INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Discourse Intonation and Second Language Acquisition: Three Genre-based Studies

by

Ann Wennerstrom

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

1997

Approved by _____
Chairperson of Supervisory Committee

_____

_____

_____

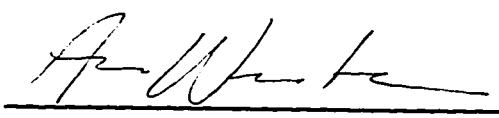
_____

Program Authorized to Offer the Degree English

Date April 8, 1997

Doctoral Dissertation

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to University Microfilms, 1490 Eisenhower Place, P.O. Box 975, Ann Arbor, MI 48106, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature 

Date 4/18/97

University of Washington

Abstract

Discourse Intonation and Second Language Acquisition:
Three Genre-based Studies

by Ann Wennerstrom

Chairperson of the Supervisory Committee:
Professor Sandra Silberstein
Department of English

This dissertation investigates intonation in the discourse of nonnative speakers of English. It is proposed that intonation functions as a grammar of cohesion, contributing to the coherence of the text. Based on a componential model of intonation adapted from Pierrehumbert and Hirshberg (1990), three empirical studies were conducted in different genres of spoken discourse: academic lectures, conversations, and oral narratives. Using computerized speech technology, excerpts of taped discourse were measured to determine how intonation associated with various constituents of text. All speakers were tested for overall English level on tests adapted from the SPEAK Test (ETS, 1985). Comparisons using native speaker data were also conducted.

The first study investigated intonation in lectures given by Chinese teaching assistants. Multivariate analyses showed

that intonation was a significant factor contributing to better scores on an exam of overall comprehensiblity in English. The second study investigated the role of intonation in the turn-taking system in conversations between native and nonnative speakers of English. The final study considered emotional aspects of intonation in narratives, using the framework of Labov and Waletsky (1967). In sum, adult nonnative speakers can acquire intonation as part of their overall language development, although there is evidence against any specific order of acquisition. Intonation contributes to coherence by indicating the relationship between the current utterance and what is assumed to already be in participants' mental representations of the discourse. It also performs a segmentation function, denoting hierarchical relationships among utterances and/or turns. It is suggested that while pitch can be a resource in cross-cultural communication to show emotion and attitude, the grammatical aspects of intonation must be acquired gradually.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CLS	Computerized Speech Lab
ESL	English as a second language
H	Hearer
Hz	Hertz (a unit of pitch)
ITA	International teaching assistant
L1	First language
L2	Second language
NP	Noun phrase
NS	Native speaker
NNS	Nonnative speaker
S	Speaker
SPEAK	An on-site version of the TSE test
TA	Teaching assistant
TOEFL	Test of English as a Foreign Language
TSE	Test of Spoken English

ACKNOWLEDGMENTS

Based on a highly unscientific discourse analysis of the acknowledgments of various dissertations and books, it has come to my attention that in the former, the writer is allowed to be more emotional and effusive than in the latter. Never one to waste an opportunity, I submit the following acknowledgments of those who have made contributions to this dissertation.

First of all, I would like to say that I cannot imagine a more wonderful dissertation adviser than Sandra Silberstein in terms of personal support, enthusiasm for my project, academic savvy, and knowledge of the discipline. I am also hopelessly enamored of the rest of my committee: Heidi Riggenbach for many long discussions of conversation analysis, as well as for personal friendship, book exchanges, and curry and kale recipes; Ellen Kaisse for her thorough and good-humored overseeing of the phonological aspects of this dissertation and her continued encouragement of and interest in my work; Michael Toolan for introducing me to Halliday and Hasan, for helping me develop the ability to argue, and for very often thinking that data is interesting for the same reasons I do; and Alan Klockars for two very painless and useful statistics courses.

Second, I would like to thank my two department heads, Bill Harshbarger and Tom Lockwood, for cooperating to allow me

a flexible job schedule as well as conference support during the completion of my doctoral work. Bill, moreover, has gone out of his way to set up a work environment which has encouraged me to experiment and develop in a variety of directions over the years.

Several other friends and colleagues also deserve my heartfelt thanks not only for their personal support but also for comments and suggestions pertaining to my work. I am grateful to Suzy LePeintre, Phil Gaines, Alice Taff, Kathy Ewing, Guy Modica, Pat Sheahan, Laurie Stephan, Glen Zimmermann, and Lucy Pickering, fellow graduate students who have generally "hung out" with me at various crucial times during my graduate career and contributed bursts of creative energy to my work. I also wish to thank Patty Heiser, Anita Sokmen, and Cherie Lenz-Hackett for their unflagging interest in my progress and friendship at my workplace. An extra special thanks goes to Kathy Mork for both personal and professional guidance in the dissertation process. Finally, I would like to thank Jim Tollefson, Gerry Philipsen, Gail Stygall, George Dillon, and Fritz Newmeyer, professors at the University of Washington whose ideas have influenced my thinking.

Next, let me express my appreciation to the many students in the University of Washington English as a Second Language Programs for volunteering their language for my

investigations. I am also grateful to the University of Washington Linguistics department for allowing me access to the computerized speech equipment in the Phonetics Lab.

On a more personal level, I'd like to thank several close friends and family members. May I extend my deepest, life-long thanks to: Paulette Booker for helping me raise my children; Steve Teig, for frequently bringing a breath of intellectual fresh air into my home; my parents-in-law, Armand and Mildred Siegel, for upholding the highest intellectual standards while being one heck of a fun couple; my mother, Jeanne Loomer, who, trusting me to do things well, did not push or intervene; my daughters, Bonnie for being a brave and sensitive person, and Clara for being so unfailingly upbeat; for Andy Siegel, my husband, consultant, and role model for academics, who always kept the home fires burning; and for my son Mikie, little though he is.

Remembering Papa, working on his thesis

Chapter 1

Introduction

1.1 The Role of Intonation in Discourse

It may be readily observed that intonation plays a central role in the coherence and organization of discourse in spoken English, well beyond the sentence level. For example, in a conversation about wines, we might imagine a high pitch associated with the words *white* and *red* to show that these two categories were cohesively linked in a comparison/contrast relationship. In such a discussion we would not be surprised to find the pitch peaks on *white* and *red* maintained throughout an extended piece of discourse, even across speakers. Moreover, if one speaker went off on a tangent about his travels in France, another told a joke, and the third excused himself to answer the telephone, we might still find the same high pitch associated with *white* and *red* upon the speakers' return to the topic of comparing wines; indeed the high pitch would be one of the key indicators that the topic was still salient and that cohesion had been maintained in spite of the interruptions. This observation, obvious, yet fundamental in scope, serves as a starting point for this dissertation, which investigates in detail the contributions that intonation makes to coherence in discourse.

Intonation, which is defined simply as the pitch of

speech, has both universal and language-specific functions. Universally, there is a relationship between intonation and emotion, in that high pitch is often a signal of excitement or alarm (Bolinger, 1986; Chafe, 1994; Ladd, 1980; Tench, 1996). On the other hand, language-specifically, pitch associates with particular constituents in discourse in a rule-governed and meaningful way. In this dissertation, I consider both universal and language-specific functions of intonation. With respect to the latter, I will apply an adapted version of Pierrehumbert and Hirschberg's (1990) interpretational model of intonation to identify a system of pitch "morphemes" which contribute meaning to English discourse. At the phrase level, the utterance level, and the topic level, the intonation associated with the boundaries of constituents signals a hierarchy of interdependencies among those constituents; at the level of information structure, pitch patterns indicate the roles of the lexical items with which they associate (new, given, contrastive, structural, etc.).

These intonational units are all important to the cohesion and coherence of a text at the discourse level. For example, boundary intonation links constituents together just as lexical conjunctions do, indicating how they are interrelated; low pitch sends a message that the associated lexical item should already be available in the participants' mental representation of a discourse; and as we saw in the wine example above, high pitch may be used to make a contrast

over a long sequence of text. In sum, I hypothesize that there is an English-specific intonation system which functions as a grammar of cohesion, a morphemic system of tones that conveys meaning beyond what is provided through lexical, grammatical, semantic, and logical systems. This system can be studied separately from the more universal, emotional aspects of intonation.

In the context of cross-cultural communication, the role of intonation becomes particularly important. While the universal ability to convey emotional states through intonation is likely to enhance understanding between speakers of different languages, the existence of language-specific systems may present a barrier. For English, the second language (L2) learner faces a challenging situation. Intonation is so intuitive to native speakers that it is only marginally indicated in orthography and seldom taught in ESL classes. Since the system is little understood and instruction sporadic, the main access an L2 learner has to intonation is through induction. If we consider there to be a natural language acquisition process for adult L2 learners, independent of the memorization of grammar rules (Krashen, 1981), we would expect intonation to be a part of that acquired system. In this regard, the interlanguage of nonnative speakers provides interesting data for the study of intonation as a naturally acquired linguistic system.

In this dissertation, I present three empirical studies

of nonnative speaker discourse, focusing on intonation in three different oral genres: academic lectures, conversations, and oral narratives. I take the term genre rather liberally, in line with Hymes' (1974) definition, as a category of communicative event with a set of culturally defined norms. "The notion of genre," says Hymes, "implies the possibility of identifying formal characteristics traditionally recognized" (p. 61). This is similar to what Biber (1994) has referred to as "register" or Selinker and Douglas (1985) as "discourse domain". To Hymes' definition may be added Swales' (1990) insight that genre is community based; participation in a community depends on being fluent in its genres. It is not uncontroversial that I have included conversation as a full-fledged genre, rather than "unmarked" speech or "pregenre" (Swales, 1990). While the familiarity of casual speech may make its generic features less recognizable to insiders, these features become more obvious when casual speech occurs in an unexpected context (Hymes, 1974). Moreover, it has clearly been shown that many norms and rules of conversation vary depending on the speech community (Hymes, 1974; Saville-Troike, 1982; Schiminoff, 1980; Tannen, 1984a) and are therefore not universal. Indeed, Selinker and Douglas (1985) and Mason (1994) cite cases of international students whose English was more proficient in formal academic genres than in informal conversational genres. For these reasons, I place the conversation study in this dissertation as much

within a genre framework as I do the lecture and oral narrative studies.

A genre approach to second language study is similar to an English for Special Purposes (ESP) approach -- instead of studying "general English" learners can begin to identify the discourse communities to which they wish to belong and to become familiar with the genres in use within those communities. Since, as Biber (1994) has shown, different genres consistently involve distinctive clusterings of linguistic patterns, a focus on differing genres in language study allows for concentration on those linguistic patterns most prevalent in each genre. From my own point of view as researcher, a genre approach affords a similar opportunity to analyze a variety of intonation patterns. Specifically, academic lectures, which usually display an overtly marked organizational structure at the macrolevel (see Chaudron and Richards, 1986; Hansen, 1994), afford access to the intonation associated with the organizing language of higher order constituents. The conversation genre provides a venue for investigating intonation's role at the microlevel in interaction, conversation management, and turn-taking (Riggenbach, 1991; Sacks, Schegloff, & Jefferson, 1974). Finally, oral narratives allow for the investigation of both macrolevel structure of narrative components (Labov & Waletsky, 1967) and emotional aspects of intonation. Thus, since these three genres are so diverse in their structure and

scope, each highlights a different aspect of intonation in discourse.

There are four main research questions addressed in this investigation:

1. How does intonation in English contribute to coherence at the discourse level?
2. Can nonnative speakers acquire the intonation system of English?
3. How does L2 intonation affect judgments of comprehensibility by native speakers?
4. Is there evidence for emotion-based intonation universals?

The methods of assessment of these questions are both quantitative and qualitative. Using computerized speech analysis equipment, it is possible to make precise measurements of particular pitch levels of words and phrases from audiotapes of discourse. Thus, by taking multiple measurements of recurring intonation patterns, and by performing various statistical analyses on these, I am able to compare the intonation of groups of speakers as well as to determine how individual intonation variables interact with each other and with English test scores. These kinds of analyses are used to establish general trends in L2 intonation. In a more qualitative mode, the intonation patterns of individual words and utterances of particular speakers provide insight into how these contribute to, or

detract from, coherent communication in specific contexts.

The remainder of this chapter surveys the literature of intonation in second language acquisition. In Chapter 2, I provide a justification for my choice of Pierrehumbert and Hirschberg's system as a theoretical model, along with my own adaptations of that model. In Chapter 3, I discuss the role of intonation in cohesion and coherence. Drawing on the work of Chafe (1994) and Sperber and Wilson (1995), I show how intonation guides conversants to integrate each new utterance into their mental representation of the discourse in progress.

Chapters 4, 5, and 6 present the studies themselves. Chapter 4 is a study of academic lectures given in English by native speakers of Mandarin Chinese. The lectures have been rated in comprehensibility by assessment specialists. Multiple regression analysis shows that certain aspects of intonation contribute to better test scores. Examples of native speaker lecturers' intonation are also included. Chapter 5 is a study of informal conversations between native and nonnative speakers from a variety of language backgrounds. Each nonnative speaker has been rated for fluency. The study is a detailed conversation analysis of ways in which the intonation of the highest rated speakers contributes to conversational fluency, while problems exist for those at the lower level. Chapter 6 is a study of the role of intonation in oral narratives solicited from both native and nonnative speakers. Although data indicate language-specific

differences in intonation at structural junctures, this study focuses mostly on the emotional, language-universal functions of intonation. In particular, pitch maxima were found to accompany evaluation mechanisms (in Labov and Waletsky's, 1967 terms) which I argue are emotionally laden regardless of one's language background.

Finally, Chapter 7 provides a discussion and synthesis of the results of all three studies, considering both the common threads and the genre-specific differences among them. In short, the fact that certain patterns of intonation uttered by nonnative speakers are shown to influence interactional dynamics and test scores lends a new kind of empirical support to the fact that a grammatical system of intonation exists in the minds of native speakers. Moreover, this system is acquirable by nonnative speakers as indicated by comparisons in the use of intonation among speakers at different levels of English.

1.2 Intonation, Coherence, and Nonnative Speakers

In a useful review article, Chun (1988) outlines the many discourse functions of intonation, advocating a more elevated status for its role in second language research and pedagogy. While some ESL materials writers have attempted to cover intonation, producing pronunciation textbooks with a strong emphasis on prosody (Gilbert, 1993; Grant, 1993; Hagen & Grogan, 1992; Morley, 1993; Wennerstrom, 1991), there has been

only a modest quantity of empirical work on the intonation of nonnative speakers of English at the discourse level. In sociolinguistic studies of interaction, Gumperz (1982) emphasizes the importance of prosody in "contextualization cues" in cross-cultural communication. His most famous examples involve Indian English speakers who are perceived as angry by British English speakers, a fact which he attributes in part to certain intonational forms in the Indians' speech. In a more quantitative mode, Derwing and Munro (1997) found a modest correlation between prosody and perceived comprehensiblity in a study of Japanese speakers engaged in a story-telling task. Looking at specific intonational features in a discourse context, Wennerstrom (1994) found differences between how intonation is used in English oral reading and free speech tasks by intermediate speakers of Thai, Japanese, and Spanish as compared with native speakers. While native speakers made statistically significant pitch distinctions at the discourse level on all the intonation features measured, the other groups made fewer, or in the case of the Thai speakers, none of the same distinctions. Hewings (1995) analyzed the intonation of Indonesian and native speakers of English reading scripted dialogues. He found that Indonesians used falling tones at junctures where native speakers used rising tones, which he attributes to lack of knowledge of social conventions of English. All of these findings reveal differences in intonation at the discourse level between

native and nonnative speakers.

Focusing more on the acquisition process, Juffs (1990) measured word stress and sentence-level pitch accent in an oral reading task by Chinese speakers of English, where he found that the majority of errors were associated with word stress. He also found errors due to strategies such as the placement of a heavy stress on every word and consistent sentence-final pitch accent assignment. Wenk (1985), in a study of French speakers acquiring English, found developmental stages of the acquisition of stress patterns. At an intermediate stage, the subjects produced interlanguage patterns that were between French and English. Thus, both Juffs and Wenk contend that transfer alone does not explain nonnative speakers' acquisition of intonation since these intermediate stages reveal patterns characteristic of neither the native nor the target language. Gilbert (1980) and Neufeld & Schneiderman (1980) investigated the role of practice in the acquisition of native-like intonation patterns. Both studies found a positive result from oral and aural practice. In Gilbert's study, the language was filtered so that only the intonation was accessible; similarly, in Neufeld and Schneiderman's study, subjects practiced intonation itself without knowing the meaning of what they were uttering. Other studies of second language intonation which have focused on the sentence level (Lepetit, 1989; Shen, 1990b) are less relevant in determining the discourse level

functioning of intonation, relying as they do on short utterances out of context.

Within the International Teaching Assistant (ITA) literature, we find a general interest in cohesion in oral discourse, where attention has focused on identifying multiple factors contributing to coherence (Rounds, 1987; Williams, 1989). Intonation has been identified among these factors in certain studies: Davies and Tyler (1994), in a microanalysis of an ITA dialogue with a native speaker in an office conference, document the confusion generated by mismatches between lexico-syntactic and prosodic cues. Their example (p. 215), is taken from an exchange in which the conjunction *because*, which would indicate the beginning of a dependent clause, was preceded by a low intonational boundary and a pause, both of which would indicate "finality" of the previous clause. Indeed, at this juncture a native speaker attempted an interruption. In a similar analysis of the classroom discourse of a Chinese and a native speaker TA, Tyler (1992) considered discourse markers, lexical specificity, syntactic factors, and prosody as contributors to coherence. Her analysis provides another example of an intonational sequence from the Chinese speaker which detracted from coherence: Again, a low intonational boundary, produced in utterance-medial position, made it difficult for a native speaker to parse the utterance, since this low boundary signals finality.

In sum, although there is clearly an interest in issues

of intonation in the second language acquisition community, studies have tended to focus narrowly on particular intonation patterns or individual examples. Nor have there been systematic attempts in these studies to follow a particular theoretical model of intonation. With this dissertation, I hope to contribute a framework within which to integrate intonation research into the broader context of studies of coherence, and to provide a quantitative methodology for the measurement and assessment of a speaker's intonation in context. I have chosen to maintain a focus on extended segments of discourse instead of constructed examples in order to study long-range cohesive links made through the intonation system as people are actually speaking in context. Finally, I hope to provide evidence that intonation universally expresses emotion, which, I submit, can enhance cross-cultural communication regardless of one's language background.

Chapter 2

The Phonology of Intonation

In this chapter, I explain the background concepts related to intonation as developed in the phonology and phonetics literature. In an outline of three major theories of intonational meaning, those of Halliday (1967), Pierrehumbert and Hirshberg (1990), and Brazil (1985), I argue in favor of a componential approach which will then serve as the framework for the studies reported in Chapters 4-6. The chapter concludes with a section on computer measurement of intonation.

2.1 What is Intonation?

Intonation, or pitch sequence during linguistic production, plays a fairly complex role in communication. To understand how it works, it will be useful to divide intonation into three main subcomponents and discuss each one separately. The first aspect of intonation, and the most important one for this dissertation, is its ability to convey linguistic meaning. While tone languages such as Chinese or Thai are examples of languages for which pitch differentiates lexical meaning, English is believed by many scholars to be a language in which pitch is used to distinguish meaning in discourse[1] (Bing, 1985; Bolinger, 1986; Brazil, 1985; Brown,

1977; Chafe, 1994; Chomsky & Halle, 1968; Chun, 1988; Couper-Kuhlen, 1986; Cruttendon, 1986; Crystal, 1969, 1975; Gussenhoven, 1984; Halliday, 1967; Ladd, 1980; O'Connor & Arnold, 1973; Pierrehumbert & Hirschberg, 1990; Pike, 1945; Tench, 1991, 1996; Traeger & Smith, 1951; Wennerstrom, 1994; Werth, 1984). For example, high pitch is often associated with new information and low pitch with given; rising pitch at the end of an utterance may indicate a yes/no question; a flat, continued pitch on a word may signal the speaker's (S's) intention to continue. The focus of study in this regard has mainly been language-specific: how do certain pitch configurations associate with text in particular languages to convey distinctive meanings. The majority of this chapter centers on how this question has been answered for English, after first describing the other two aspects of intonation.

The second aspect of intonation, which I am inclined to believe is universal, is its role in conveying emotion and attitude. Although some scholars view expression of emotion and attitude as the most central function of intonation (Brown, 1977; Brown, Currie, & Kenworthy, 1980; Crystal, 1969; O'Connor & Arnold, 1973; Pike, 1945; Tench, 1991, 1996), many others agree that a speaker's association of pitch to text is a separate issue from the relative degree of pitch height he or she uses to convey emotion (Bolinger, 1978, 1986; Brazil, 1985; Chafe, 1994; Ladd, 1980; Pierrehumbert, 1980; Shen, 1990a). As Shen, who works on Chinese intonation, states,

"Emotional, expressive, or emphatic nuances (anger, irony, astonishment, apprehension, etc.) can be superimposed upon almost any utterance" (p. 9). Ladd (1980) explains this as the "gradient" aspect of intonation versus its phonemic, "all-or-none" aspect, drawing a useful analogy with expressive lengthening in segmental phonology: the meaning distinction between the words *big* and *bi-i-ig*, he points out, is not phonemic (as would be *big* and *beg*), but rather gradient, for an expressive purpose (p. 113). Similarly, a word may have a particular pitch accent, but the height of its pitch relative to that of other pitch accents of the same type will be a matter of degree, depending on speaker attitude, emotion, expressiveness, and the like.

Bolinger (1986) and Chafe (1994) draw an iconic connection between primitive cries of arousal and high pitch as the evolutionary origin of systems of intonation, of which certain aspects have then become conventionalized in the development of human languages. In Bolinger's opinion, ". . . we have a mixed system, expressive at base but with adaptations that differ from culture to culture" (1986, p. 198). In other words, we would expect a universal tendency for speakers of any language to associate pitch extremes with emotionally charged text, regardless of what system of intonation that language might have developed to govern pitch placement.

Similarly, pitch may be increased or decreased for

pragmatic reasons; for example, to be heard by someone at a distance requires a higher pitch and louder voice than does murmuring at close range. As with emotion, these paralinguistic aspects of intonation are available to speakers regardless of language background, although the appropriate contexts may be culturally influenced and quite ritualized. For example, in many cultures, members of audiences are allowed to scream at particular junctures in a sports event but may be chastised for doing so in a theater production.

In sum, we find language-universal, emotional, attitudinal, and pragmatic aspects of intonation which can be explored independently of language-specific, grammatical intonation systems. These universal uses of intonation are taken up in Chapter 6.

The third aspect of intonation, which is usually independent of meaning, is its interaction with the metrical systems of languages. It is posited as a universal that rhythmic systems underlie the phonologies of the languages of the world (Goldsmith, 1990; Halle & Vergnaud, 1987; Hayes, 1985, 1995; Liberman, 1975; Nespor & Vogel, 1986, 1989). Hayes (1995) discusses metrical rules which tend to create a regular alignment of beats through time in speech. Using data from a range of languages, he argues convincingly that languages can be characterized as either iambic or trochaic in structure. One manifestation of this alternation in English is the Rhythm Rule (Hayes, 1984; Kaisse, 1987; Kiparsky,

1979), the tendency for speakers of English to shift word stress to the left to avoid what Nespor and Vogel (1989) have called a stress clash. For example, *California* is normally stressed on the third syllable, but in the phrase *California dreaming*, speakers tend to shift stress, and therefore higher pitch, to the first syllable (see Figure 2.1). Other phonetic principles also influence the shape of a pitch contour. Catathesis (Pierrehumbert, 1980), or downdrift (Hyman, 1975) is the tendency for pitch to gradually lower within an intonational phrase. Such factors belong in a description of intonation but cannot be said to contribute to the meaning of an utterance. They are not the focus of study of this dissertation.

2.2 Contending Models of Intonational Meaning

While few would doubt that intonation in English carries meaning, there is a good deal of disagreement in the literature about how to define the basic unit of meaning. Should intonational tunes be viewed as holistic contours or as sequences of smaller units within those contours? In other words, what minimal units of intonation can be identified that are both meaningfully and phonologically distinctive?

Starting on the meaning side, scholars have focussed on two main interpretive functions of intonation in English. First, intonation keeps track of the status of given and new information at the lexical level (Brazil, 1985; Chafe, 1994;

Couper-Kuhlen, 1986; Cruttendon, 1986; Halliday, 1967; Ladd, 1980; Pierrehumbert & Hirschberg, 1990; Wennerstrom, 1994; Werth, 1984), and second, it indicates the hierarchical relationships among units of discourse, such as phrases, clauses, turns, topic units, etc. (Brazil, 1985; Brown, 1977; Chafe, 1994; Couper-Kuhlen, 1986; Couper-Kuhlen & Selting, 1994; Gussenhoven, 1984; McLemore, 1991; Pierrehumbert & Hirshberg, 1990; Wennerstrom, 1994; Yule, 1980). Therefore, the most explanatory models of intonation will be those which can express these meaning distinctions phonologically. If the unit is too large, these two functions may be conflated; if too small, too many possible contours will be generated.

What follows is an argument in support of a componential approach to intonational meaning, along the lines of Pierrehumbert (1980) and Pierrehumbert and Hirschberg (1990), in which the function of sorting information structure is expressed through a system of pitch accents, and the function of showing relationships among larger constituents is expressed through a system of boundary tones. While this approach is not incompatible with there being intonation contours with holistic interpretations, I will show that having the tone as the minimum component allows for more versatile and economical analyses. First I will introduce the holistic contour models, mainly in the work of Halliday (1967), then I will explain Pierrehumbert and Hirshberg's model along with arguments for its explanatory power, and

finally, drawing from the work of Brazil (1985), Ladd (1980), and Brown (1977), I will indicate modifications which will expand Pierrehumbert and Hirshberg's model to accommodate certain phenomena within longer stretches of discourse.

2.2.1 Holistic Contour Models

There is general agreement in the literature that something close to a syntactic clause tends to correspond to an intonation contour. Indeed, psycholinguistic evidence based on processing experiments (see Levelt, 1989) indicates that a speaker preplans clause-sized intentional units prior to speech which correspond to these contours. A number of scholars have therefore attempted to define inventories of meaningful shapes for intonation contours. For example, Liberman and Sag (1974) and Sag and Liberman (1975) have identified "the contradiction contour," and "the surprise/ redundancy contour" corresponding to particular clause-sized contour shapes. Similarly, several other accounts of intonational meaning (Bing, 1985; Cruttendon, 1986; Halliday, 1967; Ladd, 1980; and others), identify sections of contours as meaningful.

It will be useful to begin with Halliday's (1967) system, since it precedes many later treatments of intonation. Although this model was developed for British English, its influence is reflected in the work of many scholars, including North Americans such as Robert Ladd. Halliday identifies five

contours, which govern the shape of the utterance from the "tonic" or nuclear pitch accent to its end. Below is a list of Halliday's contours and a summary of the main sentence functions with which they associate (pp. 25-28). It should be stressed that Halliday does not consider these meanings to be rigidly fixed; instead, they interact with other grammatical systems of English in context.

1. Falling tone ends low and associates with statements, answers, imperatives, exclamations, and canonical WH-questions.[2] This contour has a neutral mood with no special challenge or contradiction.
2. Rising or falling-rising tone ends high and associates with neutral yes/no questions, as well as statements that are challenging, aggressive, defensive, etc., and contradictory WH-questions.
3. Rising tone ends at a mid level and associates with noncommittal replies which are disengaged or discouraging, as well as statements which are dependent, confirmatory.
4. (Rising-)FALLING-rising tone ends at a mid level and conveys reservation or conditionality. (The capital letters in this and #5 indicate the area of greatest intensity.)
5. (falling-)RISING-falling ends low and conveys involvement, assertion, "I'm telling you".

In general, those contours which end with a falling pitch indicate that the "polarity" (whether the utterance is negative or positive) is known; rising pitch indicates that the polarity is unknown. Halliday goes on to show how these contours can combine together in longer sequences of discourse. It is noteworthy that each contour is normally said to contain a single tonic. For longer utterances where more than one strongly accented item is manifest, Halliday posits a "prehead," an additional contour in utterance-initial position. He also suggests that in marked cases, a contour with a "double tonic" can exist (p. 22). However, he does not provide any means to distinguish between a contour with a double tonic and two separate contours.

Several other systems reflect the influence of Halliday. Ladd (1980) has a system of four categories, conflating Halliday's contours 1 and 5. Bing (1985) contributes the "class O" contour, a low-rise used in parentheticals, vocatives, and other peripheral types of clauses. Crystal (1969) conflates Halliday's second and third categories, but distinguishes an additional rise-fall-rise. Gussenhoven (1984) posits three contours, a fall, a rise, and a fall-rise. The common thread in this body of work is that the last pitch accent of the contour has a special status as the "nucleus" of the utterance (Halliday's "tonic"). Also, the rightmost segment of an utterance, from the nucleus to the end of the contour is considered to be a meaningful unit.

Although I find Halliday's model to be explanatory for many cases, I believe that it can be improved upon in two major areas. First, as Pierrehumbert and Hirshberg (1990) show, it is possible to identify smaller units below the contour level which share common meaning with similar units across a variety of contours. Second, certain longer utterances pose problems for the claim that there is only one nuclear pitch accent (tonic) per intonational contour. At this point we will turn to models which recognize meaningful pitch units below the contour level, which are referred to as "componential" models.

2.2.2 Pierrehumbert and Hirschberg's Componential Model

In a discussion of smaller units of intonation, it should be mentioned at the outset that Pike (1945), Traeger and Smith (1951), and Chomsky and Halle (1968) developed what are often referred to as "level" systems which assign numeric values to degrees of stress. While descriptively useful, this body of work is not relevant to the current discussion because no claims are made about the meaning of the levels. Instead, I will consider the work of those authors who assign an independent interpretation to the boundary of an utterance, separate from the rest of the contour (Beckman & Pierrehumbert, 1986; Brazil, 1985; Pierrehumbert, 1980; Pierrehumbert & Beckman 1988; Pierrehumbert & Hirshberg, 1990).

Pierrehumbert proposed a generative model of intonation in her dissertation (1980) which was later used at Bell Labs to improve the quality of synthesized English speech for the telephone service. Subsequently, Pierrehumbert and Hirshberg (1990) developed an interpretational model within the same theoretical framework. According to Pierrehumbert and Hirshberg, an intonation contour is made up of a bare-bones sequence of discrete meaningful tones, which are linked together in the same breath group through principles of phonetic implementation to form a surface pitch contour. The tones in this model fall into two basic categories, pitch accents and boundary tones. Pitch accents associate with lexical items which a speaker (S) wishes to make salient and indicate the status of these items in the information structure of a text; that is, how does S's contribution coincide with knowledge, ideas, and beliefs assumed to be already shared with hearers (H). This shared information will be referred to as the "mental representation" of the discourse (a notion which is explored in detail in Chapter 3). There are six possible pitch accents in the model, including both individual and contour pitch accents (the * indicates the part of the pitch accent which associates with the stressed syllable of the word in question):

H* - associated with items which are being added into the mental representation of the discourse as new.

H*+L - associated with items about which one is making a

predication which should already be inferable from the mental representation.

L+H* - associated with items on a scale which should be mutually believed instead of some alternative item on the scale (often used for contrast).

L* - associated with items that are salient but about which no predication is made; not to be added to the mental representation.

H+L* - associated with items which are already inferable from the mental representation and about which no predication is being made (sometimes used to confirm what is already known).

L*+H - associated with items which are part of a scale, but about which one is uncertain.

In general, any pitch accent containing a H* associates with items introduced into the mental representation of the discourse as new; any L* pitch accent associates with those items which S "intends to be salient but not to form part of what [S] is predicating in the utterance" (Pierrehumbert & Hirschberg, p. 291). The interpretation of these pitch accents will be discussed in more detail in Section 2.3, particularly with regard to the relationship between low pitch and inferable information.

The second category of pitch morpheme in Pierrehumbert and Hirshberg's model is the set of tones which associate with the ends of phrases and utterances. In general, a high pitch

at a boundary indicates nonfinality, or S's intention for the hearer to interpret what comes after the tone with respect to what has come before. Pierrehumbert and Hirshberg recognize two levels of hierarchy for these tones, the level of the "intermediate phrase" and the level of the utterance. Intermediate phrases within an utterance end with "phrase accents" which may have high (H) or low (L) pitch, indicating their relationship to a subsequent phrase within the same utterance. For example, in the following conjoined intermediate phrases, they claim that we are likely to perceive (2.1) as a closed-choice question, whereas in (2.2) we perceive a more open-ended choice. This is because a high phrase accent conveys the meaning that the two phrases in (2.1) are interdependent:

(2.1) Do you want apple juice^{H} or orange juice?$_{LL\%}$

(2.2) Do you want apple juice$_{L}$ or orange juice?$_{LL\%}$ (p. 302)

At the next level, an utterance of one or more intermediate phrases ends with a "boundary tone" (L% or H%) This governs the utterance as a whole, indicating its relationship to the subsequent utterance. According to Pierrehumbert and Hirschberg, " . . . choice of boundary tone conveys whether the current intonational phrase is 'forward looking' or not -- that is, whether it is to be interpreted with respect to some succeeding phrase or whether the direction of interpretation is unspecified" (p. 305). The

subsequent material may be *given by S or anticipated from H.* Thus, a H% will be characteristic of several types of utterances such as these examples from Pierrehumbert and Hirshberg:

(2.3) George likes cake$_{L}^{H\%}$
He adores pie$_{L}^{H\%}$
He'll eat anything that's sweet and calorific$_{LL\%}$. (p. 306)

(2.4) a. Does it snow a lot in New Jersey$^{HH\%}$?
b. It does this year$_{LL\%}$. (p. 306)

(2.5) (Man walks into dentist's office) My name is Mark Liberman$^{HH\%}$? (p. 290)

(2.6) (woman asked how she liked a movie)
I thought it was good$^{HH\%}$? (p. 290)

In (2.3) the H% boundary tone at the end of the first two clauses, uttered by the same speaker, indicates that the whole sequence of three clauses is to be interpreted as a unit. In (2.4), we have a yes/no question anticipating the subsequent answer from another speaker. In (2.5) and (2.6) the utterance anticipates not the answer to a question, but a backchannel, or more specifically, confirmation of the man's appointment and validation of the woman's opinion. The H% boundary tone captures the commonality of meaning in all four utterances in that the clause before the boundary tone anticipates subsequent material to complete its interpretation.

Phrase accents and boundary tones occurring in

combination yield an inventory of four possible boundaries shapes: HH%, HL%, LH%, LL%, depending on S's intentions about the interdependencies of the utterances produced. Each component of the boundary carries its bit of meaning, independent of the meaning conveyed by the pitch accents within the utterance. The meanings of the four shapes are summarized as follows:

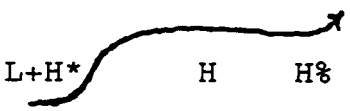
HH% - a high rising boundary, indicates that the material within the utterance requires subsequent discourse for its interpretation, and this link is often cross-speaker.

LH% - a low rising boundary indicates that the idea within the utterance is to be continued in the next one, often by the same speaker.

HL% - a plateau boundary, indicating that the material within the utterance is to be continued or elaborated upon.

LL% - a falling boundary, indicates that the material within the utterance concludes a thought or turn.

Certain phonetic implementation rules influence the shapes of these boundaries. The phrase accent affects all the material between the final pitch accent and the final boundary tone. For example:



(2.7) Do you always do this?

The last syllable of *always*, though normally unstressed, is uttered with a high pitch, as are the unaccented words *do this*. This is because the effect of the H phrase accent is interpolated across the entire final sequence.

Another phonetic principle that affects these boundary shapes is upstep, which raises a L% boundary tone following a H phrase accent. The result is that the HL% configuration is pronounced as a mid-level, flat boundary or "plateau" from the last pitch accent to the end. Here is an example from Pierrehumbert and Hirshberg, (p. 291):

(2.8) Wally: Mostly they just sat around and knocked stuff. You know.

H* H L%

The school

H* H L%.

Other people

2.2.2.1 Arguments for the Componential Model

Pierrehumbert and Hirshberg make three main arguments in favor of their system (p. 283). I have paraphrased these below, adding data of my own in some cases:

First of all, by taking advantage of the work done by phonetic principles, their theory can automatically account for certain details of intonation contours which other

theories must attempt to explain phonemically. One such principle is catathesis, the gradual declination of pitch within the domain of the intermediate phrase. Beckman and Pierrehumbert (1986) illustrate this with a list of fruits, uttered with a "descending staircase" of tones (p. 273), reprinted in Figure 2.2.

H* H* H* H*

(2.9) Blueberries, bayberries, raspberries, mulberries, and

H*

brambleberries.

A system which attempts to assign phonemically or morphemically distinct tones to each of these levels would be obliged to include an unlimited number of levels since, in principle, the list of fruits could always be expanded. However, Pierrehumbert and Hirshberg's system simply assigns H* to each fruit, capturing the commonality of meaning (all the fruits are being introduced into the discourse as new), and letting the catathesis rule explain the declination.

Second, Pierrehumbert and Hirshberg's system makes no particular distinction between nuclear and prenuclear tones; any lexical item in an utterance can have any of the six possible pitch accents, depending on its role in the information structure. Therefore, there is no need to develop a separate inventory of tones to describe the prehead as in Halliday's system. This is also illustrated in the fruits example above. As far as their discourse role, the fruits

have equal status, providing no justification for the division of this list into prehead fruits and nuclear fruits. Halliday's system, as we saw, posits only one main nuclear pitch accent per contour in the unmarked case, and at maximum two in the marked case. Thus, long utterances with many accents such as the fruits example defy interpretation in his system.

Moreover, it is not the case that multiple pitch accents are confined to listing sequences. Cruttendon (1986), who, as Halliday, argues in favor of a single nuclear pitch at the right end of utterances, is forced to admit that certain utterances contain several new ideas, none of which stands out as nuclear. He considers the following two "out-of-the-blue" utterances in which all the information is considered new.

H* H* H*

(2.10) It's not quite the right shade of blue.

H* H* H*

(2.11) Her face used to be much fatter (p. 51).

He notes that these are likely to be uttered with a descending series of pitch accents and therefore do not conform to a theory of right-end nuclear pitch. He states:

> Some analysts indeed think that problems of this sort argue for abandoning the whole notion of nucleus and just settling for a series of pitch accents; this is an extreme view and one which does not take into account the very large number of

> cases where the nucleus assignment is straightforward (pp. 51-52).

Rather than extreme, I would say that this view is simply more explanatory, especially since in the data from the current studies, I discovered numerous utterances of this sort. This is particularly true in the academic lecture genre where a series of new ideas is often presented in a single utterance. We see this in the following introduction to a lecture given by a native speaker (Wennerstrom, 1991, p. 189), shown in Figure 2.3 as a series of downward "stair steps":

```
                H*    H*          H*                H*
(2.12)  We're all intuitively familiar with the idea of a
                  L*     LH%
        derivative.
```

Within this single intonational phrase, four of the content words contribute a significant portion of the new idea (*derivative* has already been given as the topic of the lecture). For Pierrehumbert and Hirshberg this poses no problem; there are simply four H* pitch accents. Within Halliday's system, the claim would be that the last prominent words, *idea* and *derivative* have a special significance as the nucleus of a contour 4 (rise-FALL-rise), which is hard to maintain in the face of this evidence.

The final argument against contour systems is that by decomposing Halliday's "tail" of an utterance into smaller units of meaning (a final pitch accent, phrase accent, and

boundary tone), finer interpretive power is achieved. For instance, in Halliday's terms, H*HH% and H*LL% are completely distinctive contours, being respectively a Contour 2 (rising) versus a Contour 1 (falling). However, for Pierrehumbert and Hirshberg, the common meaning "new" of the H* pitch accent can be extracted from both, regardless of the direction of the boundary. In (2.13) is an example from Pierrehumbert and Hirshberg (p. 286), which is compared to my own (2.14):

H* H* H* LL%

(2.13) The train leaves at seven.

H* H* H* HH%

(2.14) Does the train leave at seven?

In both examples, the content words *train, leaves,* and *seven* are associated with H* pitch accents, which tells us that these words are being newly introduced into the discourse. However, the boundaries differ and this affects the interpretation of how the utterances as wholes fit into the discourse. The first ends with a low boundary, indicating that it is an independent statement, whereas the second with a high boundary elicits a response to complete its interpretation. Within the componential approach, it is possible to discuss the role in discourse of the lexical items, as indicated by the shape of the pitch accents, separately from the role of the utterances themselves, as indicated by the boundary tones. A contour approach cannot capture this common interpretation of the role of the lexical

items, since rising and falling contours are taken as morphemically distinctive.

To summarize at this point, Pierrehumbert and Hirshberg's componential model of intonational meaning is efficient because it identifies a skeletal sequence of meaningful tones to define the shape of an utterance, while relying on phonetic principles to fill in the details of the contour shape. By allowing for more than one pitch accent per intonational phrase, the model explains many utterances with sequences of several accented lexical items. In contrast, Hallidayan-based systems cannot account for such sequences, since they posit a single nuclear pitch accent at the right end of the utterance. Finally, by analyzing pitch accents, which associate with lexical items, separately from phrase accents and boundary tones, which associate with phrases and utterances as wholes, a componential model is able to tease out finer distinctions of intonational meaning than can a contour model.

It should be noted that Pierrehumbert and Hirshberg's componential system does not preclude the existence of "idiomatic" intonational contours. For example, Sag and Liberman's (1975) "surprise/redundancy" contour has a consistent form and a dual meaning. It can either show surprise:

(2.15) (Professor walks into a room where he plans to teach; sees that the blackboard is painted orange):

H+L* L+H* LL%

Prof: The blackboard is painted orange! (p. 491)

Or it can show redundancy or obviousness, as in the same situation continuing below:

(2.16) Student: Well, why don't you start the class?

Prof: (Isn't it obvious, you fool!)

H+L* L+H* LL%

The blackboard is painted orange!

This sentence and its contour are easily accounted for in Pierrehumbert and Hirschberg's system of tones as L* H* LL%. *Blackboard* is salient, but already obvious in the context of the classroom so it gets a H+L*; *orange* is contrastive to the expectation of a black blackboard, so it gets a L+H*; and the boundary falls because there is no subsequent discourse needed for the interpretation of the utterance. The point is that participants in a speech community becoming familiar with common contour shapes as "idioms" is not incompatible with there being smaller grammatical components underlying these contours.

2.2.3 Brazil's Model

Finally, there remains one more comparison to be made: between Pierrehumbert and Hirshberg (1990) and Brazil (1985). These two models, though developed independently, appear to have remarkably similar configurations. The following

similarities exist between the two:

1. Both have a binary distinction at final utterance boundaries: for Pierrehumbert and Hirshberg, H% and L%; for Brazil "referring" tones and "proclaiming" tones. These describe the end point of the intonation contour.
2. Both have a binary distinction prior to the final boundary: for Pierrehumbert and Hirshberg, H or L phrase accent; for Brazil, the + or -. These describe the shape of an intonation contour between the final pitch accent and the end of the intonation contour.
3. Both have tones which associate with lexical items that contribute to the information structure. These are "salient" for Pierrehumbert and Hirshberg, or "prominent" for Brazil. As discussed, there are six possible pitch accents for Pierrehumbert and Hirshberg; for Brazil, the choice is among high, mid, or low.
4. Neither model regards function words or unstressed syllables as participating in the pitch accent system. Pierrehumbert and Hirshberg regard these as simply not having pitch accents and note that their pitch will be governed by phonetic rules of

interpolation; Brazil calls them "enclitics" and "proclitics" and does not discuss their pitch in particular.

In summary, both models take a componential approach, regarding an intonation contour as a series of pitch choices that contribute to a surface contour, which, as we saw in earlier examples, provides superior interpretive power over holistic contour models.

There are, however, a few distinguishing characteristics between these two models. First of all, Brazil insists that there is a maximum of two prominent, meaning-bearing tones per intonational phrase, the "key" and the "termination", while Pierrehumbert and Hirshberg place no such restriction on the number of pitch accents per phrase. As was discussed in the previous section, an approach that allows for multiple pitch accents is preferable.

Second, Brazil's model is more developed in its treatment of the beginnings of utterances, especially in their interactional value. This is probably due to the fact that he works with real conversation data, while Pierrehumbert and Hirshberg's examples are largely constructed. His notion of key refers to the starting pitch level of the first prominent word in an utterance, which may be high, mid, or low, and which expresses the relationship between the current utterance and the previous one. We will return to this idea in Section 2.3.2 where I argue that key can be said to function as an

initial boundary tone in Pierrehumbert and Hirschberg's terms.

2.3 Adjustments in the Componential Approach

So far in this chapter I have stressed the importance within Pierrehumbert and Hirshberg's model of there being meaningful units below the level of the intonation contour. At this point I will suggest two areas where I believe Pierrehumbert and Hirshberg's model can be expanded to explain certain additional aspects of the role of intonation in cohesion. First I will discuss the L* pitch accents and suggest that these need to include "deaccent" on given items as part of their interpretation. Second, I will recommend that initial pitch boundaries, Brazil's key and a higher level, topic-transitional pitch morpheme, the "paratone" (Brown, 1977; Yule, 1980) are necessary as well. With these modifications, I submit the model used in the studies in Chapters 4-6, which I maintain is still very much in the spirit of Pierrehumbert and Hirshberg's componential approach.

2.3.1 Deaccent

One important aspect of intonation is "deaccent," a low pitch which occurs when an item is given or inferable from previous discourse or context (Chafe, 1970, 1994; Halliday, 1994; Ladd, 1980, 1984; Werth, 1984). Chafe (1994) considers the lower pitch to be S's signal that the associated item is already active in consciousness, and requires minimal

cognitive effort for its interpretation. Here are two examples of deaccent from Ladd (1980, p. 52) (deaccent is indicated with small print):

(2.17) A: Has John read Slaughterhouse-Five?

B: No, John doesn't read books.

(2.18) Harry wants a VW, but his wife would prefer an American car.

"In each of these examples," says Ladd, "the deaccented noun [*books; car*] has somehow been referred to or alluded to earlier in the discourse" (p. 52). He also points out that deaccent is not restricted to sentence-final position. This is illustrated by a comparison between the following two examples (p. 55):

(2.19) A: What's the matter?

B: There's nothing to make French Toast out of.

(2.20) A: Why don't you make some French Toast?

B: There's nothing to make French Toast out of.

In the latter case, *French Toast* has already been introduced into the discourse and is therefore deaccented, in spite of the fact that it comes not after but before the accented word *out*. Ladd's position is that, "the deaccenting of a syllable can best be understood as a *relative weakening of its hierarchical rhythmic position*" (p. 56).[3] This is important in explaining why deaccented items may not always have low pitch in absolute terms; deaccent interacts with the underlying rhythmic structure of an utterance which affects

the associated word's position relative to its neighboring constituents. Thus, a deaccented word may actually be higher in pitch than another word further along in the same utterance.

I propose to represent deaccent as one instance of Pierrehumbert and Hirshberg's L* pitch accent, since their general description of L* is that it is used for items which are salient, but are not being added to the mental representation of the discourse. They list four reasons that S may not wish to add an item to the mental representation, and would therefore associate it with a L*. First, it may be part of a yes/no question in which S is not predicating anything about the utterance, but hoping that H will do so as a response. Second, L* may convey incredulity, that is, that S thinks that what is predicated is incorrect. Third, an item may already be believed to be part of the mental representation, or S may think that it should be mutually believed. Fourth, L* may associate with material that is "extrapropositional, such as greetings, vocatives, and so-called cue phrases" (p. 293) and other conventional statements. In line with this characterization, deaccent would coincide with the third instance of L*: associated with items which S wishes to make salient, but which are already given, understood, or inferable from the text or context and thus already in the mental representation. However, in personal communication, Hirschberg (May 16, 1995) maintains

that deaccent means not having a pitch accent at all and cannot be equated with L*. This would put deaccented items in the same category as function words which have no pitch accents and whose pitch is therefore interpolated from the pitch of surrounding tones according to phonetic principles.

Nevertheless, I submit that a model of intonation needs to account for deaccent as a pitch accent morpheme rather than merely a lack of pitch accent. I propose that speakers associate this morpheme with material which is salient but given or inferable from previous text or context. Despite Hirschberg's personal communication, cited above, I do not find deaccent in any way incompatible with Pierrehumbert and Hirshberg's (1990) characterization of L* pitch accent. Thus, additional morphemes need not be added to Pierrehumbert and Hirshberg's model; the claim is that L* includes deaccent, associated with givenness or inferability among its several interpretations.

What follows are three arguments in favor of a deaccent morpheme being equivalent to L*, involving the notion of "salience." I suggest that the salience of an item should include its psychological (and therefore its phonological) prominence for S, as well as its importance in interpretation for H. In the examples, the crucial deaccented items are shown in small print.

First, evidence of the psychological salience of deaccent comes from the fact that speakers usually maintain deaccent

during a repair, just as they would other pitch accents. The following example of this occurred in a lecture given by a native speaker:

(2.21) We're all intuitively familiar with the idea of a derivative. It is simply a rate of change. We'll be talking about the mathematics of how one thing changes as

L+H* L* L* L+H* L* L*

another things change, as another thing changes, we will define this concept . . . (Wennerstrom, 1991, p. 189).

In this excerpt, the lecturer states that "one thing changes," which he then wants to contrast with "another thing changes." This contrast is appropriately made by associating L+H* pitch accents with the words *one* and *another* and then by deaccenting the second utterance of the words *thing changes* since this is given in previous text. However, he misstates the second part of the contrast as "as another things change," and quickly repairs to "as another thing changes." The crucial point is that he maintains the deaccent during the repair (shown in Figure 2.4 where the curser marks the repair). This is noteworthy because although we might expect the lecturer to accent the corrected words for added clarity, he preserves the low-pitched signal of givenness as part of the rearticulated form. This is an indication that the speaker considers the deaccent necessary for a coherent interpretation. It appears

that other pitch accents maintain their integrity in repairs as well. In the same excerpt, the word *another* maintains its contrastive L+H* pitch accent in the repair; similarly, in the next example we see a H* pitch accent maintaining its shape during a repair in a conversation (see Figure 2.5):

(2.22) H* H*
. . . the only way she could get avenge- or revenge for the fact that her father had been unrightly accused . . . (Wennerstrom, 1991, p. 207).

As a second type of evidence for its salience, it may be observed that deaccent often takes effect across speakers. In other words, if one speaker introduces a referent into the mental representation of the discourse, another speaker may assume it is still accessible and deaccent it. Here is an example from a lecture on marketing:

(2.23) Kate: Also the disposable income of the market that you're selling to.

Daphne: OK. Can you think of any ways that you can, um, if you're, if you're trying to market something, you as a, somebody in the United States trying to market something in, uh, China.

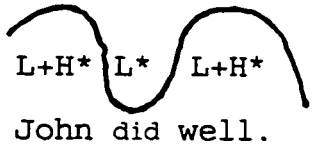
L* L*
How would you find out about disposable income? Where would you go to find that information? (Wennerstrom, p. 200, 202).

In the first line of this text, Kate introduces the idea of

"disposable income." Several lines later, Daphne deaccents the term as she repeats it (shown in Figure 2.6). This gives an indication that both participants are aware of the structure of the mental representation of the discourse, and that its contents remain psychologically salient.

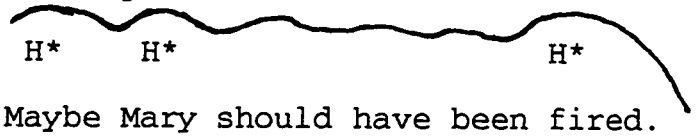
Turning to phonological salience, there are cases where deaccented items are lower in pitch than they would be with a simple lack of accent. Of course, relative pitch value depends on many factors (catathesis, value of surrounding pitch accents, distances between pitch morphemes, etc.), but the following constructed minimal pairs show cases where a deliberate pitch lowering on the part of S is the only explanation for how low the pitch falls relative to surrounding accented items.

(2.24) a. Tom did poorly and John as well.



b. Tom did poorly and John did well.

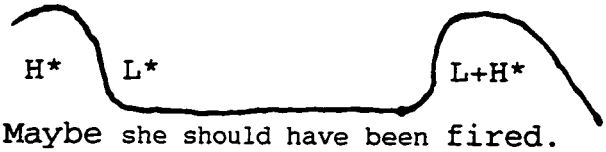
The first example shows a simple contrast between the function word *as* and the discourse-linked, deaccented *did*, which is uttered with a lower pitch.



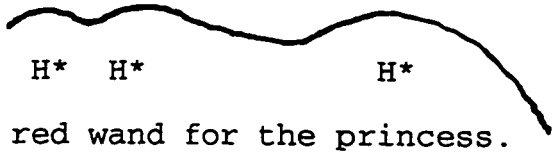
(2.25) a. Maybe Mary should have been fired.

H* H* H*

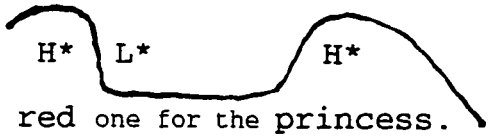
b. Maybe Mary should have been promoted, but who knows?



Maybe she should have been fired.



(2.26) a. A red wand for the princess.



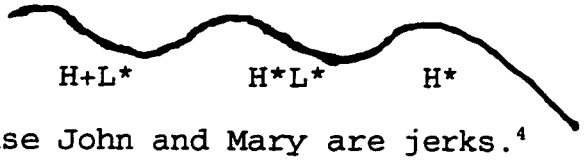
b. A red one for the princess.

In examples (2.25) and (2.26), attention goes to the function words, which follow H* pitch accents in the (a) sentences and follow deaccents in the (b) sentences. In the (b) cases, the function words fall to the bottom of S's range, which I maintain can only be due to a L* pitch accent preceding them on *she* and *one;* in the (a) cases, the function words are uttered at a higher pitch because their pitch is interpolated between that of the surrounding H* pitch accents on *Mary* and *wand*. In sum, unless we take deaccent to be equivalent to L* pitch accent, there is no explanation for the sharp drop on *did, she* and *one;* their pitch would be the same as that of the function words.

The next example shows a use of H+L* pitch accent which Pierrehumbert and Hirschberg contend can be used to confirm

that which is already known since the accented items are inferable:

(2.27) A: Why haven't John and Mary sent us a thank you?

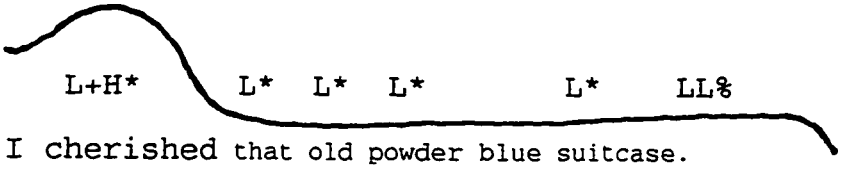


B: Because John and Mary are jerks.[4]

This is also a special sense of given information in the terms we have been using. It conveys the meaning that A and B have previous experience stored in memory of dealing with rudeness from John and Mary.

A further phonological fact is that it is possible for deaccent to be maintained throughout fairly long phrases. Consider the following constructed example.

(2.28) My grandmother had an old powder blue suitcase which I inherited. Although it was pretty beat up,



I cherished that old powder blue suitcase.

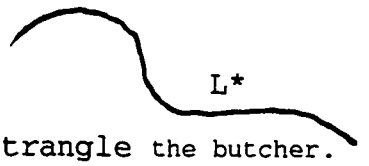
The word *cherished* is contrastive in the context of the suitcase being "beat up," so it has a L+H* pitch accent. The final boundary is LL% since this is an independent statement. There is nothing to explain the sudden drop to a low pitch on the word *that* unless we admit the effect of deaccent; with no accent at all, the tail of the utterance would drop gradually toward the bottom of the speaker's range at end of the phrase, as in:

(2.29) *I cherished that old powder blue suitcase.

The final argument refers to the importance in discourse interpretation of deaccented items. The following examples illustrate cases where the deaccent instigates the search for an antecedent on the part of H (Ladd, 1980, p. 64-65):

(2.30) a. A: How did your operation go?

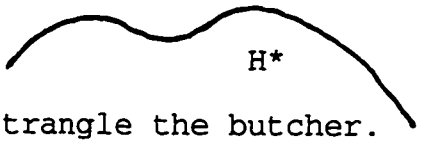
B: Don't talk to me about it --



I'd like to strangle the butcher.

b. A: How did your operation go?

B: Don't talk to me about it --

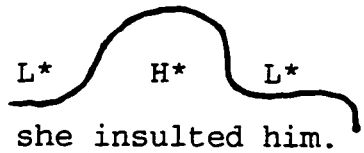


I'd like to strangle the butcher.

In (a), where *butcher* is deaccented, A takes it as inferable and searches for an antecedent in the mental representation of the discourse, in this case, the doctor who performed the operation. In (b), where *butcher* has a high pitch accent, A may attempt to construct a context in which a person who chops meat in a grocery store is involved. The point of this example is that deaccent conveys a specific meaning in English, namely that the interpretation of the deaccented item must be made with respect to what is already in the common ground between S and H. Here is another similar example

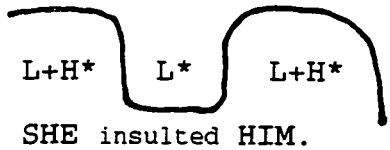
(Lakoff, 1971):

(2.31) John called Mary a republican and then



she insulted him.

b. John called Mary a republican and then



SHE insulted HIM.

Upon hearing the low pitch associated with *insulted* in (b), H searches for a relevant antecedent. The most likely possibility is the act of calling someone a republican. The point of these examples is that since the deaccent is the only factor that allows us to distinguish between the two interpretations, it must be considered a meaning-bearing unit.

In view of these facts, I propose to include deaccent within the current intonation model as a L* pitch accent (or either of its variations, L*+H and H+L*), instead of it being simply a lack of pitch accent. This adjustment, along with a second one described in the next section, is reflected in Figure 2.7, a diagram of the theoretical model of English intonation used for the remainder of this dissertation.

2.3.2 Initial Pitch Boundaries

No discussion of discourse intonation is complete without a word about pitch changes which come into play at the

beginnings of discourse units. I propose that previous scholars' treatments of initial pitch boundaries can be generalized and incorporated into the componential model of intonation under discussion. In this regard, I will discuss "paratones" (Brown, 1977) and "key" (Brazil, 1985), which I believe to be parts of the same phenomenon.

Paratones, so named because they correspond to written paragraphs in having a single topic, have been noted by several scholars (Brazil, Coulthard & Johns, 1980; Brown, 1977; Brown & Yule, 1983; Couper-Kuhlen, 1986; Wennerstrom, 1992, 1994). In phonetic terms, Beckman and Pierrehumbert (1986) discuss this phenomenon as an increased pitch range at topic-initial boundaries and a compressed pitch range at the ends of topics. However, they do not provide examples of extended discourse to illustrate the details. In two independent empirical studies of paratones, (Wennerstrom, 1992, 1994) I found statistically significant differences in pitch range between sentences in paragraph-initial position and those in paragraph-medial position in oral readings by native speakers of English. In both studies, when native speakers perceived a topic shift, marked orthographically by the paragraph indentation, they increased their pitch range over the first sentence of the new topic. When a sentence of the same structure was embedded within a paragraph, it was uttered with a narrower pitch range by the same speakers. In my 1992 paper, I characterized the paratone as a floating tone

because instead of reliably associating with the very first word of the utterance, its effect was spread over several words, even across an intermediate phrase boundary. This spreading behavior is similar to that of Pierrehumbert and Hirschberg's (1990) phrase accent and boundary tone, which also affect the shape of the entire tail of an utterance. I suggested then, and I continue to maintain, that by simply extending the scope of how much material a boundary governs to higher levels of discourse we can retain Pierrehumbert and Hirshberg's notion of boundary pitch as a signal of hierarchical relationships in discourse without adjusting the substance of their model.

Scholars have further distinguished initial pitch boundaries which associate with more minor units of discourse. Yule (1980) discusses "major" and "minor" paratones, the latter being associated with subtopics within the former. Beckman and Pierrehumbert (1986) and Couper-Kuhlen (1986) also suggest that there may be a hierarchy of discourse-level intonation structures. I propose that Brazil's (1985) notion of key, also an utterance initial pitch, can be regarded as the lowest level in a hierarchy of initial pitch boundaries. In addition, Brazil's notion of tone concord between the terminal pitch of one constituent and the onset pitch of the subsequent can be expanded to a higher level of discourse to include the paratone.

According to Brazil, the pitch of the first content word

of an utterance (in terms of our model, the first pitch accent) is the key of that utterance, and it may be high, mid, or low. Likewise, the end of an utterance has a "termination" level, also high, mid, and low. Key and termination refer not to the shape of the contour itself, but to its relative location in the speaker's range. Thus, even a yes/no question with a high rising boundary (HH%) could possibly end in a low termination, that is, low relative to the speaker's range. In conversation, key and termination capture the phenomenon of "tone concord" (also discussed by Schegloff, 1995) by which a speaker indicates his stance with respect to the previous speaker's contribution. In order to show agreement or rapport, the speaker can pitch the new utterance to coincide with the termination of the previous utterance; on the other hand, to present some kind of opposition, the speaker can choose a different key to initiate the new utterance. Here is an example from Brazil (p. 85-6) of three variations on this theme. It should be noted that all of A's utterances end with the rising pitch of a yes/no question (HH%), and all of B's end with a falling (LL%) pitch. The first case is a simple clarification pair with tone concord. There is no contrast or contradiction expected:

(2.32) A: Do you underSTAND (mid)? B: YES (mid)

In the second case, A uses a high termination to imply that perhaps B will not have understood. B returns with the same high key to assert that he absolutely understands. The high

pitch indicates that both speakers are making a contrast with the proposition of misunderstanding, and their tone concord indicates that they are in agreement on this:

(2.33) A: Do you underSTAND (high)? B: YES (high)!

In the final case, A uses a mid termination to convey the expectation that B will have understood. However, B has not understood, so he uses a high key on *no* (lack of tone concord) to indicate a contrast with the wrongly assumed agreement:

(2.34) A: Do you underSTAND (mid)? B: NO (high)!

Although Brazil's discussion centers on key in dialogues, while paratones have mainly been observed in monologue genres (newscasts and oral readings), they have in common the fact that they are both intonational choices at the initial edges of various-sized textual constituents. In monologue, a high paratone can be regarded as lack of tone concord on a grand scale, to indicate complete shift of topic. Although the discussion of paratones in the literature has mainly focused on high pitch at topic shifts, we can also readily imagine situations where a speaker would choose a low paratone, just as there is low key; for example, going off on a tangent, providing some quick flashback material in a story, or an "aside" in conversation might all be distinguishable by low-pitched onsets.

Moving down the hierarchy, minor paratones mark smaller scale rhetorical junctures and key marks junctures at the level of the utterance. In dialogue, similar sorts of

intonationally marked junctures may occur from the topic level down to the utterance level; however, these are more likely to be interactional in their interpretation, indicating the speakers' stance toward each other's contributions to the text. In sum, the generalization can be made that major and minor paratones and key are all intonational choices available to speakers as they initiate a new unit of text, of whatever size, to distinguish it from previous text and to provide information about how it is related to that text.

With little adjustment, we can smoothly incorporate these initial pitch phenomena into a more general hierarchical theory of boundary tones by making two simple additions to the current model: key, which governs the utterance (represented by %H and %L) and paratones, which govern higher-order constituents of discourse (represented by &H and &L), can both be added to the existing inventory of boundary tones. Both are initial boundary tones which provide information about how the constituent with which they associate is related to previous discourse.[5]

2.3.3 A Synthesis of Intonation Models

At this point, we are in a position to construct an intonation model which combines the essential features of previous models. The result will be a model which is most similar to Pierrehumbert and Hirshberg's in having pitch accent and boundary tone components, but includes certain

aspects from Ladd (deaccent), Brazil (key), and Brown (paratones).

The resulting model of English intonation which underlies the studies done in this dissertation includes the characteristics summarized as follows:

1. The fundamental meaningful unit of intonation is the tone rather than the contour.
2. The pitch accent has six possible forms (H*, H*+L, L+H*, L*, L*+H, H+L*) with six possible meanings, and provides an indication of how the associated lexical item fits into the information structure of the discourse. There is no theoretical upper limit to the number of pitch accents in an intonational phrase, nor is there a nuclear pitch accent with special status.
3. Deaccent is considered a L* pitch accent, (or L*+H, H+L*), used to make salient an item already assumed by S to be in H's mental representation of the discourse.
4. Phrase accents (H, L), boundary tones (H%, L%), key (%L, %H), and paratones (&H, &L) associate with boundaries of phrases, utterances, and even larger, discourse-level constituents, to provide information about hierarchical relationships among these constituents.
5. Both final and initial boundaries have significance: final boundaries indicate how S intends a constituent to be interpreted with respect to the subsequent constituent; initial boundaries indicate how S positions

the new constituent with respect to the previous. Lack of concord between these two indicates contrast, disagreement, or change.

6. Not every word needs to have a pitch accent. The pitch of function words is interpolated from surrounding pitch accents and boundary tones.

Figure 2.7 gives a schematic representation of the components of the revised intonation model, indicating the hierarchical relationships among its components. Solid lines represent hierarchies within the text; dotted lines show the associations of intonation to text. This configuration establishes a more complex network of discourse-level relationships than does the current form of Pierrehumbert and Hirschberg (1990). It captures the notion that large chunks of organizational structure are marked intonationally in English.

2.4 Measuring Intonation by Computer

The technology used to analyze data[6] in this dissertation is a Computerized Speech Lab (CSL machine[7]), which takes a speech signal, either live or from a tape, and provides graphs of pitch and intensity measured through an adjustable increment of time. The machine also has a playback function so that all or part of the pitch graph shown on the screen can be replayed. A small segment of speech can be expanded in order to observe finer levels of detail. By placing cursers

at specific points along the pitch graph, a reading of the exact pitch at that point can be obtained. This is an invaluable tool for speech study.

In the current studies, taped speech was played into the CSL machine, usually in 4- to 8-second segments. On the screen appeared a frozen graph of the pitch for that segment of speech, which could then be printed out if desired. Using cursers, measurements were made of the pitch of the whichever discourse items were under study. The resulting data were numeric, the pitch measurement unit being in Hertz (Hz).

Unfortunately, several unexpected problems presented themselves at the measurement stage. Since there are so few studies of intonation in natural discourse using computers, I was obliged to make my own decisions regarding questions of measurement specificity, statistical methods, and the identification of intonational patterns. While I was not always able to solve every problem, I hope to show that I dealt with them in a consistent manner throughout the data measurement. The first problem was that the machine itself was so accurate and sensitive to sound that it picked up background noise and tape buzz, making it difficult to get a good reading of the actual speaker's voice in the case of noisy tapes. This was particularly acute in certain student-made tapes which had been recorded in noisy areas or with poor equipment. If this problem only involved the occasional utterance, I simply omitted those utterances from study; but

if interference reached the point that I was not able to get accurate measurements in the majority of that subject's utterances, I excluded the subject. Figures were also affected by background noise, and in some cases a clear illustration of a speaker's pitch track could not be provided at all. In several of the figures of this dissertation, I have erased some of the interfering noise marks so that only the speaker's pitch is illustrated.

Another related factor was that although the machine has a statistical calculator which gives readings for mean, median, and other summary statistics, it did not prove to be reliable because it takes all background noise into the calculation. Therefore, it was not possible to get an accurate mean or median pitch for a given segment of speech. It was possible, however, to collect range data, using cursers to measure the highest and lowest points of speech.

The next problem was to decide at which point to measure a word's pitch, given a gradually sloping, continuous pitch stream. The decision was made to consistently take the measurement at the center of the stressed syllable which was judged to be the point of highest amplitude (loudness)[8]. Consistency of the measurement point was especially important at utterance boundaries where the tail of the boundary is considered morphemically distinct from the final pitch accent. In Figure 2.8 the vertical line going through the center of the word *give* is at the point of the syllable's highest

amplitude, shown in the upper panel. The measurement for the word was recorded at this point as 200 Hz, although the pitch continues to fall into the boundary where the amplitude trails off.

Similarly, at the beginnings of utterances, the effect of key, combined with the onset of catathesis, often produced a very high pitch in phrase-initial position, even if the first word(s) was a function word. In the chapter on lectures, where I was attempting to average the pitch of many function words for each speaker, this interfered with the average so much that I decided to consistently exclude the first function words of an utterance from measurement.

The final issue involved the fact that speakers' ranges vary significantly by gender and on an individual basis. In order to compare intonation among subjects with different ranges, the actual pitch measurements in Hz would distort the comparison. For example, for a speaker with a narrow range, a 20 Hz pitch increase to make a contrast is in my perception quite noticeable, whereas for a speaker with a large range, 20 Hz is barely perceived as a shift. To solve this problem, I adjusted the raw measurements for pitch range by converting them into percentage values of each speaker's range within each intonational phrase. This adjustment allowed speakers' pitch values to be compared on the same scale regardless of gender or individual range variations.

As a final word, it should be noted that I have ventured

into largely uncharted methodological territory, especially since the speech technology involved is quite new. As outlined in Chapter 1, the combination of natural discourse data, machine measurement, and large numbers of data points for statistical analysis, is seldom found in the intonation literature. For this reason, I have not had the luxury of being able to model my methodology on previous studies. It is my hope that others will build on and improve this methodology as computerized measurement devices become more widely available and as applied studies of discourse intonation continue to develop.

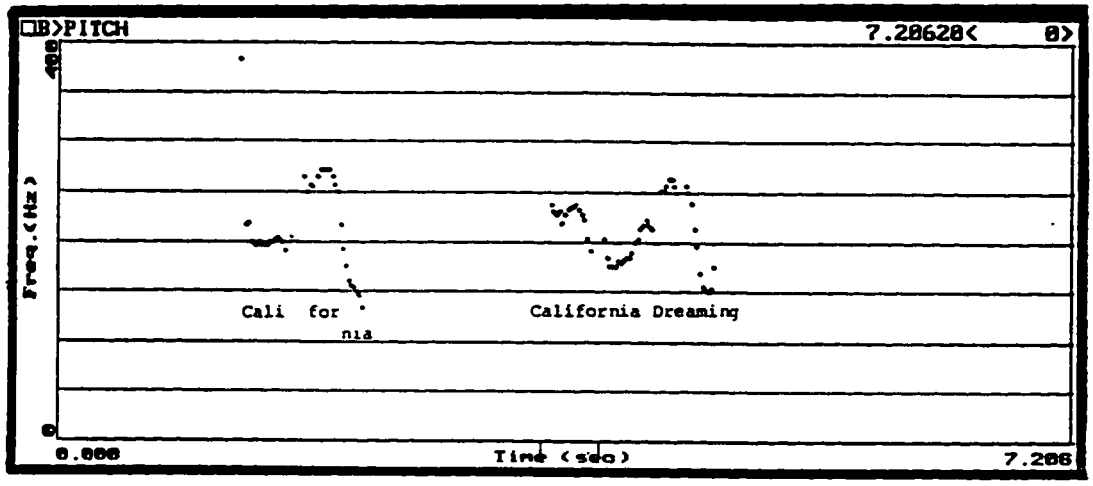


FIGURE 2.1 The Rhythm Rule shifts stress to the first syllable of *California* before *dreaming*.

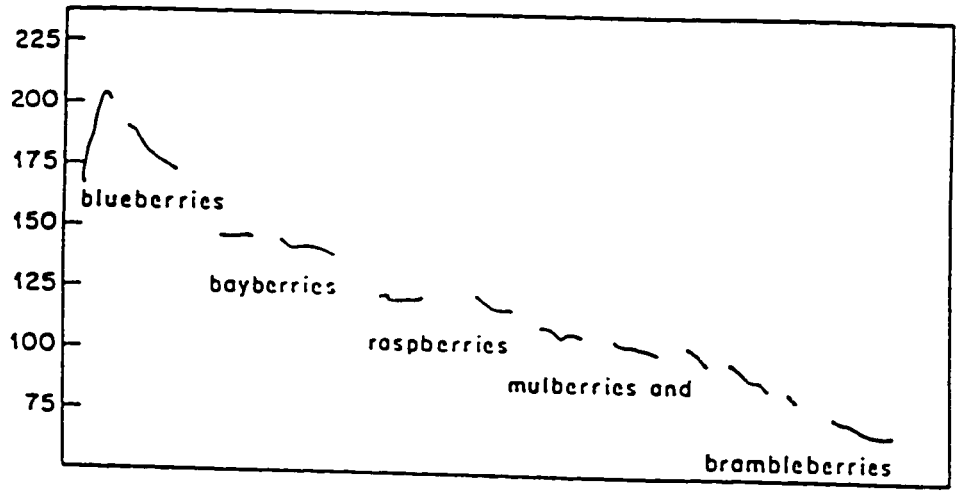


FIGURE 2.2 The list of fruits illustrates the phonetic principle of catathesis (reprinted from Beckman & Pierrehumbert, 1986, p. 273).

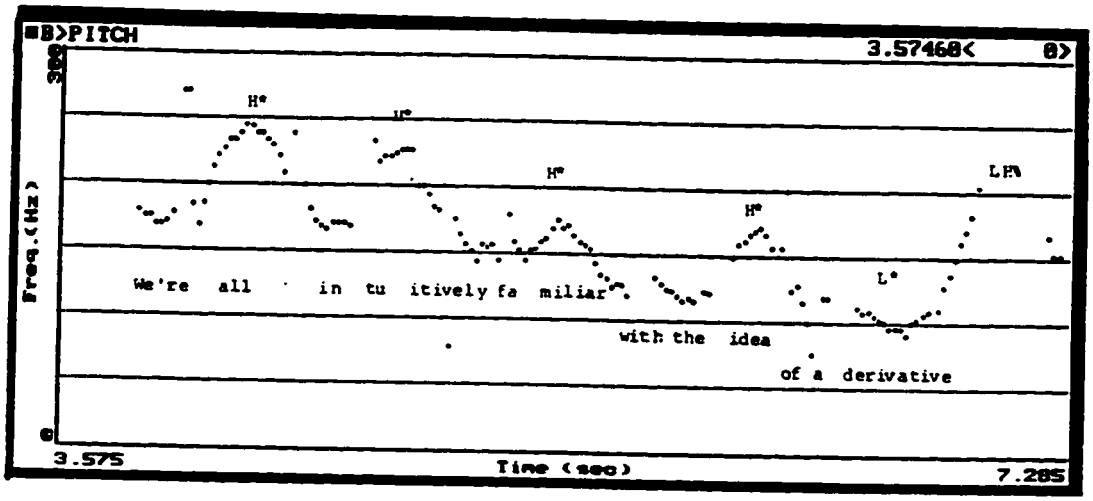


FIGURE 2.3 This series of descending H* pitch accents provides counterevidence to the notion of a single nuclear pitch accent.

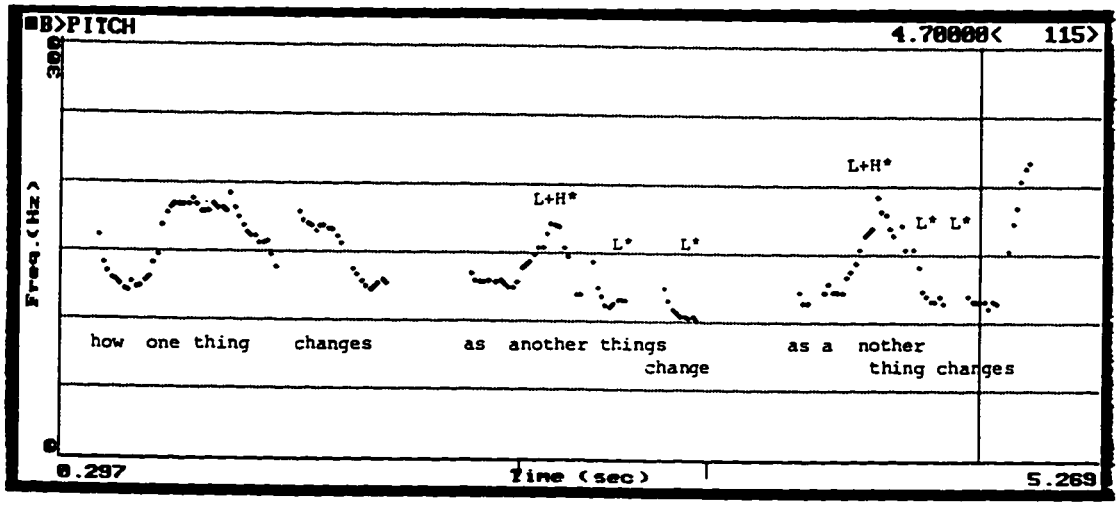


FIGURE 2.4 Deaccent is maintained on the repaired words *thing changes* (at the curser), which shows its psychological salience.

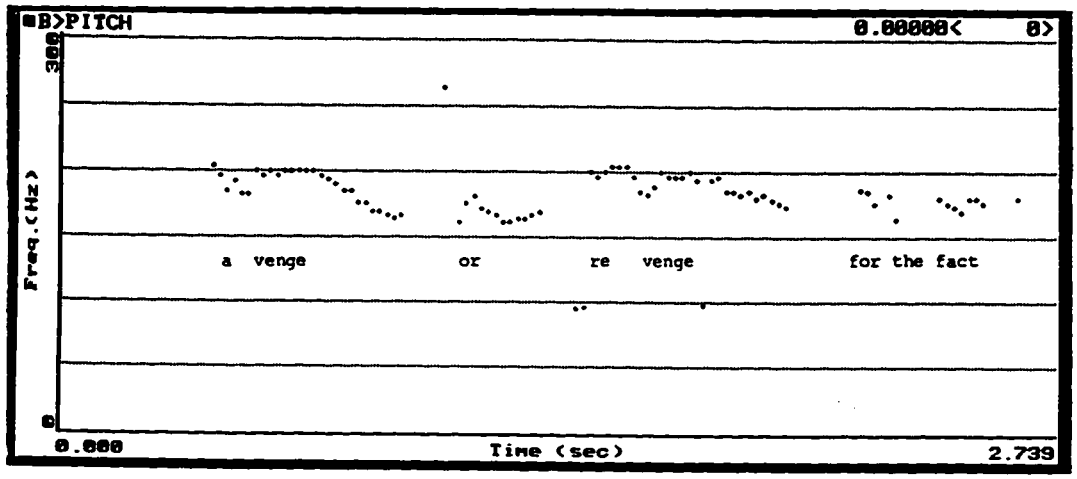


FIGURE 2.5 H* pitch accent is maintained during a repair.

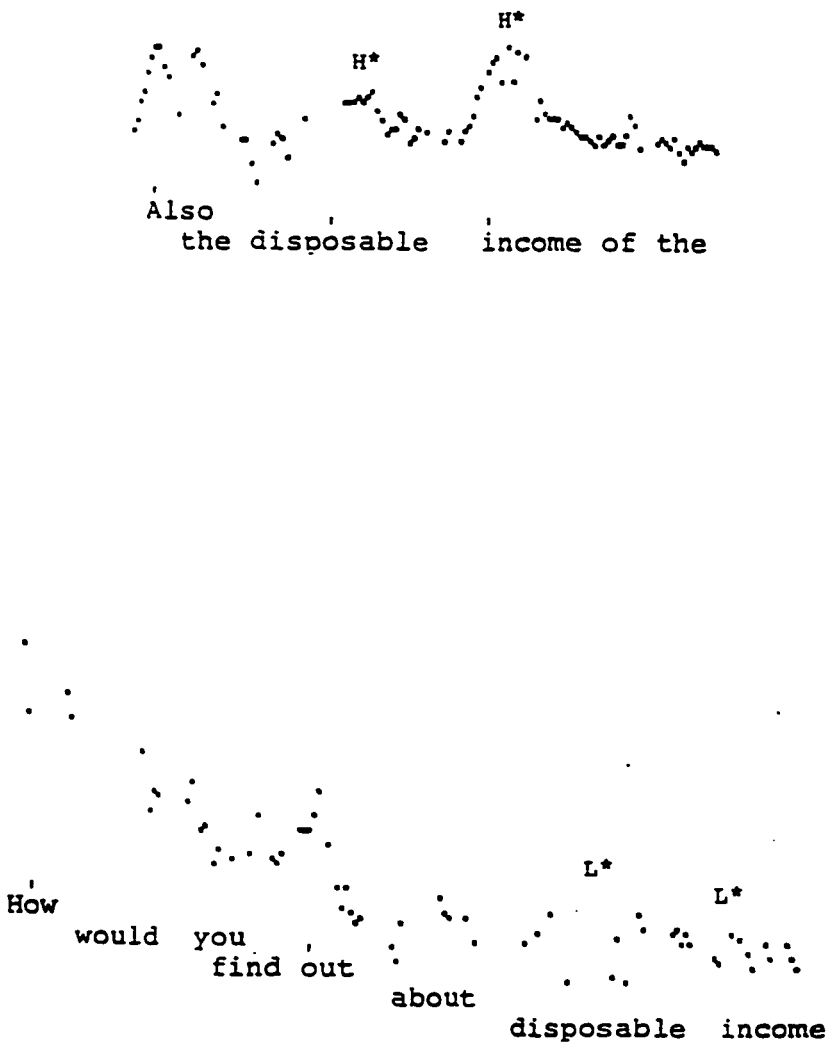


FIGURE 2.6 *Disposable income* is introduced by one speaker and deaccented by a second speaker.

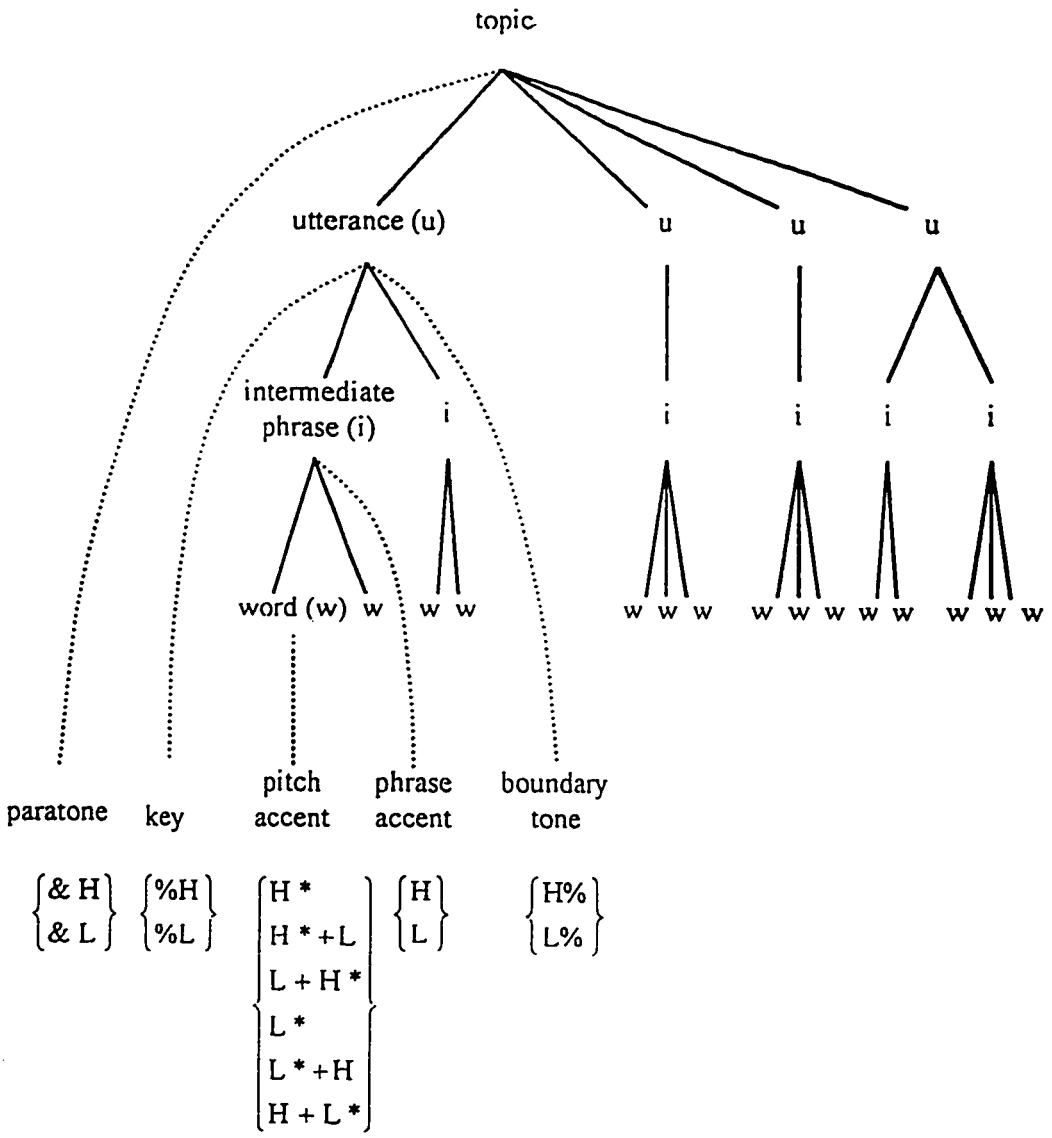


FIGURE 2.6 *Disposable income* is introduced by one speaker and deaccented by a second speaker.

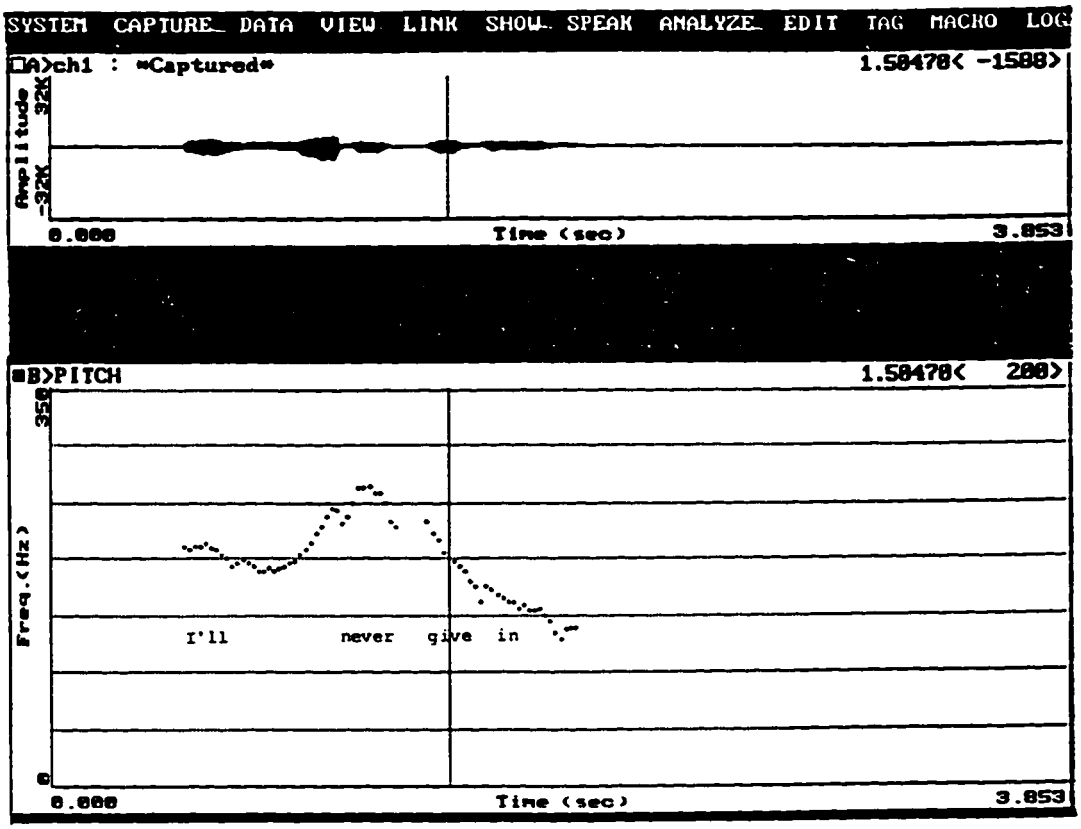


FIGURE 2.8 The pitch measurement of a word (bottom graph) is taken at the point of highest amplitude (top graph).

Notes to Chapter 2

1. This is not to say that tone languages do not have discourse-level intonation in addition to lexical tones.

2. I say "canonical" WH-questions because there are many cases where a WH-question has rising intonation.

3. I differ with Ladd in his next line of argument, that deaccent can trigger a "default accent" on it's rhythmic sister constituent. Although I agree that rhythmic adjustments occur (as Hayes', 1985 rules of eurythmy), I believe that Ladd's examples of default accent can all be explained independently by discourse-driven reasons for associating a pitch accent with the accented item.

4. Thanks to Phil Gaines for this example.

5. I am not prepared to take a position on whether these utterance-initial boundary tones might have contour counterparts as the final ones do. Nor am I certain of the status of the mid-level initial tones. Therefore I will be conservative and leave these as binary choices, (%H and %L for paratones; &H and &L for key), to coincide with the final boundary tones in Pierrehumbert and Hirschberg's original model.

6. All data were obtained from subjects voluntarily with permission.

7. Computerized Speech Lab, Kay Elemetrics, Model 4300.

8. This was done consistently, although there is no precedent in the literature that I know of.

Chapter 3

Intonation and Coherence

In the last chapter I argued in support of a model of intonation in which tonal units below the level of the intonation contour are recognized as bearers of meaning. At this point, I will continue the discussion with a more detailed look at how the tones within this model work together to provide cohesion in a text, and thus contribute to overall coherence in discourse. It is not unexpected that intonation plays a role in coherence, since we have already seen its linking functions in the last chapter: boundary tones can serve to link constituents together; low pitch accents can send the message "this should already be known or believed". In other words, these morphemes help communicators draw connections between what is uttered and what is represented in the mind where discourse links are made. The chapter begins with a discussion of mental representation, followed by an analysis of a sample text to show some of the linking functions that intonation performs.

3.1 Mental Representation in Communication

Many scholars agree that as people communicate, they construct a short-term mental representation of the discourse as it progresses. As we saw in the previous chapter,

Pierrehumbert and Hirschberg call the knowledge shared by participants in a discourse situation "the mutual belief space"; Werth (1984) refers to this as "the common ground"; Sperber and Wilson (1995) use the term "mutual cognitive environment"; Chafe (1994) discusses ideas recently introduced into the discourse as "active" or "semiactive" in consciousness. What all of these terms have in common is that they refer to a short-term memory system, constructed for the purpose of the current discourse, which can interact with other cognitive systems, such as perception and long-term memory. However, as Toolan (1996) points out, the idea of "mutual" knowledge is problematic. It is more accurate to say that participants in discourse have individual mental representations which will resemble each other to some extent if understanding is to take place. Each participant must make judgments about how the other participants will have constructed their mental representations.

According to Sperber and Wilson, there are four sources of input to a person's mental representation at any given point in a speech situation: The ideas currently being verbalized in the discourse, information from "encyclopedic" or long-term memory, ideas available from perception of the immediate environment, and new ideas, entailments, and inferences, derived from combining any of the information in the first three categories. Within the mental representation, Chafe (1994), Sinclair (1994), Levelt (1989), and others also

grant a special status to the utterance currently being verbalized by the speaker (S) and processed by the hearer (H). Sinclair points out that only in the utterance currently being processed are the exact words salient. Once we process text, it simply becomes mental representation along with the rest of our background knowledge. As evidence, he cites the difficulty one has in quoting someone verbatim, even a very short time after they have spoken, although the ideas can easily be remembered. Even in reading, maintains Sinclair, where we may exercise the option of rereading previous text, the norm is to process in a forward direction just as in spoken discourse. From the speaker's side, Levelt maintains that preplanning of speech takes place in clause-sized units, immediately prior to verbalization. Chafe comes to a similar conclusion, but considers thought groups to be prosodically determined; the unit of thought is the intonational phrase rather than the syntactic clause.

As the current utterance is parsed by H, one resource H has to bring to the interpretation is knowledge stored in long-term memory. There seems to be general agreement that so called "common" knowledge is stored in an organized way by situation, topic, setting and the like. Such clusters of associated concepts have been called "frames" (Minsky, 1975; Fillmore, 1985), "schemata" (Bartlett, 1932; Carrell, 1982), "scripts," and "scenes" (Schank & Abelson, 1977), based on repeated past experiences. Carrell (1982) emphasizes the

cultural influence on how we store information and interpret text. People from similar backgrounds or cultures are likely to have built similar schemata throughout their lives and are able to make similar associations and inferences with reference to their schemata. Carrell offers this picnic example (p. 484):

(3.1) The picnic was ruined. No one remembered to bring a corkscrew.

According to Carrell, those who are able to interpret this text as coherent do so because they have a schema for picnics which includes wine as an essential beverage and a schema for corkscrews which involves opening bottles of wine. The interpretation is made by parsing the utterances, retrieving the relevant schemata from long-term memory into the current mental representation, and drawing the picnics-need-corkscrews connection.

At this point, coherence may be defined quite simply: a coherent discourse is one for which a well-formed mental representation can be constructed. Coherence is a matter of degree rather than an all-or-nothing proposition, depending on the perspective of the person making the interpretation. Suppose, returning to the corkscrew example, that H is a 12-year-old child who has a schema for picnics that does not include wine as an essential beverage. This child may be able to construct a mental representation of the utterances in which it is clear that the picnickers were unable to open a

wine bottle, but may have trouble integrating the word *ruined* into the mental representation. For that child, this discourse would be only partially coherent.

3.2 Intonation's Role in Coherence

We are now in a position to show how intonation interacts directly with the short-term mental representation of a discourse in progress. By associating particular intonational morphemes with the phrase currently being verbalized, S signals to H how he or she expects each new utterance to fit into the mental representation which H is assumed to have constructed so far during their interaction. For example, as discussed in Chapter 2, a H* pitch accent on an item would indicate that that item is to be newly added to the mental representation; L+H* would indicate that the item is in contrast to an idea already in the mental representation; L* would indicate that the item is not be to added to mental representation, because it is already there, it is being questioned, or it extraneous to the topic. In short, intonation provides crucial information to help H draw connections between each new utterance and the mental representation in progress in order to derive a coherent interpretation.

Those familiar with the work of Sperber and Wilson will recognize that my characterization of coherence is similar to their definition of relevance. One main difference, however,

is that my focus is on linguistic communication, while theirs includes any ostensive behavior as potentially having relevance. Despite this difference, they also acknowledge the importance of intonation in the determination of relevance. One of their examples is quite interesting (p. 211) (pitch accent markings are mine):

(3.2) a. Sorry I'm late. My CAR (H*) broke down. (L*)

b. Sorry I'm late. My CAR (H*) was BOOBY-trapped. (H*)

In (a) Sperber and Wilson say that H makes an "anticipatory hypothesis" upon hearing the apology followed by the H* pitch accent on the word *car* that something must have gone wrong with S's car. One schema likely to be evoked involving "cars causing a delay" would be of a car breaking down, a frequent occurrence in modern culture. Thus the L* pitch accent on *broke down* confirms this hypothesis: *broke down* should be easily inferable within the mental representation. As they point out, one could even say "My damned car!" in the same context and H would still be likely to infer that car trouble had been the reason for the delay. In (b), there is a H* pitch accent on *booby-trapped* because this is very unlikely to be in an average person's schema about things that go wrong with cars, and must therefore be added to the mental representation. Notice how humor could be derived from these facts in a society where there had recently been a rash of

terrorist car bombs. By uttering the (b) sentence with a low pitch on *booby-trapped*, the implication is that car booby-traps have become an everyday occurrence and should therefore be available in the schema.

Other scholars have discussed the relationship between intonation and coherence. Werth (1984), whose book title, "Focus, Coherence, and Emphasis," bespeaks this connection, has also treated the subject of how intonation contributes to the identification of cohesive links in a text. His theory is based on semantic relationships among lexical items and the propositions associated with them. Gumperz (1982, 1992) also identifies prosody, including intonation, as one of the main "contextualization cues" used by conversants to make inferences about each others' meanings and attitudes. He provides examples of cross-cultural miscommunications based on misreading of these cues. Chafe (1994) has a processing account of the role of intonation in coherence. Since given and new information are respectively associated with low and high pitch, processing effort is minimized on the part of a listener who naturally attends to the more audible, high-pitched signal. Low-pitched signals send the message that the associated item is readily available in active consciousness, and therefore does not require the expense of much cognitive effort for its interpretation.

Within the framework of the mental representation outlined so far, I submit that intonation contributes to

coherence in the following ways:

1. Intonation associates systematically with cohesive language within the text.
2. Intonation provides connections between the text and information available through perception of the environment.
3. Intonation links the text to previous knowledge schemata.
4. Intonation contributes to the construction of new knowledge schemata.

In order to illustrate these four points, I present an analysis of a sample text and show how the intonation supports and supplements the lexical and grammatical cohesion.

3.3 Intonation Analysis of a Sample Text

The following text, an excerpt of an academic lecture in Statistics, is given by a female native speaker of English, from the Northern United States. She is describing some graphs on a handout which her students have already received. The subscripted numbers indicate the pitch in Hz. of those lexical items used in the examples to follow. The sub- and superscripted letters (LH%, LL%, HH%, and LH%) at the ends of phrases indicate the boundary direction. This speaker's overall range is 222 Hz., from 178-400.

Ducks Text

Let's begin$_{358}{}^{HL\%}$, I think the$_{270}$ best$_{313}$ way$_{218\ L}{}^{H\%}$

to begin$_{208\ L}{}^{H\%}$ is by looking at some examples$_{LL\%}$.

So refer first of all to the upper right-hand plot on your handout. This particular plot depicts data from the results of a study on hybridization of two different closely related species of ducks, the mallard and the pintail. And the question that was being addressed by this particular study is whether crossbreeds, so ducks that have a one mallard parent and one pintail parent, if you look at them, and you notice that a particular duck looks more like the pintail parent than it does like the mallard parent, is it also true that its behavioral charact- behavioral characteristics will be more like the pintail .

So what we have here is we have a scale, where the appearance of the duck is rated on a scale from zero actually all the way up to 20 , (we only observe values between 4 and 16), and we've also got observations on the behavior of ducks, scaled in the same way; smaller values correspond to being more like a mallard , larger values correspond to being more like a pintail and we have this particular scatter plot (data from

Wennerstrom, 1991, p. 190).

My purpose is not to do an exhaustive analysis of this text, but to exemplify various ways in which the intonation of the items in the text provides information about how it is to be interpreted.

3.3.1 Intonation and Textual Cohesion

My first pass at this text will consist of an analysis of how intonation interacts with items that are actually verbalized in the text. A useful model to identify such items is Halliday and Hasan's (1976) taxonomy of cohesion. They define cohesion as follows:

> Cohesion is a semantic relation between an element in the text and some other element that is crucial to the interpretation of it. This other element is also to be found in the text; but its location in the text is in no way determined by the grammatical structure (p. 9).

Halliday and Hasan's notion of "element" in this definition is a lexicogrammatical one: they identify categories of words which contribute to the cohesion of a text in consistent ways.[1] Their taxonomy includes five major categories which are described briefly below:

1. Reference includes personal and demonstrative pronouns, which substitute for full nouns, and determiners.
2. Substitution is a small class of words such as *one*, *same*, and *do* which can stand in place of semantically richer

lexical items.

3. Ellipsis is considered simply substitution by zero.
4. Lexical cohesion refers to full lexical items which are to be interpreted as semantically connected with other items, through reiteration, antonymy, collocation, etc.
5. Conjunction (words such as *and, but, however*) links constituents of text together while describing the relationships between those constituents.

Halliday and Hasan have been criticized for limiting cohesion to the text, which many authors see as simply a surface representation of the ideas in the mental representation (Brown & Yule, 1983; Carrell, 1982; Levinson, 1983; Steffensen, 1986; Werth, 1984). Indeed, as indicated in Halliday and Hasan's definition above, in order for a semantic relationship to "count" as a cohesive tie, both the anaphor and the antecedent must appear in the text itself. Carrell considers this level of cohesion to be superficial and merely an artifact of the underlying coherence. She cites examples of text which can be considered coherent even though there is no surface cohesion among the lexical items, such as the picnic-corkscrew utterances (3.1).

To be fair to Halliday and Hasan, however, it should be stated that their goal is in fact modest: to categorize lexicogammatical items that serve a linking function in texts. They themselves plainly state that cohesion is only a part of what contributes to coherence:

> Cohesive ties, especially those with the immediately preceding text, are only one source for the information that the reader or listener requires. Both situational and more remote textual information are necessary components (p. 303).

For the purposes of the current discussion, I will take the term *cohesion* to refer to endophoric (within the text) connections, and use Halliday and Hasan's taxonomy as a convenient tool to categorize such connections. Since verbalizations have been identified as one source of input to the mental representation, and since it is within the text that intonation resides, this category is not to be ignored. After all, arguments (such as Carrell's) against Halliday and Hasan based on the importance of underlying ideas do not claim that schema appear in the mental representation out of the blue -- they are activated by their association with items in a text in a particular discourse. Intonation will be shown to support and supplement the links provided by Halliday and Hasan's categories, in some cases providing more information than the lexical items alone.

Returning now to the Ducks Text, for each of Halliday and Hasan's five categories of cohesion, I will show one or more examples from the text and discuss how intonation interacts with the cohesive element. In general, we can expect low pitch to play a role in this interaction. As Halliday and Hasan point out, "anaphoric items in English are

phonologically non-prominent . . . in other words, they are 'reduced'" (p. 271). As we saw in Chapter 2, this often corresponds to the L* pitch accent associating with items which are already assumed to be in the mental representation. As we will also see, boundary tones serve to conjoin constituents, as do Halliday and Hasan's conjunctions.

Reference: There are several instances of reference in the Ducks Text, and these items tend to associate with low pitch. Examples of pronouns which refer back to full noun phrases (NPs) for their interpretations are *them*, in line 13, which refers back to *crossbreeds* in line 10; and *it* and *its* in lines 15 and 18, which refer back to *duck* in line 14. As shown in the subscripts, the pitch of these items is appropriately low, relative to neighboring new items, indicating that they are not being newly added to the mental representation; their antecedents are available in the text itself in this case.

The determiner *the* is used frequently and has a relatively lower pitch than the associated noun. In the first few lines, we observe this in *the best way*, *the upper right hand plot*, and *the results*. The demonstrative determiner *this* in line 5 is part of the NP *this particular plot*, which is to be singled out in contrast to the other plots being presented. It's pitch is lower than the focus word *particular*, but still relatively high in S's range, being at the beginning of a structural shift (minor paratone). In sum, although in principle, a reference item could be used in contrast, in this

text S supports the anaphoric sense of the reference items intonationally by giving them a lower pitch than neighboring new items.

Substitution: This text only contains one instance of substitution, the word *does* in line 15. It derives its interpretation from an association with the previous verb *looks*, and the low pitch supports this link. Figure 3.1 shows this utterance with the curser line going through the center of the word *does*.

Ellipsis: Halliday and Hasan have called ellipsis substitution by zero, so of course no pitch can possibly associate with an ellipted item. It is useful, however, to point out the connection between zero pitch and extremely low pitch. Chafe (1994) argues that given items in discourse have low pitch because they are so accessible in H's mental representation that they require little expense of cognitive effort to interpret. By extending this idea, we might say that ellipted items are so accessible as to not need to be articulated at all. Returning to the sample text, there are two instances of ellipsis in lines 26-28, where S says, "smaller values correspond to [ducks] being more like a mallard, larger values correspond to [ducks] being more like a pintail." In each case the word *ducks* (added in brackets) is ellipted. If this ellipted word were articulated, it would have a L* pitch accent; in fact, the sense of the utterance would be confused if *ducks* were uttered with a high pitch.

Lexical Cohesion: We can expect lexically cohesive items to have a variety of relationships with their antecedents, any of which can be indicated through the associated pitch accent: they could be used in contrast to a previous item (L+H*), they may be simple reiterations about which nothing is predicated (L*), or they may be reiterated as part of a new point (H*). Whatever the case, the intonation provides more information about the anaphor-antecedent relationship than the mere fact of lexical cohesion alone. The word *begin* in line 2, *plot* in line 5, *study* in line 10, and *pintail* in line 18 are all examples of repetitions of words that have occurred already, about which nothing new is being predicated, and which are supported by a low pitch. The word *pintail* is repeated several other times, but in these cases it is referred to in contrast to the other species of duck, so it has a higher pitch accent (this is discussed further in the Section 3.2.2). Thus, by specifying the type of relationship with the antecedent, the intonation supplements the information provided by the mere fact of lexical reiteration. The relationship of synonymy is illustrated in lines 14-18 where the word *characteristics* refers back to *looks like* and is also low-pitched. In fact, it can be said that the low pitch on the word *characteristics* (which S maintains throughout a repair) helps us recognize the synonymy. If the pitch were high, it might not be so obvious that *looking like* is one type of *characteristic*. Figure 3.2 illustrates the sequence.

Finally, we can see more examples of contrast in this text with the appropriate L+H* pitch accent. The numbers in lines 22-23, *zero* and *twenty* can be said to be contrastive since they are at opposite ends of an ordered set (Halliday and Hasan, p. 285). We also find contrast in lines 26-27 in the words *smaller* and *larger*, which have a high pitch (see Figure 3.3). In Halliday and Hasan's model, these items are said to be contrastive "by virtue of the English language"; that is, they would be contrastive in almost any context.[2] Other contrasts in this text are made instead by virtue of ideas and organization structures introduced by S and will be discussed in the section on schema.

I will also defer a discussion of collocation to the section on schema, since Hasan (1984) actually excludes this category from the cohesion taxonomy. In her analysis of children's compositions, she states that her set of cohesive categories, "did not include collocation. It proved remarkably difficult to operationalize this category sufficiently to ensure consistent analysis." The problem, in other words, is that there is no objective way to decide what is collocative and what is not based on the text alone.

Conjunction: In this category, we can begin by noting the fact that there is a parallel between intonational boundary tones and lexical conjunctions (I use the term *lexical conjunction* to refer to words such as *so, also, moreover,* which actually appear as lexical items in the text): Boundary

tones are located at the periphery of constituents, providing information about how those constituents are related to each other; similarly, lexical conjunctions are usually located at the beginning of a constituent, also providing information about the relationship between that constituent and the previous. It appears that intonation supplements the information about relationships between constituents provided by lexical conjunctions.

In the Ducks Text, we find the conjunctions *first of all, also, and* and *so*. *First of all* in line 4 and the two occurrences of *also* in lines 16 and 24, are fairly straightforward: all have a contrastive sense, supported by a high pitch; and all are embedding within an intonational phrase rather than at the boundary. More central to the discussion of utterance periphery are the other two conjunctions *and* and *so*. *And* conveys that what follows is an elaboration of what has come before.[3] Correspondingly, in all four cases of *and* (in lines 9, 13, 24, and 28), the boundary shape of the end of the previous utterance is LH%, supporting the meaning of interdependency between one utterance and the next elaborating one.

So tends to have a more transitional meaning. According to Schiffrin (1987), it is used to shift to a new organizational level in discourse. This is the case in line 4 where it is used to move from the introductory statement to the first example; in line 11 it is used to introduce an

aside, a quick definition of *crossbreeds*; and in line 20, it makes a larger transition from the description of the study to the actual plots on the handout. In lines 3 and 19, the boundary shape preceding the *so* is LL%. This is appropriate because it indicates that the former rhetorical structure is finished and separate from the subsequent. In line 20, we see clear evidence of a paratone, or range increase, accompanying the rhetorical shift, since S's maximum pitch jumps from 333 in the preceding intonational phrase, to 400 in the one following. In line 11, where *so* introduces an embedded parenthetical, the intonation is also supportive (see Figure 3.4): the LH% boundary after *crossbreeds* indicates that there is a thought to be continued; the compressed range of the whole definition following *so* indicates that it is not meant to be instantiated into the main line of discourse -- it is an aside. In Brazil's (1985) terms, we could call this a change of key. The utterance is to be taken as a shift of stance from the main organizational thread. After the parenthetical, on the word *if*, S resumes tone concord back up to the level of *crossbreeds*, indicating that the main thread is to be continued after the aside. It is noteworthy that this same effect could have been achieved by the intonation alone, without the conjunction *so*.

A more general observation can also be made about conjunctions in this text: every phrase has an intonational boundary providing information about how that phrase is to be

conjoined with the subsequent one; whereas, not every phrase has a lexical conjunction. Indeed, in this text there are 28 intonational phrases, all with boundaries. Yet, as we saw, only 7 of those begin with conjunctions. This is the basis of my statement that the intonation actually provides more information about how phrases are conjoined than do the lexical conjunctions alone. For example, although there is no lexical conjunction at the end of line 8, LH% boundary conjoins the general statement that the study deals with species of ducks to the naming of the specific species in line 9. This is shown in Figure 3.5.

To summarize, the purpose of this section on textual cohesion has been simply to show that intonation supports and enhances cohesion of a text by supplementing the information provided by the lexicogrammatical items. Taking Halliday and Hasan's model as a starting point to identify very obvious, overt instances of cohesion, I have tried to show how intonation works as part of cohesion. Thus, L* pitch accent tends to associate with anaphors, L+H* pitch with contrasts, and boundary tones with conjunctions. Since pitch accents are generally not represented in orthography, except in the case of special italics, their role in cohesion is easy to lose sight of. However, as we have seen they work in tandem with lexicogrammatical items to achieve cohesion in text by providing information about an item's position in mental representation.

3.3.2 Intonation and Schemata

Looking further at the "Ducks Text," it is evident that intonation helps listeners to make connections between the text and information perceived in the environment and retrieved from long-term memory. Examples of how the intonation interacts with the text to build new knowledge structures will also be shown.

As noted above, one nonlinguistic source of input to mental representation is what enters through perception. In this text, we see examples of this in the personal deictic *we*, which refers to the lecturer and the members of the audience. This is low-pitched in all cases, being an evident part of the context. The phrase in lines 4-5, *upper right hand plot on your handout*, makes exophoric reference to the handout which the students have in from of them. The word *handout* is therefore low-pitched, as is the word *plot* because the handout has visible plots. However, the specific location on that handout, *upper righthand*, is high-pitched, since it is being singled out as the focus of attention (shown in Figure 3.6). Finally, the place deictic *here* in line 20 is another example of givenness via perception, also low-pitched; it refers to the plot on the handout to which S is pointing. In sum, we see S associating low pitch with items which she judges not to need to be added to H's mental representation because those items should already be there through perception of the environment.

Next, I will discuss cases where the intonation signals listeners to draw on schemata stored in memory for an interpretation. In lines 11 and 12 we noted that S introduces a parenthetical definition of *crossbreeds*. Within this definition, the word *parent*, repeated twice, has a low pitch. This is due to the fact that S assumes the listeners have schema for breeding and genetics of which an offspring having two parents is an obvious part. This sequence is shown in Figure 3.7. Another example comes from the collocation among several words related to the plotting of points on graphs. The word *values* in lines 23, 26, and 27 is low-pitched in all three instances (refer back to Figure 3.3 for the latter two). This can be attributed to the schema associated with *scale* introduced in line 21, and *rated* in line 22, of which the notion of *values* a part. It is probably safe to say that the visual representations of the plots on the handout interact with the text as well to evoke "math" schemata of how plots are constructed and read, how numbers are sequenced, and other knowledge brought to the interpretation of the lecture.

Finally, I consider cases in which new schematic structures are being created. Many of the items in the first part of the lecture are introduced as new with H* pitch accent. Listeners learn what kind of study is being discussed, what the object of study was, and what the research question was. All these items become part of the mental representation being created for this lecture, along with

whatever schemata they may evoke involving similar studies: what graphs consist of, facts about ducks, breeding, reproduction, and the like. Information perceived visually from the handout is also added to the mental representation. In other words, the listeners are constructing a mental representation for this discourse situation, within which, inferences may be drawn that lead to new knowledge. Consider the contrast between *mallard* and *pintail*, the two species of ducks (shown in Figure 3.8). Listeners may never have heard of these types of ducks, or given any particular thought to how much variety among duck species there might be. Yet, the second species is signaled by L+H* pitch accent (and by lexicogrammatical means as well) to be in contrast, for the purpose of this lecture. This is different from Halliday and Hasan's notion of contrasts that are recognized by virtue of being part of the English language such as *larger* or *smaller*, discussed earlier. There is no reason to suppose that nonornithologist speakers of English have stored schemata for these species of ducks. Only upon hearing them contrasted here, do they construct a new mental representation that includes them as distinct species which can interbreed. Those who consider these ideas important enough may store them in long-term memory and be able to access this new schema about species of ducks later, the next time the topic rolls around in conversation, (or on the exam for that matter!)

A similar point could be made about the contrast between

the idea of how a duck looks and what its behavioral characteristics are like, in lines 14-18. Again, listeners may have had no previous schema to tell them that looks and behavior are factors worth comparing in a study of species. However, the L+H* pitch accent on *behavioral* and the L* pitch accent on *characteristics* could enrich pervious schema about animal breeding to include such factors as comparable (refer back to Figure 3.2).

3.4 Conclusion

The purpose of this chapter has been to show that intonation plays an important role in coherence. The pitch accents associated with items in text indicate how S wishes H to interpret those items. As S and H both construct mental representations of the discourse, the intonation of each new utterance provides "signposts" to indicate how the items in the utterance are to be regarded with respect to the mental representation. H* pitch indicates that an item is being added to the mental representation; L+H* pitch accent signals that it contrasts with something already in the mental representation; low pitch often signals that that item should already be available in the mental representation, or that for other reasons it is not to be added; boundary tones signal how a constituent is related rhetorically to neighboring constituents.

Through the sample analysis of the lecture segment, I

have provided examples of relationships between intonation and four different sources of input to mental representation. First of all, intonation was shown to support and enhance the cohesion among lexicogrammatical items in the text itself, in some cases providing additional information about the nature of the link. Boundary tones were also shown to be a much more frequent indicator of conjunction than were lexical items. Second, intonation was shown to interact with what enters the mental representation through perception of the environment, in this case visual information on a handout. Third, there were examples where background knowledge of math, science, and graphs provided schemata in listeners' mental representation, within which certain lexical items were assumed to be accessible. And finally, I suggested that information provided by intonation contributed to the formation of new knowledge schemata. The point of the sample analysis was simply to illustrate ways in which the intonation system of English contributes to coherent communication.

It is now possible to consider the task of a person learning English as a second language in light of the interaction between mental representation and the intonation system. In cross-cultural communication, two challenges are immediately obvious. First, although the construction of mental representations during a speech situation is claimed to be universal, people from different cultures are less likely to have common schemata stored in memory from past experiences

to bring to interpretation[4]. To the extent that this is the case, the participants' assumptions about the contents of each others' mental representation is also less reliable. Indeed, miscommunications are probably based most often on one person's assuming that the other has knowledge to bring to an interpretation which in fact he or she does not. It is for this reason that beginning language classes often focus on pictures and objects in the environment: this helps learners construct mental representations for new vocabulary without drawing on a lot of previous culturally based schemata.

Second, intonation's signalling of relationships between the text and the mental representation is claimed to follow an English-specific system.[5] Thus, in communication between L1 and L2 speakers of English, some of these signals may be missed: S's intonation may not signal certain links between the text and the mental representation; H may not recognize and interpret such intonation if it is signaled. This may be especially problematic for NNSs at lower levels of language development who have fewer of other linguistic resources to rely on. What is especially interesting is that most ESL and EFL classes do not focus on intonation as part of English grammar. Therefore, NNSs of English mainly acquire their intonation intuitively or by transfer from the native language without learning specific rules. Even if they were taught specific rules, it would take time to integrate these patterns into their speech.

To investigate these issues more precisely, we will move on to consider three empirical studies of intonation in cross-cultural communication. The evidence indicates that NNSs do acquire the intonation system of English as they become more advanced. Conversely, it can be said that part of becoming more proficient in a language entails gaining control of the intonation system. The studies in the upcoming chapters focus on specific aspects of this acquisition process, beginning in Chapter 4 with a statistical study of intonation in academic lectures given by Chinese speakers.

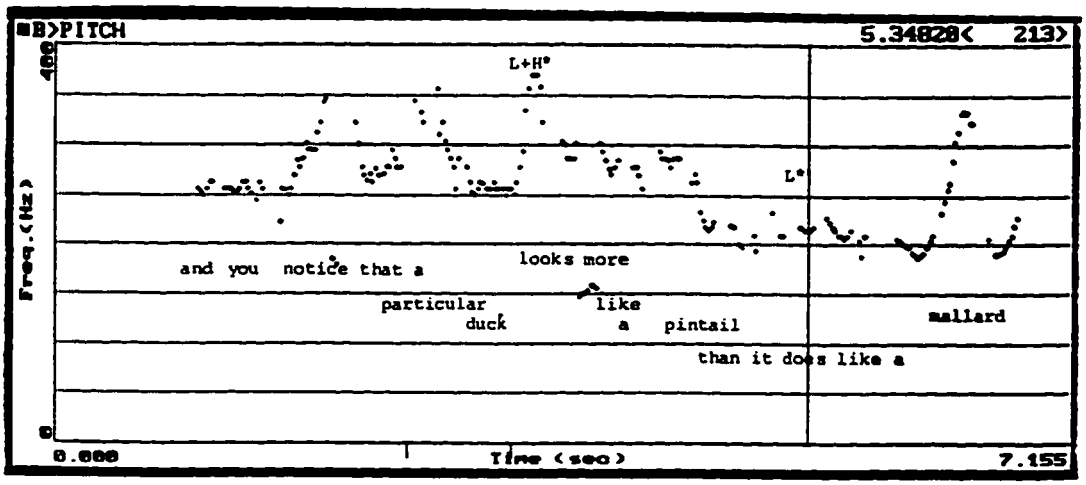


FIGURE 3.1 The curser is aligned with the low-pitched word *does*, a substitution.

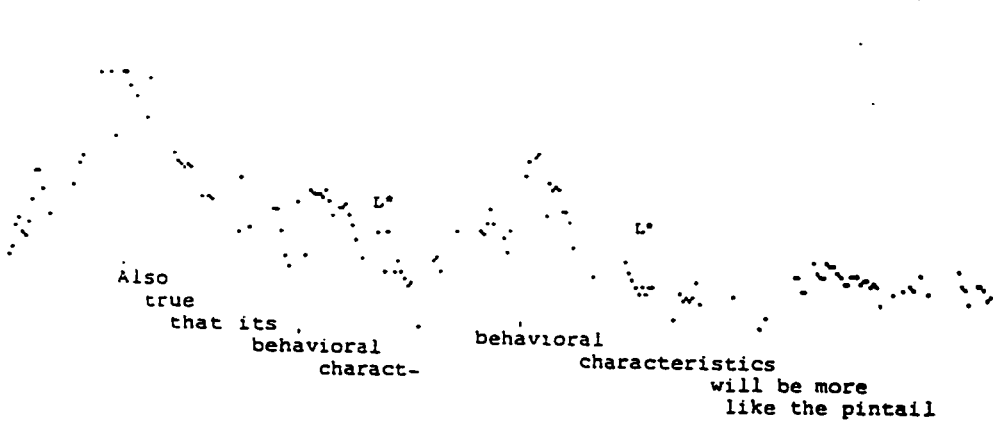


FIGURE 3.2 The word *characteristics* is low-pitched, being synonymous with the previous *looks like*.

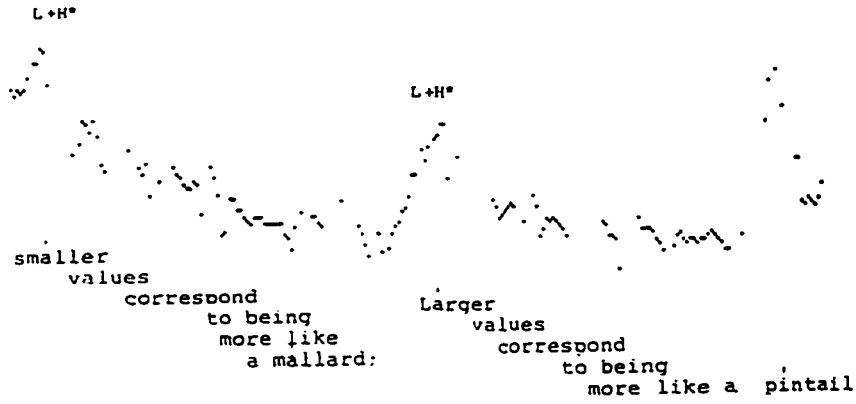


FIGURE 3.3 A contrast is shown between the words *smaller* and *larger*.

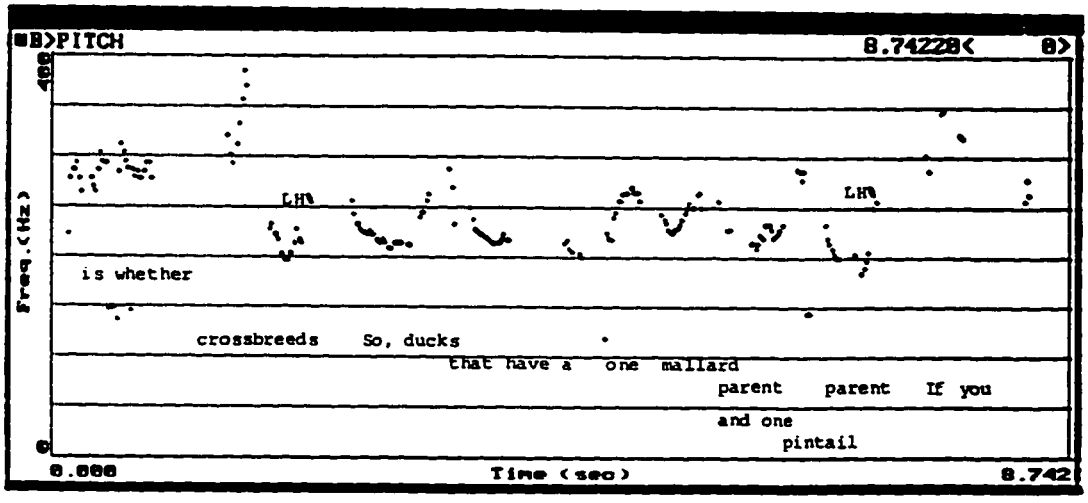


FIGURE 3.4 The conjunction *so* introduces an aside, while the LH% intonational boundary preceding it supports the transition. A shift to a lower key for the aside is also evident.

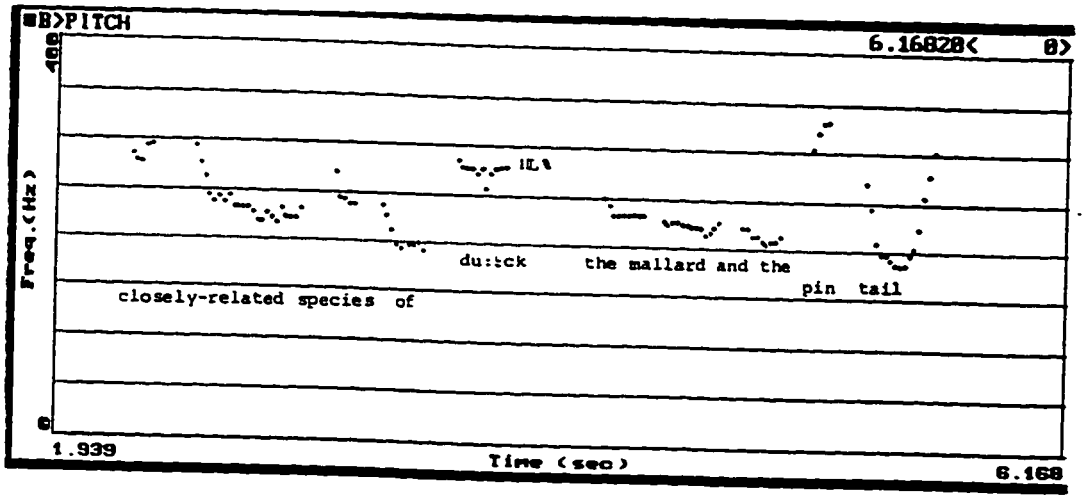


FIGURE 3.5 Although there is no lexical conjunction between these phrases, the HL% (plateau) intonational boundary conjoins them.

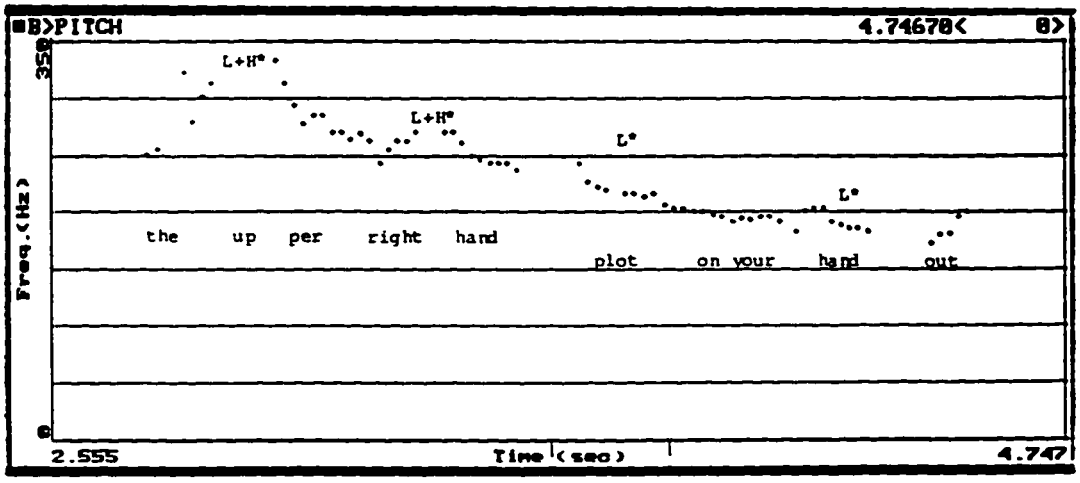


FIGURE 3.6 The words *plot* and *handout* have low pitch being perceivable in context; *upper righthand* has a L+H* contrastive pitch accent because it needs to be singled out in contrast to other parts of the handout.

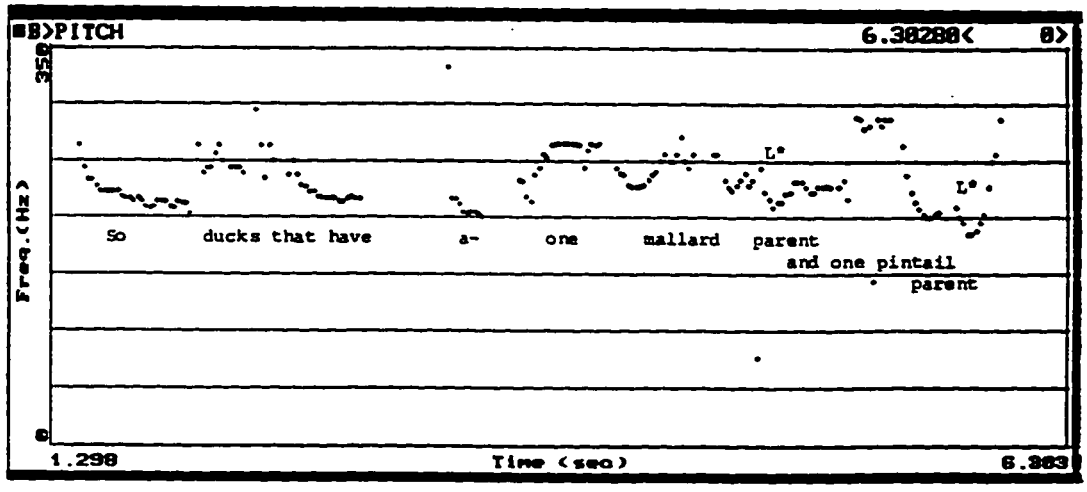


FIGURE 3.7 The low-pitched word *parent* is assumed to be part of a schema about breeding and genetics.

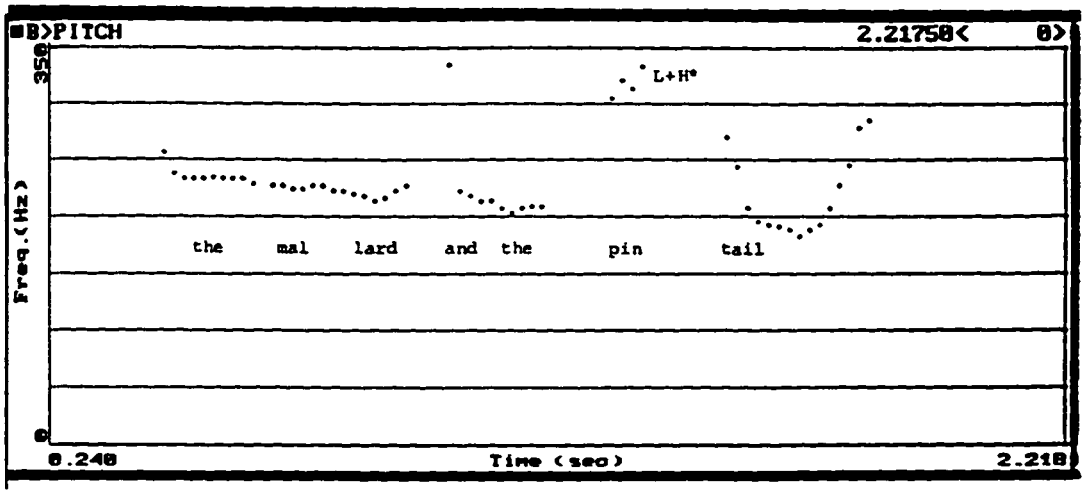


FIGURE 3.8 The intonation on *pintail* indicates that the two breeds of duck are meant to be in contrast.

Notes to Chapter 3

1. Hasan (1984) attempts to expand this model to include syntactic relationships between chains of cohesive lexical items.

2. Although they did not occur in this particular text, this category includes words whose prefixes form the semantic locus of the contrast and the domain of the L+H* pitch accent, as in *smoking* versus *NON-smoking*, or *INcreasing* versus *DEcreasing* (Wennerstrom, 1993).

3. I do not include a discussion of *and* between NPs since this type of conjunction takes place below the level of the intonational phrase.

4. Of course, people from different cultures can also have experiences in common and the context itself contributes to mutual understanding.

5. It is not claimed that other languages do not have intonation systems; only that these systems are language-specific. German, for example has a system similar to English in many ways (Selting, 1996).

Chapter 4

Study I: Academic Lectures

This chapter reports on the first of three empirical studies of how L2 speakers of English use intonation in discourse. Generally speaking, the aspects of intonation chosen for investigation in these studies are those that I have heard learners of English struggle with repeatedly in ESL classes over the years.

Study I is a quantitative study of the intonation of Chinese speakers giving academic lectures in English. These speakers were graduate students at a major U.S. university, under pressure to improve their spoken English before assuming positions as teaching assistants (TAs). All were enrolled in an ESL course which emphasized pronunciation and presentation skills. In order to pass this course, they had to pass a final exam for which they were required to give a short lecture in their own field. Their scores on this exam were used as the response variable in a multiple regression analysis to determine whether certain meaningful aspects of their intonation contributed to raters' perception of overall comprehensibility.

The study also includes data from two native speaker (NS) lecturers for the sake of comparison. These speakers, one male and one female, were asked to contribute a 30-minute

lecture for the level of a general audience as part of a videotaped materials development project for the ESL class described above (Wennerstrom, 1991). They were chosen because they had a reputation for being clear and well-organized presenters. Although not part of the statistical analyses, their data is useful as a source of examples of speakers who use the English intonation system effectively. Measurements of the NSs' intonation were collected according to exactly the same procedures as for the NNSs.

4.1 The Genre of the Academic Lecture

Throughout their ESL class, subjects had been made aware of the norms of typical academic lectures in the United States. They had seen videos of exemplary lectures and had been assigned to observe university classes in their fields. In their own presentations, they had been explicitly encouraged to incorporate some of the common organizational components of academic lectures, such as introductions, conclusions, summaries, examples, definitions, and what Chaudron and Richards (1986) refer to as "macro-markers," explicit organizational statements to announce rhetorical transitions. This strong, explicit emphasis on organization in the training of these subjects makes this genre ideal to investigate the relationship between intonation and constituent structure. First, it was possible to observe how pitch range would be altered at junctures between major

organizational units as paratones, and within in those units, it was possible to observe how cohesion was intonationally reinforced by boundary tones, linking related phrases and utterances together.

Another unique feature of this genre that would be expected to manifest itself intonationally has to do with information structure. Since it is not as feasible in lectures as it is in conversation to check with every member of an audience to determine their exact level of understanding, lecturers often restate information. This affects the intonation since in the restatements, H* pitch accent is associated with the repeated concepts, as if they were being newly introduced. It is as if the lecturer were saying, "I am not making the assumption that these are part of your mental representation yet."[1] Here is an example of a deliberate restatement in a lecture given by a NS. The restatement begins in line 13:

1(4.1) That derivative is defined as notation f
2 prime of x equals, the limit, as an
3 auxiliary variable that's just in there
4 temporarily, goes to zero, of the ratio
5 of how much f changes, that's f of (x
6 plus h) minus f of x, with respect to how
7 much x itself is changed by moving from x
8 to (x plus h). That ratio, in the limit,
9 for very very small h, that gives us the

> instantaneous rate of change , defines the derivative. That concept is important so let me write it down for you and say it again; this derivative gives us the instantaneous rate of change of the function f with respect to the variable letter x (Wennerstrom, 1991, p. 191).

The content words in the restatement are uttered with a high pitch although they are "given."

This is also convenient for the researcher, who is afforded a similar perspective to that of the audience with respect to what is given versus new in the mental representation being constructed. In personal conversation, the prediction of intonation is more difficult to the extent that the researcher is positioned as an outsider, not privy to previous shared experiences stored similarly as schema between the participants.

Another characteristic of this genre is that there are fewer interactional features than in other genres. As Goffman (1981) points out, part of the lecture genre includes an understanding that the lecturer has an institutionalized license to keep the floor for an extended turn. Thus we would not predict that the lecturer would have to negotiate frequently to take the floor. Nor would we expect interruptions, backchannels, overlapping speech, and other

features of more interactional genres to be as frequent in lectures. To the extent that these moves are supported by intonational features, this will affect the tune of this genre.

A final intonational characteristic of this genre is entirely pragmatic: because a lecturer is deliberately trying to project his or her voice over the distance of a large room, lecturers tend to speak loudly and to exploit a greater pitch range. This makes measurement a much easier task in this genre than in other genres since pitch contours tend to be exaggerated.

4.2 Methodology

4.2.1 Subjects

The participants who volunteered for this study were 18 speakers of Mandarin Chinese attending graduate programs at a major research university in the U.S. As part of their stipend, all were expected eventually to work as TAs, conducting lectures, quiz sections, or lab sessions in English for undergraduates. Before assuming this position, however, they were required to pass an English class whose focus was on presentation, interactive teaching techniques, and pronunciation, including some explicit instruction about intonation. All the students in this class had scored less than 230 on the 1985 version of the TSE or SPEAK tests, which is described in the Educational Testing Service training kit

(ETS, 1985) as comprehensible despite some systematic errors. All had scored above 500 on the TOEFL upon their entrance to the university, having studied English in China or Taiwan since elementary school. During the 10-week class, they worked intensively on oral language skills, spending 5 hours a week in class, and from 2 to 8 hours a week outside of class with tutors, in study groups, in the language lab, and/or on individual homework assignments. The pressure to pass the course was high since, for many, continued financial support from their departments depended on their success on the final exam.

To say a brief word about Chinese, the native language of the subjects, Shen (1990a) finds evidence of an intonation system above the level of the lexical tone. In a study of native speakers of Mandarin Chinese in oral reading of sentences in their native language, distinctive intonation contours were identified for statements and two different types of questions. According to Shen, these utterance types were distinguished, not by the shape of the boundary at the end of the intonation contour as in English, but by the initial pitch and the utterance maximum. Shen also makes some interesting points about stress. First, although stress in Chinese does involve raised pitch, it is mainly duration that adds prominence. This includes contrastive stress; that is, in Chinese, a contrasting lexical item is not perceived so much by its higher pitch as by its longer duration. Also,

stress accumulates throughout the sentence resulting in a tendency for final stress, unless the final word is an enclitic. This coincides with the finding by Juffs (1990), that sentence-final stress occurred in oral readings in English by Chinese speakers. As will be shown, these patterns appear to be transferring into English speech for many of the subjects in the current study.

4.2.2 Discourse Data

An discourse analysis was made from the subjects' final exams in the ESL course described above. The exam procedure was designed especially for this ITA ESL class, and had been in place for eight years. It consisted of a 10-12 minute lecture in each subject's field, delivered spontaneously using notes or handouts to an audience composed of their classmates, their ESL teacher, and the exam raters. All exams were videotaped in order to document the results for departmental record-keeping purposes.

The exams were rated in four categories by three independent raters, all of whom had been trained in both SPEAK Test rating and a in-house training procedure for this final exam. The rating levels were designed to be similar to those of the SPEAK Test scores, ranging from 0 to 3. The score chosen to be used in this study was the "production" score which most closely reflects overall language ability, independent of factors such as organization of the lecture,

teaching techniques, use of visuals, etc. which were subsumed under other categories of scoring. The scoring levels are described in the rater guidelines as follows:

0 = unintelligible except for short phrases; necessitates asking for repetitions and strained listening (not passing).

1 = generally comprehensible but with frequent errors; concentrated listening is required most of the time (not passing)

2 = comprehensible despite certain systematic errors; concentrated listening is not required (passing)

3 = effortlessly intelligible despite isolated occurrences of errors (passing)

Raters also had the option of assigning plus and minus scores in between the whole-numbered scores. These were coded into the statistics as .3 and .8 (for example, a 2+ was converted to 2.3). Scores of 2.0 and above were considered passing. Interrater reliability was .76 among the three raters for the 18 subjects in this study. This is slightly low, but not completely out of line with others' findings for ratings of spoken discourse. Upshur and Turner (1995, p. 9) found an interrater reliability of .81 for the rating of "communicative effectiveness" of oral summaries; Hatch and Lazaraton (1991) also cite an example of between .75 and .76 interrater reliability in an assessment of conversational fluency by Riggenbach (p. 534).[2]

For the study, the subjects' videotapes were converted to audiotape and transcribed by a research assistant. Thereafter, a discourse analysis was performed on the transcripts to predict intonational characteristics which are described in the next section. The sample text below shows a transcript at the prediction stage of analysis. Each lexical item is marked for its role in the information structure, as indicated in subscripts, utterance-medial phrase boundaries are predicted as "boundary" in superscripts, and paratones are to the left of new topics, also in superscripts. Uncoded words include certain conjunctions, hesitation sounds, pronouns, and the second part of compound terms (which tend to be low pitched due to compound stress).

Sample Lecture with Intonation Predictions

Good$_{content}$ morning$_{content}$ everybody$_{given}$. My name$_{content}$ is$_{function}$ XXX$_{content}^{boundary}$ and I'm from$_{function}$ the$_{function}$ department$_{content}$ of$_{function}$ Physics$_{content}$. Welcome$_{content}$ to$_{function}$ the$_{function}$ course$_{content}^{boundary}$ Introduction$_{content}$ to$_{function}$ Optics$_{content}$. ^{paratone}And last$_{contrast}$ week$_{given}^{boundary}$ ah we talk$_{given}$ about$_{function}$ the$_{function}$ focus$_{content}$ of$_{function}$ focal$_{content}$ point of$_{function}$ a$_{function}$ concave$_{content}$ mirror$_{content}^{boundary}$ and the the$_{function}$ principle$_{content}$ axis$_{content}$ of$_{function}$ a$_{function}$ concave$_{given}$ mirror$_{given}$. And I claim$_{content}$ that$_{function}$ I will$_{function}$ use$_{content}$ this$_{given}$ concept$_{given}^{boundary}$ to$_{function}$ do$_{content}$ the$_{function}$ image$_{content}$ formation for$_{function}$ a$_{function}$ concave$_{given}$ mirror$_{given}$. Uhm before$_{contrast}$ doing$_{content}$ that$_{given}^{boundary}$ let$_{content}$ me give$_{content}$ you a$_{function}$ brief$_{content}$ review$_{content}$.

$^{paratone}Before_{contrast}$ uhm $last_{contrast}$ $week_{given}{}^{boundary}$ we $defined_{content}$ the $focal_{content}$ point $as_{function}$ $the_{function}$ $point_{content}{}^{boundary}$ $where_{function}$ $all_{content}$ $reflected_{content}$ $rays_{content}$ $from_{function}$ $incident_{content}$ $rays_{given}{}^{boundary}$-- $rays_{content}$ $parallel_{content}$ $to_{function}$ $the_{function}$ $principle_{given}$ $axis_{given}$ $intersect_{content}$.

Finally, the pitch of the predicted features was measured on a CSL machine to test the predictions against actual behavior. The NS transcripts had already been transcribed (Wennerstrom, 1991). The discourse analysis and measurement procedures were conducted in the same manner for the NSs as for the NNSs.

4.2.2.1 The Intonation Variables

There were four intonation measures considered for analysis, chosen because they had been consistently observed as problematic in the speech of this population. Each intonation variable will be discussed individually in terms of its discourse function, in order on a scale from micro- to macro- discourse functions. At the micro- end of the scale are intonational phenomena occurring within the intonational phrase: the difference in pitch level between content and function words, and the shape of the phrase boundary itself. At the macro- end are those phenomena which convey long-distance cohesive links within larger constituents of the lectures: the difference between contrasting words and given words, and paratones, the increase in pitch range to mark

topic shift. In addition, information is also given as to how measurements were taken for each variable.

Variable 1: Content versus Function Words

This variable was examined in order to determine whether speakers were using intonation to distinguish key items being added to the discourse as new, in contrast to grammatical items or "function words." As noted in Chapter 2, function words receive no pitch accent at all unless they participate in the information structure of the discourse (as in the case of contrasting prepositions, for example) and should therefore be lower in pitch relative to neighboring content words. However, I have observed NNSs stressing all words equally in order to articulate them clearly, resulting in confusion as to which words are to be added to the mental representation as new, and which should not be worthy of special attention.

For this study, function words included articles, prepositions, auxiliary verbs, and the copula, excluding any contrastive uses of these, and excluding prepositions participating in phrasal verbs (as *stand up*). Pronouns were not considered function words, but given words, because they are either cohesively linked to a referent in previous text or obvious from the deictic system perceivable in the environment. Content words, on the other hand, included nouns, adjectives, main verbs, and adverbs, excluding those words which fell into the categories of given or contrast (see Variable 3). In other words, all content words chosen for

study were introduced by the speaker as new in the discourse and were predicted to have a high (H*) pitch accent. This variable was measured as a percentage value between 0% and 100%, representing the average pitch difference in a speaker's range between function words and new content words. For each subject, the first 20 pairs of function and content words were measured, and the function mean was subtracted from the content mean. Measurements were made for only one function-content pair per phrase. In the case of co-occurring function words within a phrase, such as preposition-article, the function word having the lowest pitch in the cluster was chosen. In the case of co-occurring content words, such as adjective-noun, the word with the highest pitch was chosen. The decision to take only one content and one function word per phrase was methodological; the pitch of co-occurring words tends to be similar, so taking clusters of measurements would distort the mean. Finally, function words at the periphery of intonational phrases were excluded since their pitch might be affected by boundary tones at their ends and the onset of catathesis or paratone effects at their beginnings. The NS's scores on this variable were 41% and 38%; in other words, content words averaged between one-third and one-half of speaker range higher than function words.

Variable 2: Phrase Boundaries

This variable addresses the issue of how intonation was used at the end of a phrase to show local cohesion between

that phrase and the subsequent phrase. The variable was chosen because I had observed a tendency among some Chinese speakers for pitch to drop to the bottom of the range at the end of a phrase, regardless of the relationship between that phrase and the subsequent. The environment measured was the utterance-medial position where the phrases on either side of the boundary were closely connected in their rhetorical purpose. This corresponds roughly to where a writer would place a comma instead of a period. The boundary shapes in Pierrehumbert and Hirschberg's (1990) model most likely to be used in this situation are HL% (plateau), LH% (low rise), or simply H (a high phrase accent), all of which convey a sense of connectedness to the phrase that follows. For example, in the following utterance (reprinted from the example of a NS lecture in Chapter 3), the first three boundaries are nonlow and utterance-medial; the final one is LL% to complete the utterance:

(4.2) Let's begin $^{HL\%}$, I think the best way ^{H} to begin^{H} is by looking at some examples$_{LL\%}$.

Such nonlow boundaries are as expected in the lecture genre within rhetorical units such as examples or explanations where a speaker's purpose is to make the unit as a whole cohere by linking closely related phrases. The next two examples show NNSs using boundary tones, illustrated respectively in Figures 4.1 and 4.2. The first example comes from a speaker who scored high (2.3) on the final exam (brackets indicate the

portion of the utterance shown in the figure).

(4.3) [A simple pendulum^{H} is a heavy mass^{H} at the end of a string^{H} that swings back and forth.$_{LL\%}$]

The H boundaries at the ends of the first three phrases appropriately indicate that all four phrases together form a larger unit, in this case, the definition of a pendulum. The second example, from a low-scoring (1.3) NNS, illustrates the opposite case. The low boundaries after the words *light,* and *focus* do not signal an interdependency among the three phrases, in spite of the fact that rhetorically they appear to be closely related.

(4.4) Any lens any lens that is thicker at the center than at its- than at its edges is called converging lens. It can [make parallel light$_{L}$ focus$_{L}$ to a point$_{LL\%}$.]

In the measurement stage, boundaries were identified by the presence of prepausal lengthening (Klatt, 1975), an extension of the intonation contour at the end of the last word; if a word was cut off in the middle, it was not considered to have a fully produced intonational boundary. For each speaker, a tabulation was made of the first 30 phrase boundaries in the mid-utterance position. Utterance-final position, identified syntactically and semantically, was not included for boundary measurement because a falling pitch in this environment would be expected rather than problematic. Thus each subject received a number between 0 and 30 for this

variable, indicating the number of nonlow (high or mid) boundaries in the first 30 utterance-medial environments. A number close to 30 represents a consistent use of intonation to signal an interrelationship between phrases within an utterance. The two NSs scored 29 and 25 on this variable, indicating a strong use of this conjoining function of intonation.[3]

Variable 3: Contrast versus Given

This variable is intended to convey information about how each speaker used intonation to signal cohesive relationships between lexical items assuming a L+H* pitch accent (steeply rising) for contrast and a L* (low) pitch accent for given as the NS norm. In the terms discussed in Chapter 3 of how a discourse is constructed in mental representation, both given and contrast items form a link between the current utterance and the mental representation. Contrast items are in a sense given, but their link with the mental representation is one of antonymy rather than reiteration. NNSs who assign equal pitch to all items are not always able to convey which items are meant to be contrasting, or whether certain items are already accessible in the listener's mental representation. If given items are associated with a high pitch, a NS listener may take them to be new.

Both Halliday and Hasan (1986) and Werth (1984) have discussed contrasting items as being part of a subsuming category or set, sharing features of the set with something

previously given, but being alternative members of that set. In this genre, many instances of locally defined sets occurred both within the lecture itself and within the lecturer's field. For example, in the Ducks Text analyzed in Chapter 3, we were told that two types of ducks, the mallard and the pintail were to be contrasted for the purpose of that lecture. At the measurement stage, such locally established contrasts were counted in the tabulation. Given items were selected for measurement using a conservative criterion: only those items which were not in theme position[4], and for which proforms could reasonably be substituted were chosen as given. Personal pronouns themselves were excluded because it could not be categorically predicted whether they were used contrastively.

The next example illustrates a speaker from the study, whose score on the final exam was very high (2.8), using intonation to distinguish contrast from given items in this manner, with L+H* pitch accent on contrastive items *larger, smaller, scattered,* and *closer* and L* pitch accent on reiterated *points*. The highest pitched items are indicated in capital letters.

(4.5) Exactly. Um, we want this measurement [to be LARGER when the points are more SCATTERED] [and to be SMALLER when the points are CLOSER togeth]er.

Figure 4.3 shows the bracketed portions of the above contrast.

Conversely, the following example shows a speaker who

scored low (1.0) on the exam making a contrast lexically but not supporting it intonationally. The contrasting words *less, faster, greater,* and *smaller* are not distinguished intonationally, while the reiterated words *accelerate* and *mass* have a relatively higher pitch as shown in capital letters.

(4.6) . . . so if the mass is greater when you push on this object it with the same with with the same force, then it [will ACCELERATE less if the mass is greater.] [And it will ACCELERATE faster if the MASS is smaller.]

Figure 4.4 shows the words in brackets from the above example.

This contrast-given variable was calculated by subtracting the mean pitch of the first 20 given items from that of the first 20 contrast items for each speaker. Again, raw values were converted to percentage of speaker range. For this variable, a higher percent indicates that that speaker was using a greater pitch differential to distinguish given items from contrasting items. The NSs showed an extremely large distinction between these categories, a 73% difference and a 67% difference.

Variable 4: The Paratone

This variable was identified to measure the change in a speaker's pitch range to mark rhetorical shifts. There was a good deal of regularity in the types of rhetorical chunks occurring in these short lectures. With a few exceptions,[5] lectures were organized as follows, parentheses indicating

optional chunks:

- introduction of self
- (review of background)
- introduction of topic
- main point
    (subpoints)
    (examples)
- main point
    (subpoints)
    (examples)
- (more main points, etc.)
- conclusion

For all speakers, regardless of level of English, the junctures themselves involved complex structure. Borrowing the convenient terms of *coda*, a final closing utterance, and *abstract*, an initial introductory utterance, from Labov and Waletsky's (1967) narrative analysis, I have listed below the typical rhetorical components of the transitions that occurred in the current study. I assume that the components prior to the transition, associated with the end of a topic, will have a narrower range than those after the transition, where a paratone would associate with the new topic. Combining these two sets of facts, I propose the following structure of a transition (optional parts are in parentheses):

Rhetorical structure	**Prosodic structure**
- end of content of topic t	range is compressed

- (coda to topic t)	range is compressed
- (solicitation of questions)	(range varies)
	pause
- (abstract to topic u)	range may be expanded
- beginning of content of u	range is expanded

It is possible to move from one topic to the next with a great deal of lexically displayed organizing material or with only prosodic cues to mark the transition. The following is an example of a full-fledged transition with all the optional parts, uttered by a speaker who scored 2.0 (passing) on the exam. Ranges are in parentheses in Hz:

(4.7)	end of topic t	. . . we call this term *heat conduction*, expressed by QC. (141)
	coda to topic t	Therefore, there are two main processes in uhm heat energy balance at the sea surface. (152)
	sol. of questions	Are there any questions? (54) (answers 2 questions) Clear? (21)
	abstract to u	Ah, so let's move on. (89)
	beginning of u	Now, the question is, question is whether not these three processes can balance the he- heat input . . . . (194)

For this excerpt, the range expansion was considered to be 42 Hz, or a 28% increase, the difference between the range of the

end of topic t's coda (152) and that of the beginning of topic u (194), regardless of intervening ranges. The more a speaker did used paratones in this manner, it easier it would presumably be for listeners to realize that he or she was moving on to a new topic.

For all subjects, the ranges before and after the first 10 major rhetorical junctures were measured.[6] The percent increase of each shift was then averaged, resulting in a single percentage value for each speaker. For NSs, the average range shift at transitions was large: 62% and 51%.

To summarize this section on intonation variables, each speaker in this study had four measurements associated with his or her short lecture: a percentage value indicating the difference in pitch between function and content words; the number of nonfalling utterance-medial phrase boundaries for the first 30 such environments; a percentage value indicating the difference in pitch between contrast and given items; and a percentage value indicating range increase (or paratone) at organizational junctures.

4.2.3 Multiple Regression Analysis

In addition to these four intonation measures, the data from each subject also included a final exam score. These numbers were used in a multiple regression analysis to determine whether intonation (the independent variables) was a predictor of final exam scores (the dependent variable).

This is a statistical method which looks at each intonation variable individually to determine its effect on exam score, holding all the other variables constant. In other words, the question was whether the subjects who had more nativelike intonation scored higher on the exam, and which aspects of intonation had the most effect. As noted in the chapter introduction, the NS data included only the four intonation measures and no scores, and these measurements were not used in the statistical analyses.

4.3 Results

4.3.1 Native Speaker Averages

The averages for the intonation measurements of the two NSs are summarized in the following table (percentages refer to the average differences in pitch converted into a percentage of speaker range):

Table 4.1: Native Speaker Intonation

Variables	Female NS	Male NS
Content - Function	41%	38%
Boundary (out of 30)	29	25
Contrast - Given	73%	67%
Paratone	62%	51%

These results indicate that the NSs both did exactly as

predicted on all four measures: They distinguished content from function words; they had a high number of utterance-medial boundary tones (close to 30); they made a large distinction between contrast versus given lexical items; and they differentiated rhetorical units through paratones at major organizational junctures. Also according to prediction, the two speakers appear to be quite consistent, having similar numbers in each category.

4.3.2 Correlation Coefficients

Returning to the NNS data, it can be noted in the following correlation matrix that the strongest relationship among the four variables is between the contrast variable and the paratone variable with a correlation coefficient of .44, not too strong.[7]

Table 4.2: Pearson Correlations among Intonation Variables

	Content	Boundary	Contrast	Paratone
Content	1.000	.028	.235	.058
Boundary		1.000	-.275	-.019
Contrast			1.000	.444
Paratone				1.000

These correlation coefficients support the notion that these intonation variables are not being used consistently by the group of NNSs as a whole.

4.3.3 Multiple Regression Analysis Results

As shown in Table 4.3, the multiple regression analysis was significant (p=.011) with the paratone variable highly significant (p=.004). This means that on average the more a subject increased his or her range to mark topic shifts, the better score that person received on the exam. None of the other variables were significant in this analysis.

Table 4.3: Multiple Regression Analysis

Variable	Coeff	St. Error	p-value
Intercept	.078	.626	.902
Content - Function	-.016	.011	.176
Boundary	.016	.015	.309
Contrast - Given	.002	.007	.838
Paratone	.014	.004	.004**

* = p < .05; ** = p < .01

R square = .61 F = 5.037 St. Error = .38 p = .011*

To explore the data further, it will be useful to separate the subjects into 3 levels and look at the means for each level and for each variable. Six subjects scored less than 2.0, six exactly 2.0, and six greater than 2.0 on the final exam.[8] Since the passing score on the exam was 2.0, the latter two groups passed. For the paratone variable, which was significant, we can see the linear trend in Table 4.4.

Table 4.4: Paratones and Scores by Level

Score Level	Paratone Mean (SD)
< 2.0 (low)	20% (20%)
= 2.0 (medium)	52% (18%)
> 2.0 (high)	56% (25%)

Although this table shows a general increase by level, it is the lowest (failing) group which shows the striking difference from the top two groups, with only 20% pitch increase at topic shifts. It is also noteworthy that all groups increased their range to some extent.

The contrast-given variable resulted in the following means, by level:

Table 4.5: Contrast minus Given and Scores by Level

Score Level	Contrast-Given Mean (SD)
< 2.0 (low)	28% (20%)
= 2.0 (medium)	36% (8.9%)
> 2.0 (high)	31% (7.3%)

The graph in Figure 4.5 displays the data. From this graph a general problem with the data is evident. Although the graph reveals a modest linear trend, individual outliers interfere with the statistical significance of the result, due to the

fact that the final exam scores are extremely variable. The global production score is influenced by a wide range of factors, such as grammar, vocabulary, pronunciation of sounds, communicative competence, and perhaps even personality factors such as charisma and enthusiasm of delivery. Thus we see outliers such as those indicated in the graph. It is indeed possible to get a high score, in spite of problems in intonation, as did outlier A, based on other aspects of production, such as clear articulation of sounds, expressive vocabulary, etc.. Conversely, outlier B makes effective use of intonation, but raters perceived other language features, such as unclear sounds, grammar problems, vague vocabulary, etc., that prevented him from receiving a top score on the exam.

The phrase boundary means by level were as follows:

Table 4.6: Phrase Boundaries and Scores by Level

Score Level	Phrase Boundary Mean (SD) (out of possible 30)
< 2.0 (low)	19 (7.2)
= 2.0 (medium)	22 (4.4)
> 2.0 (high)	25 (7.3)

Again there appears to be a promising trend, but there is too much variability in the data for a statistically significant

result.

Finally, the content-function variable looks like this:

Table 4.7: Content minus Function Words and Scores by Level

Score Level	Content-Function Mean (SD)
< 2.0 (low)	50% (8.3%)
= 2.0 (medium)	46% (7.6%)
> 2.0 (high)	51% (10.8%)

Here we can say with confidence that all subjects used pitch to distinguish content from function words without regard to level of English.

4.4 Conclusion

The most salient result of this study is the significance of the paratone variable in the regression analysis. The NNSs who scored 2.0 and above on the exam were similar to the NSs on this variable: the former group averaged a difference of just above 50% before and after topic shifts, and the NSs averaged a difference of 62% and 51%. For NNSs who scored below 2.0 on the exam, however, the average pitch increase for paratones was only 20%. Thus, higher level NNSs were using intonation at the discourse level to distinguish rhetorical units in a similar way to NSs, and on average, this helped increase their exam scores. It is noteworthy that the

variable having the most weight in the multiple regression analysis was the most macro-level variable. This represents a use of intonation at the discourse level to differentiate the major organizational components of the lecture. Since I know of no study of Chinese that considers the paratone at topic shifts, it is not possible to comment on a transfer effect, but I would speculate that such discourse intonation varies from language to language.

That the other intonation variables were not significant in the regression analysis shows how variable the data is. As we saw in Figure 4.5, "outlier" subjects whose overall English level was assessed as high might have low intonation scores, while others whose intonation measurements were high received low scores in overall level of English. Moreover, it is clear that subjects' were using intonation variables with different degrees of effectiveness. For example, a high-scoring speaker might have dramatically increased his pitch to make contrasts, but inserted a number of misleading low boundary tones in mid-utterance position, while another speaker at the same level of English might have used the boundary tones effectively but not distinguished the contrasts. Since statistical methods derive average trends, the behavior of such individuals in the study has been obscured. In sum, there appears to be a good deal of variation in how these aspects of intonation are acquired, relative to each other and relative to other language skills. To counteract this variability, it would be necessary to

conduct studies with larger groups of subjects.

In a more qualitative mode, I will discuss the nonsignificant intonation variables of the study, comparing individual NSs and NNSs. The variable showing the greatest cross-linguistic difference was the contrast-given variable, for which the two NSs made a huge distinction (about 65%) between contrast words and given words. The NNSs' averages hovered around 30% for this variable with no significant difference found by level in the multiple regression analysis. There were certain individual subjects who scored as high as 40-60% on this variable, but in general, the NNSs did not use pitch extremes to support contrasts made in their lectures. Returning to Shen's (1990a) assertion that contrastive stress in Mandarin Chinese is achieved more through duration than pitch, it would seem that transfer could lead to this result and that this cohesive use of intonation is acquired late in the language development process.

As for utterance-medial intonation boundaries, we saw that NSs had close to 30 (29 and 25) appropriately nonfalling boundaries. The high-scoring group of NNSs was on par with this, averaging 25, while the lowest group only averaged 19. The speaker illustrated in Figure 4.2, who received one of the lowest exam scores (1.3) only had 15 nonfalling intonation boundaries in this environment. This is consistent with Shen's finding that Chinese intonation is more distinctive at initial than final boundaries.

Finally, the content-function variable showed no cross-linguistic differences at all; if anything, the NNSs made a higher distinction between content and function words (above 50%) than did the NSs whose numbers were closer to 40%. Again, this coincides with Shen's comment that enclitics in Chinese are not stressed. In other words, for both languages, content words are more salient than function words and this shows up as a slightly higher pitch.

To summarize, there appears to have been a transfer effect from Chinese intonation into the English of these speakers. There is also evidence of interlanguage development in progress since speakers showed varying degrees of accuracy with respect to the target model. Indeed, certain individual subjects' intonation was very close to the model's predictions, although no one was exactly nativelike. However, there is no evidence for a lock-step order of acquisition, since there appeared to be a good deal of variability in how speakers at the same level of English used intonation, as well as the sequence in which the intonation variables were acquired.

In general, it can be concluded that intonation, as a linguistic system, matters in the communicative competence of NNSs. This study shows that one of the factors that could actually prevent an international graduate student from obtaining a job as a TA was the effect of intonation in this gatekeeping situation, the oral English exam. Chapter 7

contains a broader discussion of these results, along with those of the other two studies. Finally, this study makes no claims about the value of emotional and attitudinal uses of intonation (to be taken up in Chapter 6), although these are certainly important in effective lecturing. Before discussing these, we move first to a study of intonation in a more dialogic genre, informal conversation.

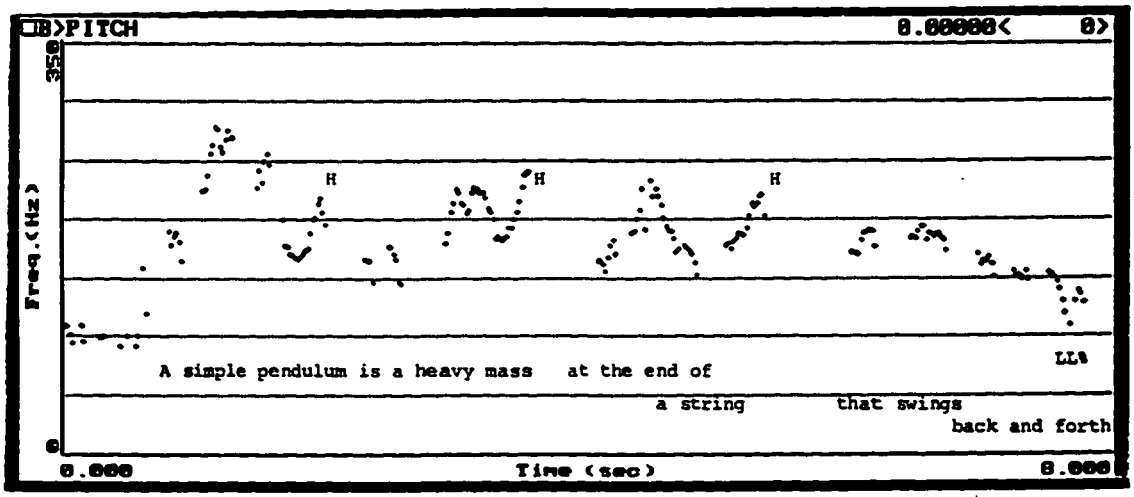


FIGURE 4.1 A high-level speaker's appropriate use of high phrase boundaries in utterance-medial position indicates the interdependency of the phrases.

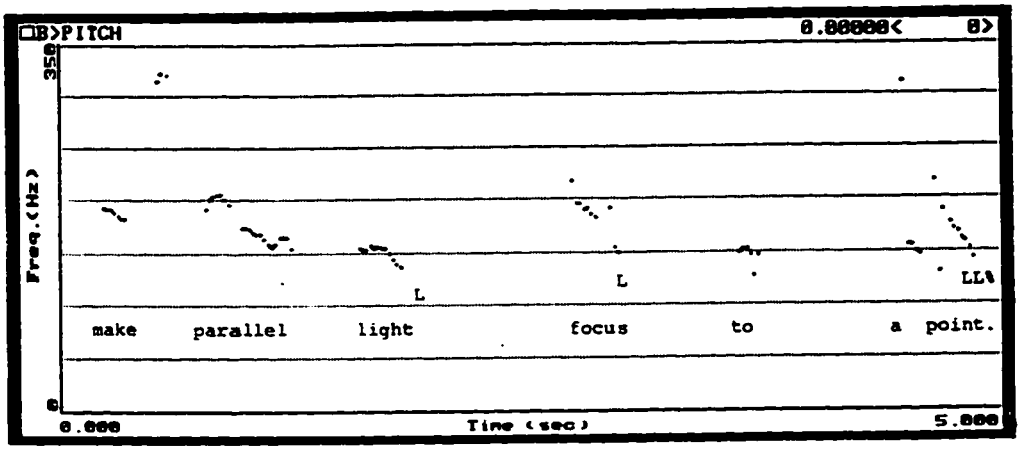


FIGURE 4.2 A low-level speaker inserts low phrase boundaries in utterance-medial position which separates the phrases into independent constituents.

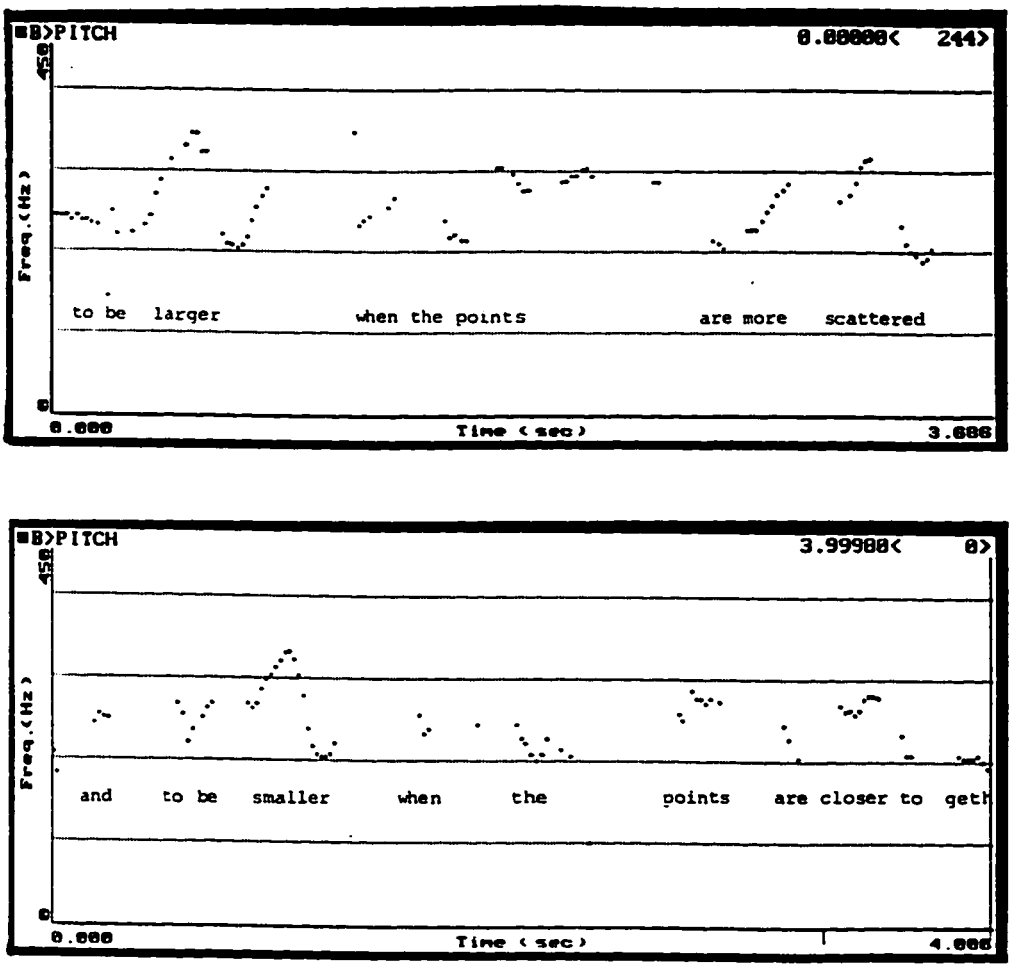


FIGURE 4.3 A high-level speaker's sharply rising pitch on the words larger, smaller, scattered, and closer indicates contrast.

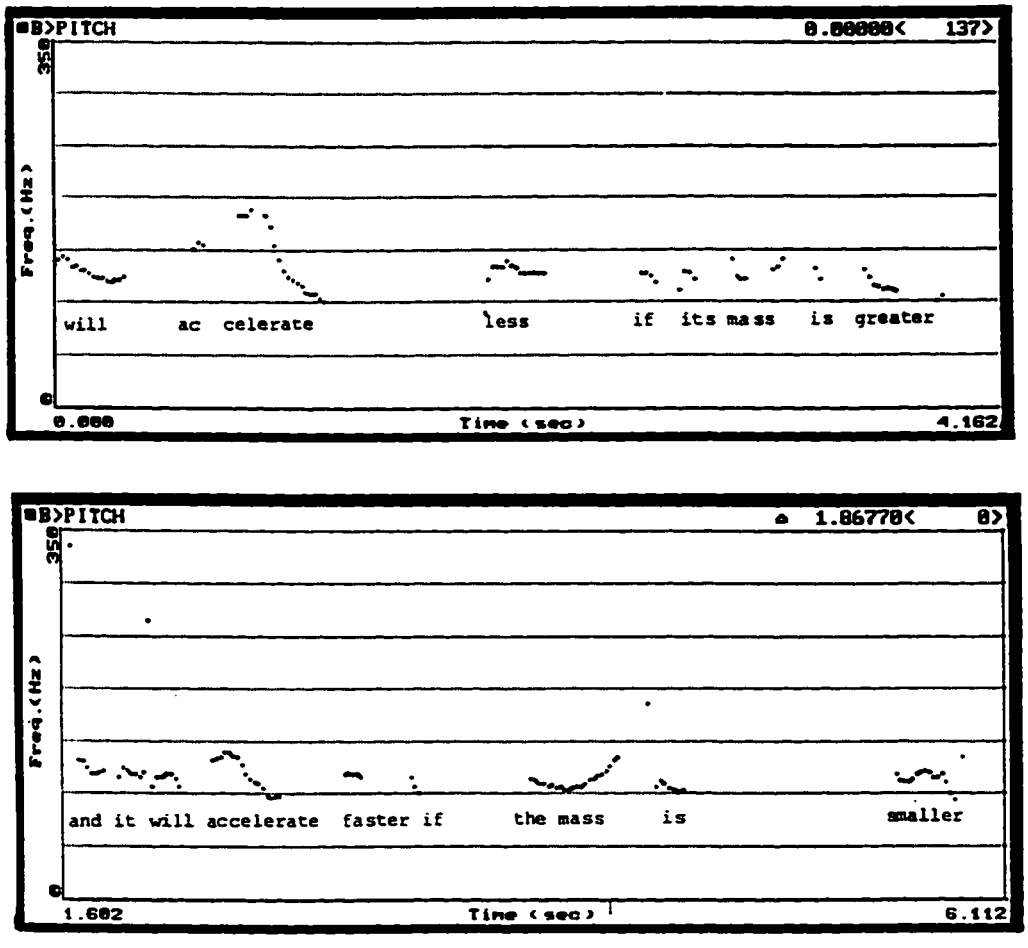


FIGURE 4.4 A low-level speaker does not support his contrasts intonationally since the given words accelerated and mass have higher pitch than the contrasting words less and faster.

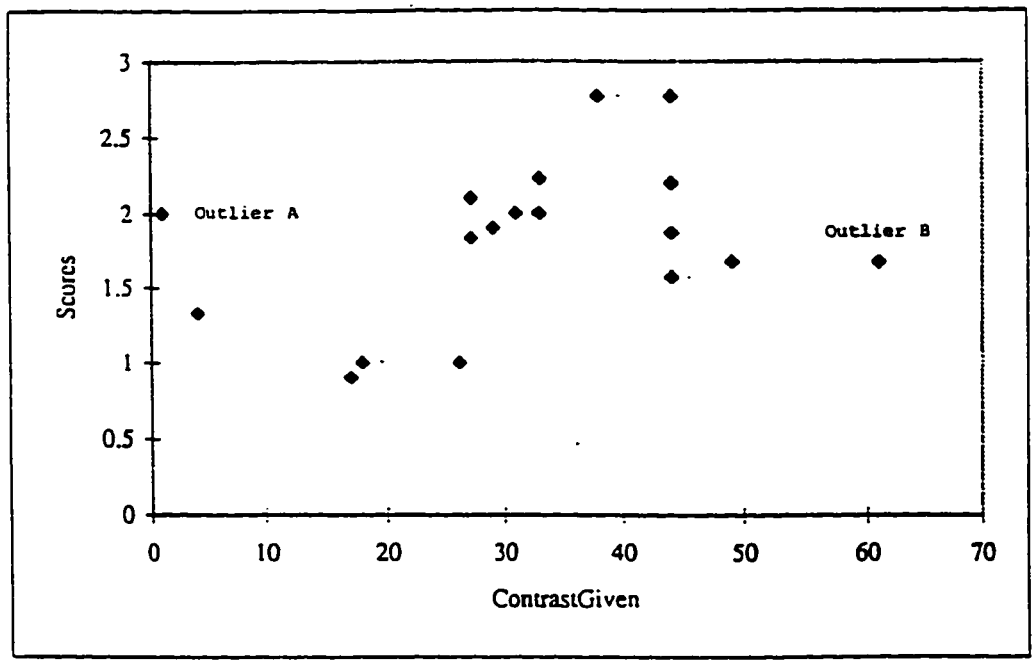


FIGURE 4.5 The contrast-given variable is plotted against exam scores to show how outliers obscure the linear relationship.

Notes to Chapter 4

1. In fact, I would speculate that what we commonly refer to negatively as a "pedantic" tone is one with the over-prevalence of H* pitch accents by a speaker when the information associated with it seems obvious or "given" to the hearer. When this occurs in conversation, it may sound "pompous."

2. Interrater reliability was calculated according to a formula provided by Hatch and Lazaraton (1991), p. 533. It is based on the average correlation of all three raters.

3. The NSs did not get perfect scores because there are certain reasons to insert a LL% boundary in utterance-medial position (for example, if one thinks one is finished, but then has an after-thought). To avoid circularity, I consistently identified the environment structurally (as utterance-medial) and did not try to speculate about the speaker's intention.

4. Werth (1984) discusses how given items in theme position tend to have high pitch.

5. One exception was a lecture by a law student who covered a particular legal case and whose organization was chronological.

6. In fact, four subjects had fewer than 10 such transitions, the fewest being seven for one subject. For these, the totals were averaged over all transitions.

7. A correlation matrix indicates the strength of association between individual variables. Values close to 1.0 and -1.0 indicate a strong relationship; values close to 0 indicate little or no relationship.

8. It was purely coincidental that the scores were aligned in this manner.

Chapter 5

Study II: Conversation

Having focussed in the last chapter on a genre that is mainly monologic, I move now to investigate the role of intonation in dialogue. While in lectures, a speaker's task involves maintaining coherence in structuring a long sequence of text, the conversation genre introduces a somewhat different set of communication issues. Since a coherent discourse structure must be built in cooperation with others, it becomes important for participants to react spontaneously to each other, bringing knowledge of the norms of conversational dynamics, many of which are language-specific, to the interpretation. The framework of conversation analysis (Sacks, Schegloff, & Jefferson, 1974) emphasizes the importance of minute linguistic details in such maneuvers as holding the floor, initiating and relinquishing turns, and resolving miscommunications. I will borrow Riggenbach's (1991) term "conversational fluency" to refer to the ability to accomplish such moves effectively in conversation, and provide analyses of how intonation contributes to or hinders fluent conversational interaction in L2 speakers of English.

5.1 Intonation and the Genre of Conversation

Clearly, all the genres investigated in this dissertation

have information structure and constituent structure. Thus, in conversation as in the other genres, high pitch associates with new and contrastive items and low pitch associates with cohesive material. We would also predict boundaries of constituents to have intonational characteristics, from the phrase level, where intonational boundaries at right edges of phrases indicate whether they are to be interpreted with respect to a subsequent unit, up to the topic level, where a conversant might use a paratone in changing the subject.

A striking feature of this genre, however, is that the construction of the information and constituent structure is collaborative. Referents entered into the discourse as new by one speaker may thereafter be in the mental representation of the other, and thus associated with a L* pitch accent in that person's next utterance. Similarly, two speakers may collaborate with their boundary tones within a topic unit, keeping phrase boundaries at high- or midlevel to indicate cohesion within a larger constituent. Another unique feature of this genre is the interactional potential of intonation. With regard to turn-taking, Sacks et al. (1974) identify the "transition-relevance place"; that is, the point at which the floor may potentially shift to another speaker. Such places can occur at the sentential, clausal, phrasal, and even lexical level (p. 702), and are thus not reliably distinguished by syntax alone. Sacks et al. state:

> Clearly, in some understanding of "sound

> production" (i.e., phonology, intonation etc.), it is also very important to turn-taking organization. For example, [in real time processing - AKW] discriminations between *what* as a one-word question and as the start of a sentential (or clausal or phrasal) construction are made not syntactically, but intonationally. When it is further realized that any word can be made into a "one-word" unit-type, via intonation, then we can appreciate the partial character of the unit-types' description in syntactic terms (p. 721-2).

The intonation to which Sacks et al. refer can be more finely described within the intonation model presented in Chapter 2: it is the boundary shape at the right edge of a constituent (of whatever size) which signals the transition-relevance place. The role in the turn-taking system of the four utterance boundary shapes described in Chapter 2 is summarized as follows.

First, HH% (high rise) has been described as indicating that material within that utterance requires subsequent discourse for its interpretation. There is no need to alter this characterization to accommodate an interactional situation since Pierrehumbert and Hirshberg point out that this reference is most often cross-speaker (1990, p. 306). As McLemore (1991) asserts, this boundary conveys "connectedness." Thus, in Sacks et al.'s terms, the HH%

boundary is a means for the current speaker to select the next speaker, while providing a cohesive link between his own and that person's response. HH% boundaries also contribute to the collaborative nature of conversation because they indicate the hierarchical distribution of constituents which exists above the level of one speaker's turn, structures which both speakers work together to create. Thus, in conversation, this boundary often accompanies the first part of an adjacency pair, soliciting a response, either in the form of a backchannel, an answer, or a confirmation.

Second, LL%, the falling boundary, indicates finality, the end of a thought. McLemore refers to this as the "segmentation" boundary. A LL% boundary often signals a transition-relevance place, but without specifying the expectation that an idea's completion will be dependent on a subsequent utterance or turn. LL% boundaries might occur, for example, as completions of adjacency pairs, or affirmations of another's opinion.

The third boundary shape, LH% (low rise), a "continuation rise," also signals an interdependency between the current and subsequent phrase, but in this case it is often the same speaker who continues.

The fourth boundary shape, HL% (plateau), is a continuing midlevel boundary, which can be used to keep one's turn. The relevance of this contour to the current discussion is that it does not signal the relinquishment of the turn, but rather

continuation (McLemore, 1991) and is therefore a common way to maintain the floor during hesitations. It can be articulated by extending the final stressed vowel of the last word uttered, or the vowel of a pause filler (such as *uhm, uh,* etc.) to diminish the likelihood that the speaker will lose the floor. Muller (1996) also describes "moderately falling pitch" (p. 145), which appears to be a version of this boundary, with which a speaker may make a provisory move toward topic closure, while negotiating for permission to continue with the next increment of talk.

To summarize, we have seen that our model of intonation can be applied to an interactive format with little modification. I have discussed the fact that within this genre, the information structure is built collaboratively, as is the constituent structure, and that boundary tones make distinct contributions to the conversants' perceptions and intentions about the "transition-relevance places" at which a new turn may begin. To the extent that these interactive functions of intonation are specific to English, NNSs may encounter communication problems involving both the perception and production of intonation in conversation. It is these problems that are the focus of the current study.

5.2 Conversational Fluency

Before moving on to the details of the empirical study, it will be useful to discuss the notion of conversational

fluency, a concept which is somewhat complex. Educational Testing Services includes fluency as a subscore in their TSE (Test of Spoken English) and SPEAK tests along with separate scores for Comprehensibility, Pronunciation, and Grammar. The fluency rating scale, which was used to rate the fluency of the speakers in this study, is reprinted below from the SPEAK test rater training kit (ETS, 1985, p. 10):

SPEAK Test Scoring Key - Fluency

0 Speech is so halting and fragmentary or has such a nonnative flow that intelligibility is virtually impossible.

1 Numerous nonnative pauses and/or a nonnative flow that interferes with intelligibility.

2 Some nonnative pauses but with a more nearly native flow so that the pauses do not interfere with intelligibility.

3 Speech is as smooth and as effortless as that of a NS.

These descriptions refer to less fluent speech as "halting and fragmentary" and as having "numerous pauses," whereas fluent speech is described as "smooth," "effortless" and as having "native flow." While such descriptions may be appropriate to instruct test raters to think holistically in their judgment of fluency, we are left without a clear notion of what linguistic features actually correspond to these general characteristics. Certainly a fluent speaker may have pauses or syntactically incomplete utterances under appropriate circumstances without a resulting loss of fluency or

intelligibility in the perception of listeners.

Furthermore, as Riggenbach (1991) points out, in the context of conversation there may be aspects of fluency that are not manifest in the monologic SPEAK Test format, and which involve more than just timing and flow. In a detailed study of NNSs of English engaged in informal conversations she provides us with a series of variables which together contribute to what is perceived as fluent speech in conversation. These variables include categories discussed in the conversation analysis tradition (Sacks et al., 1974): frequency and type of hesitation phenomena, frequency of repair, amount and rate of speech, "carrying weight" in the conversation through substantive comments (p. 439), anticipating turn endings and latching on appropriately, and other interactive features of conversation management. Riggenbach concludes that these interactional features are necessary for a complete definition of fluency, and that a number of components work together in fluent speech. She states, "In order for there to be fluency, then, it appears that many different conditions have to be met . . . . Nonfluency, on the other hand, can arise from a deficiency in any one of these areas." (p. 439).

In line with Riggenbach's concept, the present study takes the premise that fluency can be analyzed as having several linguistic components. It is proposed that intonation is also one of these components, playing an important role in

interaction.[1] Two aspects of intonation will be considered: The first is the way speakers use pitch accents to differentiate items in the information structure of the discourse, making it easier to follow their train of thought; the second is how speakers use boundary tones to indicate when they intend to relinquish their turn, solicit a backchannel, or keep the floor.

5.3 Method

At this point we move to the details of the study itself, a conversation analysis of the intonation of a group of NNSs of English from a variety of language backgrounds. In order to observe how intonation patterns occurred in situations of real communication, the data were taken from naturally occurring dialogues between NSs and NNSs. As in the other studies of this dissertation, the CSL machine was used to measure and graph specific details of speakers' pitch patterns in these dialogues, allowing for comparisons among subjects at various levels of fluency, as rated by NS judges.

5.3.1 Subject Data

Sixteen tapes of informal dialogues between NSs and NNSs were collected, with permission, for analysis. The NNSs were ESL students of various nationalities in an intensive English program. The conversations were recorded by the students themselves as part of an assignment in conversation analysis.

Students were free to choose any NS and to converse on any topic. They were asked to select a 1 to 3 minute segment of the conversation in which there was a lot of two-way interaction. Of the 16 tapes, 5 were rejected from the study due to poor recording quality. A sixth tape was used as a practice exercise to train raters for the study. The remaining 10 tapes contained conversation excerpts from 3 male and 7 female speakers from intermediate to advanced levels of English. There were 4 Korean speakers, 3 Japanese speakers, 1 Mandarin speaker, 1 Thai speaker, and 1 Italian speaker from Switzerland. The conversations were then analyzed on a CSL machine to measure the pitch patterns of the NNSs. As in the lecture study, it was expected that fluent speakers would use intonation to support distinctions between new (H*), contrasting (L+H*), and accessible (L*) content words while function words would not be associated with pitch accents. Likewise, it was expected that fluent speakers would have more success in managing the turn-taking task and keeping the floor if desired by using appropriate intonational boundaries at the ends of utterances.

The pitch of individual lexical items was measured at the syllable center at the point of highest amplitude. Boundary tones were measured at points where "prepausal lengthening" (Klatt, 1975, p. 130) could be identified in the final word (or hesitation sound) on the screen of the CSL machine. These were usually chosen at points of backchannel or turn shift,

and also at "trouble spots" where unwanted interruptions or overlaps by the NS appeared to be occurring, and before particularly long pauses (.5 seconds or more). In such cases, the aim was to search for evidence in the data that some aspect of the NNS's intonation was triggering the troublesome occurrence.

5.3.2 Rating

As for all the studies in this dissertation, the rating of subjects for level of English was done using a format and scale similar to the SPEAK Test. Raters had all been trained to rate the SPEAK Test and had experience in rating for at least one year. Although the original SPEAK Test has four scores, the fluency subscore was of central interest because of this study's focus on conversational fluency. In a comprehensive study of 90 SPEAK Test results conducted at three major universities, the correlation between fluency and comprehensibility scores was found to be .85 (Cascallar, Sarwark, Smith, & Wennerstrom, 1992), which indicates that the fluency score is a strong component of overall language ability in the perception of raters.

In the present study there were two raters for each tape. They were told to use their best intuitive judgment to rate the tapes according to the standards they were accustomed to using in rating SPEAK Tests. Raters were not informed of the purpose of the study other than that they were helping with a

research project on ESL conversation. No mention was made of intonation, nor were raters aware that the fluency subscore was the particular focus of study. Special "SPEAK-like" tapes were dubbed from the original student-made tapes and these included two sections for each subject. In Section I, raters heard the entire segment of conversation in context with both participants' utterances included. To these they gave the NNSs a holistic rating between 0 and 3 to the nearest tenth in each of three areas: pronunciation, fluency, and comprehensibility. In Section II, raters were presented with the first 10 of the NNSs' individual utterances in isolation, cut in segments from the original conversation. It was expected that their memory of the original dialogue would provide a context to interpret these short segments. To each utterance they assigned a rating from 0 to 3 to the nearest tenth in the same three areas and these 10 scores were averaged for the section. To assign a single mean fluency score to each subject, averages were taken of the two sections and the two raters. From these final averages, subjects were ranked according to fluency. Interrater reliability was .78 on the final scores.

Table 5.1 shows the subjects ordered from 1 to 10 by their fluency scores averaged for the two raters, 1 being the highest and 10 the lowest. Also shown are first language, gender, and fluency score for each subject.

Table 5.1: Subjects Ranked by Fluency Score

Subject	Language	Gender	Fluency
1	Italian	M	2.85
2	Korean	F	2.18
3	Korean	M	2.10
4	Japanese	F	2.06
5	Mandarin	F	2.00
6	Korean	F	1.93
7	Japanese	F	1.83
8	Korean	M	1.55
9	Thai	F	1.55
10	Japanese	F	1.45

As Table 5.1 shows, fluency scores ranged from 2.85 to 1.45 within a possible range of 0 to 3, as described in the scoring key in Section 5.2. Scores did not particularly coincide with country or gender, except that the highest level speaker, who was rated nearly nativelike, was the only European.

5.4 Results and Discussion

The results are organized to focus first on the intonation associated with lexical items to indicate their role in the information structure, and second on the function of boundary tones in the interaction. Within each of these categories, the intonation patterns prevalent among speakers with fluency ratings above 2 (Speakers 1-4)[2] are compared with those of speakers with the lowest ratings (speakers 8-10). The middle-level speakers varied in their intonation and will

not be the focus of the examples.

5.4.1 Information Structure: The Pitch of Lexical Items

First, we consider several examples of how the speakers rated highest in fluency used intonation effectively in their interactions. Thereafter, examples from lower level speakers are shown. The following codes are used in the transcripts:

$word_{185}$	pitch of this word at the center of the stressed syllable is 185 Hz.
$word_{(/)}$	pitch of this word could not be obtained due to low volume, low sonorancy, or background noise
WORD	high-pitched word singled out as relevant to the discussion
[word [name	(brackets) simultaneous speech

Speaker 1, a male speaker of Italian, rated the highest (2.85) in fluency, provides an example of nativelike control of intonation related to information structure. He discusses Swiss prisons in the following excerpt, segments of which are shown in Figures 5.1 and 5.2:

1 NNS: YOU_{124} got_{115} $EVERYTHING_{126}$ in_{92} $there_{94}$:

2 you_{104} got_{115} a_{86} TV_{117};

3 you_{95} can_{112} $have_{92}$ a_{90} VCR_{116};

4 you_{99} can_{102} $have_{94}$ $a_{(/)}$ -- $EVERYTHING_{130}$!

5 YOU_{120} $just_{133}$ y- you_{119} $DON'T_{130}$ $have_{123}$ the_{108} $FREEDOM_{114}$

6 $to_{(/)}$ go_{105} OUT_{116}.

7 NS: uh

8 NNS: $SOMETIMES_{130}$ you_{129} can_{107} $even_{109}$ $bring_{103}$ y- $your_{99}$

9 $WIFE_{107}$ in_{103} $there_{98}$.

Even though this speaker's range is very narrow (45 Hz for this section)[3], he uses pitch contrastively. In lines 1-4 each phrase adds one new item to the discourse. Although these items come near the end of their phrases where pitch naturally declines, their pitch is higher than that of the surrounding function words. It is also interesting to compare the five uses of the generic *you* in the first five lines. These are all function words, occurring in first position in structurally similar utterances. Yet only the first and fifth are relatively high in the speaker's range. I analyze the first and fifth as instances of high key (%H) (Brazil, 1985) or minor paratones (Yule, 1980) because they accompany a shift in organizational level. In line 1 the speaker is introducing a new topic of luxuries available to prisoners; the next three *you's* begin elaborations within the same topic and are thus initiated with a somewhat lower pitch; the *you* in line 5 again starts a new subtopic, what prisoners *don't* have.

Lines 5 and 6 also show appropriate high pitch on the three new items, *don't, freedom,* and *out*. The word *go* is appropriately low-pitched as the unstressed part of the verbal compound *go out*. The words, *freedom to go out* are shown in Figure 5.1.

Figure 5.2 shows line 8, where *sometimes* contrasts with implied normal times, having a high pitch. The figure also reveals a small pitch rise on the new word *wife*, higher than the surrounding function words. The word *there*, a proform referring back to the prison, is already assumed to be in the listener's mental representation in this context and appropriately low-pitched.

Speaker 3, a Korean man rated 2.1 in fluency, shows similar characteristics in many of his utterances, although his intonation is not quite as nativelike as that of Speaker 1. The following excerpt includes a strong contrast between computer brands, which are clearly distinguished by pitch accents. (Figures 5.3 - 5.4 show excerpts from this speaker.)

1 NS: I can't decide between a Powermac and a Pentium.

2 NNS: In_{133} MY_{141} $opinion_{125}$,

3 $PENTIUM_{167}$ is_{106} $better_{112}$

4 IF_{199} $I'm_{119}$ in_{104} the_{105} $other_{106}$ $university_{111}$.

5 BUT_{200} in_{140} the_{141} $case_{154}$ of_{110} $University_{114}$ of_{110}

6 $Washington_{103}$, um, $POWER_{136}macintosh_{107}$ is_{97} $better_{101}$.

Looking at the biggest contrast between *Pentium* in line 3 and *Powermacintosh* in line 6, we notice that these words have the highest pitch in their phrases as expected. In terms of discourse structure, this indicates that the NNS has taken the NS's input of the two brands of computer into his own mental representation of the discourse, and made a contrast between the items with L+H* pitch accents. The two low-pitched

utterances of the word *better* in the same lines are also appropriate since they are given in this context of deciding which computer product is superior. The speaker also utters *Macintosh* with a low pitch as expected in the noun-noun compound *Powermacintosh*. Figure 5.3 shows lines 3 and 6. The contrast between *other university* and *University of Washington* in lines 4 and 5 is somewhat less distinct intonationally. In the second half of the contrast, the speaker gives the word *university* slightly higher pitch where a NS would be more likely to accent *Washington*. Nevertheless, the phrase as a whole is distinguished from the local function word *of* by having a higher pitch. The word *but*, a discourse marker providing a key signal to the contrast, is also high-pitched.

In general, this speaker uses relatively lower pitch on function words with a few exceptions. The words *in* and *if* in lines 2 and 3 are quite high-pitched, which may be due to their being in first position. Figure 5.4 shows the utterance *in the case of* with distinguishing pitch between the new item, *case*, and the function words *in*, *the*, and *of*.

Finally, the phrase *in my opinion* in line 2, and shown in Figure 5.3, bears some discussion. The word *my* takes the highest pitch since it is contrastive within the deictic system established collaboratively in the mental representation (*you* expressed indecision, but *I* have an opinion). Since this is a set phrase in English, the word *opinion* is the expected completion of the phrase and thus

predictable rather than new (see Bolinger, 1986). In other words, the speaker has used pitch appropriately to indicate the information structure of the phrase.

The next example comes from Speaker 4, a Japanese woman rated 2.06 in fluency. Here the NS asks a question in which the city, Kobe, is proposed as an addition to the category of international Japanese cities. The NNS makes a strong contrast between this and a different city, Yokohama. As shown in Figure 5.5, this is supported intonationally with a sharply rising L+H* pitch accent reaching to the very top of her range.

1 NS: Do you think Kobe is more international than other
2 cities in Ja[pan?
3 NNS: [nyyes$_{286}$ yes$_{239}$ uh but uh YOKOHAMA$_{348}$ is$_{222}$
4 pretty$_{227}$ um
5 NS: mmm
6 NNS: international$_{208}$ I$_{175}$ think$_{179}$ but$_{201}$
7 NS: nn

Another interesting point is the appropriately large difference in pitch (121 Hz.) between the contrasting item *Yokohama* and the beginning of the given unit *pretty international*, which is a reiteration of the idea *more international*, introduced by the NS in line 1.

Another contrast comes from Speaker 2, a Korean woman, whose friend is teasing her about a test score:

1 NS: . . . that's really low!

2 NNS: No_{217}, $it's_{314}$ NOT_{335} low_{236}.

3 $It's_{(/)}$ [$NOT_{(/)}$ $low_{(/)}$.

4 NS: ['s a low score

5 NNS: Who_{205} SAY_{349} $that_{298}$?

In line 2, the highest pitch is on the word *not* to make the contrast with the first speaker's assertion. It is noteworthy that the word *low* is introduced as new by the NS in line 1 and is taken as given by the NNS in line 2 (uttered at a full 99 Hz lower than the previous word *not*). Similarly, in the last line, the word *say*, the focus of the utterance, is also high, while anaphoric *that* is appropriately low. These are examples of appropriate use of intonation in structuring a similar mental representation of the discourse across speakers.

To summarize, the four most fluent speakers used differing pitch levels to distinguish new information from function items, to deaccent items given in the mental representation, and to make contrasts. In other words, they demonstrated an ability to produce and perceive intonational patterns which contributed to the information structure being built in the mental representation.

Turning to the low-fluency group, Speaker 10, a Japanese woman rated 1.45 in fluency, exhibits the least nativelike intonation. Her pattern is to associate high pitch with all words, regardless of their function in the discourse as shown in the following excerpt:

1 NNS: BUT_{332} $THEY_{348}$ $DON'T_{(/)}$ $KNOW_{330}$ THE_{324} ah uh $actual_{204}$

2 JAPANESE$_{346}$ life$_{251}$. So$_{303}$ I$_{310}$ wanted$_{299}$ to$_{284}$ say$_{264}$
3 the$_{255}$ ah JAPANESE$_{348}$ LIFE$_{318}$ is$_{285}$ SIMILAR$_{341}$ TO$_{332}$
4 YOURS$_{348}$.
5 NS: The what?
6 NNS: JAPANESE$_{284}$ LIFE$_{239}$ [STYLE$_{285}$
7 NS: [lifestyle. Lifestyle.
8 NNS: Yeah$_{251}$, WE$_{348}$ DON'T$_{322}$ WEAR$_{310}$ KIMONO$_{314}$, WE$_{312}$ WEAR$_{314}$
9 SUITS$_{305}$
10 NS: yes.
11 NNS: OR$_{268}$ JEANS$_{289}$.

First of all, in spite of this speaker's very large range (167 Hz for this excerpt), there is relatively little distinction between the pitch of function words and that of new words. For example, in line 1, the function word *the* is within 6 Hz of preceding *know*, and 20 Hz higher than *actual*. Similarly, the infinitive marker *to* in line 2 is 20 Hz higher than its verb *say* which follows. The problem is not so much a lack of pitch accent on new words, as an obscuring of the focus due to the high pitch on every word. For example, in lines 3 and 4 there is a contrast between *Japanese* and *yours*, both of which are uttered at the top of the speaker's pitch range. However, in the same phrase *similar* and *to* are also near the top of the range so that the contrasting words are not especially distinguished. In fact, in line 5 we can see that the NS has to ask for clarification of the word *lifestyle* after this utterance. Although there are probably several reasons for

the NS lack of understanding, such as pronunciation of the sounds and the somewhat vague reference to *Japanese life*, the intonation is not helpful either. The equal pitch phenomenon occurs again in lines 8-9 where the speaker contrasts *kimono* with *suits*, shown in Figure 5.6. Although the contrasting words have a high pitch, so do the rest of the words in the utterance. In particular, the second occurrence of the words *we wear* is uttered with a high pitch on both words instead of the lower pitch predicted for this given information.

Speaker 9, a Thai woman rated 1.55 in fluency has a similar pattern of giving equal weight to each word, although the phenomenon is less pronounced since her range is much narrower. Instead, her speech sounds somewhat like a monotone and, as with Speaker 10, it is hard to sort out the focus of the phrase. The following short excerpt illustrates this (see Figure 5.7):

1 NNS: $WHAT_{240}$ IS_{221} THE_{233} PUR_{230} $POSE_{228}$ OF_{223} $YOUR_{227}$ $LIFE_{212}$?

2 NS: Um (2.0) the purpose of my life is to be happy

3 which ...

In line 1, the range is 60 Hz. We see a gradual descent throughout the phrase, with relatively little pitch distinction between function and new words. Within the word *purpose*, we also find that both syllables have similar pitch, obscuring word stress.

The next excerpt is from Speaker 8, a Korean man, whose fluency rating was 1.55:

1 NS: and um what size apartment is it?

2 NNS: uh $LARGE_{119}$ juh $size_{100}$.

3 NS: a large- is it a one bedroom or two bedroom?

4 NNS: $IT'S_{128}$ one_{112} $bedroom_{102}$.

5 NS: Ok, large, one bedroom

6 NNS: $yeah_{104}$.

This sequence shows inconsistent use of intonation in contrasts. The NS twice presents the NNS with a set from which to choose a member, the size of the apartment in line 1 and the number of bedrooms in line 3. This is a clear situation for using L+H* pitch in the answering choice. In terms of the discourse structure, the general category of size is introduced first by the NS. Thus, it should have a L* pitch accent in the reply, since it is accessible in the mental representation at that point. In line 2, the speaker does use pitch appropriately, since *size* is 19 Hz lower than *large*, the new idea and the focus of the response. However, in line 4, shown in Figure 5.8, there is a parallel situation in which the NNS does not follow this pattern. The NS has introduced the category of bedrooms, wanting to know a specific number. The NS reply, *one*, does not stand out intonationally since *it's* (a pronoun and contracted copula) is 16 Hz. higher, and *bedroom* is only 10 Hz. lower than *one*, in a sequence which should be naturally falling anyway due to catathesis.

To summarize the characteristics of the low-fluency

subjects, we've seen examples of the tendency to give relatively equal pitch to each word regardless of its role in the information structure of the discourse. This can occur as many sequential high-pitched words, creating a choppy, word-by-word effect, or as a flat monotonous string of words with undifferentiated pitch. We have also seen several cases where contrasts were not especially distinguished from surrounding words.

5.4.2 Turn-taking: Intonation at Utterance Boundaries

In this section, I will discuss examples of intonation at the boundaries of utterances, beginning with the highest rated speakers. Additional symbols and codes are used in this section:

LL%	falling boundary
HH%	high, rising boundary
LH%	low, rising boundary
HL%	midlevel, flat boundary or "plateau"
wo::rd	word whose vowel is extended
nnnn; mmm	hesitation noise whose consonant is extended
=	a latch, an immediate continuation from one speaker to another
(X.X)	a pause of X.X seconds
xxx	can't catch the word

To begin with, here is an example from Korean Speaker 2

(rated 2.18) of a pause filled by an extended syllable with a HL% boundary:

1 NNS: Oh (LL%), today we gonna talk abou:::::t (HL%)
2 (1.5)
3 I don't know! (LL%)
4 NS: The TOEFL test! (LL%)
5 NNS: Oh (LL%), my TOEFL test! (LL%)
6 NS: Yeah. (LL%)

In line 1, this speaker hesitates for 1.5 seconds while thinking of a conversation topic. She holds the floor, however, with the extended plateau boundary within the word *about* before this pause, signalling the intention to continue. Only after the falling boundary in line 2 does the NS take the floor and make his suggestion.

From the Italian-Swiss man, Speaker 1 (rated 2.85), comes a similar example of a series of long pauses during a topic nomination:

1 NNS: What do you think abou::::t (HL%) (1.6)
2 the actua::::l (HL%) (1.0)
3 politics of the Sta::tes (HL%) (.7)
4 a:::ss (HL%) (2.0)
5 guardians of the world? (LL%)
6 NS: huh, um, I see it as being . . .

From Figure 5.9, we can see a pause of 1.6 seconds after *about*. Since this speaker's flat HL% boundaries appropriately mark his intention to continue, the pauses are tolerated

without interruption as part of his thought process in formulating his question.

Another strategy to maintain a turn during hesitation is to use a pause filler with a HL% boundary, as we see from Speaker 4 (Japanese), rated 2.0 in fluency (lines 1-3 are shown in Figure 5.10):

1 NNS: And actually Kobe doesn't ha:::::ve (HL%) (.4)
2 nnnnnnn (HL%) (.4)
3 art school (LL%)
4 NS: Really? [That surprises me.
5 NNS: [mmmm hmmm
6 annnn (HL%) (.3)
7 just design school and two years school

Lines 2 and 6 show the speaker maintaining the turn with HL% boundaries on pause fillers. The pauses themselves are not especially long.

Another option comes from Speaker 3, a high-fluency speaker from Korea, who uses a LH% boundary to signal continuation and a pause filler to link two utterances, in spite of rather long pauses. This is shown in Figure 5.11 where the pitch rises on the word *Washington*.

1 NNS: But in the case of University of Washington? (LH%)
2 (1.0)
3 ummmm (HL%) (.8) Power macintosh is better. (LL%)

Turning now to the less fluent speakers, there were certainly many appropriate boundaries in short questions and

statements. However, there were also cases where inappropriate boundary shapes resulted in a loss of the floor. We begin with a long excerpt by Speaker 10 from Japan. Although the main focus will be on lines 6-8 and 25-31, the whole sequence in included to provide context:

1 NNS: yeah, we don't wear kimono
2 we wear suits
3 NS: yes
4 NNS: or jeans
5 NS: yes
6 NNS: but (.7) **not** (LH%) (.3) **only words** (LH%) (.7)
7 if (LH%) I (LH%) [showed
8 NS: [you not only wear the traditional
9 Japanese clothing but you wear modern
10 NNS: yeah
11 NS: Jap- modern clothing
12 NNS: yeah
13 NS: like in western ah America
14 NNS: yeah
15 NS: suits and [(.) hats
16 NNS: [suits and yeah and jeans
17 NS: and jeans
18 NNS: and [xxx
19 NS: [so, the Japanese people are becoming
20 westernized
21 NNS: yeah westernized and the um I guess elementary

22 school students didn't didn'ta know that fact.
23 Just
24 NS: that's right
25 NNS: So I wanted to show the pictures to the (.9) kids
26 (LL%) (.6)
27 [because
28 NS: [that's very good
29 NNS: [yeah
30 NS: [that's wonderful
31 NNS: and um so um **not only words** (LH%) (.5)
32 I can show the pictures
33 helped students to under-
34 help rr students understand the japanese culture
35 NS: That's right.

In this excerpt, we have a clue to the speaker's intention in lines 6 and 31, indicated in boldface: She plans to show elementary school students in the U.S. pictures of Japanese culture, so that they will understand from "not only words". However, she is continually interrupted by the NS in trying to express this. Although the NS may be overlapping her speech in an effort to be helpful, it nonetheless takes eight exchanges for her to get back to her point. As we have already seen in Figure 5.6, which shows lines 1-2 of the same dialogue, her pitch rises sharply after almost every word, making it difficult for an English listener to determine when her point is finished and thus when the turn should shift. In

Figure 5.12, we observe this rising phenomenon occurring after the words *not, words, if,* and *I* in lines 6-7. Although LH% can signal the intention to continue, it is used so frequently that it difficult to sort out the turn-taking cues. The vertical line indicates the interruption point after the phrase *if I.*

Another problem occurs in line 25, where a sharply falling boundary after the word *kids* leads to another interruption. We can see from the next word *because,* however, that she had actually intended to continue. This is shown in Figure 5.13.

The next excerpt comes from a conversation between Speaker 8 (Korean) and his landlady about closet space (lines 1-3 are shown in Figure 5.14):

1 NNS: ss s'enough to use the use the (.6) uh (1.1)
2 to ha:::ng hanging the hang [hang
3 NS: [hanging the clothes=
4 NNS: =the clothes
5 lot of clothes
6 NS: Oh good=
7 NNS: =yeah=
8 NS: =good, so it is enough for you.
9 NNS: Yes.

In the first 2 lines, the speaker's hesitation strategy is to repeat the lexical items *use the* and *hang.* He also uses a pause filler *uh.* However, instead of making use of plateau

boundaries to extend the words or pause fillers, he tends to cut the words off abruptly without extending the boundaries. The one exception to this is *hang* in line 2. Trying to be helpful, the landlady overlaps in line 3. Although understanding is finally reached in the sequence, the raters gave an average fluency rating of only 1.5 for this utterance.

To summarize the influence of boundary shapes on fluency, the more fluent speakers used HL% and LH% boundaries to extend words and pause fillers to signal the intention to continue. Pauses following these tended to be tolerated without interruption. For less fluent speakers, we saw examples where inappropriate boundaries, that is, those that were too frequent, cut short, or of the wrong type, all resulted in a loss of the floor.

5.5 Conclusion

This analysis has drawn attention to certain patterns of intonation of conversationally fluent NNSs of English which differ from those of speakers rated as less fluent. In sum, the fluent speakers in this study were better able to both perceive and produce pitch patterns to signal relationships among words and phrases, and to segment their speech into turns in collaboration with others. High pitch was used to indicate which lexical items were to be added to the mental representation as new or to be contrasted with information already assumed to be there. Low pitch associated with given

items, even when those items had been introduced into the discourse by another speaker. Function words were distinguished from content words by having a relatively lower pitch as well. Finally, fluent speakers were able to use boundary tones to indicate their intention to hold or relinquish the floor.

Thus it is not longer utterances or shorter pauses per se that lead to a perception of fluent speech; instead it is the ability to speak phrasally rather than word-by-word, focusing the main idea of each utterance in a coherent manner and collaborating in the turn-taking process. It should not be surprising that these are the same grammatical intonational components that occur in the other genres; however, this genre affords an opportunity to observe how the system works in an interactional setting.

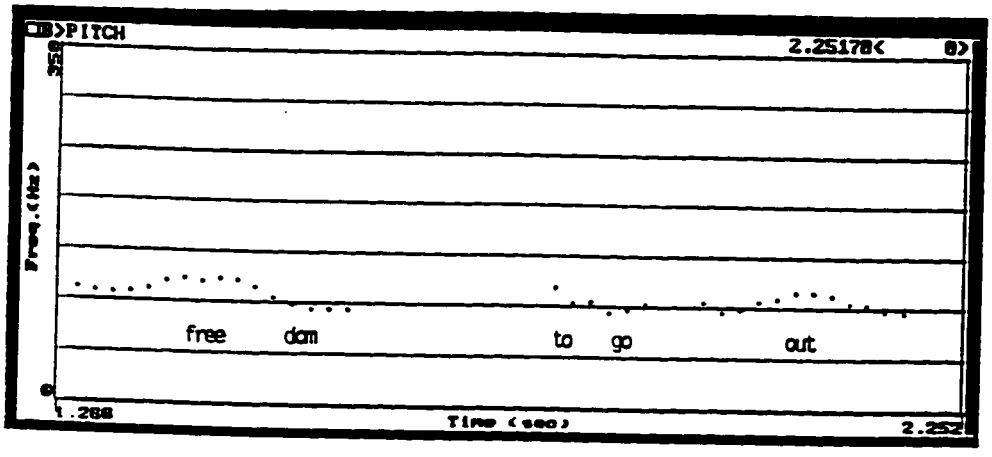


FIGURE 5.1 The speaker associates high pitch with the new ideas, *freedom* and *out*.

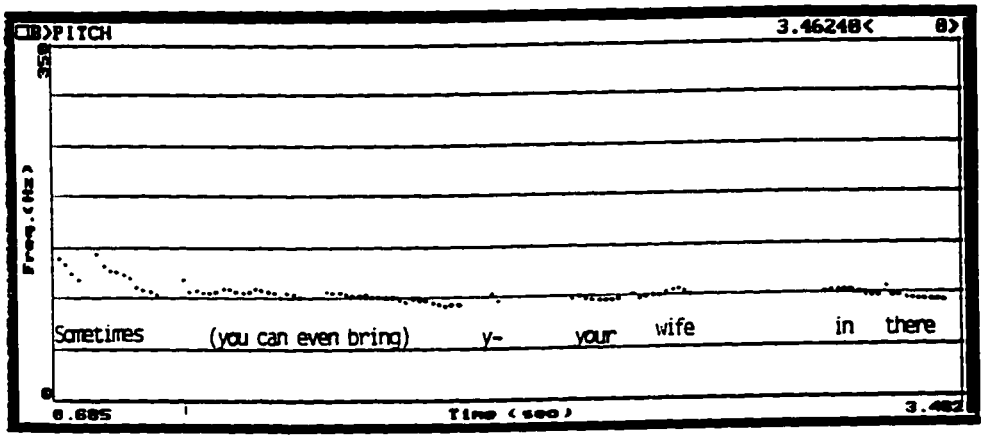


FIGURE 5.2 The speaker associates high pitch with contrasting *sometimes* and new *wife* while given *there* has a low pitch.

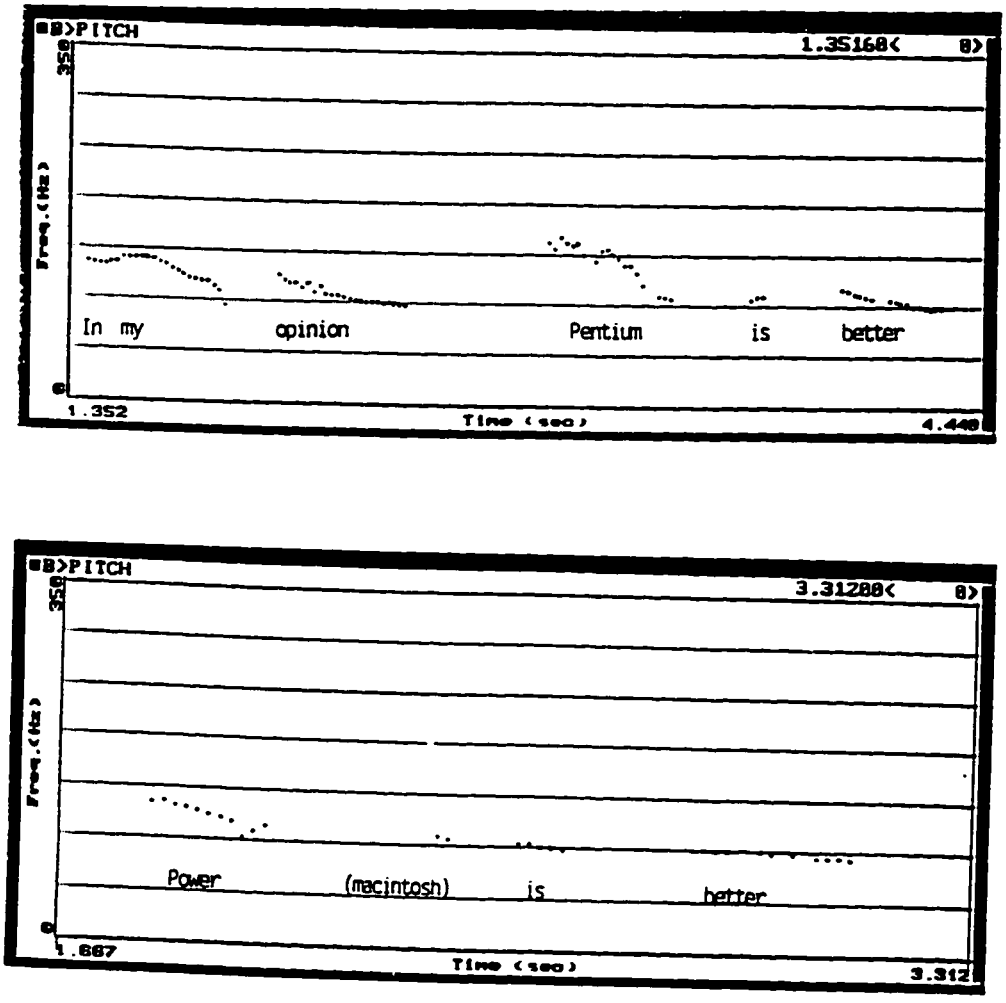


FIGURE 5.3 The speaker supports a contrast between *Pentium* and *Powermacintosh* with high pitch while given *better* has low pitch.

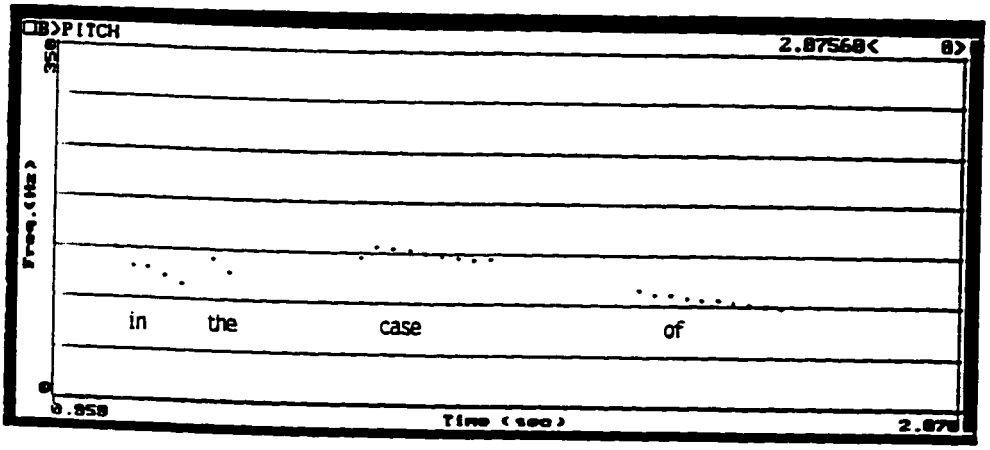


FIGURE 5.4 The content word *case* has higher pitch than surrounding function words.

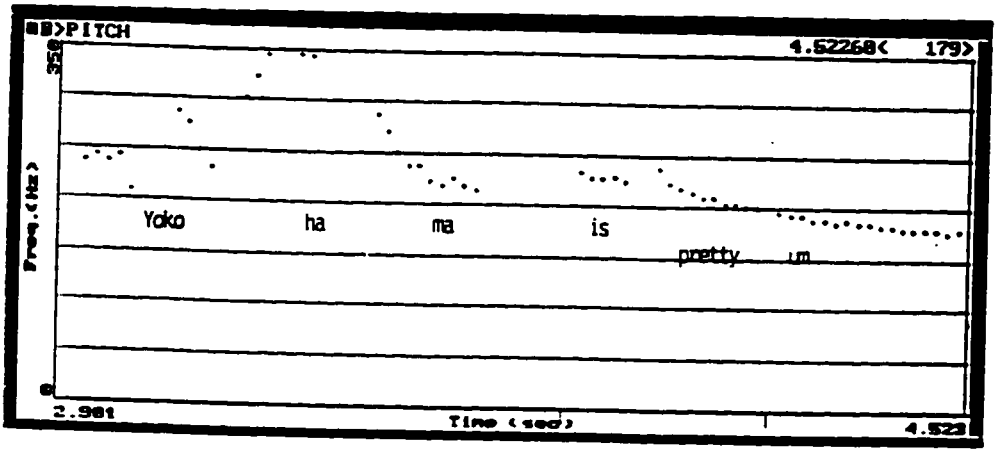


FIGURE 5.5 The speaker makes a strong contrast on *Yokohama*.

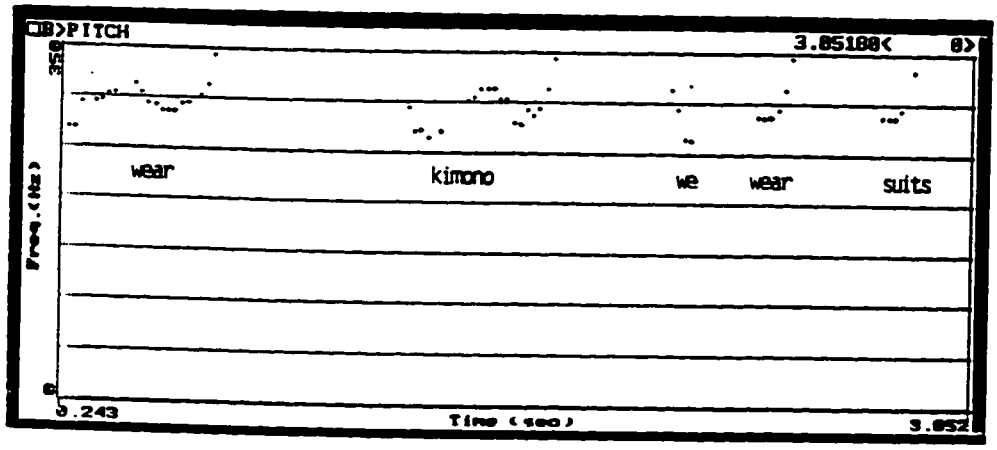


FIGURE 5.6 The contrast between *kimono* and *suits* is obscured since every word has high pitch.

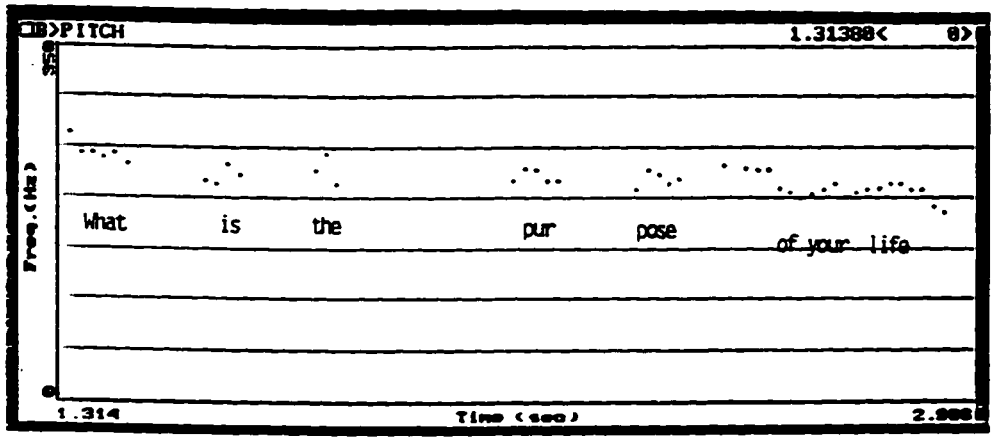


FIGURE 5.7 Information structure is obscured when all the words have equal pitch.

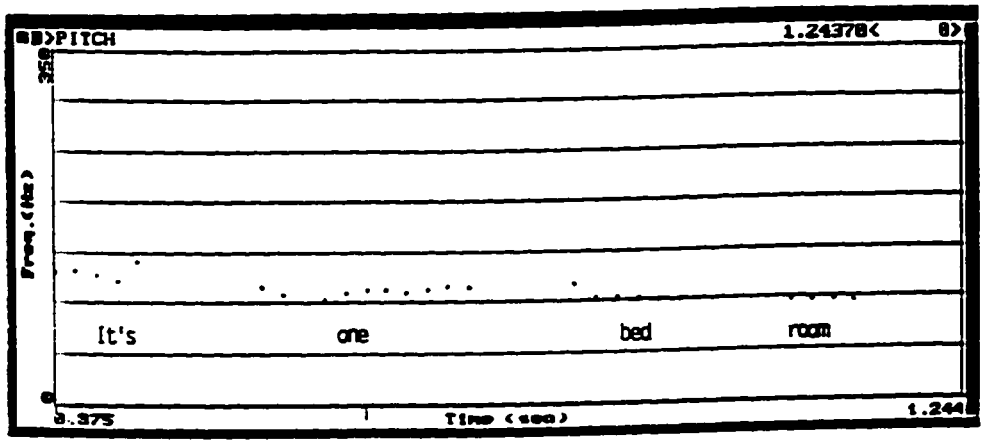


FIGURE 5.8 The focus *one* is not intonationally distinguished.

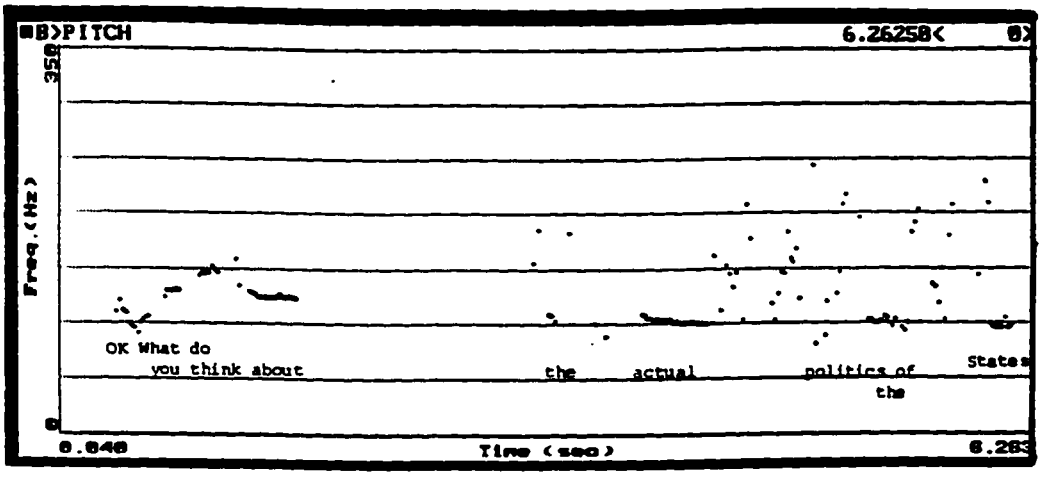


FIGURE 5.9 The speaker holds the floor by extending the words with HL% boundaries.

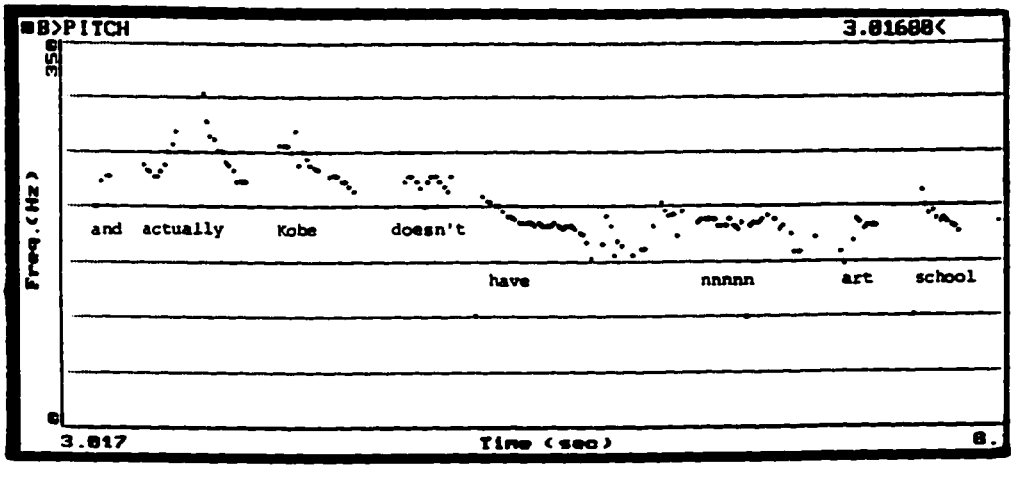


FIGURE 5.10 The speaker holds the turn with a HL% boundary on a pause filler.

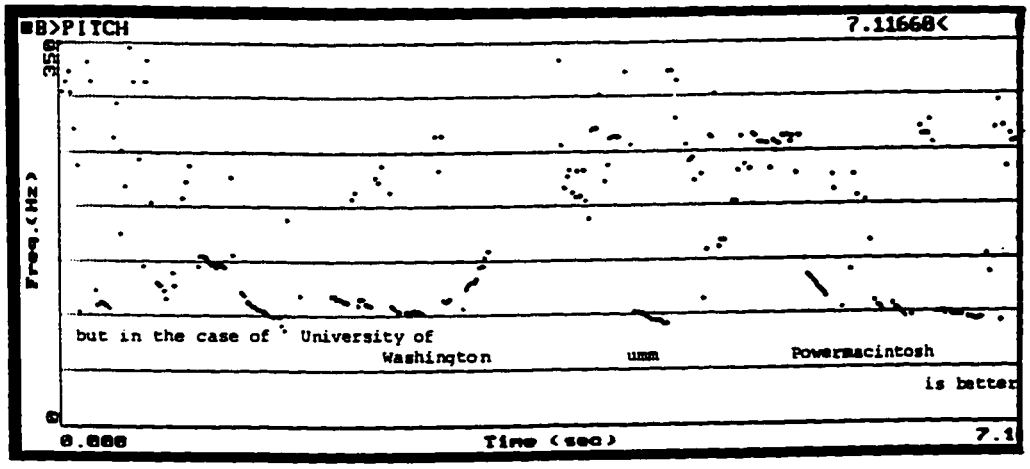


FIGURE 5.11 The speaker uses a LH% boundary on *Washington* to indicate that a related constituent will complete the utterance.

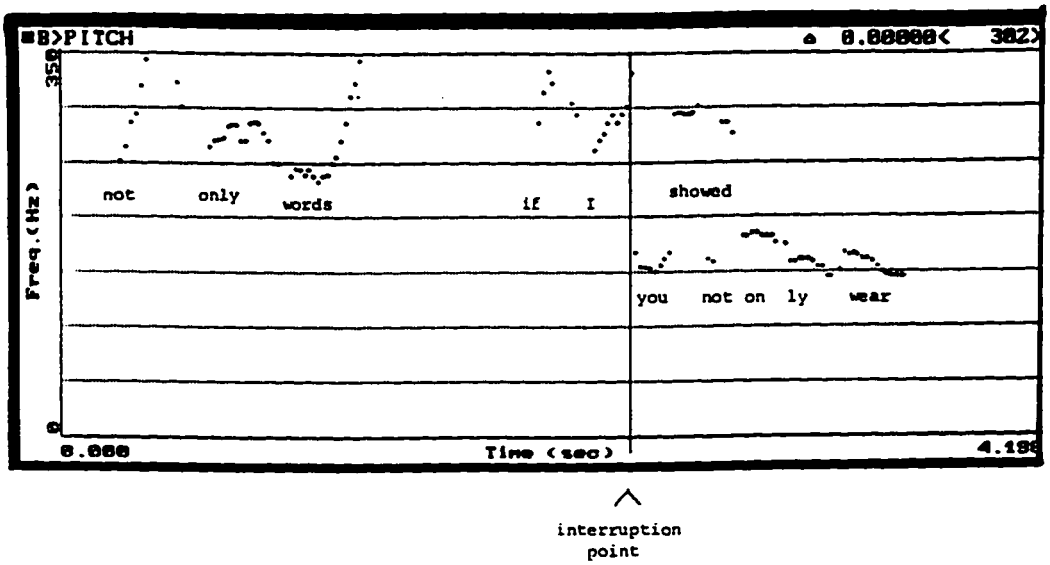


FIGURE 5.12 LH% boundaries after almost every word leads the NS to interrupt.

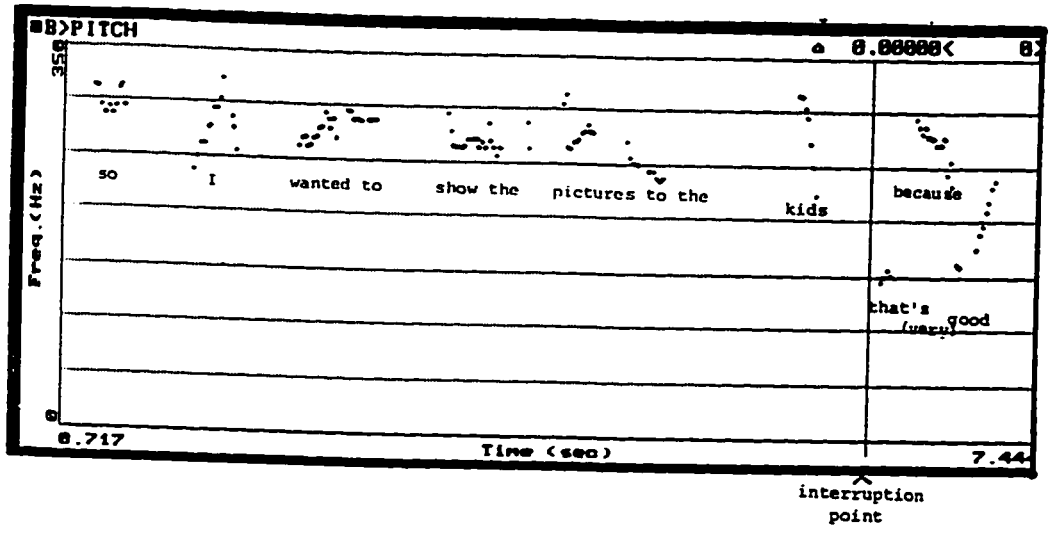


FIGURE 5.13 The NNS is interrupted after a LL% boundary on *kids*.

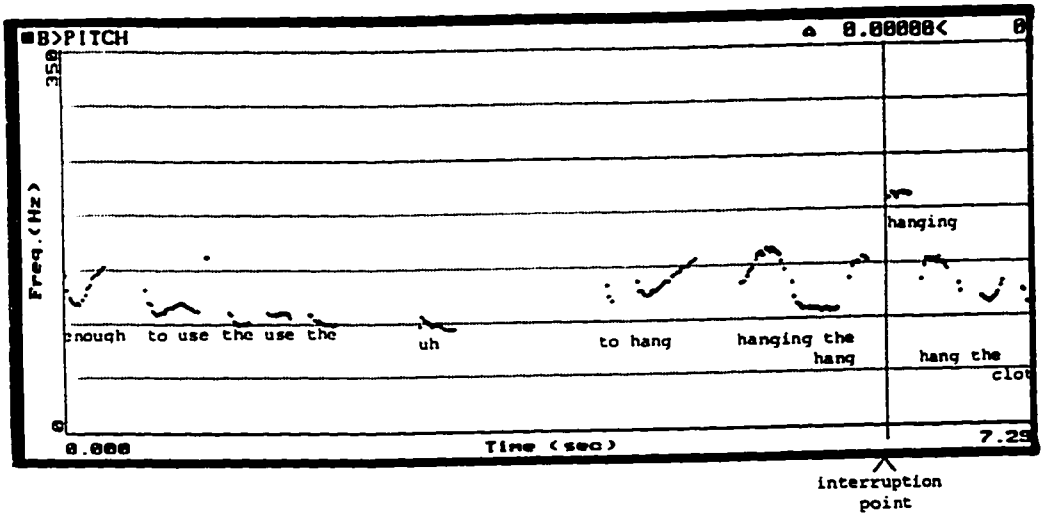


FIGURE 5.14 The speaker's boundaries are cut short before the pauses and the NS breaks in.

Notes on Chapter 5

1. I do not mean to say that fluency is the only role intonation plays.

2. Several speakers' scores were clustered around 2 so this cutoff was somewhat arbitrary.

3. Ranges reflect more than just the pitch measurements of the individual words, but also include rises and falls that occur at boundaries. Also, since individual words are measured at the syllable center, higher or lower pitch within the same word to the left or right of the measurement location may be reflected in the range. Therefore, the reader will see ranges that differ from a subtraction of maxima and minima available in the subscripts.

Chapter 6

Study III: Oral Narratives

In this chapter I consider the role of intonation in oral narratives told by both L1 and L2 speakers of English. Of interest is the contribution intonation makes to cohesion over extended stretches of text by a single speaker. Gumperz, Kaltman, and O'Connor (1984) cite the importance of prosody to the coherence of oral narratives, noting that prosodic features not only help keep the theme of the narrative in focus, but also provide information about the segmentation of the narrative and the time sequence[1]. Indeed, the data in the current study confirm the association of intonation with plot and deictic structure, while revealing additional functions as well. As will be shown, intonation also plays a key role in the teller's dramatization of the story, the evaluation of the story events, and the maintenance of the floor as story-teller. By and large, both NSs and NNSs of English used intonation for the purposes outlined above, although some minor differences did appear in the distribution of pitch maxima, especially in the areas of structural shifts and quoted speech.

6.1 Theoretical Considerations

6.1.1 Pitch Extremes and Global Cohesion

In previous chapters I have discussed intonation in terms of morphemic structure, defending the position that there is an English-specific system of pitch accents and boundary tones whose placement coordinates with discourse structure. Although the question of structure will be addressed in this chapter, I will mainly consider the role of intonation in expressing emotion and attitude. As discussed in Chapter 2, a speaker's placement of a particular pitch accent or boundary tone in a phrase is a separate issue from the relative degree of pitch height he or she uses. Therefore, in the current study, the differences between NS and NNS intonation are expected to be of a different nature than in the other two studies of this dissertation, which involve more local questions of pitch placement. There is reason to expect that NSs and NNSs alike would highlight certain key parts of the text with their highest pitch for emotional and attitudinal reasons, while structural use of intonation might show more NS/NNS differences.

The narrative genre is especially well suited for the investigation of the role of intonational extremes since narratives are likely to include emotional influences, such as dramatizations, sound effects, and other features which indicate the teller's own stance toward the story being told (Labov, 1972; Labov & Waletsky, 1967; Polyani, 1985; Toolan, 1988; Wolfson, 1982). I assume that intonation provides a guide to the level of emotional involvement of the speaker

with respect to the text, which listeners are intuitively able to follow. My investigation is focused on how these extreme pitch maxima map to the story at the global level rather than on how pitch links to lexical items from phrase to phrase at the local level. In other words, given that every intonational phrase has a pitch accent, where are the very highest pitch accents in the story, and what sort of global cohesion does this provide?

The research questions for this exploratory study, then, are the following: How do speakers associate extreme pitch maxima with story texts? Are there structural reasons for this association? Are there emotional/attitudinal motivations for this association? And are there differences between NSs and NNSs?

6.1.2 Structure in Narrative

As a convenient tool to analyze narratives I refer to Labov and Waletsky (1967) and Labov (1972), who provide a model of six components commonly found in oral narratives:

ABSTRACT:	an announcement or summary of the narrative about to be told.
ORIENTATION:	the setting and/or background for the narrative.
COMPLICATING ACTIONS:	the events of the narrative's plot.
RESOLUTION:	the final result of the narrative; the ending.

CODA: a final segment which links the narrative back to the present interaction.

EVALUATION: the teller's assessment of the narrative events.

In this section, I will discuss the interaction of intonation with the first five of these components, which refer mainly to the structure of the narrative. In Section 6.1.3, I discuss the Evaluation category which refers mainly to the teller's attitude. Finally, I include a brief section on interactive functions of intonation.

In considering how intonation maps to Labov and Waletsky's components, I hypothesize that paratone effects would occur as a speaker shifts from one component of narrative to the next. For example, as a story-teller moves from the orientation of the tale to the complicating action component, he or she makes a rhetorical shift, and would be expected to mark this shift with an increase in pitch range at the beginning of the new component.

We can further consider the substructure within Labov and Waletsky's components. As Young (1991) notes, Labov's stories are fairly simple, whereas his own data have much more complex episodic structure. Young finds that in his own narratives, Labov and Waletsky's components tend to recur for each episode, beginning with a new orientation and ending with an evaluation. Although the data herein do not fall out quite so

neatly, we will see that episodic shifts, usually deictic shifts, tend to be associated with pitch maxima.

Other than Gumperz et al.'s (1984) general findings on the importance of prosody in discourse segmentation, work on formal differences among narrative components has largely centered on syntax. Labov (1972) discusses the use of simple past tense in complicating actions, and Wolfson (1982), the shift into the historical present at episodic breaks in American English. Longacre (1983) provides evidence from several diverse languages of grammatical devices, such as verb tenses, or lexical particles, that routinely appear in certain components of narratives to distinguish the complicating actions from background or other material. My hypothesis regarding narrative structure is simply that intonation serves as an additional mechanism to signal componential or episodic shifts.

6.1.3 Evaluation in Narratives

Many scholars have discussed the importance of evaluation in narrative as a way of turning a mere series of events into a story that reflects the teller's personal and cultural values and point of view (Labov & Waletsky, 1967; Labov, 1972; Linde, 1993; Mishler, 1986; Polyani, 1985; Tannen, 1984b; Toolan, 1988; Wolfson, 1982; Young, 1991). Polyani (1985) illustrates this aptly by stripping one of Labov's story-teller's highly evaluative narratives down to a one sentence

paraphrase of the plot. The story "The Baddest Girl in the Neighborhood," which many will be familiar with from Labov (1972) (the full story is reprinted in the Chapter Notes[2]), is paraphrased by Polyani as follows:

> The narrator hit "the baddest girl in the neighborhood" because he could not tolerate being bullied for having no money (p. 26).

Polyani arrives at this rendition of the story by extracting the propositional content of the main clauses of the complicating actions, and selecting those which are most heavily evaluated to be paraphrased. The point of this exercise is to show that this skeletal plot could be recast into a wide variety of different stories depending on what the narrator wished to emphasize. The bare facts do not in themselves make much of a story.

Evaluation can take a variety of forms. Labov (1972) identifies both external and internal types of evaluations as ways for the teller to manipulate story material to make a point. He defines external evaluations as separate clauses from the actual story line in which the teller lets the hearer know his stance toward the events in progress. For example, from a woman's story about a frightening airplane trip, Labov cites the following as external evaluations:

- and it was the strangest feeling because you couldn't tell if they were really going to make it

  (material omitted)

- But it was really quite terrific
- But it was quite an experience" (p. 371).

In contrast, internal evaluations are defined as grammatical, lexical, and phonological mechanisms embedded within the clauses of the story events themselves which indicate the teller's perspective. For example, tellers can use superlatives, expletives, highly charged lexical items, sound effects, and the like to add a special evaluative status. Although Labov does not discuss intonation directly, I submit that high-pitched words act as what Labov refers to as "intensifiers", signalling to the hearer those parts of the story that are considered particularly salient. Since pitch is carried simultaneously with lexical content, intonation affords a speaker the possibility of adding intensity to an internal story clause to render it evaluative, or of intensifying an external evaluation even further.

One particular type of internal evaluation, at the top of Labov's list of intensification mechanisms, is quotation. By quoting another's words, the teller tries to make the story world in the hearer's mental representation more realistic. It is not always expected that quotations represent the exact words originally spoken; their purpose is often to combine story action with a strong evaluative function (Young, p. 45). Wolfson (1982) also discusses quotation as an evaluation device, enabling the hearer to see through the eyes of the teller and thereby better support his moral judgment. She

refers to this type of dramatization in story-telling as a "performance feature", along with other options such as gestures or sound effects. Further support for the link between quotation and evaluation comes from Bauman (1986). In a study of humorous anecdotes told by rural West Texans, he found that the crucial punch line of a story was often rendered through a direct quotation.

Thus we have two reasons to expect pitch changes to associate with quotations: first, the highly evaluative role that quoted speech often assumes will mean that it is emotionally charged; and second, the theatrical, mimetic quality of quoted speech will trigger an altered pitch range. Bauman provides us with a third reason for the association of pitch shift and quotation. From a practical standpoint, pitch changes are one mechanism to assist in keeping track of who is speaking. In Bauman's words:

> Reported speech, especially quoted speech, involves special problems of communicative management, because the narrater is actually speaking for other people in addition to himself. Accordingly, there is a need for ways of marking the difference between the voice of the narrator in the present storytelling context and the reported speech of the action in the original event being reported (one of whom can be the person who later tells the story, but in a different voice), and of marking speaker

> change within the conversational dialogue that is the core of the narrated event (p. 66).

To summarize, given the widely recognized importance of evaluation in narrative, its emotional and often dramatic nature, and its frequent reliance on quoted speech as an intensification device, it is predicted that intonation will play a role in evaluation: the highest pitched words will tend to cluster at highly evaluative points in narratives. Since this category is emotional and attitudinal rather than structural, the association between evaluation and high pitch is predicted to be universal rather than language-specific.

6.1.4 Interaction in Narratives

Labov and Waletsky's model has been critiqued for underestimating the interactional aspects of oral narrative. As Jefferson (1978) points out, other participants' responses and expectations directly and locally influence the direction a story may go even as it is told. Mishler (1986) suggests that even in elicited, largely monologic narratives such as Labov's, the effect of an interviewer is felt. Mishler states:

> . . . the interviewer's presence and form of involvement -- how he or she listens, attends, encourages, interrupts, digresses, initiates topics, and terminates responses -- is integral to a respondent's account. It is in this specific

sense that a "story" is a joint production (p. 82). Accordingly, a third area of investigation in this study will be to consider the alignment between intonation and interactional features. As in Chapter 5, several of the interactional features from the conversation analysis tradition (Couper-Kuhlen & Selting, 1996; French & Local, 1986; Sacks, Schegloff & Jefferson, 1974) will be considered. Of interest for investigation will be the role of intonation in floor-maintenance, repairs, and reactions to co-participants' contributions, such as interruptions and backchannels. The stories in the current corpus include varying amounts of such interaction "at the edges" of the narratives: beginnings, ends, tangents, and responses to others' moves. Overall, however, these narratives are mainly monologic, having many fewer actual turn shifts than the conversation genre. Therefore, in the analysis, the interaction category is somewhat smaller than the others.

6.2 Methodology

6.2.1 Narrative Data

Narratives from eight NSs and eight NNSs were taped and transcribed for study. Six of the NS stories were collected in a graduate seminar on discourse analysis in which students in the class were asked to tell about a mistake their parents had made in raising them. These stories were told in small groups while being tape recorded. The purpose of the

assignment was to generate data in order for students to learn about different methods of discourse analysis. The final two NSs narratives were taken from a naturally occurring conversation among a group of five friends.[3]

The NNS stories were collected in an advanced ESL conversation class by the students themselves. Volunteers included 7 Japanese students and 1 Korean student who were assigned to tell a story from their own life that was either embarrassing or scary. They told these stories to the class and finally submitted them to the instructor on cassette tapes. The purpose of this assignment was for students to practice speaking English for an extended period of time and to become more aware of the role of story-telling in conversation. These tapes were rated for level of English by two SPEAK Test raters in a process similar to that of the conversation study in Chapter 5. The score used this time was the "comprehensibility" score as a global measure of English ability. Raters listened to each story and assigned a holistic score according to the levels described by ETS as follows (1985, p. 6):

0 = unintelligible except for short phrases; necessitates asking for repetitions and strained listening

1 = generally comprehensible but with frequent errors; concentrated listening is required most of the time

2 = comprehensible despite certain systematic errors;

concentrated listening is not required

3 = effortlessly intelligible despite isolated occurrences of errors

All speakers scored between 1.8 and 2.4 on the test, indicating an intermediate to advanced level of spoken English.

6.2.2 Measurements

Measurements of pitch were made on a CSL machine to calculate each speaker's range in Hz, and to determine the location of pitch maxima. An early discovery in the data which influenced further methodological decisions was that all speakers had a tendency to speak within a certain comfortable range which they exceeded only occasionally. I will refer to this as the "normal range", reflecting a speaker's range within the majority of phrases. Since the main focus of this study is the association between pitch extremes and text, those words whose pitch exceeded the normal range were of particular interest. Although the normal range was originally identified somewhat impressionistically, word counts subsequently revealed that roughly 10% of the total words in the narrative had pitch peaks which lay outside of this range for NSs. Therefore, to obtain a consistent measure of pitch maxima across all subjects, the highest pitched 10% of all the words in each narrative were selected for study. These words were then categorized as to their function in the narrative,

as will be described in detail in the next section. In the examples given in this chapter, the normal range is given along with each maximum in order to provide the reader with a basis on which to judge the relative extremity of outstandingly high-pitched words. In addition to the top 10% of the speakers' ranges, a second measurement was taken of the top 3 pitch peaks for each speaker. This was done in order to determine where "the extremes of the extremes" were associated: for what aspects of the narratives did speakers reserve the absolute top of their pitch ranges? Since stories ranged from 160 to 700 words, three words represents roughly one-half to two percent of the total.

To summarize, connections between intonation and the discourse of oral narratives were analyzed in three areas: structural, attitudinal/emotional, and interactional. Moreover, with regard to NS/NNS differences, because the study assesses the role of pitch extremes rather than phrase-by-phrase pitch placement, the focus is as much on commonalities among speakers regardless of language background as it is on language-based differences. Finally, since the number of subjects is small and the data quite variable due to differing conversational styles and topics, I hope to maintain the spirit of a case study, suggesting new directions for research and supporting previous findings.

6.3 Results

In the following sections, intonation is discussed in detail as it functions in the areas of structure, evaluation, and interaction. Similarities and differences between NS and NNS are discussed and examples are given from the narratives themselves.

"Pitch maxima" refers to those 10% of the words in the text which have the highest pitch in a speaker's range. In the examples, these maxima are represented with a superscripted value in Hz, indicating the actual pitch. These are assumed to represent something extraordinary: the top emphasis, the key content, or the highest emotional charge. The reader is encouraged to "hear" these as high-pitched. Words with no superscript can be assumed to be within the speaker's normal range, which is given in parentheses at the beginning of each example. The gender and language group (NS=native speaker; NNS=nonnative speaker) is also given for each example.

6.3.1 Narrative Structure and High Pitch

There are many examples of paratones occurring at the boundaries between Labov and Waletsky's categories at the points where a new component begins. Although intonation performed this segmentation function to some extent for all speakers, it occurred more regularly among NSs. To obtain an overall measure of what percentage of the pitch maxima were

used in this structural manner, a tabulation was done of pitch maxima at componential shifts, at smaller episodic shifts, and on any other words related to a deictic shift. This resulted in a value for each speaker which indicated the percentage of total pitch maxima allotted to the structural function. Averages were then taken for each language group, and a t-test showed a significant difference between these means, as shown in Table 6.1:

Table 6.1: Percentage of Pitch Maxima for Structural Purposes

	NS Mean (SD)	NNS Mean (SD)	p
Structural Function	24.1% (9.1%)	13.6% (8.1%)	.028*

*p < .05

This indicates roughly a 10% higher allotment of pitch maxima to the structural function of intonation on the part of NSs. This is a modest difference, but in some cases the missing intonational marker for NNSs was conspicuous, especially when it involved the transition to the beginning of the complicating action, after all the introductory and background material had been completed. In the NS data, all eight speakers had pitch maxima at the initial boundary of this crucial transition, whereas only five of the NNSs did. The following are two examples of pitch maxima associated with major transitions from orientations to complicating actions. The abbreviations in brackets indicate the initial boundary of

each component (O=orientation; CA=complicating action) underlined material is to be discussed in the text:

1 6.1 (NS#2, male, 105-170) [O]Uh, they were just
2 really[237] strict and very traditional and there was
3 this big element of respect[172]. And, y'know, you
4 followed the rules of the home and you followed the
5 rules of your parents?[176] [CA]And then[188] when I
6 turned about sixteen[204], suddenly those rules were
7 lifted[204] and I no longer had to follow these really
8 strict and disciplinary type of rules . . .

In this first excerpt, the teller gives an orientation describing his parents' traditional attitude. Then in line 5 comes a shift to the complicating action component when the teller turns sixteen. The transition is shown in Figure 6.1. We notice the high-pitched words related to time, *then*, *sixteen*, and *lifted* which give additional thrust to the transition, a paratone effect. The second example is given by a NNS:

1 6.2 (NNS#4, male, 85-152) [O]And, ski[159] ground has a
2 parking[197] lot below there. It was really- there was
3 really tiny ski park? [CA]And we[154] downhilled[157] and
4 I won. But I forgot to stop at the bottom.

In line 3, the high-pitched material is located at the transition between the orientation and the complicating action.

Next we turn to an example of a NNS who did not mark the

major transition from orientation to complicating action with an pitch maximum:

1 6.3 (NNS#7, female, 170-285) [O]It was long time ago
2 when I was a high school student. [CA]I and my
3 best friend decided to eat out for some spaghetti[294],
4 and I decided to make phone call for my mother.
5 And I automatically dialed my best friend's phone
6 number [O]because I used to call her nearly
7 everyday, [CA]and I didn't notice my mistake. So,
8 I started to talk with her mother, [O]and her
9 mother and my best friend and two of them some time
10 had a bad arguments[312], [CA]and this time I didn't
11 notice she was her mother. I thought she was my
12 mother[287]. And she[286] thought I was her daughter ...

This story has no abstract. It begins in line 1 with an orientation, all delivered within the teller's normal range. In line 2, the teller introduces the first complicating action, recognizable by the past tense verb *decided* in line 3, but there is no pitch maximum until the end of the second intonational phrase of the complicating action on the word *spaghetti*. What follows are interspersed clauses of orientation and complicating actions, none of which are marked by pitch maxima at their initial boundaries. The omission is most conspicuous in line 10 where there is a specific deictic marker (*this time*), but no pitch maximum occurs until the end of the next sentence on *mother*.

The data also included many instances of pitch maxima at episodic, deictic, or other more minor structural shifts within one of Labov and Waletsky's larger components. Examples such as the following, where the %H symbol is used to indicate a structural environment for a minor paratone, were in this category:

1 6.4 (NS#3, male, 87-114) It was a very[124] frightening[125]
2 film. It had (.) very[122] graphic[123] images of how
3 horrible the world was going to be under the
4 Antichrist; how anyone who refused to submit to the
5 (.) the rule of the Antichrist was going to be
6 horribly persecuted. Had all sorts of oppressive
7 helicopters [%H]and uh also[128] it had a long a long[115]
8 bit about the rapture and how when (.) how at some[123]
9 certain[130] point in in very[118] near future history
10 they were they were telling us all all saved people
11 were gonna be- were gonna disappear be- they were
12 gonna go to heaven and everyone else was gonna be,
13 was gonna stay behind . . .

This speaker's story is about seeing a film as a child and how it frightened him and otherwise affected his life. This excerpt describes a film world embedded within a story world. In line 7, the word *also* marks the shift to a new installment in the teller's description of the film world, associated with a paratone.

Here is another example of a deictic shift, this time

from a NNS:

1 6.5 (NNS#5, female, 160-237) And, um, [%]at that
2 night[263] I put[249] a bottle of my favorite juice[252] in
3 the refrigerator, [%]and the next[262] morning[263] I woke
4 up[238], and I went to refrigerator- refrigerator to
5 get my favorite juice . . .

This speaker uses high pitch twice at time shifts: first, in line 2 on *night*, and second in line 3 on *next morning*. Both are embedded within the complicating action component of her narrative.

To summarize this section on the relationship between intonation and the structure of narratives, a small language-based difference was found in how pitch maxima were used at the boundaries of components and of embedded substructures to make transitions: NSs allotted a higher percentage of their pitch maxima to this function than did NNSs. This result is in line with the findings of Chapter 4, where a structural use of intonation (paratones) to support topic transitions in lectures led to better scores on an English test.

6.3.2 Evaluation and High Pitch

This section begins with a few summary statistics on the distribution of pitch maxima among evaluative portions of the narratives. Thereafter, examples are given of both external and internal evaluations as they associate with pitch extremes. A special subsection is also included on quotation

since this form of internal evaluation was found to be so consistently high-pitched that it deserves its own commentary. Finally, the section ends with examples of utterances where high pitch alone appears to provide an evaluative, intensifying element.

Table 6.2 shows the percentages of total pitch maxima associated with internal and external evaluations which were identified in the text by their lexical or syntactic form, independent of anything phonological. In other words, this table, and the next three as well, reflect the cases where there was a clustering effect of both a pitch peak and some other textually manifest evaluative element.

Table 6.2: Percent of Pitch Maxima Associated with Evaluation

	NS Means (SD)	NNS Means (SD)
External	23.3% (15.9)	18.9% (18.0)
Internal	28.9% (10.5)	37.6% (18.0)
Total Evaluative	52.2%	56.5%

This table indicates that a high percentage (over half for both language groups) of pitch maxima were associated with text which was coded as evaluative by Labov's (1972) criteria.

These averages go up when we consider only the top three highest-pitched words in each narrative. This second tabulation is shown in Table 6.3.

Table 6.3: Three Highest Pitched Words and Evaluation

	NS Top 3 Maxima (n=24 words)	NNS Top 3 Maxima (n=24 words)
External	25% (n=6)	8% (n=2)
Internal	46% (n=11)	63% (n=15)
Total Evaluative	71% (n=17)	71% (n=17)

This table can be interpreted to mean that speakers associated the top three highest pitched words with an evaluative aspect of their narratives in 71% of cases for both groups. NNSs appear to have a higher percentage associated with internal evaluations, while NSs appear to have more external evaluations. This is due to the high use among NNSs of quoted speech which is one category of internal evaluation. Tables 6.4 and 6.5 show the types of internal evaluative devices preferred by each language group for the top 10% of words and the top three words, respectively.

Table 6.4: 10% Pitch Maxima and Internal Evaluations

Types of Internal Evaluation	NS Means (SD) (% of all internal evals)	NNS Means(SD) (% of all internal evals)
Quotations	26.5% (35.0)	55.6% (33.1)
Lexical Devices	57.1% (30.9)	32.9% (24.3)
Syntactic Devices	15.0% (14.4)	11.5% (17.0)

Here it appears that, on average, the NNSs were more likely to associate pitch maxima with quotations than were NSs. In the next table, the same breakdown is shown for the top three maxima only, when those fell within the internal evaluation category.

Table 6.5: Three Highest Pitched Words and Internal Evaluation

Types of Internal Evaluation	NS top 3 maxima (n=11)	NNS top 3 maxima (n=15)
Quotations	64% (n=7)	88% (n=13)
Lexical Devices	27% (n=3)	0% (n=0)
Syntactic Devices	9% (n=1)	13% (n=2)

This table makes it clear that for both groups, within internal evaluations, the top three words in the narrative are frequently reserved for quoted speech.

To synthesize the information in these tables, it may be said that both groups of speakers had a high degree of association between pitch maxima and evaluative devices, and that this was even more true when only the highest three items in the speakers' narratives were considered. This result is in line with the thesis of this chapter, which is that pitch extremes universally reflect emotion and attitude, and that therefore, they are expected to associate with evaluative text, regardless of a speaker's language background.

Nevertheless, among the types of evaluative devices associated with high pitch, some different tendencies could be seen between language groups: NSs showed a higher association with external evaluations, and among internal evaluations, NNSs relied more heavily on quoted speech in comparison with NSs who used more lexical devices. In the next sections, the various types of evaluation are discussed in more detail and examples of each are given.

6.3.2.1 External Evaluation

To review, external evaluations are separate utterances in which the teller steps out of the main story line to express an attitude or assessment of the story events. As was shown in the previous tables, NSs associated 23.3% of their pitch maxima with external evaluations, and NNSs 18.9%. These percentages are quite high, considering that external evaluations are asides from the actual story and therefore not an especially large component of the text as a whole. The following examples show speakers from both groups giving external evaluations. The point to notice about the relationship between pitch and evaluation is that there tends to be a clustering of evaluative devices. The high pitch adds intensification to text that is already rendered evaluative by virtue of lexical, syntactic, or other discourse features. In the first example, a NS discusses his relationship with his parents:

1 6.6 (NS#1, male, 105-125) . . . Uh, that[145] arrest was
2 very- it- it opened a completely[139] new[127] channel of
3 communication between me and my parents (an') I
4 think[132] they realized that they couldn't just sit
5 back and think that I was uh- as good as they
6 thought I was.

This evaluative material comes at the end of a tale of bad behavior leading to an arrest, after which his parents finally take the teller more seriously. In line 1, he singles out the event of the arrest as worthy of additional comment with high pitch on the cohesive marker *that*. In line 2, he uses high pitch to emphasize the special nature of the subsequent relationship with his parents on the words *completely new*, which already have a lexically strong evaluative meaning. The pitch peak on *think* in line 4 is also evaluative, being a verb which projects the teller's opinion. This is discussed further under the section on projection verbs.

Another interesting aspect of this speaker's narrative is that his normal range is quite narrow and his intonation is fairly flat. Therefore the evaluative items stand out all the more, in view of the fact that many of the complicating actions of the narrative are delivered in a deadpan style with no particular pitch distinction. For example, earlier in the same narrative he introduces the arrest:

1 6.7 (NS#1, male, 105-125) . . . Uh, I remember[150] my
2 getting- having- my parents having complete trust

3 in me an' an- as a result I became progressively
4 more um demonic (hh) or what(hh)ever (others laugh)
5 uh, 'ts <u>almost</u>[143] to the- <u>well</u>[139] to the <u>point</u>[149] where
6 I was arrested at 'bout eight- no[161] it wasn't[137].
7 Let's see I was not- I was still a juvenile.

The words in superscripts, *remember* in line 1, and *almost, well,* and *point* in line 5 are pitched well above the speaker's normal range. In line 6, however, the complicating action verb *arrested* is uttered within the normal range at 120 Hz. A segment of this is shown in Figure 6.2 where the curser indicates the center of the word *arrested.* Thus, the intonation serves to emphasize the evaluation more than it does the event itself, which is rather understated. The next set of examples come from NNSs:

1 6.8 (NNS#7, female, 170-285) . . . and she was <u>so</u>[285]
2 surprised and she was <u>so</u>[311] embarrassed.
3 6.9 (NNS#8, female, 190-220) . . . so it made me
4 <u>really</u>[278] embarrassed and ashamed

Both of these examples show speakers associating high pitch with external evaluation statements of their and others' emotional reactions to the story events. Again, a clustering of evaluative devices occurs, since the text itself in these excerpts is evaluative, containing lexical items referring to emotions such as *ashamed, embarrassed,* and *surprised.* In the last example, also shown in Figure 6.3, the teller associates the highest pitch with the intensifying lexical item *really.*

One syntactic structure commonly found to introduce an external evaluation was the projection verb (see Halliday, 1994), such as *think, guess, know, believe,* etc. which often associated with extremely high pitch in this corpus. Projection verbs provide a way for a speaker to express the degree of commitment to the certainty of what is projected. In some cases, this has the effect of a hedging device, as when the teller is less than completely certain of his evaluation will be acceptable to the audience.

Speaker NS#1 has three pitch peaks associated with projection verbs: *Think* as we saw in line 4 of example 6.6, *remember* as shown above in line 1 of example 6.7, and *think* in the teller's introduction to his entire narrative as follows:

1 6.10 (NS#1, male, 105-125) My parents- I don't know.
2 My p- I think[167] the mistake my parents made was they
3 thought I was too good of a kid.

Here the word *think* in line 2 is the second highest pitched word of the speaker's entire story. Although part of the function of this high pitch is probably interactive: to extract the story from the preceding false starts and get back on track, it simultaneously conveys evaluation.

Next we see another example of a projection verb with a hedging effect in which the speaker ends a story about his parents' mistake with a coda introduced as follows:

1 6.11 (NS#2, male, 105-175) So I guess[188] that's kind of
2 a problem that they- a mistake that they might have

3 made with me.

The verb *guess* projects the coda, which also contains the other hedging words *kind of* and *might*. This word is extremely high pitched, as indicated by the curser in Figure 6.4.

The next excerpt shows a NNS using a projection verb with a high pitch. The story is about how her father drank vinegar by accident, and later denied having done so. She says:

1 6.12 (NNS#5, female, 160-237) So, um, yeah but my
2 father never admit he drank vinegar. But, I
3 believe[264] he drank a vinegar- he drank vinegar at
4 that night.

The entire excerpt is spoken within her normal range except for the word *believe*, shown in Figure 6.5 at the curser. Rather than a hedge, this emphasizes her strong commitment to her certainty about the episode.

Another construction which by its very structure renders the material that follows evaluative is the cleft. Here are examples of clefts used in conjunction with pitch maxima by a NS and a NNS:

1 6.13 (NS#7, male, 85-130) But what[143] I didn't[139] realize
2 at the time was I had always[164] been thinking[152], well
3 all anything[147] alcoholic has been distilled and is
4 OK.

This is a highly evaluative sequence in which the teller explains and critiques his reasons for thinking it would be safe to drink a strange Nepali beverage. It is this error in

judgment that later causes him to become violently ill. A clustering of evaluative devices is evident in the sequence: Pitch maxima associate with the word *what* in the cleft construction, the negative word *didn't*, the projection verb *thinking*, and the two words *anything* and *always* which convey the absolute nature of his mistaken beliefs.

1 6.14 (NNS#5, female, 160-237) . . . and I opened the
2 refrigerator, and what[238] made me surprised was um um
3 there[244] was a um another similar bottle of vinegar[244]
4 um next to my favorite juice[244], and that bottable uh
5 that uh bottle that bottle of vinegar was almost
6 gone!

In this segment, the cleft is initiated in line 2 with a high pitch on *what* within the construction itself. Once again, the teller has used the cleft along with pitch maxima to emphasize her own reactions to a key part of the plot, namely, the point at which she discovers that her father must have drunk vinegar instead of juice in his sleep the previous night.

6.3.2.2 Internal Evaluation

Internal evaluations are similar to external evaluations in expressing the teller's attitude toward the story events, but they occur within the actual story clauses. Labov (1972) discusses a number of devices for internal evaluation. In this section, I consider three of the most commonly used devices to be associated with pitch maxima: loaded lexical

items, syntactic structures, and negatives. Quotations, a fourth device, will be discussed in a separate section. As before, there tends to be a clustering of evaluative devices, with intonation used as an intensifier along with other devices.

LEXICAL ITEMS: One common device among NSs was to choose "loaded" lexical items with an evaluative force. As we saw in Table 6.4, these made up 57.1% of all internal evaluation devices associated with pitch maxima while for NNSs the figure was only 32.9%. Here are some examples of NSs associating pitch maxima with lexical items from various parts of speech:

Adverbs:

6.15 (NS#2, male, 105-175) Uh they were just really[237] strict and very traditional and there was this big element of respect (see Figure 6.6).

6.16 (NS#8, female, 130-170) . . . the guy who brought them back, knew exactly[173]. He and I went out immediately[185] while [man's name] sort of sprawled on the bed, immediately[208] and got three[173] different things that they always use when they're sick . . .

Adjectives:

6.17 #7 (NS#7, male, 85-130) . . . well it's in[147] these big (1.0) vases[147], you know, huge[133].

A: Vats.

B: Yeah. Yeah.

6.18 (NS#3, male, 87-114) I also[118] remember[115] the night

I returned home[115], not being able to (.) not being able to find my father in the house and k- being convinced[120] that he[116] he was gone. That was it. He had been taken up by the rapture and I was left alone . . .

Verbs:

6.19 (NS#8, female, 130-170) Next day, he was escorted[172] home . . .

6.20 (NS#5, female, 135-180) . . . I did[233] this thing[233] where I kind of leaped[286] down[233] the hill[272] . . .

While the adverbs have inherently evaluative meanings (*really, exactly, immediately*) the adjectives and verbs are evaluative by virtue of having been chosen in place of more common words. Examples of this are *huge* (v. big), *convinced* (v. sure), *escorted* (v. brought), *leaped* (v. jumped). NNSs' transcripts did not have as many examples of this type. I would speculate that NNSs were less inclined to use this evaluative device due to smaller vocabulary and/or facility in manipulating lexical meaning. Here are two examples, however, using fairly common words with inherently evaluative meanings:

6.21 (NNS#2, female, 170-300) In the reception, I won a dictionary as a prize. So I went back home with happiness[322].

6.22 (NNS#7, female, 170-285) And then after that I started to be more careful[311] when I make phone call (see Figure 6.7).

Comparators: A lexical category that is inherently evaluative is what Labov refers to as comparators. These include what Halliday and Hasan (1976) call "comparative referents," words that form cohesive links to some previous but opposite referent. Also within this category are included conjunctions which stand outside of the clause while initiating comparisons or contrasts. In the present data, pitch maxima were often associated with these, particularly when they were used in some central contrast, crucial to the outcome of the story. Here are examples from both NSs and NNSs.

1 6.23 (#NS4, female, 145-193) It was a school photo (and
2 my parents actually bought it which was the s-
3 their second mistake) but their first[232] mistake was
4 sending me to school and never combing my hair.

This speaker emphasizes the word *first* in line 3 by contrasting it to its antecedent *second* in the same line. The whole intonation structure serves to assign a higher value to the first mistake than to the second: the second mistake, indicated in parentheses, is actually uttered as an aside with a lowered pitch so that *first* is dramatically higher than *second*. Here is a NNS:

1 6.24 #1: (NNS#1, male, 90-140) The shop clerk[179]
2 showed[147] us the area[182], uh, and when we went to
3 there, we couldn't find an eraser.
4 A: mm hmm
5 #1: Instead[145] of eraser[145]?

6 A: h

7 #1: uh, we found many many condoms.

In this sequence, the word *instead* in line 5 forms a key link, contrasting their expectation of finding erasers to the reality of finding condoms. This is the punch line of the story which is about two young men asking for rubbers (the British term) instead of erasers in an American store.

In the next example, another NNS assigns a pitch maximum to a key comparison in a story about having a nosebleed in a travel agency in front of a handsome travel agent (see Figure 6.8).

1 6.25 (NNS#6, female, 170-240) . . . and what happened

2 next was my- the <u>other</u>[256] side of my nose started to

3 bleed.

The high-pitched word *other* links this event to the previous incident when the first side of her nose started to bleed. This is the climax of her story; she gives up and leaves the travel agency out of embarrassment at this point.

<u>SYNTACTIC STRUCTURES</u>: Labov points out that certain syntactic structures can render the material within clauses of the main story components evaluative. These accounted for only a small portion of the internal evaluation devices associated with pitch maxima: 15% for NSs and 11.5% for NNSs, as was shown in Table 6.4. Here are four examples of speakers manipulating their syntax in various ways to highlight an event or situation, while compounding the evaluative intensity by using

a pitch maximum as well.

1 6.26 (NS#2, male, 105-175) . . . so I made this quick
2 shift from being totally[186] under my parents' guard
3 to being totally not under their guard.

This first excerpt shows the use of parallel word order to emphasize the "night versus day" nature of the change that took place in the teller's life. The high pitch on the value-laden lexical item *totally* heightens the contrast even more.

1 6.27 (NS#5, female, 135-180) I have no idea why, or
2 what happened, but I did[217] go to bed and I cried
3 myself to sleep.

In this example, the teller uses the auxiliary *did*, articulated with a high pitch, as opposed to using the unmarked simple past (I went to bed).

The next example shows a tense shift at a highly evaluative point. In this story, a woman explains how her parents let her have her school picture taken with what she considered to be unkempt hair and later made fun of her for it.

1 6.28 (#NS4, female, 145-193) . . . They sent me to
2 school without my hair combed. And I looked like
3 an orphan. An' um, they bought[195] the picture[212] and
4 bring[196] it out now [others' laughter] during the
5 holidays[212] and stuff so they can show[200] people.
6 Like[204] it was my fault that I looked this way.

The excerpt begins with complicating actions of the story with

verbs in the simple past tense. Then, in line 4, the teller switches to present tense and describes what her parents do now using several high-pitched words. The three clauses -- "they bought the picture and bring it out now during the holidays and stuff so they can show people" -- are made evaluative mainly via the intensification provided by the high pitch and the tense shift; none of the words with superscripts are themselves especially value-laden. That it is her intention to induce incredulity is supported by the others' laughter in line 4: she has won them to her point of view.

Here is an excerpt from a NNS showing another marked syntactic form:

1 6.29 (NNS#7, female, 170-285) And then suddenly[360] I
2 noticed, it's[333] not- it doesn't[316] sounds like my
3 mother. Who[385] is she?

This excerpt shows a question format, which Labov includes in his list of marked, and therefore evaluative, syntactic forms in story-telling. The question is initiated with a high pitch on *who*. The word *doesn't* in line 2 is an example of a negative construction associated with high pitch which is discussed in the next section.

Negatives: One evaluative device used by both groups of speakers and commonly associated with high pitch was negatives. In the tabulations at the beginning of this section, these are coded as syntactic since they usually involve some word order manipulation and/or auxiliary

addition. According to Labov (1972), a negative is a marked construction, or not the norm. The following three examples are from NSs.

6.30 (NS#2, male, 105-175) . . . suddenly those rules were lifted[204] and I no longer[217] had to follow these really strict and disciplinary type of rules[209] ...

6.31 (#NS4, female, 145-193) . . . and neither[217] of my parents will admit actually doing it.

6.32 (NS#5, female, 135-180) . . . and I wanted to um I wanted to play with my friend Pam who could never[205] come out to play.

In the above examples, the underlined negative words all convey that the teller is somehow evaluating the events described as oppositional. NNSs associated negatives with high pitch as well. The following are examples of this:

6.33 (NNS#2, female, 170-300) I looked for the number, but I couldn't[343] find[366] it[354]!

6.34 (NNS#3, female, 185-295) "Oh, where is my shoes! Where are my shoes[304]!" But I couldn't find[303]! (omitted material) It[304] was so embarrassing to me, but I couldn't do anything[313]!

As lines 2 and 4 illustrate, the association of high pitch with a main verb whose inflection was negative was included in this category. In the last line, the word *anything* was counted because it is the negative of doing "something".

6.3.2.3 Quoted Speech

One heavily used category of internal evaluation in this corpus was quoted speech. A rather surprising phenomenon was the fact that speakers consistently raised their pitch in direct quotations, even when quoting themselves or people with deeper voices.

In Tables 6.4 and 6.5, it was shown that NSs and NNSs distributed their pitch maxima differently within the category of internal evaluation. Respectively NSs and NNSs associated 26.5% and 55.6% of their total pitch maxima with quotations. Considering only the top 3 pitch peaks, those statistics were 64% and 88%. Conclusions to be drawn from these numbers are, first, that for both groups, quoted speech was an important internal evaluation device which consistently associated with high pitch, and second, that speakers in both groups were frequently reserving the top 3 pitch peaks of their entire stories for quoted speech; for NNSs this was almost exclusively the case.

One issue to consider with regard to quoted speech is the degree of change in pitch range, especially considering the gender of the story-teller and that of the person quoted. There is no evidence in these data that range shift in quotation was based on an attempt to directly match another speaker's range. Pitch was usually raised regardless of who was quoted. In fact, there was not a single example in the data of a quotation where range was lowered to match that of

another speaker[4]. Instead, it appears that speakers tended to raise their pitch in quotations in order to express intensity of emotion and to create a more vivid story world in hearers' mental representations.[5]

What follows is a series of examples of quoted speech from both NSs and NNSs. The thrust of the argument is that examples of speakers raising their pitch to quote a lower-voiced person count as evidence that it is an emotional, evaluative emphasis that motivates the speaker, not an attempt to accurately represent the original utterance by matching the pitch. In the examples, the reader is encouraged to notice how far many of the Hz values given in the superscripts exceed the normal range given in parentheses.

In the first example, the teller, a woman, makes fun of herself for agreeing to let her husband go camping even though she knew he was sick:

1 6.35 (NS#8, female, 140-170) I knew[244] he was gonna go on
2 this thing an' an' I (.) an' he was really sick but
3 I didn't want to be like (.) the mama and say don't
4 go. "Go[346] ahead[335]! You'll have a good time."

Here the teller hits an extremely high pitch in the quotation in line 4 while remaining in her normal range for most of the rest of the utterance. Clearly this is not a case of imitating herself by matching of range since it is her own normal range that she exceeds.

Next, a Japanese woman quotes herself. This is also

shown in Figure 6.9:

1 6.36 (NNS#3, female, 185-295) . . . so I was not
2 surprised and I said "OK[313], I don't mind."

Similarly, a Korean woman quotes her own thoughts at the time of the story events. This is shown in Figure 6.10:

1 6.37 (NNS#8, female, 190-220) . . . so I thought, "Hmm.
2 Does Super Salad mean huge[286] and big salad?[331]"

The next three examples illustrate speakers quoting their parents. First, a NS woman quotes her mother:

1 6.38 (NS#5, female, 135-180) . . . and the next morning
2 when I went to get up to get out of bed I fell on
3 the floor and my ankle was about as big as my
4 thigh, approximately, and I had to crawl out on all
5 fours and my mom was on the phone[196]. I remember
6 distinctly. And she said, "Well I gotta go, Carol.
7 My daughter just crawled out on all fours[264]."

Again, while most of this sequence is delivered within the normal range, the woman reaches to 84 Hz above that range within the quoted portion. The next example shows a man quoting his parents (given in Figure 6.11 where the curser highlights the word *way*):

1 6.39 (NS#1, male, 105-125) I wrecked[139] the car[130] at age
2 sixteen and they sort of laughed it off and said,
3 "Well, that's the way[164] you[164] learn[154] and grow,
4 son," or something like that.

Here, the quotation stands out above the normal range, being

even higher in pitch than the words associated with the complicating actions of wrecking the car in line 1.

Finally, a Japanese man quotes his father with a higher-pitched voice (shown in Figure 6.12 with the curser at the word *who*):

1 6.40 (NNS#4, male, 85-152) . . . and after that my
2 father came to- came to me[159] and said, "I[172] thought
3 who[175] was the stupid stupid boy[156], but it was you,
4 my son."

In all of these examples, pitch is raised in the quotations, in spite of the fact that we might expect a lowered voice attributed to parental authority figures, especially since the tellers are recounting incidents from childhood.

Among the NS data, there was no case of a woman quoting a man, but we do find this in the NNS data. Here a Japanese woman quotes "a handsome guy" in a London travel agency:

1 6.41 (NNS#6, 170-240) so, um, it was- I didn't[250] get
2 used[250] to seeing such a handsome guy, with blond
3 hair and blue eyes, and to me[258] he looked like a
4 movie[251] star. And he[240] said to me um, "Can I help
5 you?[302]" or something like that.

Here the teller hits 302 Hz in the boundary tone at the end of the quotation of the handsome guy (whose words subsequently induce her to have a nosebleed in the story). This is in fact the maximum pitch level of her entire narrative.

In the final example, another Japanese woman quotes a

male friend, and then herself:

1 6.42 (NNS#3, 175-295) Oh, and one of one of my friends
2 came to the car and asked me, "What happened[343] to
3 you? We[303] are waiting[311] for you. Let's[302] go to
4 the- to the swim[343]." And I said to him, "Oh[303], I
5 can't[317] get out of here because I have no shoes[303].
6 I left[343] my shoes behind . . . ."

Here the pitch is noticeably higher within the quotations for both the man and herself, with no gender difference in evidence. In lines 2, 4, and 6 she hits 343 Hz (her maximum for the whole narrative is 345) on the words *happened* and *swim* from the man, and on the word *left* from herself. The point is that she does not appear to be attempting to replicate the speakers' pitch ranges, but rather to be dramatizing the story with high pitch.

To summarize this section, it is very common to find increased pitch range in quoted speech. This does not necessarily appear to be an attempt to match a particular speaker's range, especially given cases of speakers quoting themselves, women quoting men, and adults representing themselves as children and quoting parents. Instead, as discussed in the introductory sections of this chapter, pitch change functions to create a dramatic effect, enhancing the story world so that the hearer feels more to be "at the scene," and as Wolfson (1982) says, be more empathetic toward the story-teller's evaluation of the events. Because quoted

speech is highly evaluative, singling out the most remarkable of events for dramatization, it is not surprising that story-tellers' speech tends to hit a very high level of pitch at these points. Finally, I speculate that the reason NNSs used this device more frequently than NSs was that it is an easy way to express evaluation without resorting to complicated syntax or vocabulary.

6.3.2.4. High Pitch Alone as Intensification

So far, we have seen situations where pitch maxima intensify evaluative material already manifest in the text. It is my conjecture that in some cases, the intonation itself adds an evaluative element in the absence of any other evaluative devices. For example, in the following excerpt, the teller conveys her excitement at going to a friend's house as a child and being delighted that her friend is free to play. She says (as shown in Figure 6.13):

1 6.43 (NS#5, female, 135-180) So I went[217] over and uh she
2 was at[238] the door and I said, "Ask your mom." and
3 she did and she could come <u>out</u>[348].

The word *out,* not in itself a value-laden word, is the highest-pitched word in the entire narrative (marked with the curser in Figure 6.13). Indeed, the context dictates that this phrase should be heavily evaluated since it sets up the impetus for the entire story: Because of the extreme excitement the teller feels that her friend who is never free

can come out to play this time, she runs and jumps too high, falls, and breaks her ankle. Instead of choosing some other device such as a lexical item -- she could *actually* come out! -- the teller intensifies the phrase with the pitch maximum alone.

The second example comes from a speaker who has just described a terrifying film about the Antichrist he saw as a child. Then he says:

1 6.44 (NS#3, male, 87-114) And uh, I[149] was about eight[116]
2 and this[118] was extraordinarily[139] frightening to me.

In the first clause, the lexemes themselves carry only the information of the teller's age. Were it not for the extremely high pitch peak at the beginning of the phrase, it might seem as if he were merely providing this orientation material as an aside. The high pitch contributes to the idea that the hearer is meant to evaluate this phrase in particular: eight is an impressionable age at which to be viewing such films. The second clause is clearly an external evaluation describing the teller's emotions at the time, where we see the usual clustering of evaluative devices.

To summarize, evaluation has been discussed in the literature as a category that involves expression of tellers' emotions and attitudes toward a narrative in progress. The pitch measurements in this study provide support for this view, since at the top of speakers' ranges, where they are assumed to be at their most expressive, they associate a high

percentage of pitch peaks with evaluative material. In particular, among measurements taken from the top three pitch maxima, there is an even higher association with evaluation. Furthermore, examples show a clustering effect whereby high pitch intensifies material which is already made evaluative through other devices such as marked lexical content, syntactic structure, or quotation. There is little evidence of cross-linguistic differences within this category except for the choices of evaluative devices used, which would vary anyway due to topic and story-telling style. Therefore, I submit that these data support an iconic relationship between pitch extremes and evaluation in story-telling.

6.3.3 Interaction and High Pitch

For many speakers, pitch maxima were associated with text material that linked the narrative to the interaction going on in the present situation: to take the floor; to interact with others present; or to make repairs. Although this category represents but a small percentage of this largely monologic data, it is worth including because such bridging text was consistently associated with pitch maxima. The most consistent use of high pitch for this purpose was to associate with the beginning of the narrative, either within an abstract, or with bridging material connecting the story to the present situation. Very often tellers began their stories with a high pitch on a word that was relatively light in

semantic content. Here are two examples of pitch maxima associated with story beginnings:

1 6.45 (NS#5, 135-180) OK, let's[185] see. My story. Um.
2 I[196] was about, oh gosh, maybe six[232] years[218] old[209]?
3 And I had gone next door . . .
4 6.46 (NS#2, 105-170) Now[174]. As far[197] as my personal
5 narrative and the mistake my parents made with me,
6 I guess[286] it has to do with growing up in a very
7 traditional uh- strongly disciplined family?[174]

These examples show the use of high pitch on words which in themselves do not make especially significant semantic contributions to the stories' content: *let's, now,* and *far.* (6.46 can be seen in Figure 6.4 which was used to illustrate an earlier point.) Thus, a case can be made that this use of high pitch is separate from the association between pitch and information structure; otherwise there would be no reason to associate high pitch with semantically light words. Instead these first phrases form bridges between the previous self-introductions that the tellers have given to the group, and the narratives themselves. I have included this in the interaction category because it is a way for tellers to establish their right to an extended turn of talk from among the others present, with the bridging language augmented by the intonation.

In the next example, the speaker stops his narrative to adjust the tape recorder and then continues (shown in Figure

6.14):

6.47 (#1, 105-125) My p- I think[167] the mistake my parents made was they thought I was too good of a kid. [I think (xxx) shooting way up there. Sorry. I don't know if that's loud enough or not.] Umm. They thought[172] I was too good of a kid. I wrecked[139] the car[130] at age sixteen . . .

In lines 3-4, in the bracketed portion, the speaker is engaged in checking the volume gauge on the tape recorder. All this talk takes place within his normal range. As he begins again, he uses the highest pitch of his entire narrative on the word *thought* at 172 Hz, even though he has already used the same word in line 2 with a much lower pitch, 133 Hz. The sequence presented in Figure 6.14 shows both utterances of *thought*, with the curser through the high one. If he had considered the word contrastive (they THOUGHT I was a good kid, but they were wrong), we would expect to see this contrastive use of pitch accent in the first utterance of the word in line 2, but we do not. It is also unlikely that he was speaking up due to the tape recorder's volume problem because he immediately goes back down into his normal range after the juncture. This suggests that this is a move to reestablish his turn status as narrator rather than an attempt to speak more audibly for the machine's sake.

The next set of examples come from speakers #7 and #8, who, the reader may recall, are engaged in a naturally

occurring conversation among friends. Thus, these two narratives involve much more interaction with others than do the first six. For speaker #7, his highest pitch in the entire narrative comes during an interaction with another conversant, as follows:

1 6.48 #7: (NS#7, male, 85-105). . . they have these
2 big[117] jars full of this (.) mm- (1.0) 'ts like
3 fermented[121] wine er- fermented[118] rice. 'ts
4 like they (.) [y'know ri-
5 A: [Rice wine? (.3)
6 #7: It's <u>kinda</u>[196] like wi-[127] I mean it's it's kinda
7 like wine. It doesn't taste[116] like wine but
8 it's . . .

In line 6 the word *kinda* is far above speaker #7's normal range. This can be explained by the interaction with A who tries to supply a guess in line 5 for the word Speaker #7 seems to be searching for to describe the beverage he drank. He politely rejects the guess as being "kinda" accurate but not really correct. The high pitch, then, serves to contradict A and at the same time reestablish the right to the floor.

The next segment shows the transition point at which speakers #7 and #8 compete to tell the final episode of the story. Both speakers exceed their normal range in this brief negotiation for the floor. The sequence begins with Speaker #7 describing horrific details of his intestinal disease:

6.49 #7: (NS#7, 85-105) Yeah, [amoebas y- (hh-hh) th-
whole thing.
A: [Oh.
#8: (NS#8, female, 130-170) Hm-mm
A: Oh[: [Gh:od. Oh: [man's name].
B: [Oh: [
#7: [It[130] was the [nasty-
#8: [he went[222] on a camping
trip
#7: Yeah, and I didn't[163] [discover[127] it til I was
half-
#8: ['t was[223] like on Friday[193]
#7: [halfway[141] in
#8: I knew[244] [he was gonna go on this thing an'
an' I- an' he was really sick but I didn't
want to be like the mama and say don't go.

After line 13, #7 makes no further attempt to regain the floor. In lines 7 through 14, we see a series of high-pitched words uttered by both speakers. Two points are of interest: the fact that several of these are rather low in semantic weight (*it, went, didn't, was*) would indicate that it was not the information structure alone that influenced the pitch. Also, the fact that speaker #8 goes immediately back into her normal range as soon as she has the floor in lines 14-16 indicates as well that this is a special, turn-management phenomenon. This is consistent with a study by French and

Local (1986) where high pitch was found to be involved in floor maintenance. They found that turn-competitive interruptions by another speaker were met with increased pitch and loudness by the current speaker if he wished to retain the floor, which persisted until the competition was resolved.

The final examples show high pitch associated with repairs.

6.50 (NS#1, male, 106-122) I was arrested at 'bout eight- no[161] it wasn't[137]. Let's see I was not- I was still a juvenile.

6.51 (NS#2, male, 105-175) And of course I abused that um that right, and it led into me getting into a lot of problems[222] and a lot of um, just a lot of disciplin- no- not[207] a lot. But some discipline problems and just totally rebelling I would guess.

These two examples have pitch maxima associated with negative words, but I would distinguish this from the category under internal evaluations where some event of the story is being negatively evaluated. Instead, this is a more local phenomenon whereby tellers correct the immediately preceding assertion. The first speaker's pitch is high on the words *no* and *wasn't*, a self-correction of his statement of his age (18) at the time of his arrest. (This appears in Figure 6.2 as part of a different point.) In the second segment, the speaker is giving an external evaluation, embedded within which is a self-correction where the speaker reassesses the

seriousness of his problems, associating a high pitch with the word *not*.

To generalize, the examples in this section can be said to involve cases where speakers are at the boundaries between their narratives and other types of interaction. Speakers use intonation, sometimes in conjunction with other linguistic devices, to indicate a bridge or transition between what is narrative and what is not. We have seen examples of this in starting the narrative, in coming back to the narrative from a tangent or interruption, in competition for the floor, and in repairs. Evidence that this use of intonation is independent of its role in information structure comes from the location of the pitch maxima (at the boundaries in question), the fact that the words associated with the pitch maxima often did not in themselves carry much semantic weight, and the fact that speakers often reverted to their normal ranges after the interactive sequence.

6.4 Conclusion

This chapter has focused on the role of pitch extremes in oral narratives. Assuming an iconic relationship between pitch extremes and emotion, I have attempted to show that a higher order cohesion network is established by story-tellers beyond the local, clausal level. By following the thread of the very highest pitched words, listeners identify the parts of the text with the highest emotional priority. The three

categories that have emerged in the texts as having the most consistent degree of association with pitch extremes are evaluative, structural, and interactional.

Of these three categories, the evaluative provides the strongest case for the argument that pitch extremes reflect emotion, because it is by definition an attitudinal, emotional category. Also, the tabulations show the most consistent association between pitch maxima and evaluative material, along with the fewest cross-linguistic differences.

In the structural category, the high pitch at transition points may be a manifestation of a more language-specific paratone phenomenon, similar to what was found in the lecture data in Chapter 4. Indeed, there did appear to be some language-based differences in this category. The results in the structural category support the notion that deictic shifts are to be attended to in narratives, as are component and episodic structures.

The third, interactive category is probably the least conclusive in terms of cross-linguistic comparisons, since the NNS data were not collected in such a way as to allow for a thorough investigation of this. Nevertheless, I included the category because for the NSs there is evidence of a systematic connection between pitch maxima and points where the narratives interface with the larger discourse context.

Finally, I would like to reiterate the relationship between this chapter and the dissertation as a whole. There

is no incompatibility between the notion that pitch extremes express a universal, iconic relationship with text, and the view that intonation has a systematic, language-specific structure, along the lines of the model proposed in Chapter 2. The two occur simultaneously. The latter speaks to the question of how pitch aligns with text phonologically, while the former looks at the extremes to determine the speakers' emotional priorities at the global level of the entire text. It is for this reason that the narrative genre provides an ideal setting for this type of study: it is emotionally engaging and long enough for the teller to establish an intricate hierarchy of priorities. Indeed, we have seen that speakers do not distribute their pitch ranges evenly throughout the text. Instead, intonation is used by story-tellers to emphasize the pace and timing of events, their emotional priorities, their values, and their status as social participants.

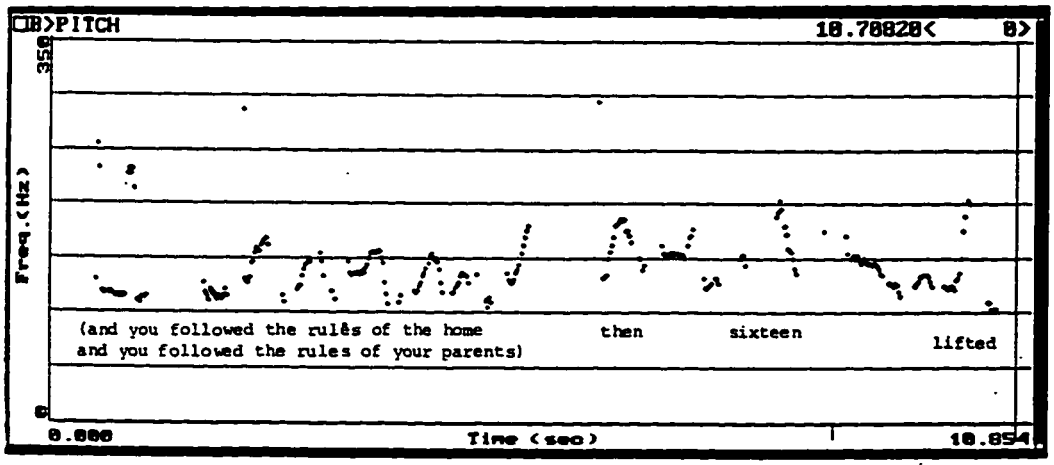


FIGURE 6.1 A paratone occurs at the shift from orientation to complicating action.

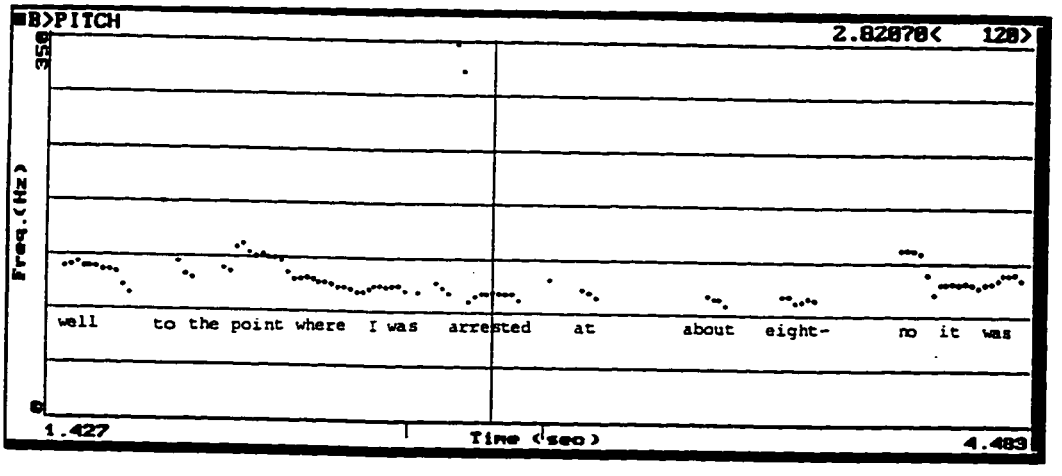


FIGURE 6.2 The action verb *arrested* has lower pitch than the preceding evaluative material.

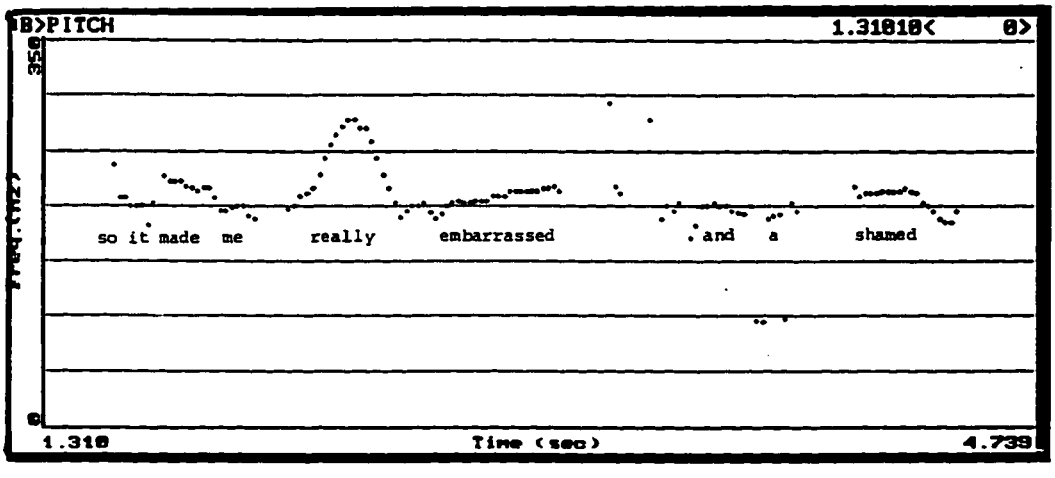


FIGURE 6.3 This external evaluation contains a pitch maximum.

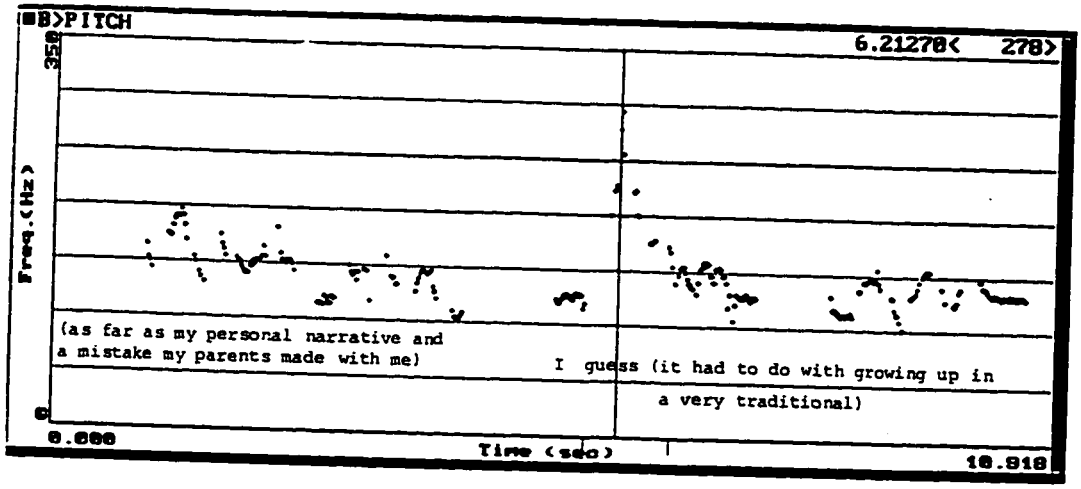


FIGURE 6.4 The evaluative projection verb *guess* has a pitch maximum.

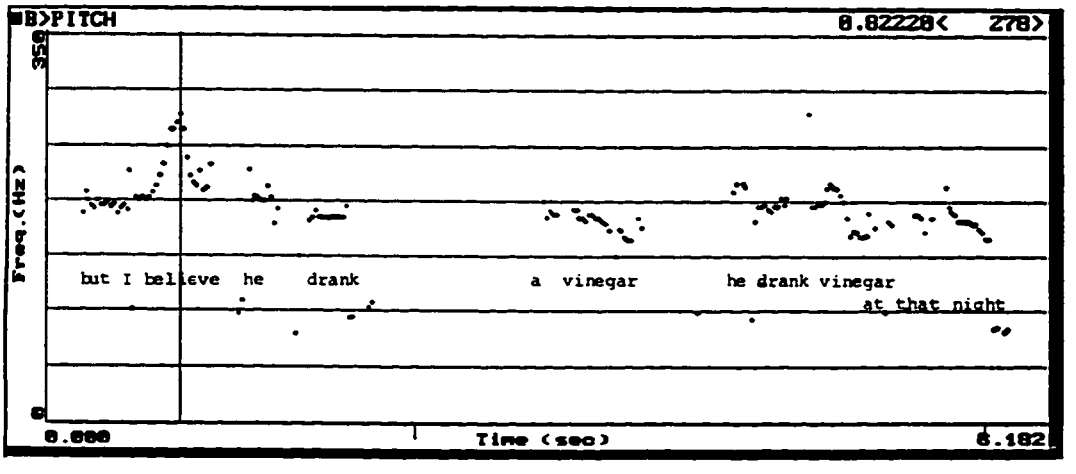


FIGURE 6.5 The evaluative projection verb *believe* has a pitch maximum.

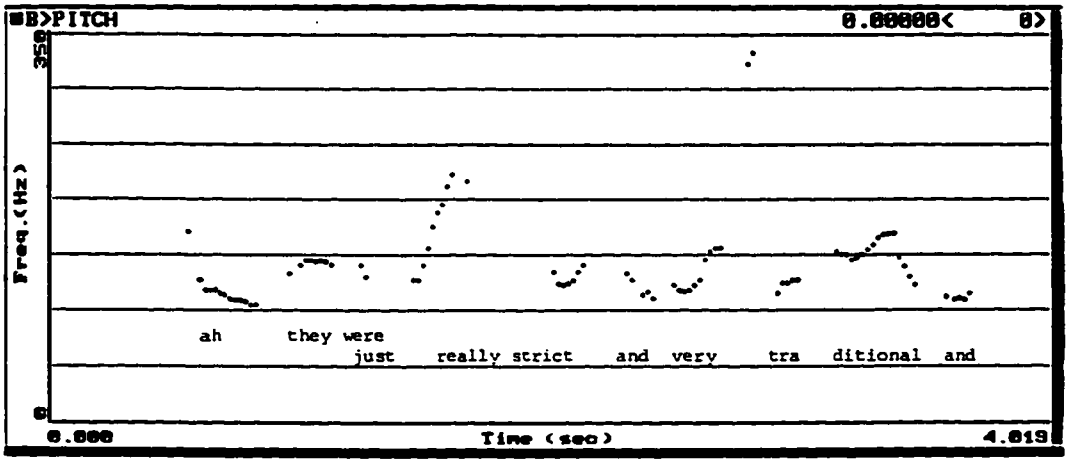


FIGURE 6.6 The evaluative adverb *really* has a pitch maximum.

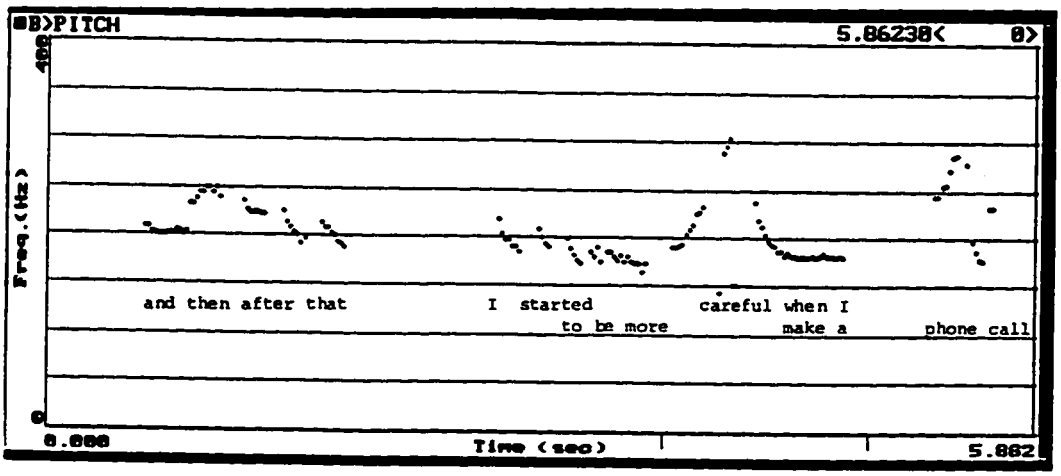


FIGURE 6.7 The inherently evaluative adjective *careful* has a pitch maximum.

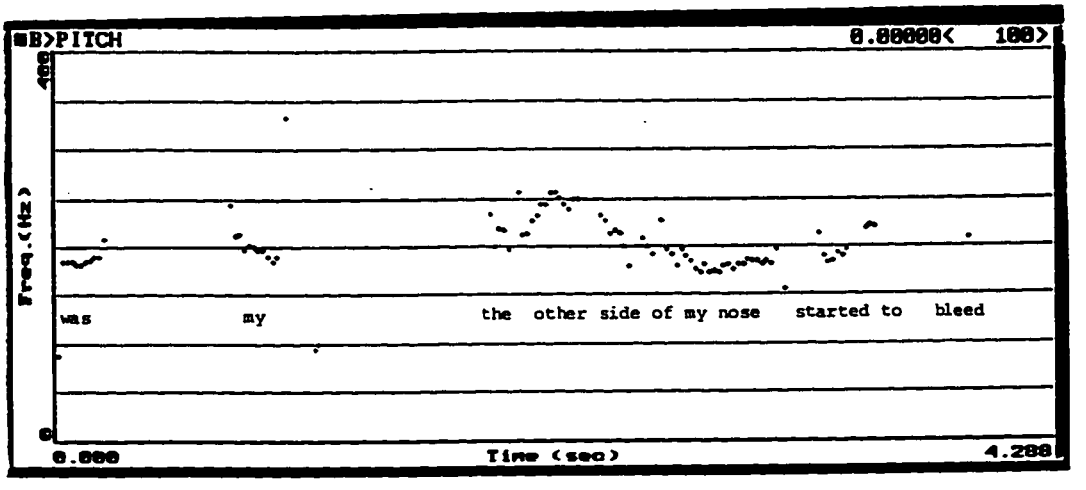


FIGURE 6.8 The comparator *other* has a pitch maximum.

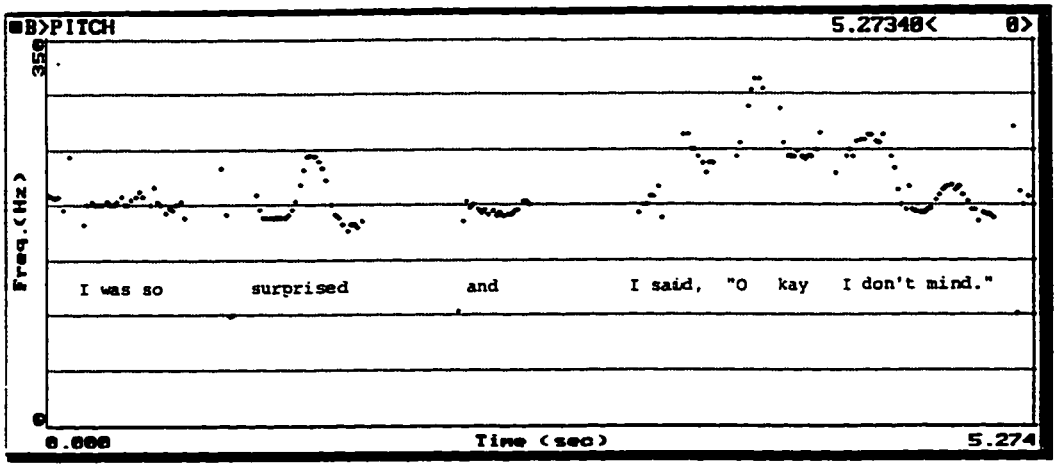


FIGURE 6.9 A woman, quoting herself, hits the top of her pitch range.

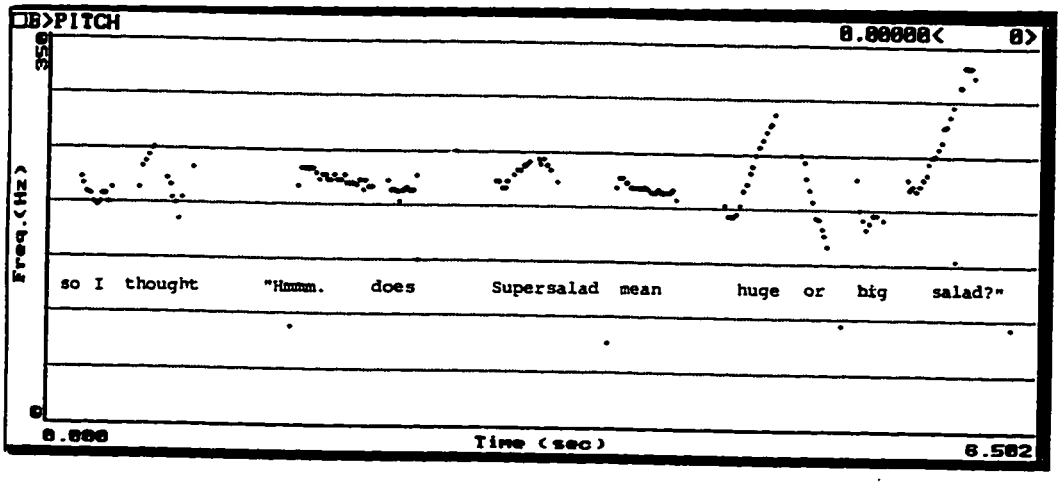


FIGURE 6.10 Another woman, quoting herself, hits the top of her pitch range.

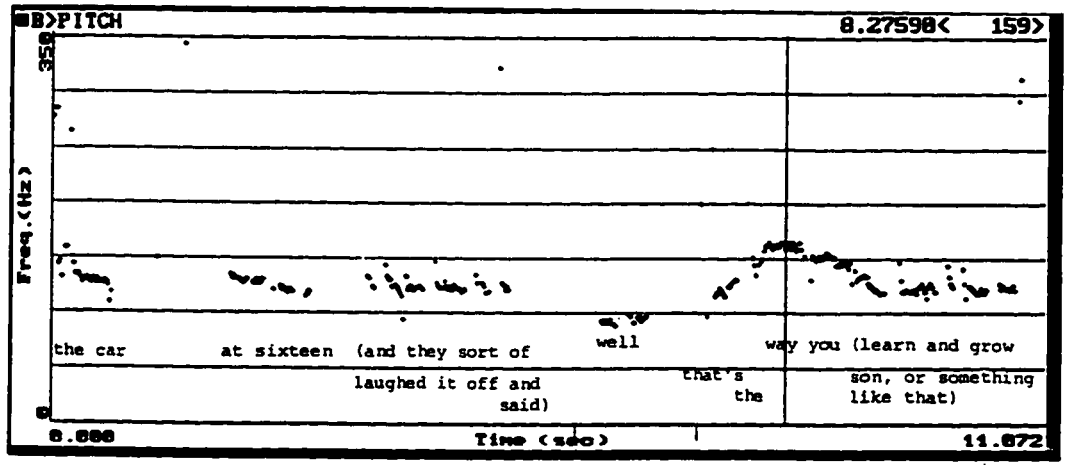


FIGURE 6.11 A man, quoting his parents, hits the top of his pitch range.

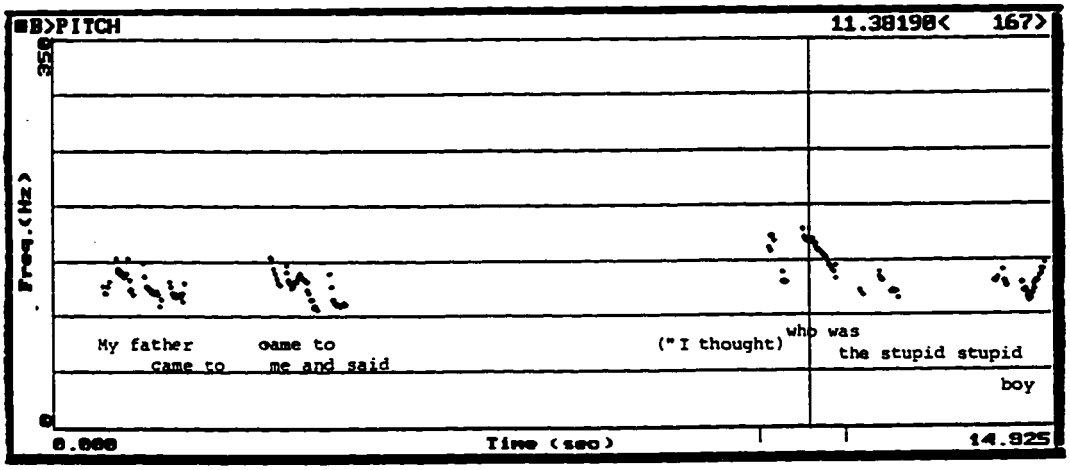


FIGURE 6.12 A man, quoting his father, hits the top of his pitch range.

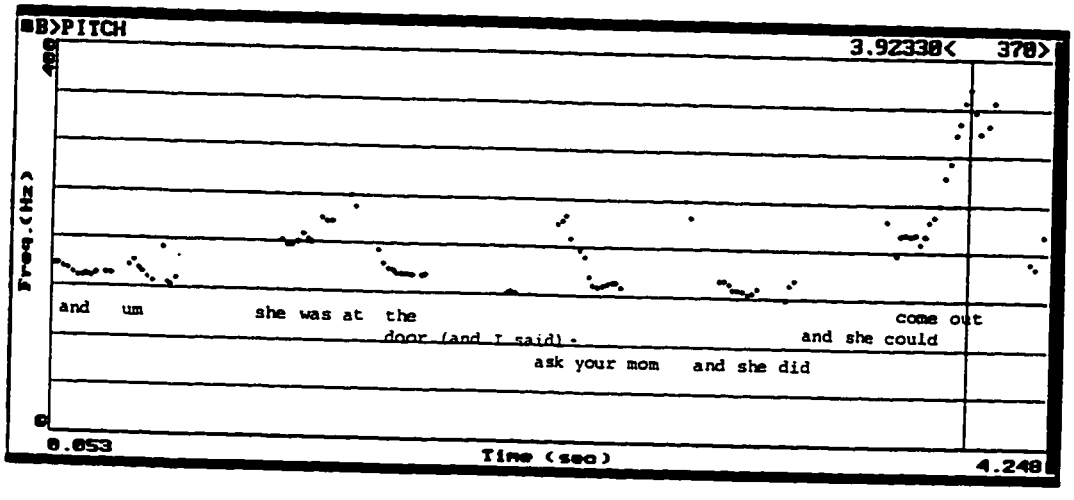


FIGURE 6.13 Excitement that her friend could come out is expressed with the maximum pitch of this speaker's entire narrative.

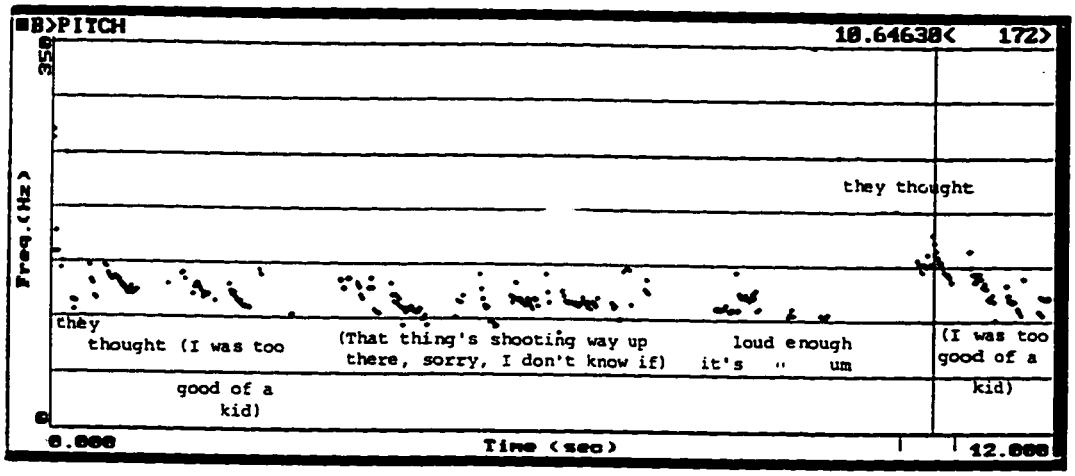


FIGURE 6.14 After an aside, the speaker reestablishes his status as story-teller with a pitch increase.

Notes to Chapter 6

1. Gumperz, et al. argue that since prosodic features must be replaced by lexico-grammatical features when oral narratives are transferred into written form, those who come from oral traditions which rely the most on prosody for coherence will have the greatest difficulty with writing.

2. Here is the text from "The Baddest Girl in the Neighborhood" (Labov, 1972, p. 383), a story about a fight:

> Well, one (I think) was with a girl.
> Like I was a kid, you know,
> And she was the baddest girl, the baddest girl in the neighborhood.
> If you didn't bring her candy to school, she would punch you in the mouth;
> And you had to kiss her when she'd tell you.
> This girl was only about 12 years old, man,
> but she was a killer.
> She didn't take no junk;
> She whupped all her brothers,
> And I came to school one day
> and I didn't have no money.
> My ma wouldn't give me no money
> And I played hookies one day,
> (She) put something on me.
> I played hookies, Man,
> so I said, you know, I'm not gonna play hookies no more

'cause I don't wanna get a whupping.
So I go to school
and this girl says, "where's the candy?"
I said, "I don't have it."
She says, powww!
So I says to myself, "There's gonna be times my mother won't give me money
because (we're) a poor family
And I can't take this all, you know, every time she don't give me any money."
So I say, "Well, I just gotta fight this girl.
She gonna hafta whup me.
I hope she don't whup me."
And I hit the girl: powwww!
and I put something on it.
I win the fight.
That was one of the most important.

3. Data source: Riggenbach, unpublished.

4. Although there were a few cases where no change in range occurred at all.

5. Couper-Kuhlen (1996) found certain cases of quoted speech in her data where mimicry of absolute register was used derisively and concludes that there are various functions of quoted speech. I have no disagreement with this conclusion; for example, (unfortunately) one can imagine that a high, nasalized pitch might be used to render

a nagging woman's voice; or the giant's "Fee Fi Fo Fum" in Jack and the Bean Stalk might be uttered with a lowered pitch by many readers.

Chapter 7

Conclusion

With these three empirical studies I have investigated the question of how NNSs at various levels of English ability use intonation. By analyzing extended stretches of discourse, I hope to have obtained a realistic picture of what happens in actual situations of communication. Each study has been conducted within a different genre of spoken discourse in order to bring particular aspects of intonation into focus. The use of computerized speech technology and statistical methods has allowed me to quantify the intonation of the speakers in these studies.

The analysis has followed a conception of intonation that includes both language-specific and language-universal functions. On the language-specific side, a model has been presented in Chapter 2 in which intonation is seen as a system of meaningful component parts, or pitch morphemes. These associate with units of text to indicate how those units are to be integrated with previous text and context in speakers' mental representations of the discourse situation at hand. The results of the studies support the position that NNSs at higher proficiency levels of English, as determined by independent assessment measures, use the intonation system in closer accordance with the intonation model than those at

lower levels. I have also considered language-universal, emotional functions of intonation: regardless of language background, speakers tended to use their highest pitch at emotional high points in narratives.

It should be evident that my theoretical interest is broader than mere curiosity about what NNSs do with intonation. I am interested in describing a set of grammatical structures within the English language whose interpretive power reaches beyond the sentence level to indicate discourse relationships in extended texts. Examples of this have been long-range contrasts among items in lectures associated with L+H* pitch accents, rhetorical shifts of topic size associated with paratones, and items assumed by a speaker to be accessible in the hearer's mental representation associated with a L* pitch accent. Intonation makes a significant contribution to the coherence of a text in its interaction with other linguistic systems such as syntax and the lexicon. This investigation of how intonation patterns manifest themselves in the speech of those who are learning English as adults provides indirect evidence that the grammar of intonation exists as stipulated by the componential model. The fact that NNSs come closer to the patterns of intonation in the model as their English reaches a higher level reflects on the model's accuracy.

7.1 Summary of Results

7.1.1 Characteristics of the Genres

At this point I shall attempt to synthesize the results of all three studies in nontechnical terms. The first two studies, within the genres of lectures and conversation, have in common that the analysis centers on the language-specific, grammatical aspects of intonation; that is, how pitch associates with various constituents to indicate their role in the discourse. Both studies have focussed on the pitch accents of individual words to determine whether intonation is performing its role of sorting items which are new, contrastive, or given in the discourse, or function words. They have also focused on intonational boundaries of various-sized constituents to determine how these are being used to indicate hierarchical relationships in the discourse organization. However, these first two genres differ in important dimensions: lectures are largely preplanned, monologic, and the organization is explicitly displayed, while conversation is dialogic and more spontaneous. The preplanning feature of the lecture genre afforded the easiest access to large rhetorical chunks which were often identifiable by being explicitly introduced with macro-markers (Chaudron & Richards, 1986) (i.e. "Next, I will move on to discuss X"). By measuring the pitch range before and after these topic shifts, I could easily determine whether speakers had increased their range (used a paratone). On the other

hand, within the conversation genre the discourse structure was more emergent. In other words, each participant's contribution directly influenced the next, resulting in a collaborative text. In order for this collaboration to proceed, the intonation at phrase and utterance boundaries was crucial in the turn-taking system, which was a strong focus of this analysis.

The third genre, oral narratives, was similar to the lecture genre in being mainly monologic and having overtly identifiable components. There was also an element of preplanning since the NNSs had practiced their stories as part of an assignment. Also, as Linde (1993) points out, important stories about one's life tend to be told and retold and may become a bit stylized even if they are told spontaneously. This genre had in common with conversation the fact that it was informal (nonacademic), and for some of the stories, the listeners were fairly active, offering backchannels, questions, and even interruptions. I chose this genre, for which the data I collected seemed the most expressive of the three, to be a venue for the investigation of intonation and emotion. Participants were actually assigned to talk on emotional topics: the NNSs were asked to tell a scary or embarrassing story; six of the NSs were told to describe a mistake their parents made in their childhood, while the last two told of a serious illness. Therefore the stories all had highly emotional content. While conversation can certainly be

an emotional genre as well, the excerpts I investigated in Chapter 5 were on fairly impersonal topics such as the weather or Japanese cities. Overall, I would say that the focus on different genres proved successful in highlighting different characteristics of discourse and the association of intonation patterns to text.

7.1.2 The Lecture Study

Moving now to a summary of more specific results, the lecture study had the largest and the most homogeneous group of subjects (18 Chinese graduate students) so it was possible to analyze statistical trends. This was useful in allowing me to make claims about what Chinese speakers in general are likely to do with their intonation. The research question was whether the people using intonation according to the componential model were obtaining higher English scores. Averages were taken of a sample of pitch values of items with certain functions in the information structure of the discourse: new items, given items, contrast items, function words, boundary tones at the phrase level, and paratones at the topic level. Regression analysis was chosen to assess these various categories relative to each other as predictors of English exam scores. The significant finding was that those who used paratones effectively, that is, who consistently made large range increases at rhetorical junctures, scored better on the English exam.

This study also included two NS lecturers whose intonation was measured in the same manner as was the NNSs'. It was reassuring to find that these speakers' measurements were consistent with each other as well as with the predictions of the intonation model: both speakers made large pitch distinctions between contrasts and given items; they increased their ranges to indicate new topics; they seldom used falling boundaries in midutterance position; and they were more likely to associate high pitch with new content words than with function words. Although this NS sample is too small to count as evidence in the statistical sense, these facts combine with those of another statistical study with a larger NS sample (Wennerstrom, 1994) to suggest that the intonation model represents an accurate characterization of standard English of Northern U.S. speakers.

7.1.3 The Conversation Study

The conversation study was a case study which focused on individual examples rather than overall averages. The intonation of the speakers who were rated the highest in fluency was compared to that of the speakers rated lowest. Particular examples of fluent uses of intonation were examined and compared to problematic ones. The results showed that while fluent speakers spoke more phrasally, less fluent speakers were more likely to assign equal pitch to every word. I argued that this word-by-word speech tends to obscure the

intonational signals about the relationship of the current utterance to what is assumed to be in the mental representation of the discourse in progress. For example, one low-level speaker associated a high pitch with almost every word in an utterance containing a contrast, so that the contrasting items were not especially distinguished from other content words, or even from the articles and prepositions. This was compared with the speech of higher level NNSs where contrast words were distinguished by being much higher in pitch than the surrounding words in the phrase.

Another result involved the use of boundary tones to indicate one's intention to maintain or yield a turn. Examples were given of more fluent speakers who used LH% (plateau) or LH% (low rising) pitch boundaries to extend the vowel of a word or hesitation sound before a pause in order to maintain the floor. In contrast, we saw examples of lower level speakers who lost the floor after using a falling boundary pitch. A final problem for low-fluency speakers was that boundary tones were inserted so frequently that it might be difficult for a NS to segment the utterance into thought groups. The general conclusion of this study was that intonation was one important aspect of conversational fluency, especially due to the collaboration involved in the turn-taking system and in the joint construction of similar mental representations of the discourse.

7.1.4 The Oral Narrative Study

The narrative study focussed on the emotional resources of intonation which speakers were assumed to have in common regardless of language background. Whereas the other two studies considered the placement of pitch accents with regard to information structure, this study looked at the association of pitch extremes with text to determine at what point in their narratives speakers hit the very top of their ranges. It was found that speakers consistently associated their highest pitch maxima with items in three main categories: structural, evaluative, and interactional.

In the first category, both NSs and NNSs associated their highest pitch with boundaries between the various components of the narrative (as defined by Labov & Waletsky, 1967) and at deictic transitions, although NSs were somewhat more likely to do so. This was regarded as a paratone and is consistent with the finding in the lecture study that speakers increased their pitch range at rhetorical shifts. Among the NNSs, the average increase was taken for all speakers regardless of level because the number of subjects was too small and the stories too diverse to obtain statistical trends by level as was done in the lecture study. It would make sense then, since the narrative study included both high and intermediate level NNSs, that their overall average on this paratone effect would be modestly lower than for NSs, as was found.

The second category with which pitch maxima associated

was what Labov and Waletsky (1967) call evaluation. For both groups, pitch maxima associated with external evaluations (separate utterances in which the speaker steps back to assess the events of the story) and internal evaluation devices embedded within the story line itself (such as loaded vocabulary, certain syntactic structures, and quoted speech). A second analysis, done with only the top three highest pitch maxima for the entire narrative showed an even more dramatic consistency in pitch peaks associating with quoted speech. This was particularly true for the NNSs. These evaluation findings are the strongest result of the narrative study. Since evaluation is by definition an expression of attitude and emotion, this finding lends support to the idea that intonational extremes reflect emotion regardless of language background. I would therefore suggest that this aspect of intonation is a resource in cross-cultural communication since speakers can convey their attitudes in this way, regardless of their level of English language development.

The third category, which I called "interaction," was comprised of those instances where pitch maxima associated with points where the narrative interfaced with some other interaction. This occurred at the beginnings of narratives as speakers claimed the floor, at points of questions or reactions by others after which tellers reestablished their right to the floor, at repairs, and at outside distractions. In a situation where the right of a story-teller to an

extended turn was challenged by an interruption or other distraction, tellers used pitch assertively to reclaim that right. This category was smaller than the others since only the NSs' stories were recorded in an interactive context, while the NNSs' stories were recorded in isolation (with the one exception of a NNS who recorded himself telling a story to a friend).

7.2 Second Language Acquisition

While attitudinal and emotional aspects of intonation can be helpful in cross-cultural communication, the English-specific system of intonational morphemes appears to be acquired more gradually. We have seen adult speakers of various background languages use intonation morphemes in English in close accordance with the model which indicates that at least some aspects of the system had been acquired; we have also seen lower level speakers from some of the same language groups whose intonation patterns were farther from the model. These facts indicate that this system is not universal. If it were, we would expect to see everyone associating pitch with text in a similar way regardless of level or language background.

In addition, the trends indicate that the acquisition of intonation takes place along with the development of other language skills. The SPEAK Test-based measures used to assess the subjects' level of English in all three studies were

constructed to measure overall language ability. In the lecture study where these scores were used in statistical procedures, the assessment included grammar, vocabulary, pronunciation, and fluency taken as a whole. That speakers' scores on this test tended to correlate with the intonation measures indicates a developmental progression of acquisition. However, the fact that the multiple regression analysis did not yield significant results on three of the variables is also an interesting finding. The reason for a result like that can only be that some speakers were using one or more intonation variables effectively but not others, while other speakers at the same level were using the others but not the first. Furthermore, we saw examples of "outlier" subjects whose overall English level was assessed as high despite low intonation scores, or vice versa: subjects whose intonation measurements were high despite low overall English level. Thus it is possible to conclude that there is no lockstep order of acquisition of intonational morphemes. I would also suggest that people with different learning styles acquire the intonation system differently. Perhaps some speakers "tune in" and pick up aspects of the intonation system holistically, whereas others' acquisition route might be, for example, more analytical.

The intonation system of subjects' background languages is also a factor in their acquisition of the English system. For the Chinese speakers, there was evidence of transfer of

specific native language stress and intonation patterns directly into English, especially at lower levels of proficiency. One example of this was the use of mid-utterance falling boundary tones (LL%), reflecting the fact that in Chinese it is the beginning of an utterance rather than the end which is intonationally distinctive. The L1s represented in the conversation and narrative studies were Japanese (10), Korean (5), Chinese (1), Thai (1), and Italian (1). Since I do not speak these languages, I recommend that contrastive analyses be conducted between the intonation of each of these languages and English. For example, it is my observation that Romance languages have contrastive pitch accents, but that they are less extreme than in English. Instead, speakers are more likely to use syntactic constructions that place the contrasting item at the end of a phrase, where stress naturally falls. In English the following utterance has a typical contrast intonation (capital letters indicate the L+H* contrastive pitch accent):

A: Where are the keys? B: I thought YOU had them.

However, in French, this intonation would sound odd:

A: Oú sont les clés? B: *J'ai pensé que TU les avais.

Where are the keys? I thought that YOU had them.

Instead, the contrastive pronoun would be inserted in a cleft structure as follows:

A: Oú sont les clés? B: J'ai pensé que c'était TOI qui les avais.

I thought that it was YOU who had them.

It would be ideal to conduct these types of contrastive analyses in other languages in naturally occurring speech in order to see what speakers actually do in context, and also to look at long-range intonational effects.

Contrastive analyses notwithstanding, there is certainly a precedent for the study of interlanguage forms, regardless of language background (Gass, Madden, Preston, & Selinker, 1989; Ioup & Weinberger, 1987; Selinker, 1972, 1992; Tarone, 1983). The premise behind these interlangauge studies, which have centered on syntax and phonology acquisition, is that there are developmental, intermediary stages between the L1 and the target language which are internally consistent but which resemble neither language exactly. The findings of the current intonation studies support this view of language acquisition. While lower level speakers may be most likely to simply transfer their native language intonation patterns while speaking English, in these studies, speakers at higher levels came close to the English model. However, none of the NNSs in the studies exhibited completely native-like intonation, evidence of interlangauge development in progress.

It is impossible to come to definitive conclusions about pedagogy based on these results. On the one hand, the findings do not contradict the argument that intonation can be

taught. Adults can grasp the idea of using pitch to distinguish items in discourse structure and can be given an outline of the grammar of intonation and exercises to practice it. In the lecture study, intonation had been an explicit pedagogical focus in the subjects' ESL class, and the results of the study could be interpreted as showing that those who responded to this teaching did better on the English exam.

On the other hand, neither do findings contradict the position that intonation *cannot* be taught. Despite the fact that none of the speakers in the conversation study had had any explicit instruction in intonation, several of the subjects attained an effective use of the system. Moreover, in the lecture study, many subjects did not use intonation according to the model despite having had instruction. Those who did use intonation effectively might have done so not as the result of being taught, but due to a natural acquisition process as their overall command of English improved. In either case, an ESL class can enhance the natural acquisition process by allowing an opportunity for communication and practice, while providing rules for those with more analytical learning styles.

7.3 Suggestions for Further Research

As noted in the last section, I would be interested in analyzing the intonation systems of the subjects' background languages to see how their native patterns interacted with

English patterns. I did not choose to do so in this dissertation because I believe that such an analysis would require collaboration with linguists who are native or near-native speakers of the languages in question. This line of inquiry could lead to findings not only about language transfer, but also about what aspects of intonation that are language specific versus language universal.

Regarding language universals, I would be interested in further pursuing the emotional/attitudinal aspects of intonation. In previous papers and presentations I have tended to dismiss these factors and move on to the grammatical aspects. In fact, when I collected the narrative data for this study it was my intention to focus on long-range contrasts, a structural aspect of intonation. The finding that emotional, evaluative narrative material correlated with pitch peaks for both NSs and NNSs came as a surprise, in spite of my best efforts to search for statistical differences between groups. However, these results are satisfying because they coincide with a recent groundswell of interest in performance and other oral traditions among anthropologists and scholars of folklore. While authors such as Bauman (1986) and Gumperz, Kaltman, and O'Connor (1984) have stressed the importance of prosody in jokes and story-telling, there appears to be room for a good deal of additional analysis in this direction.

Another area that I have had to restrain myself from

covering in this dissertation and could not completely avoid was that of NS intonation. In order to limit the scope of the dissertation, I chose to focus mainly on NNSs. However, since I have a strong theoretical interest in how intonation works, I was most excited during the analysis phase by the highest level speakers who exhibited intonation patterns *in accordance* with the intonation model, rather than low-level speakers who deviated from it. It would therefore be interesting to do a similar study of NSs to determine the subtleties of intonation in consistently facile users.

A related area in this regard is dialect differences in English intonation. It has been my impression that since the model describes fairly gross features of the intonation system of English, it covers a variety of dialects with the need for only minor adjustments for subtle differences. For example, whether an utterance boundary is high or low is a very large feature of intonation. Low-level Chinese speakers in my lecture study routinely used low boundaries in midutterance position, as in:

In my opinion$_{L}$ this topic is important.

A NS would be more likely to use a low rising boundary:

In my opinion$_{L}^{H\%}$, this topic is important.

I am not familiar with any dialect of English where a NS would use the former pattern (unless the context involved some special circumstances). However, details of how sharply the boundary rises or how long it is extended might vary from one

dialect to another, and these sorts of nuances would be interesting to investigate.

As a final word, in zeroing in on intonation as I have, devoting almost three-hundred pages to the topic, it is nonetheless my hope that the reader will not lose perspective on the larger picture, which in this case is oral communication. Cross-cultural communication depends on many factors: the relationships among the participants, how much common experience they have to bring to interpretation, the amount of experience in cross-cultural communication they have, the genre, the topic, the setting, the purpose of the interaction, whether the participants are tired or sick or resentful or bored or intimidated, as well as their level of English grammar, size of vocabulary, pronunciation of vowels and consonants, and yes, their intonation. When I have made statements in this dissertation about intonation being "effective" or "native-like" versus "problematic," I have done so with respect to a model of intonation which represents an ideal. Obviously intonation is just one of the many factors at work in communication; I have singled out intonation for special scrutiny because I believe it has received too little attention to date in applied studies of Linguistics. It is my hope that its role in communication will come to be more widely recognized through efforts such as this. It is in the spirit of this broader perspective that I submit this dissertation.

References

Bartlett, F. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.

Bauman, R. (1986). *Story, performance, and event*. Cambridge: Cambridge University Press.

Beckman, M. & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook, 3*, 15-70.

Biber, D. (1994). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.

Bing, J. (1985). *Aspects of English prosody*. New York: Garland.

Bolinger, D. (1978). Intonation across languages. In J. Greenberg (ed.), *Universals of human language: Phonology* (pp. 471-524). Stanford: Stanford University Press.

Bolinger, D. (1986). *Intonation and its parts*. Stanford: Stanford University Press.

Brazil, D. (1985). *The communicative value of intonation*. (Discourse Analysis Monograph 8, 2nd ed.). Birmingham: University of Birmingham ELR.

Brazil, D., Coulthard, M. & Johns, A. (1980). *Discourse intonation and language teaching*. London: Longman.

Brown, G. (1977). *Listening to spoken English*. London: Longman.

Brown, G., Currie, K., & Kenworthy, J. (1980). *Questions of intonation*. London: Croom Helm.

Brown, G. and Yule, G. (1983). *Discourse analysis.* Cambridge: Cambridge University Press.

Carrell, P. (1982). Cohesion is not coherence. *TESOL Quarterly, 16*(4), 479-488.

Cascallar, E., Sarwark, S., Smith, J., & Wennerstrom, A. (1992). *To SPEAK or not to SPEAK.* Paper presented at the TESOL Annual Convention, Vancouver, Canada.

Chafe, W. (1970). *Meaning and the structure of language.* Chicago: University of Chicago Press.

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing.* Chicago: University of Chicago Press.

Chaudron, C. & Richards, J. (1986). The effect of discourse markers on the comprehension of lectures. *Applied Linguistics, 7*(2), 113-127.

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

Chun, D. (1988). The neglected role of intonation in communicative competence and proficiency. *Modern Language Journal, 72*(3), 295-303.

Couper-Kuhlen, E. (1986). *An introduction to English prosody.* Baltimore: Edward Arnold.

Couper-Kuhlen, E. (1996). The prosody of repetition: on quoting and mimicry. In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in conversation* (pp. 366-405). Cambridge: Cambridge University Press.

Couper-Kuhlen, E. & Selting, M. (Eds.). (1996). *Prosody in conversation.* Cambridge: Cambridge University Press.

Cruttendon, A. (1986). *Intonation.* Cambridge: Cambridge University Press.

Crystal, D. (1969). *Prosodic systems and intonation in English.* Cambridge: Cambridge University Press.

Crystal, D. (1975). *The English tone of voice.* London: Edward Arnold.

Davies, C. & Tyler, A. (1994). Demystifying cross-cultural (mis)communication: Improving performance through balanced feedback in a situated context. In C. Madden and C. Myers (Eds.), *Discourse and performance of international teaching assistants* (pp. 201-220). Alexandria, VA: TESOL.

Derwing, T. & Munro, M. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*(1), 1-16.

Educational Testing Service (1985). *Speaking proficiency English assessment kit* [Training manual]. Princeton, NJ: ETS.

Fillmore, C. (1985). Frames and the semantics of understanding. *Quaderni di Semantica, VI,* 222-254.

French, P. & Local, J. (1986). Prosodic features and the management of interruptions. In C. Johns-Lewis (Ed.), *Intonation in discourse* (pp. 157-180). San Diego: College Hill.

Gass, S., Madden, C., Preston, D., & Selinker, L. (1989). *Variation in second language acquisition: Discourse and pragmatics*. Clevedon: Multilingual Matters.

Gilbert, J. (1980). Prosodic development: Some pilot studies. In S. Krashen & R. Scarcella (Eds.), *Issues in second language research* (pp. 110-117). Rowley, MA: Newbury House.

Gilbert, J. (1993). *Clear speech: Pronunciation and listening comprehension in North American English* (2nd ed.) [Student's book]. New York: Cambridge University Press.

Goffman, E. (1981). *Forms of talk*. Philadelphia: University of Pennsylvania Press.

Goldsmith, J. (1990). *Autosegmental and metrical phonology*. Cambridge, MA: Basil Blackwell.

Grant, L. (1993). *Well said: Advanced English pronunciation*. Boston: Heinle and Heinle.

Gumperz, J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.

Gumperz, J. (1992). Contextualization and understanding. In C. Goodwin & A. Duranti (Eds.), *Rethinking context* (pp. 229-252). Cambridge: Cambridge University Press.

Gumperz, J., Kaltman, H, & O'Connor, M.C. (1984). Cohesion in spoken and written discourse: Ethnic style and the transition to literacy. In D. Tannen (Ed.), *Coherence in spoken and written discourse* (pp. 3-20_). Norwood, NJ: Ablex.

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

Hagen, S. and Grogan, P. (1992). *Sound advantage.* New York: Prentice-Hall.

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

Halliday, M. (1967). *Intonation and grammar in British English.* The Hague: Mouton.

Halliday, M. (1994). *An Introduction to Functional Grammar* (2nd ed.). London: Edward Arnold.

Halliday, M. & Hasan, R. (1976). *Cohesion in English.* London: Longman.

Hansen, C. (1994). Topic identification in lecture discourse. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 131-145). Cambridge: Cambridge University Press.

Hasan, R. (1984). Coherence and cohesive harmony. In J. Flood (Ed.), *Understanding reading comprehension* (pp. 181-219). Newark, DE: IRA.

Hatch, E. & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics.* New York: Newbury House.

Hayes, B. (1984). The phonology of rhythm in English. *Linguistic Inquiry, 15,* 33-74.

Hayes, B. (1985). *A metrical theory of stress.* New York: Garland.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies.* Chicago: University of Chicago Press.

Hewings, M. (1995). The English intonation of native speakers and Indonesian learners: A comparative study. *RELC Journal, 26*(1), 27-46.

Hyman, L. (1975). *Phonology: Theory and analysis.* New York: Holt, Rinehart and Winston.

Hymes, D. (1974). *Foundations of sociolinguistics: An ethnographic approach.* Philadelphia: University of Pennsylvania Press.

Ioup, G. and Weinberger, S. (1987). *Interlanguage phonology: The acquisition of a second language sound system.* Rowley, MA: Newbury House.

Jefferson, G. (1978). Sequential aspects of story-telling in conversation. In J. Schenkein (Ed.), *Studies in the organization of conversational interaction* (pp. 219-248). New York: Academic Press.

Juffs, A. (1990). Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors. *IRAL, 28*(2), 99-117.

Kaisse, E. (1987). Rhythm and the cycle. *Papers from the 23rd Annual Regional Meeting of the Chicago Linguistics Society,* 199-209.

Kiparsky, P. (1979). Metrical structure assignment is cyclic. *Linguistic Inquiry, 10,* 421-441.

Klatt, D. (1975). Vowel lengthening is syntactically

determined in a connected discourse. *Journal of Phonetics, 3*, 129-140.

Krashen, S. (1981). *Second language acquisition and second language learning.* New York: Pergamon.

Labov, W. (1972). *Language in the inner city.* Philadelphia: Univeristy of Pennsylvania Press.

Labov, W. & Waletsky, J. (1967). Narrative analysis: Oral versions of personal experience. In J. Helms (Ed.), *Essays on the verbal and visual arts* (pp. 12-44). Seattle: University of Washington Press.

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

Ladd, R. (1984). English compound stress. In D. Gibbon & H. Richter (Eds.). *Intonation, accent and rhythm: Studies in discourse phonology* (pp. 253-266). Berlin: De Gruyter.

Lepetit, D. (1989). Cross-linguistic influence in intonation: French/Japanese and French/English. *Language Learning, 39*(3), 397-413.

Levinson, S. (1983). *Pragmatics.* Cambridge: Cambridge University Press.

Levelt, W. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Liberman, M. (1975). *The intonation system of English.* Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA. [Published by Garland Press, New York, 1979].

Liberman, M. & Sag, I. (1974). Prosodic form and discourse function. *Papers from the Tenth Regional Meeting of the Chicago Linguistic Society*, 416-427.

Linde, C. (1993). *Life stories*. New York: Oxford University Press.

Longacre, R. (1983). *The grammar of discourse*. New York: Plenum.

Mason, A. (1994). By dint of: Student and lecturer perceptions of lecture comprehension strategies in first-term graduate study. In J. Flowerdew (Ed.), *Academic listening: Research perspectives*, (pp. 199-218). Cambridge: Cambridge University Press.

McLemore, C. (1991) *The pragmatic interpretation of English intonation: sorority speech*. Unpublished doctoral dissertation, University of Texas at Austin.

Minsky, M. (1975). A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision* (pp. 211-277). New York: McGraw-Hill.

Mishler, E. (1986). *Research interviewing: Context and narrative*. Cambridge, MA: Harvard University Press.

Morley, J. (1993). *Rapid review of vowels and prosodic contexts*. Ann Arbor: University of Michigan Press.

Muller, F. (1996). Affiliating and disaffiliating with continuers: Prosodic aspects of recipiency. In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in conversation: interactional studies* (pp. 131-176). Cambridge: Cambridge

University Press.

Nespor, M. & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Nespor, M. & Vogel, I (1989). On clashes and lapses. *Phonology, 6*, 69-116.

Neufield, G. & Schneiderman, E. (1980). Prosodic and articulatory features in adult language learning. In S. Krashen & R. Scarcella (Eds.), *Issues in Second Language Research* (pp. 105-109). Rowley, MA: Newbury House.

O'Connor, J. & Arnold, G. (1973). *Intonation of colloquial English* (2nd ed.). New York: Garland.

Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation.* Unpublished doctoral dissertation. Massachusetts Institute of Technology, Cambridge, MA.

Pierrehumbert, J. & Beckman, M. (1988). *Japanese tone structure.* Cambridge, MA: MIT Press.

Pierrehumbert, J. & Hirschberg, J. (1990). The meaning of intonational contours in discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in Communication* (pp. 271-311). Cambridge, MA: MIT Press.

Pike, K. (1945). *The intonation of American English.* Ann Arbor: University of Michigan Press.

Polyani, L. (1985). *Telling the American story.* Norwood, NJ: Ablex.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations.

Discourse Processes, 14, 423-441.

Rounds, P. (1987). Characterizing successful classroom discourse for NNS teaching assistant training. *TESOL Quarterly, 21*(4), 643-672.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turntaking for conversation. *Language, 50*(4), 696-735.

Sag, I. & Liberman, M. (1975). The intonational disambiguation of indirect speech acts. *Papers from the 11th Regional Meeting of the Chicago Linguistic Society*, 487-497.

Saville-Troike, M. (1982). *The ethnography of communication: An introduction.* Baltimore: University Park Press.

Schank, R. & Abelson, R. (1977). *Scripts, plans, goals, and understanding.* Hillsdale, NJ: Lawrence Erlbaum.

Schegloff, E. (1995). Reflections on studying intonation in talk-in-interaction. Paper presented at the Annual AAAL Conference, Long Beach, CA.

Schiminoff, S. (1980). *Communication rules.* Beverly Hills, CA: Sage.

Schiffrin, D. (1987). *Discourse Markers.* (Studies in interactional sociolinguistics, Vol. 5). Cambridge: Cambridge University Press.

Selinker, L. (1972). Interlanguage. *IRAL, 10*, 209-231.

Selinker, L. (1992). *Interlanguage revisited.* New York: Longman.

Selinker, L. & Douglas, D. (1985). Wrestling with "context" in

interlanguage theory. *Applied Linguistics, 6*(2), 190-204.

Selting, M. (1996). Prosody as an activity-type distinctive cue in conversation: The case of so-called 'astonished' questions in repair initiation. In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in conversation* (pp. 231-270). Cambridge: Cambridge University Press.

Shen, X. (1990a). *The prosody of Mandarin Chinese.* (University of California Publications in Linguistics, Vol. 118). Berkeley: University of California Press.

Shen, X. (1990b). Ability of learning the prosody of an intonational language by speakers of a tonal language: Chinese speakers learning French prosody. *IRAL, 28*(2). 119-134.

Sinclair, J. (1994). Trust the text. In M. Coultard (Ed.), *Advances in written text analysis* (pp. 12-25). London: Routledge.

Sperber, D. & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Oxford: Blackwell.

Steffensen, M. (1986). Register, cohesion and cross-cultural reading comprehension. *Applied Linguistics, 7*(1), 71-85.

Swales, J. (1990). *Genre analysis.* Cambridge: Cambridge University Press.

Tannen, D. (1984a). *Conversational style.* Norwood, NJ: Ablex.

Tannen, D. (1984b). Spoken and written narrative in English and Greek. In D. Tannen (Ed.), *Coherence in spoken and*

written discourse (pp. 21-41). Norwood, NJ: Ablex.

Tarone, E. (1983). On the variability of interlanguage systems. *Applied Linguistics, 4*(2), 143-163.

Tench, P. (1991). The stylistic potential of intonation. In W. van Peer (Ed.), *The taming of the text* (pp. 50-82). London: Routledge.

Tench, P. (1996). *The intonation systems of English*. London: Cassell.

Toolan, M. (1988). *Narrative: A critical linguistic introduction*. London: Routledge.

Toolan, M. (1996). *Total speech: An integrational linguistic approach to language*. Durham, NC: Duke University Press.

Traeger, G. & Smith, H. (1951). *Outline of English structure*. Norman, OK: Battenburg Press.

Tyler, A. (1992). Discourse structure and the perception of incoherence in ITAs' spoken discourse. *TESOL Quarterly, 26*(4), 713-729.

Upshur, J. & Turner, C. (1995). Constructing rating scales for second language tests. *ELT Journal, 49*(1), 3-12.

Wennerstrom, A. (1991). *Techniques for teachers: A guide for nonnative speakers of English*. Ann Arbor, MI: University of Michigan Press.

Wennerstrom, A. (1992). The paratone: An intonational marker of English discourse organization. Unpublished master's thesis, University of Washington, Seattle, WA.

Wennerstrom, A. (1993). Focus on the prefix: Evidence for

word-internal prosodic words. *Phonology, 10,* 309-324.

Wennerstrom, A. (1994). Intonational meaning in English discourse: A study of nonnative speakers. *Applied Linguistics, 15*(4), 399-420

Wenk, B. (1985). Speech rhythms in second language acquisition. *Language and speech, 28*(2), 157-175.

Werth, P. (1984). *Focus, coherence and emphasis*. Breckenham, Kent: Croom Helm.

Williams, J. (1989). Planning, discourse marking, and the comprehensibility of international teaching assistants. *TESOL Quarterly, 26*(4), 693-712.

Wolfson, N. (1982). *CHP: The conversational historical present in American English narrative*. Cinnarinson, NJ: Foris.

Young, D. (1991). Projection and deixis in narrative discourse, in W. van Peer (Ed.), *The taming of the text* (pp. 20-49). London: Routledge.

Yule, G. (1980). Speakers' topics and major paratones. *Lingua, 52,* 33-47

CURRICULUM VITAE

Ann Wennerstrom

19243 Ashworth Avenue, North
Seattle, WA 98133
(206) 546-3888 (home)
email: akw@u.washington.edu

EDUCATION

Ph.D.	**English**	University of Washington	1997
M.A.	**Linguistics**	University of Washington	1992
M.A.T.	**ESL**	University of Washington	1984
B.A.	**Scand. Studies**	University of Wisconsin (Phi Beta Kappa)	1978

Dissertation: *Discourse intonation and second language acquisition: Three genre-based studies.*
Director: Professor Sandra Silberstein

PUBLICATIONS

Refereed Articles:

Intonation as cohesion in academic discourse: A study of Chinese speakers of English. *Studies in Second Language Acquisition*.

The role of intonation in second language fluency (forthcoming). In H. Riggenbach (Ed.), *Perspectives on fluency*.

Intonational meaning in English discourse: A study of nonnative speakers. *Applied Linguistics, 14(4),* 1994: pp. 399-419.

Focus on the prefix: Evidence for word-internal prosodic words. *Phonology, 10,* 1993: pp. 1-15.

Student bias in instructional evaluation. (principle author with P. Heiser) *TESOL Quarterly, 26(2),* 1992: pp. 271-288.

Content-based pronunciation. *TESOL Journal, 1(3),* 1992: pp. 15-18.

Submitted:

Is it me or is it hot in here? Menopause, humor, and identity. *Journal of Narrative and Life History*.

Textbooks:

Techniques for teachers: A guide for nonnative speakers of English. Ann Arbor: University of Michigan Press, 1989. (214 pgs.)

Techniques for teachers: A guide for nonnative speakers of English [videotape]. Seattle, WA: University of Washington Instructional Media Services, 1988 (1 hr. 45 mins.).

Book Reviews:

Rapid review of vowels and prosodic contexts [3 volume set], J. Morley. *TESOL Quarterly, 27(4)*, 1993: pp. 761-2 (solicited).

Academic English, Rossi. *WAESOL Newsletter*, March, 1983. Reprinted in *Penn TESOL-East Newsletter, 5(1)*, 1985.

PROFESSIONAL EXPERIENCE

Teaching:

Acting Instructor, Department of English, University of Washington. Teach grad and undergrad courses in language studies and TESL. Appointed for 9/97.

Lecturer, University of Washington ESL. Teach ESL in academic, intensive, and business English programs, all subjects, all levels. 6/85 - present

ESL Instructor, Bellevue Community College. Taught developmental composition for ESL. 3/86-6/86

ESL Teaching Assistant, Academic ESL Program, University of Washington. Taught academic ESL -- intermediate to advanced Reading, Writing, Listening, Speaking; responsible for all aspects of teaching including curriculum development and placement. 9/82-6/84

EFL Instructor, English Department, Dalian Institute of Technology, People's Republic of China. Taught EFL to Chinese university students. 6/82-7/82

ESL Instructor, Princeton University League, Princeton, NJ. Taught English conversation to foreign visitors at all levels. Volunteer. 1981-82

ESL Tutor. Have tutored informally at all levels.

Administration:

Graduate Student Advisor, University of Washington MATESL Program. Advise MATESL students on academic matters, process admissions, coordinate practicum. 9/93 - present

International TA Training Coordinator, University of Washington ESL. Responsible for development and instruction of course for ITAs, and coordinating and training of SPEAK test raters. 6/87-6/91

ESL Program Coordinator, University of Washington ESL. Administered Academic ESL Program of 300 students; supervised and advised 10 - 15 instructors; assisted in project to establish curriculum for program. 9/84-6/85

Consulting:

Research Consultant, Educational Testing Services. Coordinated subject recruitment, data collection, and SPEAK rating at the University of Washington for nation-wide study of SPEAK Test validity and reliability. 3/91-9/91

Reviewer, *Applied Linguistics*, University of Michigan Press, TESOL Press, TESOL.

CONFERENCE PRESENTATIONS

"Intonation in Oral Narratives," AAAL, Orlando, 1997.

"The Role of Intonation in L2 Fluency," TESOL, Orlando, 1997.

"Intonational meaning in English discourse: a study of nonnative speakers," AAAL, Chicago, March 1996.

"Intonation in Academic Discourse," AAAL, Chicago, 1996.

"English as a cohesive device in ITA discourse," TESOL Annual Convention, Chicago, March 1996.

"A study of intonation in ESL student speech," TESOL Annual Convention, Long Beach, April 1995.

"Content-based pronunciation," TRI-TESOL Convention, Vancouver, BC., November 1993.

"Student bias in instructional evaluation," (with P. Heiser), TESOL Annual Convention, Atlanta, April 1993.

"ITA training: analysis of natural discourse," Preconference symposium on the training of international teaching assistants. TESOL Annual Convention, New York, March 1991 [invited talk in colloquium].

"To SPEAK or not to SPEAK," 3rd National Conference on the Training and Employment of Teaching Assistants, Austin, November 1991.

"Using media to enhance ITA training" (with E. Axelson and C. Madden) 2nd National Conference on the Training and Employment of Teaching Assistants, November 1989.

"Strategies for ITA language training" (with J. Bolivar and S. Sarwark) 2nd National Conference on the Training and Employment of Teaching Assistants, November 1989.

"Meet the SPEAK Test," WAESOL Annual Convention, November 1989.

"Student bias in instructional evaluation," (with P. Heiser), TESOL Annual Convention, March 1989.

"ITA training: Pronunciation on the job," TESOL Annual Convention, March 1989.

"Video materials for ITA training," TESOL Annual Convention, March 1989.

"ESL for foreign TAs: University politics, testing, and training," (with P. Dame and E. Sanchez), NAFSA Region I Conference, 1988.

"International teaching assistants: From institutional responsibility to instructional content," (with T. Delamere and B. Feetham), TESOL Annual Convention, March, 1988.

Guest Lectures:

"Adult Second Language Acquisition" in Linguistics 400, Survey of Linguistics, Fall 1991

"The T-Test" in English 574, TESL Research Methods, Spring 1995.

"The SPEAK Test", in English 576, TESL Testing, Winter 1990.

"Content-Based Pronunciation" in English 572, TESL Methods and Materials, Winter 1990, 1992.

"Discourse-based Materials Design" in English 576, TESL Testing, Spring 1992.

"Pierrehumbert & Hirschberg Model of Intonation" in Linguistics 552, Metrical Phonology, Spring 1994

"The 'ITA Problem': Prejudice Against Nonnative Accents" in English 572, TESL Methods/Materials, Winter 1979.

MEMBERSHIPS

American Association of Applied Linguistics (AAAL).

Teachers of English to Speakers of Other Languages (TESOL).

Washington Association for the Education of Speakers of Other Languages (WAESOL).