)
3. And Fred ordered steak. (double)

A.11 Target Sentence 11: Anthony met 'Sting'.

Setting:
The cousins were telling their aunt about having spent their summer vacation in London. Their aunt then told a friend about what she had just heard from the cousins:

Narrative Monologue:

It was raining almost every day, and they went to museums and galleries and concerts and stuff. In a random drawing a radio station gave away backstage passes for a concert. Anthony met 'Sting'.

A.11.1 Questions for Experiment 1:

1. What happened? (broad)
2. Who did Anthony meet? (late)
3. Who met whom? (double)
4. Who met 'Sting'? (early)

A.11.2 Continuations for Pilot Study & Experiment 2:

1. And the paparazzi took pictures. (broad)
2. And he met 'Prince'. (late)
3. And Barbara met 'Prince'. (double)

A.12 Target Sentence 12: Barbara switched jobs.

Setting:

At the family gatherings of the Smiths, as usual, everyone was gossiping. Mary was talking to Joe:

Narrative Monologue:

I am really happy to see you again. Did you hear the news? Barbara switched jobs.

A.12.1 Questions for Experiment 1:

1. What happened? (broad)
2. What did Barbara switch? (late)
3. Who switched what? (double)
4. Who switched jobs? (early)

A.12.2 Continuations for Pilot Study & Experiment 2:

1. She now makes a ton of money. (broad)
2. And she switched boyfriends. (late)
3. And Dora switched husbands. (double)

A.13 Target Sentence 13: Tiffany writes songs.

Setting:

A Bexley woman has the following to say about her neighbors:

Narrative Monologue:

I've known this family for 17 years. The parents are friendly people who work a lot and appreciate the arts. Tiffany writes songs.

A.13.1 Questions for Experiment 1:

1. What are they about? (broad)
2. What does Jessica write? (late)
3. Who writes what? (double)
4. Who writes songs? (early)

A.13.2 Continuations for Pilot Study & Experiment 2:

1. And in the summer they have parties in their garden. (broad)
2. And she writes poetry. (late)
3. And Wally writes poetry. (double)

A.14 Target Sentence 14: Mellany served fish.

Setting:

In the 1997 Martha Stewart recipe contest, the finalists were invited to the studio to prepare their dish. The proud grandmother of the winner told his chess friends:

Narrative Monologue:

The studio air was filled with exotic smells. The tables were artfully set. Mellany served fish.

A.14.1 Questions for Experiment 1:

1. What was going on? (broad)
2. What did Mellany serve? (late)
3. Who served what? (double)
4. Who served fish? (early)

A.14.2 Continuations for Pilot Study & Experiment 2:

1. And the judges really admired the garnishing. (broad)
2. And she served asparagus. (late)
3. And the others served steak. (double)

A.15 Target Sentence 15: Robertson washed cars.

Setting:

In her memoirs, one of the first female newscasters describes the time of the post-war depression.

Narrative Monologue:

Some people stayed rich but most people became poor. Whole families were starving. Nobody had secure jobs at that time. Robertson washed cars.

A.15.1 Questions for Experiment 1:

1. What happened? (broad)
2. What did Robertson wash? (late)
3. Who washed what? (double)
4. Who washed cars? (early)

A.15.2 Continuations for Pilot Study & Experiment 2:

1. And every evening he put a few dimes aside. (broad)
2. And he washed windows. (late)
3. And I washed windows. (double)

A.16 Target Sentence 16: Stephanie chopped wood.

Backround:

The aging publisher of the New York Times tells:

Narrative Monologue:

It was just really nice to get out of the office. Each year, we had a fun-filled 4th of July picnic on Martha's Vineyard. One afternoon I remember particularly well. Stephanie chopped wood.

A.16.1 Questions for Experiment 1:

1. What happened? (broad)
2. What did Stephanie chop? (late)
3. Who chopped what? (double)
4. Who chopped wood? (early)

A.16.2 Continuations for Pilot Study & Experiment 2:

1. And then she started the grill. (broad)
2. And she chopped vegetables for the BBQ. (late)
3. And Elly chopped vegetables. (double)

A.17 Target Sentence 17: Jonathan switched lanes.

Setting:

After the accident, the report read:

Narrative Monologue:

It had just stopped raining. The street was still wet when the sun came back out. The radio was blaring. Jonathan switched lanes.

A.17.1 Questions for Experiment 1:

1. What happened? (broad)
2. What did Jonathan switch? (late)
3. Who switched what?c (double)
4. Who switched lanes? (early)

A.17.2 Continuations for Pilot Study & Experiment 2:

1. And then the other car started to overtake us. (broad)
2. And he switched glasses. (late)
3. And Jerry switched glasses. (double)

A.18 Target Sentence 18: Jessica saw rats.

Setting:
At the Office of Residence and Dining at OSU, complaints were filed about the cleanliness of the commons and the courteousy of the staff. One letter from a Drackett Tower resident said:

Narrative Monologue:
We have been eating there twice a day now for a year. The food is greasy, cold when you get it slapped on your plate, and overall just really bad in quality. We are also wondering about the state of hygiene. Jessica saw rats.

A.18.1 Questions for Experiment 1:

1. What happened? (broad)
2. What did Jessica see? (late)

3. Who saw what? (double)

4. Who saw rats? (early)

A.18.2 Continuations for Pilot Study & Experiment 2:

1. But they didn't really care about it. (broad)

2. And she saw roaches. (late)

3. And Ken saw roaches. (double)

A.19 Target Sentence 19: Gideon did disco.

Setting:

Last year, all graduating seniors from OSU went to Mekka, a Columbus dance place with several rooms in which they play different kinds of music, and celebrated the end of their time in school. A member of the organizing committee reports to the Lantern:

Narrative Monologue:

We rented the entire place and decorated everything in scarlet and gray and the whole event was a total success. Gideon did disco.

A.19.1 Questions for Experiment 1:

1. What happened? (broad)

2. What did Gideon do? (late)

3. Who did what? (double)

4. Who did disco? (early)

A.19.2 Continuations for Pilot Study & Experiment 2:

1. But for sure he is no John Travolta. (broad)
2. And then he did Salsa. (late)
3. And the Dean did Salsa. (double)

A.20 Target Sentence 20: Johnathan ordered beer.

Setting:

The Other Paper reported on an open stage poetry reading at Larry's last fall during which the owner of the bar had to intervene. A Larry's patron said:

Narrative Monologue:

The room was quiet, and people were listening to the poetry. It was a really long piece and we started to get bored and thirsty. Johnathan ordered beer.

A.20.1 Questions for Experiment 1:

1. What happened? (broad)
2. What did Jonathan order? (late)
3. Who ordered what? (double)
4. Who ordered beer? (early)

A.20.2 Continuations for Pilot Study & Experiment 2:

1. And the artist got totally offended by him breaking the silence. (broad)
2. And then he wanted some peanuts also. (late)
3. And Mary ordered wine. (double)

APPENDIX B

PLOTS OF F0 MEASUREMENT IN SUBJECT, VERB & OBJECT POSITION

B.1 Descriptions

Along the x−axis are five measurement points, the first one corresponds to the initial vowel in the subject, the second to the second vowel in the subject, and the third corresponds to the third vowel in the subject. The fourth point corresponds to the F0 in the verb while the fith point is the F0 in the object.

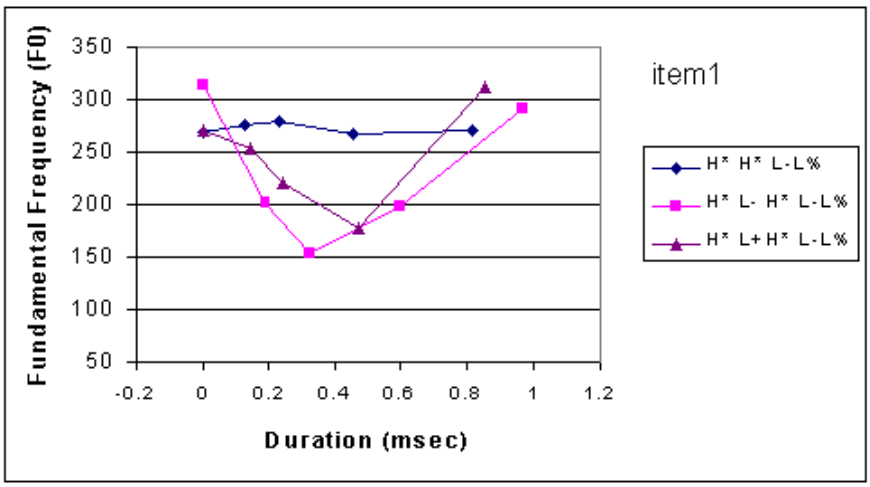

Figure B.1: Comparison of F0 in Subject, Verb and Object position in item 1, *Benjamin heard shots.*

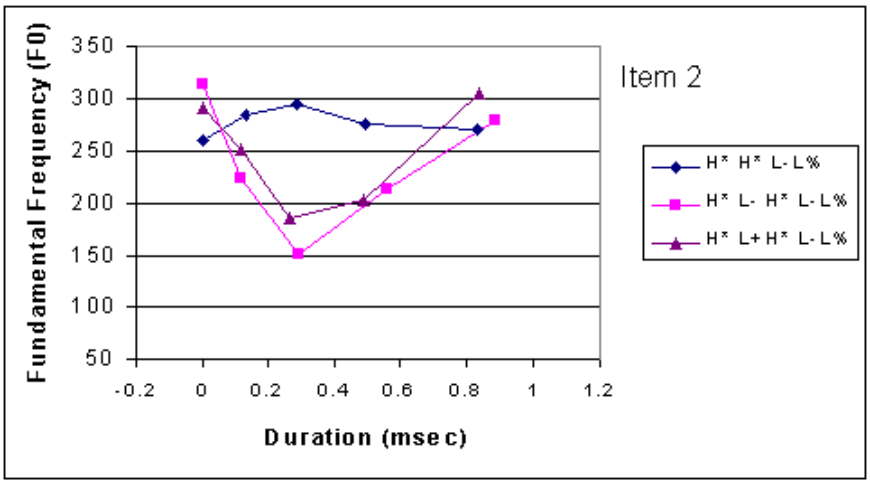

Figure B.2: Comparison of F0 in Subject, Verb and Object position in item 2, *Johnathan wore shorts.*

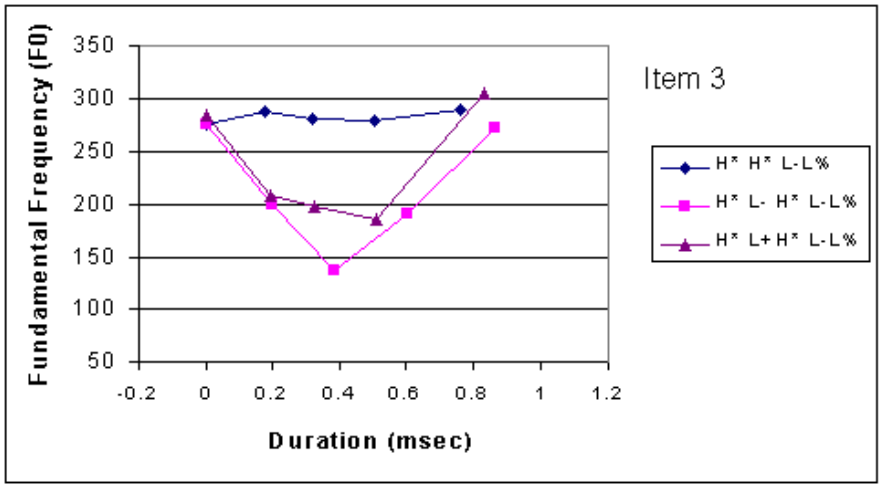

Figure B.3: Comparison of F0 in Subject, Verb and Object position in item 3, *Jessica took pictures.*

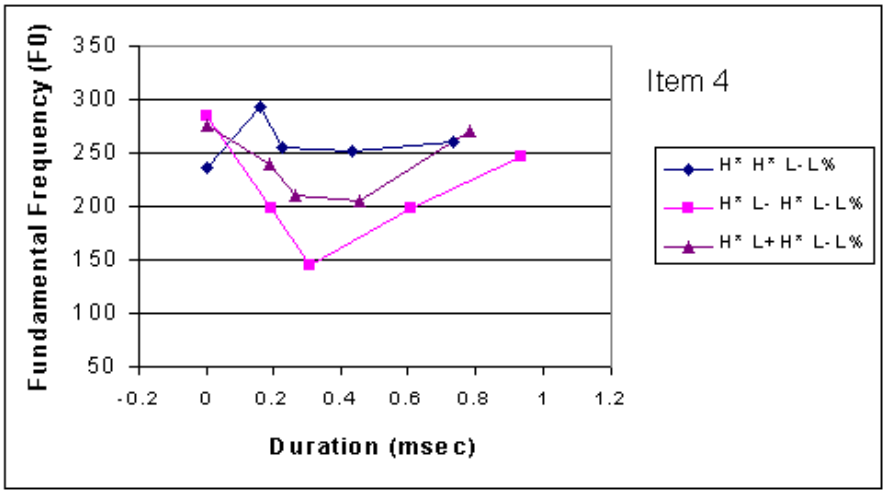

Figure B.4: Comparison of F0 in Subject, Verb and Object position item 4, *Catherine bought jello.*

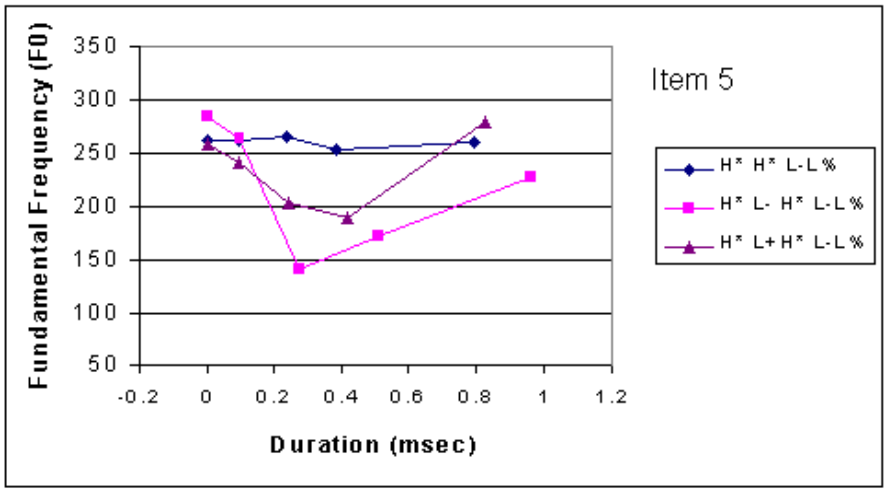

Figure B.5: Comparison of F0 in Subject, Verb and Object position in item 5, *Christopher made hummus.*

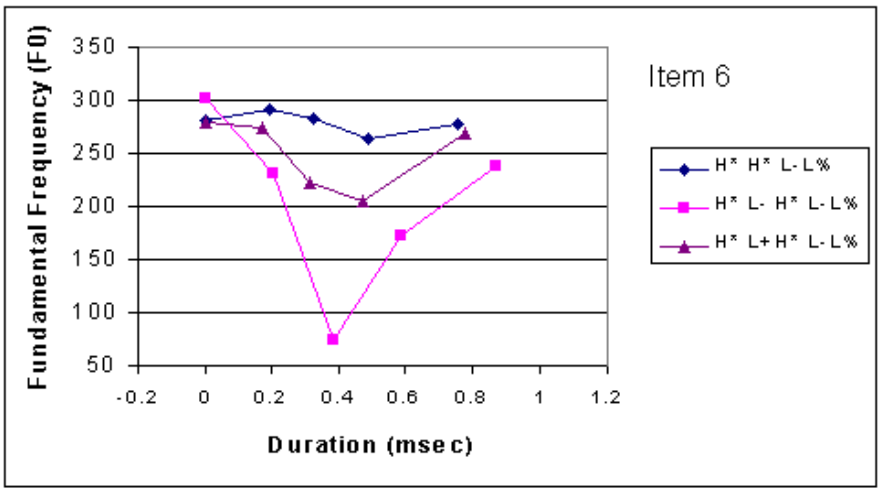

Figure B.6: Comparison of F0 in Subject, Verb and Object position in item 6, *Jennifer bought shoes.*

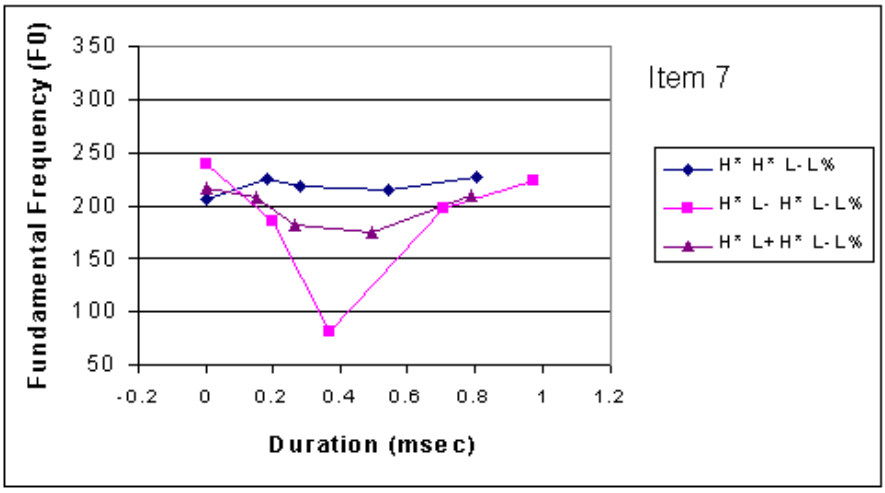

Figure B.7: Comparison of F0 in Subject, Verb and Object position in item 7, *Grandmother made dinner.*

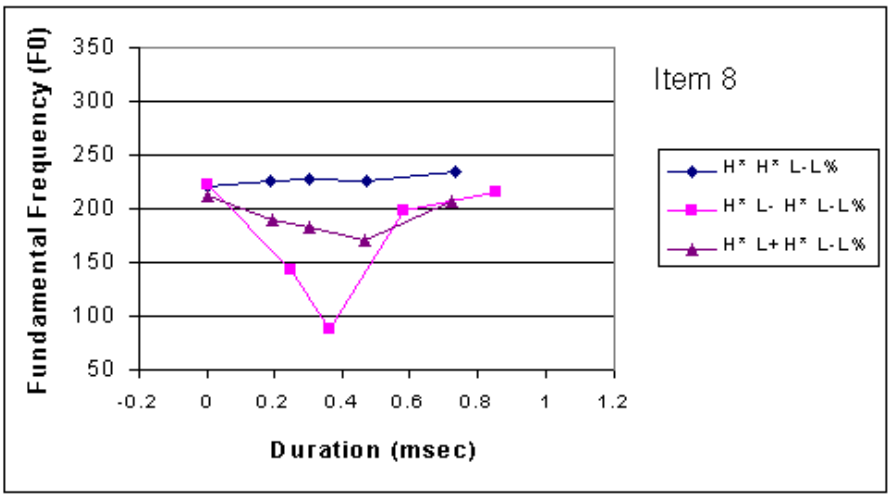

Figure B.8: Comparison of F0 in Subject, Verb and Object position in item 8, *Benjamin served lemonade.*

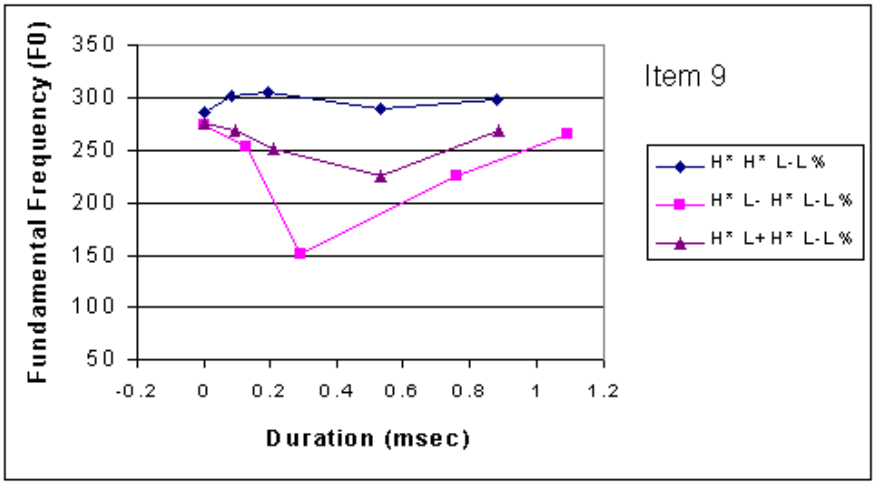

Figure B.9: Comparison of F0 in Subject, Verb and Object position in item 9, *Miller-man insulted Reynolds.*

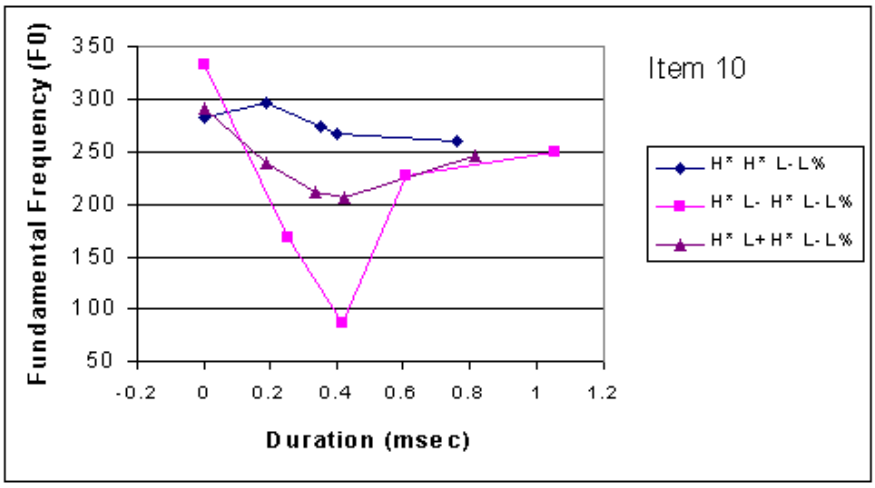

Figure B.10: Comparison of F0 in Subject, Verb and Object position in item 10, *Christopher ordered nachos.*

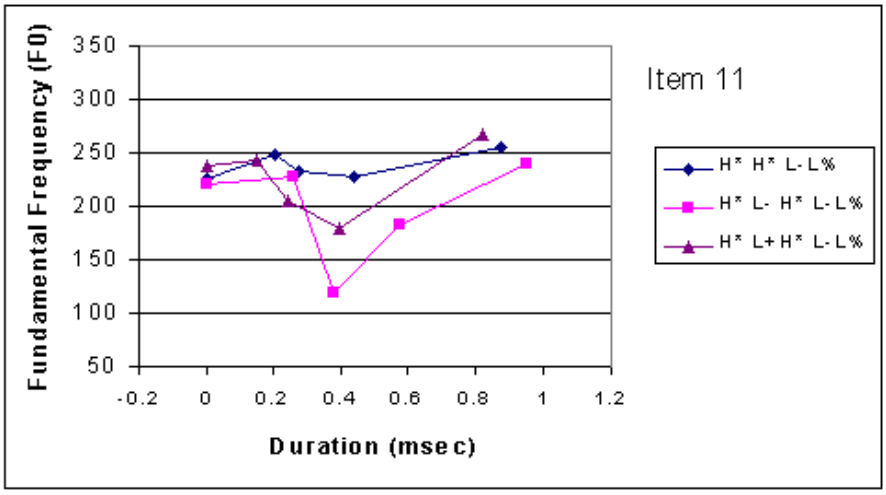

Figure B.11: Comparison of F0 in Subject, Verb and Object position in item 11, *Barbara switched jobs.*

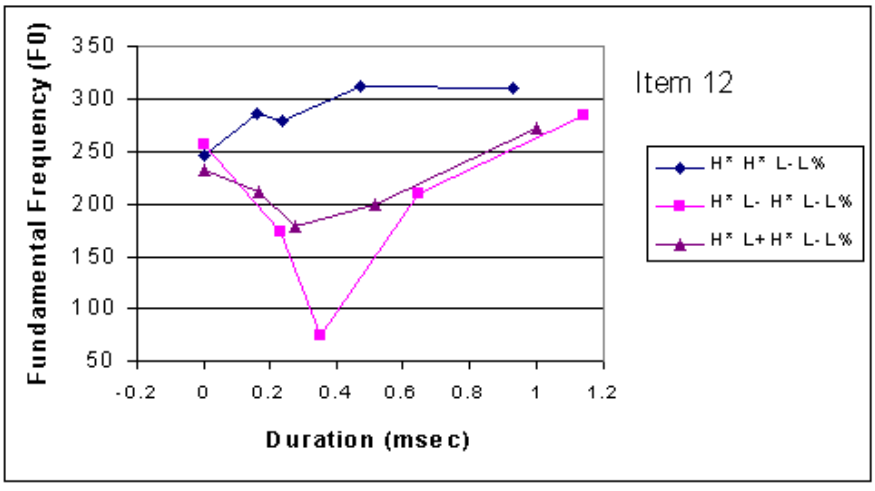

Figure B.12: Comparison of F0 in Subject, Verb and Object position in item 12, *Anthony met 'Sting'.*

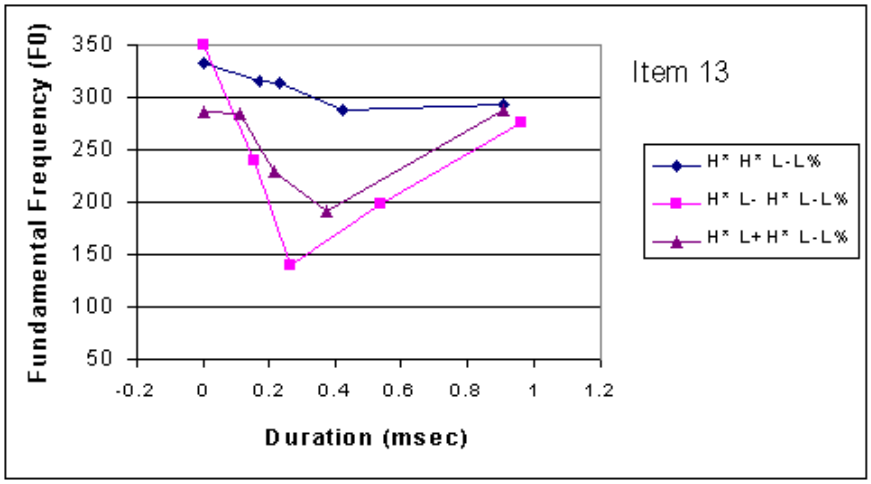

Figure B.13: Comparison of F0 in Subject, Verb and Object position in item 13, *Tiffany writes songs.*

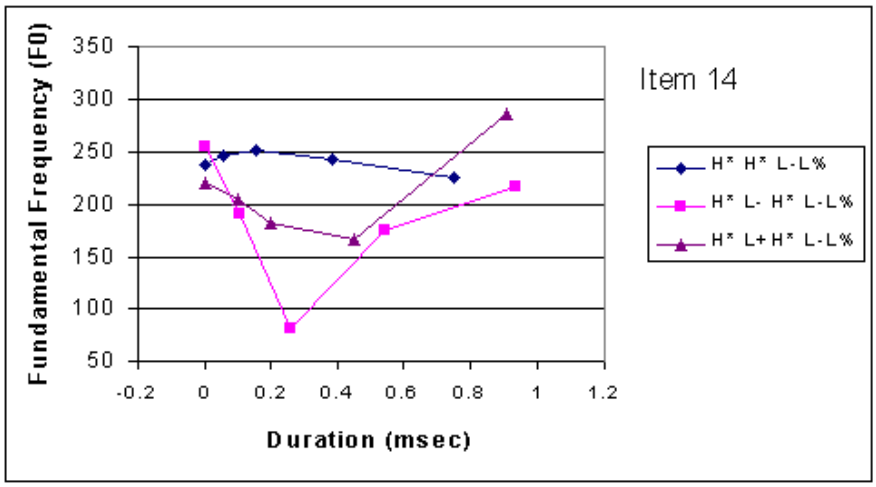

Figure B.14: Comparison of F0 in Subject, Verb and Object position in item 14, *Melanie served fish.*

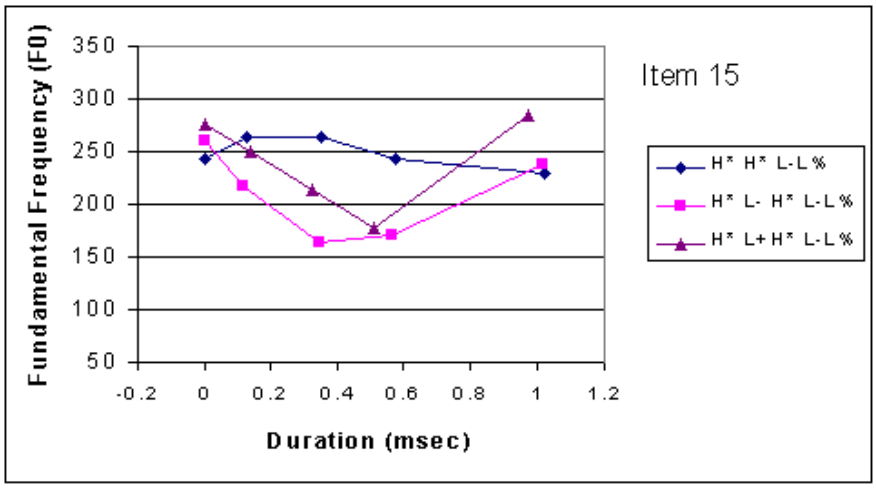

Figure B.15: Comparison of F0 in Subject, Verb and Object position in item 15, *Stephanie chopped wood.*

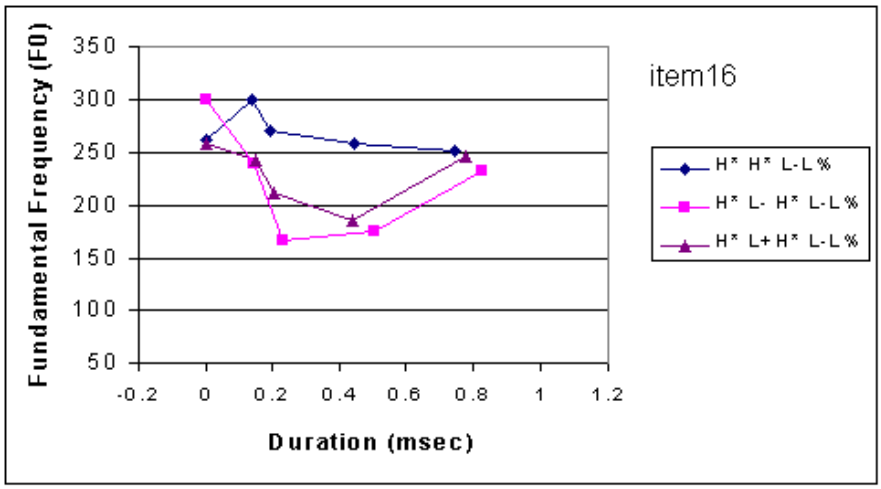

Figure B.16: Comparison of F0 in Subject, Verb and Object position in item 16, *Robertson washed cars.*

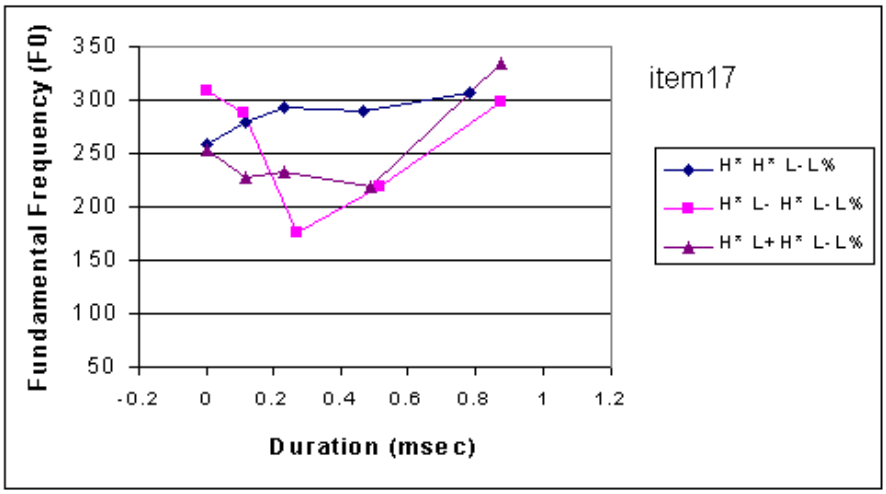

Figure B.17: Comparison of F0 in Subject, Verb and Object position in item 17, *Johnathan switched lanes.*

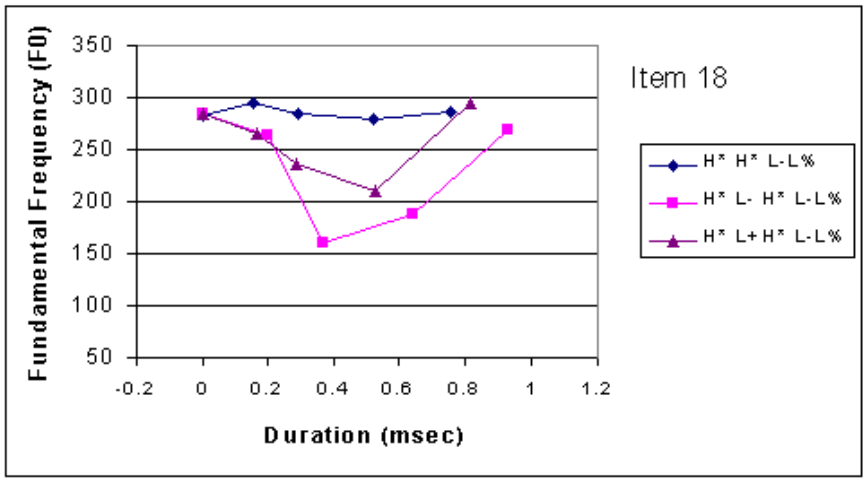

Figure B.18: Comparison of F0 in Subject, Verb and Object position in item 18, *Jessica saw rats.*

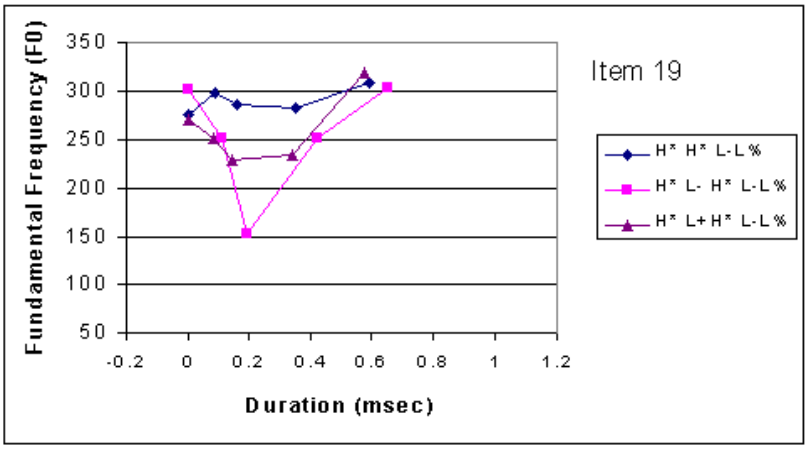

Figure B.19: Comparison of F0 in Subject, Verb and Object position in item 19, *Gideon did disco.*

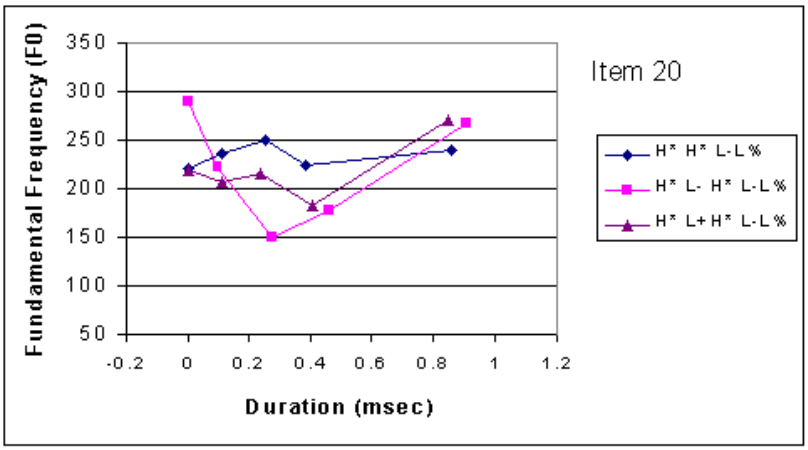

Figure B.20: Comparison of F0 in Subject, Verb and Object position in item 20, *Johnathan ordered beer.*

APPENDIX C

BY–ITEM RESULTS QUD EXPERIMENT

C.1 H* H* L-L%

	broad	late	double	early
1	36.666	50	10	3.333
2	4.761	66.666	19.047	9.523
3	4.761	66.666	14.285	14.285
4	15	55	20	10
5	23.333	63.333	10	3.333
6	23.809	47.619	23.809	4.761
7	14.285	52.380	19.047	14.285
8	15	20	45	20
9	63.333	6.666	30	0
12	40	40	20	0
13	10	73.333	16.666	0
14	28.571	71.428	0	0

	broad	late	double	early
15	19.047	52.380	14.285	14.285
16	20	50	20	10
17	66.666	23.333	3.333	6.666
18	28.571	52.380	9.523	9.523
19	19.047	47.619	23.809	9.523
20	25	70	0	5

Table C.1: Percentages of elicited broad, late, double and early question responses for the hat pattern contour.

C.2 H* L-L%

	broad	late	double	early
1	0	4.761	0	95.238
2	4.761	19.047	4.761	71.428
3	0	15	0	85
4	0	6.666	3.333	90
5	0	0	4.761	95.238
6	0	4.761	4.761	90.476
7	0	10	5	85
8	0	0	3.333	96.666
9	0	0	9.523	90.476
12	3.333	0	3.333	93.333
13	0	9.523	0	90.476
14	4.761	4.761	0	90.476
15	5	0	0	95
16	3.333	6.666	0	90
17	0	0	9.523	90.476
18	4.761	4.761	0	90.476
19	0	25	0	75
20	0	3.333	6.666	90

Table C.2: Percentages of elicited broad, late, double and early question responses for the early nuclear accent contour.

C.3 H* L- H* L-L%

	broad	late	double	early
1	28.571	28.571	4.761	38.095
2	0	60	25	15
3	6.666	50	33.333	10
4	38.095	23.809	23.809	14.285
5	38.095	38.095	19.047	4.761
6	5	30	25	40
7	23.333	16.666	43.333	16.666
8	33.333	9.523	47.619	9.523
9	57.142	9.523	23.809	9.523
12	57.142	14.285	19.047	9.523
13	0	47.619	28.571	23.809
14	20	40	25	15
15	16.666	20	43.333	20
16	38.095	19.047	23.809	19.047
17	71.428	4.761	0	23.809
18	5	55	15	25
19	36.666	23.333	13.333	26.666
20	42.857	42.857	14.285	0

Table C.3: Percentages of elicited broad, late, double and early question responses for the double nuclear accent contour.

C.4 H* L+H* L-L%

	broad	late	double	early
1	20	75	0	5
2	6.666	66.666	16.666	10
3	14.285	66.666	19.047	0
4	19.047	38.095	28.571	14.285
5	10	75	15	0
6	13.333	66.666	13.333	6.666
7	57.142	33.333	9.523	0
8	9.523	38.095	23.809	28.571
9	20	35	30	15
12	47.619	23.809	19.047	9.523
13	15	80	0	5
14	13.333	70	13.333	3.333
15	23.809	38.095	38.095	0
16	0	52.380	79.523	38.095
17	55	35	10	0
18	3.333	73.333	16.666	6.666
19	33.333	57.142	9.523	0
20	4.761	76.190	4.761	14.285

Table C.4: Percentages of elicited broad, late, double and early question responses for the late rising nuclear accent contour.

APPENDIX D

BY–ITEM RESULTS CONTINUATION PARADIGM

D.1 H* H* L-L%

H* H* L-L%	broad	late	double
1	25	75	0
2	12.5	12.5	75
3	31.25	25	43.75
4	43.75	12.5	43.75
5	6.25	25	68.75
6	43.75	18.75	37.5
7	31.25	18.75	50
8	6.25	50	43.75
9	0	87.5	12.5
10	18.75	0	81.25
11	12.5	25	62.5
12	93.75	0	6.25
13	25	31.25	43.75

H* H* L-L%	broad	late	double
14	43.75	12.5	43.75
15	62.5	6.25	31.25
16	68.75	6.25	25
17	87.5	12.5	0
18	25	37.5	37.5
19	43.75	37.5	18.75
20	50	25	25

Table D.1: Percentages of elicited broad, late and double continuation responses for the hat pattern contour (H* H* L-L%).

D.2 H* L-L%

H* L-L%	broad	late	double
1	56.25	25	18.75
2	25	31.25	43.75
3	25	37.5	37.5
4	43.75	0	56.25
5	12.5	0	87.5
6	62.5	0	37.5
7	18.75	12.5	68.75
8	31.25	37.5	31.25
9	37.5	62.5	0
10	62.5	6.25	31.25
11	50	6.25	43.75
12	68.75	6.25	25
13	18.75	12.5	68.75
14	50	6.25	43.75

H* H* L-L%	broad	late	double
15	81.25	6.25	12.5
16	50	6.25	43.75
17	87.5	12.5	0
18	12.5	37.5	50
19	50	6.25	43.75
20	50	6.25	43.75

Table D.2: Percentages of elicited broad, late and double continuation responses for the early nuclear accent contour (H* L-L%).

D.3 H* L- H* L-L%

H* L- H* L-L%	broad	late	double
1	62.5	31.25	6.25
2	0	25	75
3	31.25	6.25	62.5
4	25	18.75	56.25
5	6.25	37.5	56.25
6	43.75	6.25	50
7	31.25	6.25	62.5
8	31.25	31.25	37.5
9	25	62.5	12.5
10	43.75	6.25	50
11	25	0	75
12	62.5	25	12.5
13	12.5	31.25	56.25
14	37.5	0	62.5

H* L- H* L-L%	broad	late	double
15	68.75	0	31.25
16	50	12.5	37.5
17	87.5	6.25	6.25
18	6.25	31.25	62.5
19	43.75	12.5	43.75
20	25	12.5	62.5

Table D.3: Percentages of elicited broad, late and double continuation responses for the double nuclear accent contour (H* L- H* L-L%).

D.4 H* L+H* L-L%

H* L+H* L-L%	broad	late	double
1	25	75	0
2	6.25	31.25	62.5
3	25	18.75	56.25
4	68.75	18.75	12.5
5	25	31.25	43.75
6	31.25	12.5	56.25
7	25	18.75	56.25
8	12.5	31.25	56.25
9	31.25	56.25	12.5
10	50	6.25	43.75
11	31.25	0	68.75
12	81.25	6.25	12.5
13	18.75	25	56.25
14	43.75	18.75	37.5

H* L+H* L-L%	broad	late	double
15	50	18.75	31.25
16	68.75	12.5	18.75
17	93.75	6.25	0
18	6.25	31.25	62.5
19	31.25	31.25	37.5
20	43.75	18.75	37.5

Table D.4: Percentages of elicited broad, late and double continuation responses for the rising nuclear accent contour (H* L+H* L-L%).

BIBLIOGRAPHY

[Aye96] Gayle Ayers. *Nuclear Accent Types and Prominence: Some Psycholinguistic Experiments.* PhD thesis, Ohio State University, 1996.

[BÖ0] Daniel Büring. On D-Trees and B-Accents. University of California, ms, 2000.

[BÖ1] Daniel Büring. Discourse trees and dynamic updates. Conference on Formal Pragmatics., 2001.

[BA94] Mary E. Beckman and Gayle Ayers. ToBI labeling guide. Ohio State University ms, 1994.

[BE90] M. E. Beckman and J. Edwards. Lengthenings and shortenings and the nature of prosodic constituency. In J. Kingston and M. E. Beckman, editors, *Paper in Laboratory Phonology I: Between the Grammar and Physics of Speech*, pages 152–178. Cambridge University Press, Cambridge, 1990.

[Bec96] Mary E. Beckman. The parsing of prosody. *Language and Cognitive Processes*, 11(1/2):17–67, 1996.

[Bol58] Dwight D. Bolinger. A theory of pitch accent in English. *Word*, 14:109–149, 1958.

[Bol86] Dwight D. Bolinger. *Intonation and its parts: melody in spoken English.* Stanford University Press, 1986.

[Bol89] Dwight D. Bolinger. *Intonation and its uses: melody in grammar and discourse.* Stanford University Press, 1989.

[Cho72] Noam Chomksy. Some empirical issues in the theory of transformational grammar. In Paul Stanley Peters, editor, *Goals of Linguistic Theory.* Prentice Hall, Englewood Cliffs, 1972.

[Cli95] Stacey Birch & Charles Clifton. Focus, Accent and Argument Structure: Effects on Language Comprehension. *Language & Speech*, 38:365–391, 1995.

[CR90] Peter Culicover and M. S. Rochemont. Extraposition and the complement principle. *Linguistic Inquiry*, 21:23–47, 1990.

[DP00] Mariapaola D'Imperio-Piterman. *The Role of Perception in Defining Tonal Targets and Their Alignment.* PhD thesis, The Ohio State University, 2000.

[God00] Svetlana Godjevac. *Intonation, Word Order, and Focus Projection in Serbo-Croatian.* PhD thesis, The Ohio State University, 2000.

[GR88] Carlos Gussenhoven and Toni Rietveld. Fundamental frequency declination in Dutch: testing three hypotheses. *Journal of Phonetics*, 16:355–369, 1988.

[GS86] B. Grosz and C. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.

[Gus83] Carlos Gussenhoven. Testing the reality of focus domains. *Language & Speech*, 26:61–80, 1983.

[Gus84] Carlos Gussenhoven. *On the Grammar and Semantics of Sentence Accents.* Foris, Dordrecht, 1984.

[Gus92] Carlos Gussenhoven. Sentence accents and argument structure. In I. Roca, editor, *Thematic Structure: It's Role in Grammar*, pages 79–106. Foris, Berlin and New York, 1992.

[Gus99] Carlos Gussenhoven. On the limits of focus projection in English. In P. Bosch and R. van der Sandt, editors, *Focus: Linguistic, Cognitive, and Computational Perspectives*, pages 43–55. Cambridge University Press, Cambridge, 1999.

[Ham73] C. Hamblin. Question in Montague English. *Foundations of Language*, 10:41–53, 1973.

[Hir85] Gregory Ward & Julia Hirschberg. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language*, 61:747–776, 1985.

[Jac78] Ray Jackendoff. *Semantic Interpretation in Generative Grammar.* MIT Press, Cambridge, MA, 1978.

[Jin96] Shunde Jin. *An Acoustic Study of Sentence in Mandarin Chinese.* PhD thesis, The Ohio State University, 1996.

[Lad80] Robert D. Ladd. *The Structure of Intonational Meaning.* Indiana University Press, Bloomington, 1980.

[Lad83] D. Robert Ladd. Peak features and overall slope. In D. Robert Ladd and Anne Cutler, editors, *Prosody: Models and Measurements*, pages 39–52. Springer Verlag, Berlin, 1983.

[Lad92] D. R. Ladd. An introduction to intonational phonology. In G. J. Docherty and D. R. Ladd, editors, *Papers in Laboratory Phonology 2: Gesture, Segment, Prosody*, pages 321–334. Cambridge University Press, Cambridge, 1992.

[LM97] D. Robert Ladd and R. Morton. The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics*, 25:313–342, 1997.

[LP84] Mark Liberman and Janet Pierrehumbert. Intonational invariance under changes in pitch range and length. In Mark Aronoff and R.T. Oehrle, editors, *Language Sound Structure: Studies in phonology*, pages 157–233. MIT Press, Cambridge, MA, 1984.

[LVJ94] D. R. Ladd, J. Verhoeven, and K. Jacobs. Influence of adjacent pitch accents on each other's perceived prominence: two contradictory effects. *Journal of Phonetics*, 22:87–99, 1994.

[Nak97] Christine H. Nakatani. *The Computational Processing of Intonational Prominence: A Functional Prosody Perspective.* PhD thesis, Harvard University, 1997.

[Nak99] Christine H. Nakatani. Discourse structural constraints on accent in narrative. In Jan P.H. van Santen, Richard Sproat, Joe Olive, and Julia Hirschberg, editors, *Progress in Speech Synthesis.* Springer Verlag, 1999.

[PB88] Janet Pierrehumbert and Mary E. Beckman. *Japanese Tone Structure.* MIT Press, 1988.

[PH90] Janet Pierrehumbert and Julia Hirschberg. The meaning of intonation contours in the interpretation of discourse. In J. Morgan P. R. Cohen and M. E. Pollack, editors, *Intentions in Communication*, chapter 14, pages 271–311. MIT Press, Cambridge, MA, 1990.

[Pie80] Janet Pierrehumbert. *The Phonology and Phonetics of English Intonation.* PhD thesis, Massachussets Institute of Technology, September 1980.

[Pie01] Janet Pierrehumbert. Exemplar dynamics: Word frequency, lenition, and contrast. In *Frequency Effects and the Emergence of Linguistic Structure.* John Benjamins, Amsterdam, 2001.

[PS87] Janet Pierrehumbert and S. A. Steele. How many rise-fall-rise contours? In *Proceedings of The Eleventh International Congress of Phonetic Sciences*, volume 3, pages 145–147. Academy of Sciences of the Estonian S.S.R., Tallin, Estonia, 1987.

[Rob96] Craige Roberts. Focus, the flow of information, and universal grammar. In *The Limits of Syntax*. Academic Press, 1996.

[Roo85] Mats Rooth. *Association with Focus*. PhD thesis, Univeristy of Massachusetts, Amherst, 1985.

[Roo92] Mats Rooth. A theory of focus interpretation. *Natural Language Semantics*, 1:75–116, 1992.

[Sel84] Lisa O. Selkirk. *Phonology and Syntax: The Relation between Sound and Structure*. MIT Press, Cambridge, MA, 1984.

[Sel95] Lisa O. Selkirk. Sentence prosody: Intonation, Stress and Phrasing. In John Goldsmith, editor, *Handbook of Phonological Theory*. Basil Blackwell, 1995.

[SJE93] G. V. Louda P.R. Mueller D. W. Lotts S. J. Eady, W.E. Cooper. Acoustic Characteristics of Sentential Focus: Narrow vs. Broad and Single vs. Dual Focus Environments. *Language & Speech*, 29:233–251, 1993.

[Swe01] Emil Krahmer & Mark Swerts. On the alleged existence of contrastive accents. *Speech Communication*, 34:391–405, 2001.

[Ter91] Jacques Terken. Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America*, 89:1768–1776, 1991.

[Ter94] Jacques Terken. Fundamental frequency and perceived prominence of accented syllables. ii. nonfinal accents. *Journal of the Acoustical Society of America*, 95:3662–3665, 1994.

[tH84] J. t' Hart. A phonetic approach to intonation: from pitch contours to intonation patterns. In Daffyd Gibbon and H. Richter, editors, *Intonation, Accent, and Rhythm: Studies in Discourse Phonology*, pages 193–202. Walter de Gruyter, Berlin, 1984.

[vS] Armin von Stechow. Focusing and backgrounding operators.

[Wel01] Pauline Welby. Effects of pitch accent position, type and status on focus projection. Ohio State University ms, 2001.