# Tactics and strategies for thinking about F0 variation: comments on the papers by Kubozono and by van den Berg, Gussenhoven, and Rietveld

Mary E. Beckman
Ohio State University
Columbus, Ohio, USA

and

Janet B. Pierrehumbert*
AT&T Bell Laboratories
Murray Hill, New Jersey, USA

The two papers in this group are about downstep and its interactions with phrasal pitch range, local prominence, and the like -- things that cause variation in the fundamental frequency values that realize a particular tonal event. Sorting out this F0 variation is like investigating any other physical measure of speech; making the measure necessarily involves making assumptions about its phonetic control and its linguistic function, and since the assumptions shape the investigation whether acknowledged or not it is better to make them explicit.

Leaving aside for the moment assumptions about the control mechanism, we can classify the assumed linguistic function along two dimensions. The first is categorical versus continuous: the variation in the measure can function discretely to symbolize qualitatively different linguistic categories of some kind or it can function continuously to signal variable quantities of some linguistic property in an analogue way. The second is paradigmatic versus syntagmatic: values of the measure can be freely chosen from a paradigm of contrasting independent values or they can be relationally dependent on some other value in the context.

---------

*Present address: Department of Linguistics, Northwestern University, Evanston, Illinois, USA.

Describing downstep along these dimensions, we could focus on different things, depending on which aspect we are considering at which level of representation. The aspect of the phonological representation which is relevant to downstep is fundamentally paradigmatic; it is the tone string, which depicts a sequence of categorical paradigmatic choices. A syntagmatic aspect of downstep comes into focus when we look at how it is triggered; the pitch range at the location of a given tone is lowered, but the lowering depends upon the tone appearing in a particular tonal context. In Hausa, where downstep applies at any H following a HL sequence, this contextual dependence can be stated in terms of the tone string alone. In Japanese and English, the context must be specified also in terms of a structural feature -- how the tones are organized into pitch accents. The syntagmatic nature of downstep comes out even more clearly in its phonetic consequences; the new value is computed for the pitch range relative to its previous value. When this relational computation applies iteratively in long sequences, the step levels resulting from the categorical input of H and L tone values tend toward a continuous scale of F0 values.

These aspects of downstep differentiate it from other components of downtrend. The phonological representation of the categorical input trigger, for example, differentiates downstep from any physiological downtrend that might be attributed to a reduction in subglottal pressure over time, as in the classic model. The phonetic representation of the syntagmatic choice of output values differentiates it also from the use of register slope to signal an interrogative or some degree of finality in a declarative, as in Thorsen's (1978, 1979) model of declination, or from the use of initial raising and final lowering of the pitch range to signal the position of the phrase in the discourse hierarchy of old and new topics, as proposed by Hirschberg &

Pierrehumbert (1986). In either of these models, the declination slope or the initial or final pitch range value is an independent paradigmatic choice that is linguistically meaningful in its own right.

These definitional assumptions about its linguistic function suggest several tactics for designing experiments on downstep and its domain. First, to differentiate downstep from other components of downtrend, it is important to precisely identify the phonological trigger. This is trivial in languages such as Hausa or Japanese, where the trigger involves an obvious lexical contrast. In languages such as English and Dutch, on the other hand, identifying the downstep trigger is more difficult and requires an understanding of the intonational system as a whole. In this case, when we run into difficulties in the phonological characterization of downstep, it is often a useful tactic to wonder whether we have the best possible intonational analysis. Thus, given the admitted awkwardness of their characterization of downstep as a phrasal feature, we might ask whether van den Berg, Gussenhoven, and Rietveld (this volume) have found the optimal tonal analysis of Dutch intonation contours.

We are particularly inclined to ask the question here because of the striking resemblance to difficulties that have been encountered in some analyses of English. For example, in translating older British transcription systems into autosegmental conventions, Ladd uses H*+L for the representation of a falling F0 around the nuclear accent, precluding its use for a prenuclear H*+L which contrasts with H* phonologically primarily in its triggering of downstep. This analysis led him first (in Ladd 1983) to a characterization of downstep in English as a paradigmatic feature of H tones, exactly analogous to van den Berg, Gussenhoven, and Rietveld's characterization of downstep in Dutch as a paradigmatic choice for the phrase, specified independently of the choice of tones or pitch accents. In subsequent work, Ladd has recognized that this characterization does not explain the syntagmatic distribution of downstepped tones which is the original motivation for the downstep analysis. He has replaced the featural analysis with a phonological representation of register relationships among pitch accents using a recursive metrical tree, as in Liberman & Prince's (1977) account of stress patterns or Clements's (1981) account of downdrift in many African tone languages. We feel that a more satisfactory solution is to use H*+L to transcribe the prenuclear tone in many downstepped contours where the L is not realized as an actual dip in F0 before the next accent's peak. This more abstract analysis captures the phonological similarities and the common thread of pragmatic meaning that is shared by all downstepping pitch accents in English, whether they are rising (L+H) or falling (H+L) and whether the following downstepped tone is another pitch accent or is a phrasal H (see Pierrehumbert & Hirschberg, in press).

A similar problem arises in the transcription of boundary tones. Grønnum, (formerly Thorsen, this volume) criticizes van den Berg, Gussenhoven, and Reitveld for transcribing the contour in (1a) with a L% boundary tone, on the grounds that there is no marked local lowering of F0. But not transcribing a L% boundary tone here would mean jettisoning the generalization that all intonation phrases are marked with boundary tones -- that (1a) contrasts with the initial portion of contour (2a) as well as with contours with a H% boundary tone, realized as a clear local rise. It would also jettison the cross-language generalization that final lowering -- a progressive compression of the pitch range that reaches in from the end to lower all tones in the last half-second or so of the phrase -- is typically associated with L phrase tones (see Pierrehumbert & Beckman, 1988, chapter 8, for a review). Rejecting L% boundary tones in Dutch on the grounds that contours such as (1a) have no more marked local fall than contours like (1b) puts a high value on superficial similarities between the F0 representation and the tonal analysis, at the expense of system symmetry and semantic coherence. On these grounds, we endorse Gussenhoven's (1984, 1988) approach to the analysis of intonation-phrase boundary tones in Dutch, which he and his coauthors assume in their paper in this volume, and we cannot agree with Grønnum's argument against transcribing (1a) with a L% boundary tone. We think that the awkwardness of characterizing downstep in transcription systems that analyze the nuclear fall in English as a H*+L pitch accent is symptomatic of a similar confusion between phonetic representation and phonological analysis.

As Ladd does for English, Gussenhoven (1988) analyzes nuclear falls in Dutch as H*+L pitch accents. Adopting this analysis, van den Berg, Gussenhoven, and Rietveld transcribe the falling patterns in Figures (2a), (3a), and (5b) all as H*+L. We wonder whether this analysis yields the right generalizations. Might the first F0 peak and subsequent sharp fall in contour (2a) be instead a sequence of H* pitch accent followed by a L phrase accent? This alternative analysis would give an explicit intonational mark for the edge of this sort of prosodic constituent, just as in English (see Beckman & Pierrehumbert 1986) and in keeping with the accounts of a unified prosodic structure proposed on various grounds by Selkirk (1981), Nespor & Vogel (1986), Beckman (1986), and Pierrehumbert & Beckman (1988). Might the gradually falling slope after the first peak in contour (3a) be instead an interpolation between a H* pitch accent and the L of a following L+H* accent? This alternative analysis would attribute the contrast between absence of downstep in (3a) and its presence in (5b) to the singleton versus bitonal pitch accent, just as in similar contours in English (Beckman & Pierrehumbert

1986) and in keeping with the descriptions of other languages where a bitonal pitch accent triggers downstep, such as Tokyo Japanese (see Poser 1984; Pierrehumbert & Beckman 1988; Kubozono, this volume). Because they analyze the falls in (2a), (3a), and (5b) all as H*+L, van den Berg, Gussenhoven, and Rietveld must account for the differences among the contours instead by operations that are specified independently of the tonal representation. They attribute the step-like character of the fall in (5b) to the operation of downstep applied as a paradigmatic feature to the whole phrase. And they account for the more gradual nature of the fall in (3a) by the application of a rule that breaks up the H*+L accent unit to link its second tone to the syllable just before the following nuclear pitch accent. Elsewhere, Gussenhoven (1984, 1988) describes this rightward shift of the second tone as the first step of a two-part Tone Linking Rule, which in its full application would delete the L entirely to create a hat pattern. The partial and complete applications of Tone Linking merge the prenuclear accent phonologically with the nuclear accent, thus giving the sequence a greater informational integrity. Tone Linking and Downstep are mutually exclusive operations.

In testing for the systemic and semantic coherence of such an intonational analysis, it is a useful strategy to exhaustively determine which patterns and contrasts are predicted to exist and to not exist. One can then systematically determine whether the ones that are predicted to exist do contrast with each other and are interpretable in the expected way and whether the ones which are predicted not to exist are indeed illformed or noncontrastive. Among other things, Gussenhoven's analysis of Dutch predicts that two sequences of prenuclear and nuclear accent which contrast only in whether Tone Linking applied partially or completely should not contrast categorically in the way that contours like (3a) and the hat pattern do in

English. Also, the description of Tone Linking implies that the operation does not apply between two prenuclear accents, so that three-accent phrases are predicted to have a smaller inventory of patterns than do two-accent phrases. That is, sequences of three accents within a single phrase should always be downstepped rather than Tone-Linked. More crucially, the analysis predicts the impossibility of three-accent phrases in which only one of the accents triggers a following downstep. As Pierrehumbert (1980) points out, the existence of such mixed cases in English precludes an analysis of downstep as an operational feature of the phrase as a whole. In general, if downstep is to be differentiated from other components of downtrend, we need to be careful of analyses that make downstep look like the paradigmatic choice of whether to apply a certain amount of final lowering to a phrase.

A second tactical point relating to the categorical phonological representation of downstep is that in looking at the influence on downstep of syntax or pragmatic focus, one needs first to know whether or not downstep has occurred, and in order to know this, it is imperative to design the corpus so that it contrasts the presence of the downstep trigger with its absence. That is, in order to claim that downstep has occurred, one cannot simply show that a following peak is lower than an earlier peak; one must demonstrate that the relationship between the two peaks is different from that between comparable peaks in an utterance of a form without the phonological configuration that triggers downstep. Kubozono reminds us of this in his paper, and it is a point well worth repeating. Using this tactic can only bring us closer to a correct understanding of the relationship between syntactic and prosodic constituents, including the domain of downstep.

A third tactical point is to always remember that other things which superficially look like downstep in the ways in which they affect pitch range do not necessarily function linguistically like downstep. For example, it is generally agreed that downstep has a domain beyond which some sort of pitch range reset applies. Since the sorts of things that produce reset seem to be just like the things that determine stress relationships postlexically -- syntactic organization and pragmatic focus and the like -- it is very easy to assume that this reset will be syntagmatic in the same way that downstep is. Thus, van den Berg, Gussenhoven, and Rietveld list in their paper only these two possible characterizations:

(1) The reset is a syntagmatic boost that locally undoes downstep.

(2) The reset is a syntagmatic register shift by a shift factor that either

(a) reverses the cumulative effects of downstep within the last domain, or

(b) is an independent "phrasal downstep" parameter.

They do not consider a third characterization, suggested by Liberman & Pierrehumbert (1984) and developed in more detail by Pierrehumbert & Beckman (1988):

(3) The reset is a paradigmatic choice of pitch range for the new phrase.

In this last characterization, the appearance of phrasal downstep in many experiments would be due to the typical choice of a lower pitch range for the second phrase of the utterance, reflecting the discourse structure of the mini-paragraph.

A criticism that has been raised against our characterization in (3) is that it introduces too many degrees of freedom. Ladd (1989), for example, has proposed that instead of independent paradigmatic choices of pitch range for each phrase and of tonal prominence for each accent, there is only the limited choice of relative pitch registers that can be represented in binary branching trees. Kubozono in his paper finds this view attractive and adapts it to the specification of pitch registers for Japanese major and minor phrases. Such a phonological characterization may seem in keeping with results of experiments such as the one that Liberman & Pierrehumbert (1984) describe, where they had subjects produce sentences with two intonation phrases that were answer and background clauses, and found a very regular relationship in the heights of the two nuclear accent peaks across ten different levels of overall vocal effort. Indeed, results such these are so reminiscent of the preservation of stress relationships under embedding that it is easy to see why Ladd wants to attribute the invariant relationship to a syntagmatic phonological constraint on the pitch range values themselves, rather than to the constant relative pragmatic saliences.

However, when we consider more closely the circumstances of such results, this criticism is called into question. The design of Liberman & Pierrehumbert's (1984) experiment is typical in that it encouraged the subjects to zero in on a certain fixed pragmatic relationship -- in that case, the relationship of an answer focus to a background focus. The constant relationship between the nuclear peak heights for these two foci may well reflect the subject's uniform strategy for realizing this constant pragmatic relationship. In order to demonstrate a syntagmatic phonological constraint, we would need to show that the peak relationships are constant even when we vary the absolute pragmatic salience of one of the focused elements.

The analogy to stress relationships also fails under closer examination in that the purely syntagmatic characterization of stress is true only in the abstract. A relational representation of a stress pattern predicts any number of surface realizations, involving many paradigmatic choices of different prominence-lending phonological and phonetic features. For example, the relatively stronger second syllable of red roses relative to the first might be realized by the greater prominence of a nuclear accent relative to a prenuclear accent (typical of the citation form), or by a bigger pitch range for the second of two nuclear accents (as in a particularly emphatic pronunciation that breaks the noun phrase into two intermediate phrases), or by the greater prominence of a prenuclear accent relative to no accent (as in a possible pronunciation of the sentence The florist's red roses are more expensive.). Similarly, a weak-strong pragmatic relationship for the two nouns in Anna came with Manny. can be realized as a particular choice of pitch ranges for two intonational phrases, or as the relative prominence of prenuclear versus nuclear pitch accent if the speaker chooses to produce the sentence as one intonation phrase. As Jackendoff (1972), Carlson (1983), and others have pointed out, the utterance in this case has a somewhat different pragmatic interpretation due to Anna's not being a focus, although Anna still is less salient pragmatically than Manny.

The possibility of producing either two-foci or one-focus renditions of this sentence raises an important strategic issue. Liberman & Pierrehumbert (1984) elicited two-foci productions by constructing a suitable context frame and by pointing out the precise pragmatic interpretation while demonstrating the desired intonation pattern. If they had not taken care to do this, some of their subjects might have given the other interpretation and produced the other intonation for this sentence, making impossible the desired comparison of the two nuclear-accent peaks. A more typical method in experiments on pitch range, however, is to present the subject with a randomized list of sentences to read without providing explicit cues to the desired pragmatic and intonational interpretation. In this case, the subject will surely invent an appropriate pragmatic context, which may vary from experiment to experiment or from utterance to utterance in uncontrolled ways. The effects of this uncontrolled variation is to have an uncontrolled influence on the phrasal pitch ranges and on the prominences of pitch accents within a pitch range. The variability of results concerning the interaction of syntax and downstep in the literature on Japanese (e.g., Kubozono 1989, this volume; Selkirk 1989; Selkirk & Tateishi, forthcoming) may reflect this lack of control more than it does anything about the interaction between syntactic constituency and downstep.

The fact that a sentence can have more than one pragmatic interpretation also raises a methodological point about statistics: Before we can use averages to summarize data, we need to be sure that the samples over which we are averaging are homogeneous. For example, both in Poser (1984) and in Pierrehumbert & Beckman (1988), there were experimental results which could be interpreted as showing that pragmatic focus reduces but does not block downstep. When we looked at our own data more closely, however, we found that the apparent lesser downstep was actually the result of a single outlier in which the phrasing was somewhat different and downstep had occurred. Including this token in the average made it appear as if the downstep factor could be chosen paradigmatically in order to give greater pitch height than normal to prosodic constituents with narrow focus. The unaveraged data, however, showed that the interaction is less direct; elements bearing narrow focus tend to be prosodically separated from preceding elements and thus are realized in pitch ranges that have not been downstepped relative to the pitch range of preceding material. It is possible that Kubozono could resolve some of the apparent contradictions among his present results and those of other experiments in the literature on Japanese if he could find appropriate ways of looking at all of the data token by token.

The specific tactical lesson to draw here is that since our understanding of pragmatic structure and its relationship to phrasing and tone-scaling is not as well developed as our understanding of phonological structure and its interpretation in F0 variation, we need to be very cautious about claiming from averaged data that downstep occurs to a greater or lesser degree in some syntactic or rhythmic context. A more general tactical lesson is that we need to be very ingenious in designing our experiments so as to elicit productions from our subjects that control all of the relevant parameters. A major strategic lesson is that we cannot afford to ignore the knotty questions of semantic and pragmatic representation that are now puzzling linguists who work in those areas. Indeed, it is possible that any knowledge concerning prosodic structure and prominence that we can bring to these questions may advance the investigative endeavor in previously unimagined ways.

Returning now to assumptions about control mechanism, there is another topic touched on in the paper by van den Berg, Gussenhoven, and Rietveld which also raises a strategic issue of major importance to future progress in our understanding of tone and intonation. This is the question of how to model L tones and the bottom of the pitch range. Modeling the behavior of tones in the upper part of the pitch range -- the H tones -- is one of the success stories of laboratory phonology. The continuing controversies about many details of our understanding (evident in the two papers in this section) should not be allowed to obscure the broad successes. These include the fact that [+H] is the best understood distinctive feature value. While work in speech acoustics has made stunning progress in relating segmental distinctive features to dimensions of articulatory control and acoustic variation, the exact values along these dimensions which a segment will assume in any given context in running speech have not been very accurately modeled. In contrast, a number of different approaches to H-tone scaling have given rise to F0 synthesis programs which can generate quite accurately the contours found in natural speech. Also, work on H-tone scaling has greatly clarified the division of labor between phonology and phonetics. In general, it has indicated that surface phonological representations are more abstract than was previously supposed, and that much of the burden of describing sound patterns falls on phonetic implementation rules, which relate surface phonological representations to the physical descriptions of speech. Moreover, attempts to formulate such rules from the results of appropriately designed experiments have yielded insights into the role of prosodic structure in speech production. They have provided additional support for hierarchical structures in phonology, which now stand supported from both the phonetic and morphological sides, a fate we might wish on more aspects of phonological theory.

In view of these successes, it is tempting to tackle L tones with the same method that worked so well for H tones -- namely, algebraic modeling of the F0 values measured in controlled contexts. Questions suggested under this approach include: What functions describe the effects of overall pitch range and local prominence on F0 targets for L tones? What prevents L tones from assuming values lower than the baseline? In downstep situations, is the behavior of L tones tied to that of H tones, and if so, by what function?

We think it is important to question the assumptions that underlie these questions, particularly the assumptions about control mechanisms. We suggest that it would be a strategic error to apply too narrowly the precedents of work on H-tone scaling.

Looking first at the physiological control, we see that L-tone scaling is different from H-tone scaling. A single dominant mechanism, cricothyroid contraction, appears to be responsible for H tone production, in the sense that this is the main muscle showing activity when F0 rises into a H tone. In contrast, no dominant mechanism for L tone production has been found. Possible mechanisms include:

- Cricothyroid relaxation -- e.g., Simada & Hirose (1971), looking at the production of the initial boundary L in Tokyo Japanese; Sagart et al. (1986), looking at the fourth (falling) tone in Mandarin.
- Reduction of subglottal pressure -- e.g., Monsen, Engebretson & Vermula, (1978), comparing L and H boundary tones.
- Strap muscle contraction -- e.g., Erickson (1976), looking at L tones in Thai; Sagart et al. (1986), looking at the third (low) tone in Mandarin; Sugito & Hirose (1978), looking at the initial L in L-initial words and the accent L in Osaka Japanese; Simada & Hirose (1971) and Sawashima et al. (1973), looking at the accent L in Tokyo Japanese.
- Cricopharyngeus contraction -- Honda and Fujimura (1989), looking at L phrase accents in English.

Some of these mechanisms involve active contraction whereas others involve passive relaxation. There is some evidence that the active gesture of strap muscle contraction comes into play only for L tones produced very low in the pitch range. For example, of the four Mandarin tones, only the L of the third tone seems to show sternohyoid activity consistently (see Sagart et al. 1986). Similarly, the first syllable of L-initial words in Osaka Japanese shows a marked sternohyoid activity (see Sugito & Hirose 1978) that is not usually observed in the higher L boundary tone at the beginning of Tokyo Japanese accentual phrases (see, e.g., Simada & Hirose 1971). Lacking systematic work on the relation of the different mechanisms to different linguistic categories, we must entertain the possibility that no single function controls

L-tone scaling. Transitions from L to H tones may bring in several mechanisms in sequence, as suggested in Pierrehumbert & Beckman (1988). One of the tactical imports of the different mechanisms is that we need to be more aware of the physiological constraints on transition shape between tones; we should not simply adopt the most convenient mathematical functions that served us so well in H-tone scaling models.

Another common assumption that we must question concerns the functional control of the bottom of the pitch range. We need to ask afresh the question: Is there a baseline? Does the lowest measured value at the end of an utterance really reflect a constant floor for the speaker, which controls the scaling of tones above it, and beyond which the speaker does not aim to produce nor the hearer perceive any L tone?

Tone-scaling models have parlayed a great deal from assuming a baseline, on the basis of the common observation that utterance-final L values are stable for each speaker, regardless of pitch range. On the other hand, it is not clear how to reconcile this assumption with the observation that nuclear L tones in English go up with voice level (see, e.g., Pierrehumbert, in press). This anomaly is perturbing because it is crucial that we have accurate measures of the L tones; estimates of the baseline from H-tone scaling are quite unstable in the sense that different assumptions about the effective floor can yield equally good model fits to H-tone data alone. The assumption that the bottom of the pitch range is controlled via a fixed base line comes under further suspicion when we consider that the last measured F0 value can be at different places in the phrasal contour, depending on whether and where the speaker breaks into vocal fry or some other aperiodic mode of vocal fold vibration. It is very possible that the region past this point is intended as, and perceived as, lower than the last point where F0 measurement is possible.

A third assumption that relates to both the physiological and the functional control of L tones concerns the nature of overall pitch range variation. It has been easy to assume in H-tone modeling that this variation essentially involves the control of F0. Patterns at the top of the range have proved remarkably stable at the different levels of overall F0 obtained in experiments, allowing the phenomenon to be described with only one or two model parameters.

We note, however, that the different F0 levels typically are elicited by instructing the subject to "speak up" to varying degrees. This is really a variation of overall voice effort, involving both an increased subglottal pressure and a more pressed vocal-fold configuration. It seems likely, therefore, that the actual control strategy is more complicated than our H-tone models make it. While the strategy for controlling overall pitch range interacts with the physiological control of H tones in apparently simple ways, its interaction with the possibly less uniform control mechanism for L tones may yield more complicated F0 patterns. In order to find the invariants in this interaction, we will probably have to obtain other acoustic measures besides F0 to examine the other physiological correlates of increased pitch range besides the increased rate of vocal-fold vibration. Also, it may be that pitch range variation is not as uniform functionally as the H-tone results suggest. It is possible that somewhat different instructions to the subject or somewhat different pragmatic contexts will emphasize other aspects of the control strategy, yielding different consequences for F0 variation, particularly at the bottom of the pitch range.

These questions about L-tone scaling have a broader implication for research strategy. Work on H tones has brought home to us several important strategic lessons: Experimental designs should orthogonally vary local and phrasal properties; productions should be properly analyzed phonologically; and data analysis should seek parallel patterns within data separated by speaker. We clearly need to apply these lessons in collecting F0 measurements for L tones. However, to fully understand L tones, we will need something more. We will need more work relating linguistic to articulatory parameters. We will need to do physiological experiments in which we fully control the phonological structure of the utterances we elicit, and we will need to develop acoustic measures that will help to segregate the articulatory dimensions in large numbers of utterances.

**References**

Beckman, Mary E. (1986). Stress and Non-Stress Accent (Netherlands Phonetic Archives 7). Dordrecht: Foris.

Beckman, Mary E., & Janet B. Pierrehumbert (1986). "Intonational structure in Japanese and English." Phonology Yearbook 3, 255-310.

Carlson, Laurie (1983). Dialogue Games: An Aproach to Discourse Analysis. Synthese Language Library, 17. Dordrecht: Reidel.

Clements, G.N. (1981). "The hierarchical representation of tone features." Harvard Studies in Phonology 2, 50-105.

Erickson, Donna (1976). A Physiological Analysis of the Tones of Thai. Doctoral dissertation, University of Connecticut.

Grønnum (Thorsen), Nina (this volume). "Comments to 'Downstep in Dutch: implications for a model' by Rob van den Berg, Carlos Gussenhoven and Toni Rietveld."

Gussenhoven, Carlos (1984). On the Grammar and Semantics of Sentence Accents. Dordrecht: Foris.

Gussenhoven, Carlos (1988). "Adequacy in intonational analysis: The case of Dutch." In H. van der Hulst & N. Smith, eds., Autosegmental Studies in Pitch Accent. Dordrecht: Foris.

Hirschberg, Julia, & Janet Pierrehumbert (1986). "The intonational structuring of discourse." In Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics, 136-144.

Honda, Kiyoshi, and Osamu Fujimura (1989). "Intrinsic vowel F0 and phrase-final lowering: Phonological vs. biological explanations." Paper presented at the 6th Vocal Fold Physiology Conference, Stockholm, August 1989. (Written version to appear in the proceedings. New York: Raven Press.)

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

Kubozono, Haruo (1989). "Syntactic and rhythmic effects on downstep in Japanese," Phonology 6.1, 39-67.

Kubozono, Haruo (this volume). "On the hierarchy of intonational representation: Evidence from Japanese."

Ladd, D. Robert (1983). "Phonological features of intonational peaks." Language 59, 721-759.

Ladd, D. Robert (1987). "Metrical representation of pitch register." In John Kingston and Mary E. Beckman, eds., Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech, 35-57. Cambridge: Cambridge University Press.

Liberman, Mark, & Janet Pierrehumbert (1984). "Intonational invariance under changes in pitch range and length." In M. Aronoff & R. T. Oerhle, eds., Language Sound Structure: Studies in Phonology Presented to Morris Halle, 157-233. Cambridge, MA: MIT Press.

Liberman, Mark, & Alan Prince (1977). "On stress and linguistic rhythm." Linguistic Inquiry 8, 249-336.

Monsen, R. B., A. M. Engebretson, & N. R. Vermula (1978). "Indirect assessment of the contribution of subglottal pressure and vocal fold tension to changes of fundamental frequency in English." Journal of the Acoustical Society of America 64, 65-80.

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

Pierrehumbert, Janet B. (1980). The Phonology and Phonetics of English Intonation. Doctoral dissertation, MIT.

Pierrehumbert, Janet (in press). "A preliminary study of the consequences of intonation for the voice source." To appear in Speech Transmission Laboratory Quarterly Progress and Status Reports (Royal Institute of Technology, Stockholm).

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

Pierrehumbert, Janet, & Julia Hirschberg (in press). "The meaning of intonation contours in the interpretation of discourse." In M. Pollack & P. Cohen, eds., Plans and Intentions in Communication and Discourse (SDF Benchmark Series in Computational Lingusitics). Cambridge, MA: MIT Press.

Poser, William J. (1984). The Phonetics and Phonology of Tone and Intonation in Japanese. Doctoral dissertation, MIT.

Sagart, Laurent, Pierre Halle, Benedicte de Boysson-Bardies, Catherine Arabia-Guidet (1986). "Tone production in modern standard Chinese: An electromyographic investigation." Paper presented at the 19th International Conference on Sino-Tibetan Languages and Linguistics, Columbus, Ohio, 12-14 September 1986.

Sawashima, M., Y. Kakita, & S. Hiki (1973). "Activity of the extrinsic laryngeal muscles in relation to Japanese word accent." Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (University of Tokyo) 7, 19-25.

Selkirk, Elisabeth O. (1981). "On the nature of phonological representation." In T. Myers, J. Laver, & J. Anderson, eds., The Cognitive Representation of Speech, 379-388. Amsterdam: North-Holland.

Selkirk, Elisabeth O. (1989). "On the nature of prosodic constituency: comments on Beckman and Edwards's paper." In Kingston & Beckman, eds., Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech, 179-200 (Cambridge University Press).

Selkirk, Elisabeth O., & Koichi Tateishi (forthcoming). "Syntax and downstep in Japanese." To appear in D. Zec & S. Inkelas, eds., The Phonology-Syntax Connection (Univeristy of Chicago Press).

Simada, Z., & H. Hirose (1971). "Physiological correlates of Japanese accent patterns." Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (University of Tokyo) 5, 41-49.

Sugito, Miyoko, & Hajime Hirose (1978). "An electromyographic study of the Kinki accent." Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (University of Tokyo) 12, 35-51.

Thorsen, Nina (1978). "An acoustical analysis of Danish intonation." Journal of Phonetics 6, 151-175.

Thorsen, Nina (1979). "Interpreting raw fundamental frequency tracings of Danish." Phonetica 36, 57-78.

van den Berg, Rob, Carlos Gussenhoven, & Toni Rietveld (this volume) "Downstep in Dutch: Implications for a model."