6 The Trouble with ToBI

D. R. Ladd

6.1 Introduction

ToBI came into existence about 1990. Its original stated purpose was to serve as an agreed standard for the prosodic annotation of English speech databases. The primary aim of such annotation was to render databases easily searchable—to make it possible to look for all the phrase-final rises, or all the low accented syllables, or all the sequences of downstepped accents. Two main groups of researchers were behind the initiative: a group associated with Janet Pierrehumbert's tone-based autosegmental analysis of English intonation contours (e.g., Pierrehumbert 1980; Beckman and Pierrehumbert 1986; Pierrehumbert and Hirschberg 1990) and a group who had been transcribing hierarchical prominence and phrasing structure using numerical notations of boundary strength ("break indices") at each orthographic word boundary (Price et al. 1991; Wightman et al. 1992). Together the autosegmental tones and the numerical break indices yielded the acronym ToBI.

The early consultation meetings involved not only representatives of the two groups but a number of intonation researchers associated with the British and Dutch traditions, as well as several speech technologists not associated with any particular school of prosodic analysis. At these meetings, there was general agreement on the value of a common transcription standard, but little consensus on how to achieve it. There were long discussions of specific proposals, some details of which I summarize in the description of ToBI presented in Ladd (2008a, 104–113), and one important aspect of which I consider at greater length in the next section. Eventually a published standard resulted (Silverman et al. 1992). There were some limited evaluation studies (Pitrelli, Beckman, and Hirschberg 1994), and extensive training materials were created (Beckman and Ayers Elam 1997; Brugos, Shattuck-Hufnagel, and Veilleux 2006). The various components of a full ToBI annotation are described in more detail in the historical and theoretical summary by Beckman, Hirschberg, and Shattuck-Hufnagel (2005).

As early as the mid-1990s, however, in a kind of "mission creep," the ToBI tonal analysis took on a life of its own separate from the break-index analysis, and it rapidly became common to regard a ToBI tone sequence as the most acceptable way of reporting intonational data in published articles that had nothing to do with the labeling of speech databases (e.g., Birch and Clifton 1995). My comments in this chapter focus on this general use of the tonal notation and its claims to broader validity. To avoid any confusion about my intentions, I emphasize that my critique deals specifically with ToBI in its institutionalized form and with the theoretical justifications for that form laid out by Beckman, Hirschberg, and Shattuck-Hufnagel (2005). I remain strongly convinced of the potential of the autosegmental-metrical (AM) approach to describing

intonation, concisely yet comprehensively presented in Arvaniti (chapter 1, this volume). The trouble with ToBI, in my view, is that it precludes exploration of some of the important theoretical issues within the general AM approach.

6.2 Phonological Distinctions and Phonetic Categories

As the original ToBI tonal transcription system acquired the status of received analysis, there was a proliferation of ToBI notation systems for other languages (e.g., Japanese, Venditti 1997; German, Grice et al. 1996; Greek, Arvaniti and Baltazani 2005); more recently, there has been a move to create a universally applicable international prosodic alphabet (Hualde and Prieto 2016). This dual outcome—specific ToBI variants for specific languages, followed by a call for a transcription system that is not specific to any language—reflects the tension that surfaced during the first consultation meetings between what we might call "phonological" and "phonetic" conceptions of transcription. This is a fundamental tension in any symbolic annotation system: What is the proper basis for choosing one symbolic phonetic string over another? Even in segmental transcription with the International Phonetic Alphabet (IPA), where the issues are more clearly understood, there is no agreement; the problem is usually treated as a matter of "broad" and "narrow" transcription. In a narrow transcription, one is theoretically annotating all the phonetic detail one can perceive, but everyone acknowledges that there are degrees of narrowness. In a broad transcription, IPA symbols are used (e.g., in dictionaries) to spell the phonemically distinct segment types in any given language. Because of the phonemic principle, such broad transcriptions of individual words are often uncontroversial, but even so, the line between broad and narrow remains fuzzy.

For one thing, there is a practical issue: different varieties of the same language often have slightly different systems of phonemic contrast. In a word like *Marrakesh*, the initial vowel is systemically different for speakers who pronounce *Mary* and *marry* identically and speakers who distinguish them. Such cases are fairly well understood, and it is possible (albeit ultimately arbitrary) to establish transcription policies to deal with them. In theory, the ideal solution might be to give different transcriptions for different varieties on the basis of the phonemic distinctions they actually make, but compromises are usually adopted to avoid an impractical degree of fragmentation into multiple transcription systems. This issue is extensively discussed in the Wikipedia guidance for the phonetic transcription of English (https://en.wikipedia.org/wiki/Help:IPA/English), which takes its overall aim to be that of maintaining a unitary transcription system for all of English wherever possible. I return to this problem (and Wikipedia's solution) in the next section.

But there is another more fundamental problem for which ordinary notions of phonemic distinctiveness provide no solution. This is the problem of salient subphonemic detail, especially in connected speech. In English, these phenomena include such things as cluster simplification and place assimilation (Do we transcribe the sentence *It costs ten pounds* as /ɪtˈkɒstsˈtɛnˈpɑʊndz/ or /ɪʔˈkɒsːˈtɛmˈpɑʊnz/?), but the problem can be simply illustrated even in isolated words by the use of the glottal stop. In many North American varieties of English, orthographic T is regularly realized by a glottal stop before sonorants in names like *Atwood* and *Whitlock*, and before unstressed /ən/ or syllabic /n̩/ in names like *Clinton* and *Horton* and words like *blatant* and *important*. Should this difference be noted in transcribing? Superficially, these cases seem to be covered by the phonemic principle: the distribution of glottal stop is phonologically conditioned; glottal stop is therefore an allophone of /t/ and should not be indicated in a phonemic transcription. Somehow, though, some allophones are more noticeable than others.

When Bill Clinton was elected U.S. president in 1992, for example, I heard a British television reporter comment with some surprise on the fact that most Americans pronounced his name with a glottal stop. Given the well-known stigma attached to many uses of glottal stop in British English, it is obviously not surprising that a British reporter noticed its nonstigmatized use in American pronunciations of *Clinton*. But my point here is not sociolinguistic; it is theoretical. What kind of entity is glottal stop in English? The expectation of classical phonemic theory is that phonetic distinctions that can be used to distinguish different meanings in a given situation will be noticed and reliably reproduced by the language's native speakers, while phonetic distinctions that are completely predictable—allophonic differences—will be reliably reproduced but will not be noticed. This expectation is met in some cases, as, for example, the different aspiration of English voiceless stops in absolute syllable initial position and following /s/. This difference predictably astonishes beginning linguistics students when it is used to introduce the phonemic principle. But in other cases—English glottal stop, the German [ç/χ] distinction (Moulton 1947), the Russian [i/ɯ] (и/ы) distinction (Kiparsky 2014)—what we might call "allophonic awareness" is undeniably real. Somehow the members of these pairs constitute distinct phonetic categories for many speakers of these languages. If these categories cannot be grounded in phonemic contrast, where do they come from?

I do not propose to try to answer this question here, but simply to suggest that it lies at the heart of the difficulties with ToBI (and ultimately with the IPA; see, e.g., Pike 1943, 42–55; Ashby et al. 1995). The notion of phonological distinctiveness, and the idea that different languages make different distinctions, is well established and reasonably well understood. The goal of transcribing only the phonological distinctions in the intonation system of a given language variety, carefully laid out for ToBI by Beckman, Hirschberg, and Shattuck-Hufnagel (2005), is theoretically coherent. But there is also a kind of strictly phonetic distinctiveness that is poorly theorized yet intuitively obvious—the kind of distinctiveness that tells us that /t/ and glottal stop are different. This is what leads almost anyone who sets about transcribing intonation to question the theoretical discipline promoted by Beckman, Hirschberg, and Shattuck-Hufnagel and to try to find symbolic means to represent audible differences that have no obvious place in the official phonology.

In the following sections, I discuss reasons why I believe Beckman, Hirschberg, and Shattuck-Hufnagel's confidence in ToBI's phonological foundations is at best premature.

6.3 ToBI and the Phonological Analysis of Intonation

The first question that ToBI or any other transcription standard needs to address is what to do about differences between different language varieties. The trend in ToBI, sketched by Beckman, Hirschberg, and Shattuck-Hufnagel (2005), has been to narrow the range of varieties of English that the system claims to cover. The original idea was that ToBI was to be a suitable transcription for intonation in at least Southern British, Australian, and North American English. This list implicitly acknowledges that Irish and Northern British intonation is conspicuously different in some ways, but treats the rest of "English" as involving more or less the same system of intonational phonology. But by the late 1990s, the ToBI system being elaborated by various scholars at Ohio State, MIT, and elsewhere in the United States was renamed MAE-ToBI and was said to cover Mainstream American English.

I see at least two reasons why this narrowing is a bad idea. The first is merely practical: given the high degree of uniformity and mutual intelligibility across most varieties

of "English," there are obvious advantages to having a unified transcription. For segmental transcription, this goal is explicitly adopted by the Wikipedia guidelines for the use of the IPA to indicate pronunciation. This obviously requires some compromises, mostly in the direction of indicating phonological distinctions made by large and/or sociolinguistically important groups of English speakers even where those same distinctions are neutralized by other equally significant groups. For example, Wikipedia recommends transcribing the presence or absence of postvocalic /r/ in English words such as *doorway* and *daughter*, even though the initial syllables are pronounced identically by millions of English speakers. (In terms of the standard Wells lexical sets, that is, the recommendation is to distinguish between THOUGHT and FORCE/NORTH). This recommendation applies even to place names in nonrhotic parts of the English-speaking world, such as *Worcester* (transcribed /ˈwʊstər/) and *Melbourne* (transcribed /ˈmɛlbərn/), which would normally be pronounced locally without the final /r/. In the same way, to repeat an example used earlier, Wikipedia transcribes the Anglicized pronunciation of *Marrakesh* as /ˌmærəˈkɛʃ/, observing the distinction between the vowels of TRAP and SQUARE in transcribing the first syllable even though the majority of North American speakers of English do not maintain this distinction before /r/. While such compromises are arbitrary—for example, Wikipedia explicitly recommends ignoring the distinction between NORTH and FORCE made by a relatively small minority of English speakers in Scotland, Ireland, and parts of the United States—they offer the convenience of a uniform transcription system that is fairly easy to apply consistently and interpretably to the pronunciation of hundreds of millions of speakers. In turning away from a similar goal, the developers of MAE-ToBI ostensibly chose strict theoretical adherence to the ideal of transcription based on an analysis of the system of phonological contrasts.

Yet this adherence is illusory, not least because "Mainstream American English" is ultimately the same kind of abstraction as "English" is, namely, an abstraction over a set of idiolects. And in any case, there is a more fundamental problem with claiming that MAE-ToBI is properly based on the phonological contrasts of only one variety of English. This claim presupposes that we have figured out how intonational phonology works well enough to establish what those phonological contrasts are. In my view, our understanding of intonational phonology is actually still fairly primitive, and a standard transcription that purports to be based on a correct phonological analysis prematurely closes off avenues of investigation and theoretical debate. A clear illustration of this problem involves the analysis of phrase-final level pitch. This is seen in a variety of contexts in English, perhaps the most familiar being the stereotypical contour used for calling someone from a distance (e.g., Gibbon 1976; Leben 1976; Ladd 1978; Gussenhoven 1993). In this contour, the stressed syllable is produced on a high level pitch, and the remaining syllables are produced on another level that is often three semitones lower:

Ma—
ry—

In Pierrehumbert's (1980) original analysis of this phenomenon, the level pitch on the final syllable was analyzed phonologically as a combination of a high phrase accent (H-) followed by a low boundary tone (L%). Though this transcription might appear to indicate a contour that falls in pitch from a postnuclear high, the analysis included the proviso that the L% is "upstepped" in this specific context and is realized at the level of the phrase accent. This contrasts with the phrase-final sequence H- H%, where

the upstepping of the H% leads to a rise. This analysis works, in the sense that it can be made to generate the contrast between rising and sustained level phrase-final pitch, and it remains the official MAE-ToBI transcription of contours involving sustained final level.

However, the idea of the upstepped L% was subject to criticism during the burst of theoretical discussion and development in the 1980s and 1990s that followed in the wake of Pierrehumbert (1980). One line of theoretical criticism dealt with the status of phrase accents and boundary tones in general. In Ladd (1983), I proposed that boundary tones need not be part of a well-formed tonal string and argued that the absence of a final rise or fall in phrases ending on steady level pitch was best seen as a reflection of the absence of a boundary tone. This idea was subsequently adopted by Gussenhoven (1984) and Grabe (1998) on the basis of evidence from English, Dutch, and German; it was later incorporated into the ToDI system for transcribing Dutch intonation (Gussenhoven et al. 2003), though not into the GToBI system for German (Grice et al. 1996; Grice and Baumann 2002). Another line of theoretical criticism questioned the appropriateness of opaque phonological analyses that depend on the use of phonetic realization rules like upstep and downstep to yield surface pitch contours. For example, in Ladd (1983), I argued against Pierrehumbert's (1980) original phonological analysis of downstep, which relied on abstract tonal strings subject to rather complex phonetic realization rules (see Arvaniti, chapter 1, section 1.2.4, this volume), and I suggested instead that downstep is an independent phonological choice that can apply to any high accent. As it happened, Pierrehumbert's abstract analysis of downstep was an early casualty of the consultations that led to ToBI, being replaced in the agreed standard by a transparent "downstep feature" transcribed with the traditional Africanist device /!/, but the comparably abstract H- L% analysis of level phrase-final pitch was retained unchanged.

The most specific critique of the H- L% analysis of level phrase-final pitch (Ladd 1983; Mayo, Aylett, and Ladd 1997) focused on the fact that it leaves no obvious way of describing numerous British English intonation patterns in which the pitch gradually falls, or "slumps" after a nuclear accentual rise (the "rise-plateau-slump" contour of Cruttenden 1994). The two issues summarized in the preceding paragraph—the status of boundary tones and the abstractness of tonal strings—are general theoretical questions that it might have been useful to debate more thoroughly while the potential of the general AM approach to describing intonation was being explored. But these questions went largely unexamined, because it was possible for the Pierrehumbert/ToBI analysis to focus on the argument from the rise-plateau-slump and to shelter behind the claim that ToBI deals only with Mainstream American English. ToBI's theoreticians came to favor the view that different analyses reflect genuine differences between language varieties:

> A ToBI framework system devised for one language variety cannot be assumed to be applicable even to other varieties of the same language, without first establishing appropriate intonational and prosodic analyses for each variety. Any claim that the symbolic tags [i.e., ToBI labels] are comparable across varieties must be based on a thorough variety-specific analysis of each of the varieties involved. (Beckman, Hirschberg, and Shattuck-Hufnagel 2005, 42)

This is clearly an intellectually respectable position. The phonological distinctions of what is sometimes termed "Urban North British" varieties of English, to say nothing of Dutch and German, are obviously irrelevant to determining the phonological contrasts of Mainstream American. But they are not so easily dismissed as irrelevant to a broader theoretical understanding of the nature of boundary tones or the scope and power of the phonetic realization rules involved in intonational phonology. I believe

that the determination to hold firm to an agreed standard based on one specific autosegmental analysis of one specific variety of English has impeded the full discussion of important theoretical issues that arise from the general AM approach.

6.4 ToBI and Gradience

The question of the upstepped boundary tone is a mere detail, however, compared to what I see as the central unsolved problem with intonational phonology. This is the phenomenon that Bolinger (1961) termed "gradience." In the ordinary lexicon, word meanings are pretty much discrete, and if two words are different, we have reasonably well-established ways of describing the phonetic properties that signal the difference. The subtlety of the phonetic distinctions doesn't matter: *ankle* and *uncle* are similar enough phonetically to cause problems for many second-language speakers of English, but we know they mean something different and can confidently conclude that the phonetic difference between their initial vowels is a systematic part of English segmental phonology.

We can't proceed so confidently with intonation, because our metalanguage for describing intonational phonetics is not so well developed. And that, I believe, is due to the fact that the phonological categories of intonation, whatever they are, are subject to meaningful gradient variation, or "gradience" for short. The simplest example of this kind of variation involves pitch range. If we emphasize a word by putting a high accent peak on it, we can emphasize it even more by making the high accent peak higher. This kind of thing doesn't happen with *ankles* and *uncles*; phonetic variability in the quality of the first vowel may say something about where the speaker is from, but it doesn't affect the meaning of the words in the way that gradient pitch range modification affects the meaning of an accent.

Importantly, this kind of phonetically gradient variability has the potential to signal pragmatic nuances that seem quite distinct; for example, Hirschberg and Ward (1992) show how a distinction between "uncertainty" and "incredulity" can be signaled by gradient differences in pitch range in the phonetic realization of the English rising-falling-rising contour that they notate L* + H L- H%. It is far from clear how such sharply distinct pragmatic inferences emerge from the application of multiple dimensions of gradient variability to the same intonational category. Nevertheless, as I have argued before (e.g., Ladd 2008a, 151–156), it is important to be aware that this is part of how intonational meaning works. The fact that an intonationally conveyed pragmatic distinction seems categorical does not entail that the corresponding intonational distinction involves categorically distinct phonological elements. Until we understand this better, our phonological analyses are likely to make spurious categorical distinctions.

For example, ToBI draws a categorical distinction between a pitch accent transcribed L + H* and one transcribed H*—informally speaking, a distinction between a rising accent and an accent that is merely high. In the second case, the more or less inevitable rise preceding the peak is considered to be transitional; in the first, it counts as an essential part of the pitch accent's phonetic manifestation. There is little doubt that in some contexts, the presence or absence of a clear accentual rise yields distinct nuances in the pragmatic force of a high accent. Equally, there is little doubt that this distinction is difficult to draw consistently (the evaluation study of intertranscriber reliability in ToBI by Pitrelli, Beckman, and Hirschberg 1994 collapsed the two categories because agreement on the distinction was so poor). This particular descriptive dilemma has a long history that precedes the emergence of AM intonational phonology by several decades. In British school analyses of intonation of the sort pioneered

by Palmer (1922), some analysts (including Palmer himself) posited a single category of falling nuclear accent, while others (such as Kingdon 1958 and Halliday 1967) distinguished fall (Halliday's tone 1) from rise-fall (Halliday's tone 5).

In my opinion, the explanation for this situation is that the distinction between L+H* and H* is a matter of meaningful gradient phonetic variability—Bolingerian gradience—within a single phonological category. The extent (and probably the timing) of the rise in pitch preceding the accentual high are manipulated by the speaker to express a variety of nuances, but the basic phonological choice is in all cases "high accent." However, my point here is not to argue for one analysis over the other, but only to point out that the widespread acceptance of ToBI as a standard—together with the fact that the phonetic basis of the distinction is sometimes readily observable—means that many transcribers of English intonation take it as uncontroversial that there is a categorical distinction between L+H* and H*. The specific issue appears settled, and its potential value as evidence for how gradient phonetic variability affects meaning remains unrealized.

For a more complicated case, consider the transcription of what is often called *uptalk*—the use of rising phrase-final contours on statements in a variety of contexts of interaction. There is a substantial set of research findings showing that variation in the pitch level at the beginning of the rise is systematically related to various pragmatic factors such as whether the utterance is intended as a statement or a question (Warren 2016, 33–45 passim) and whether the utterance involves new or given information (Warren 2016, 62–64). These are empirical results, and we want to incorporate them into our understanding. However, because of the problem of gradience, it's not clear how to do this. For a number of investigators who use some version of the ToBI/AM analysis of pitch contours,[1] the obvious way to annotate this difference is to distinguish between an L* and an H* pitch accent, so that uptalk includes both L* H- H% (for statements, for given information, and so on) and H* H- H% (for questions, for new information, and so on). But any such analysis has serious problems.

First of all, it has the problem that the empirical findings are a matter of statistical generalization, not categorical distinction. It is as if we discovered that when speakers intend to talk about the brother of one of their parents, they use *uncle* 87 percent of the time and *ankle* 13 percent of the time, while references to the joint between lower leg and foot involve *ankle* in 79 percent of cases and *uncle* in 21 percent. Obviously, ordinary word meaning doesn't work this way; if intonational meaning does work like this, then we have to adjust our understanding of what a phonological notation system for intonation is supposed to tell us.

Second, in practice, the notational distinction between L* H- H% and H* H- H% results from the fact that transcribers will try to find a way to represent audible differences that convey an intonational nuance, without regard to the phonological implications. The runaway success of ToBI-style tonal transcription systems for intonation has meant that certain categorical notational distinctions (e.g., L* versus H*) are now widely taken for granted as a useful representation of phonetic reality; the notations L* H- H% and H* H- H% are used to express phonetic observations that may actually involve the extremes of a gradient continuum. Yet if we believe (with, e.g., Pierrehumbert and Hirschberg 1990 and Steedman 2014) that the elements of a ToBI string are distinct phonological categories with distinct meanings, then L* H- H% and H* H- H% are simply different strings, like *ankle* and *uncle*, and they should not be regarded as variants of a single basic phenomenon (uptalk) any more than *ankle* and *uncle* can be regarded as variants of the same word. But this is exactly what frequently happens. The theoretical contradiction between strictly interpreting the notations as a phonological

analysis and informally lumping the two types together as uptalk, which implies some sort of functional and phonological unity, goes unremarked.

The way to resolve the theoretical contradiction seems obvious to me: we need to recognize fewer phonologically distinct intonational categories, and more dimensions of meaningfully gradient phonetic variation. I'm quite sure that uptalk really is a phenomenon, and we want to be able to describe it in such a way that we can separate uptalk from non-uptalk with the same kind of confidence that we can distinguish *ankle* from *uncle*. But to do this, we cannot also give categorically distinct labels to uptalk-starting-low and uptalk-starting-high; we have to be content to describe those nuances in gradient, statistical, phonetic terms. This is hard to do in a convenient notational system. The choice of one transcription symbol over another to express an audible phonetic difference inclines us to believe that we are dealing with two different phonological categories. In segmental transcription, this belief is often justified, but in intonational transcription, it actively hinders the development of our understanding.

Mark Liberman makes essentially the same point in one of his early *Language Log* posts on uptalk:

> It's also worth mentioning that the form of final rises can vary a lot. The starting point and ending point can move around, with respect to the speaker's pitch range and also with respect to the immediately preceding material. The rate of rising and the alignment with the words of the message also vary. *It remains unclear, in my opinion, which aspects of this variation are phonetic dimensions that speakers can choose to deploy to a greater or lesser degree, like the choice of how fast or loud to talk, and which aspects involve qualitatively different alternatives, like the choice between two different words*. (2006, emphasis added)

Liberman's frank acknowledgment that "it remains unclear" which phonetic features of intonation are gradient and which categorical strikes me as exactly right. The trouble with ToBI is that it doesn't acknowledge the existence of such fundamental uncertainty.

Acknowledgments

I have discussed aspects of the trouble with ToBI in several publications, and in trying to put everything together in this essay, I have made use of various phrases, sentences, and even a few whole paragraphs from those earlier works. Specifically, I acknowledge borrowing from chapter 3 of my book *Intonational Phonology* (2008a) and from my reviews (Ladd 2008b, 2017, respectively) of Jun (2005) and Warren (2016).

Note

1. Strictly speaking, the following discussion concerns a problem not with ToBI proper, but only with the unquestioned extension of ToBI-influenced tonal analyses for use in a wide variety of contexts. I include it because it provides a clear example of the difficulty of dealing with intonational gradience in any symbolic transcription system.

References

Arvaniti, Amalia, and Mary Baltazani. 2005. "Intonational Analysis and Prosodic Annotation of Greek Spoken Corpora." In *Prosodic Typology: The Phonology of Intonation and Phrasing*, edited by Sun-Ah Jun, 84–117. Oxford: Oxford University Press.

Ashby, Michael, Patricia Ashby, John Baldwin, Frederika Holmes, Jill House, and John Maidment. 1995. "Broad Transcription in Phonetic Training." In *Proceedings of the XIIIth International Congress of Phonetic Sciences*, edited by Kjell Elenius and Peter Branderud, vol. 3, 170–173. Stockholm: Royal Institute of Technology and University of Stockholm.

Beckman, M. E., and G. Ayers Elam. 1997. "Guidelines for ToBI Labelling." *OSU Research Foundation* 3:30.

Beckman, Mary E., Julia Hirschberg, and Stefanie Shattuck-Hufnagel. 2005. "The Original ToBI System and the Evolution of the ToBI Framework. In *Prosodic Typology: The Phonology of Intonation and Phrasing*, edited by Sun-Ah Jun, 9–54. Oxford: Oxford University Press.

Beckman, Mary E., and Janet B. Pierrehumbert. 1986. "Intonational Structure in English and Japanese." *Phonology Yearbook* 3:255–310.

Birch, Stacy, and Charles Clifton Jr. 1995. "Focus, Accent, and Argument Structure: Effects on Language Comprehension." *Language and Speech* 38:365–391.

Bolinger, Dwight L. 1961. *Generality, Gradience, and the All-or-None*. The Hague, the Netherlands: Mouton.

Brugos, Alejna, Stefanie Shattuck-Hufnagel, and Nanette Veilleux. 2006. "Transcribing Prosodic Structure of Spoken Utterances with ToBI." MIT Open Courseware. https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/.

Cruttenden, Alan. 1994. "Rises in English." In *Studies in General and English Phonetics: Essays in Honour of Professor J. D. O'Connor*, edited by J. Windsor Lewis, 155–173. London: Routledge.

Gibbon, Dafydd. 1976. *Perspectives of Intonation Analysis*. Bern, Switzerland: Lang.

Grabe, Esther. 1998. "Comparative Intonational Phonology: English and German." PhD diss., Max Planck Institute for Psycholinguistics.

Grice, Martine, and Stefan Baumann. 2002. "Deutsche Intonation und GToBI." *Linguistische Berichte* 191:267–298.

Grice, Martine, Matthias Reyelt, Ralf Benzmüller, Jörg Mayer, and Anton Batliner. 1996. "Consistency in Transcription and Labelling of German Intonation with GToBI." Paper presented at the Fourth International Conference on Spoken Language Processing. https://ieeexplore.ieee.org/abstract/document/607958.

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

Gussenhoven, Carlos. 1993. "The Dutch Foot and the Chanted Call." *Journal of Linguistics* 29:37–63.

Gussenhoven, Carlos, Toni Rietveld, Joop Kerkhoff, and Jacques Terken. 2003. "Transcription of Dutch Intonation" (interactive course). 2nd ed. http://todi.let.kun.nl/ToDI/home.htm.

Halliday, M. A. K. 1967. *Intonation and Grammar in British English*. The Hague, the Netherlands: Mouton.

Hirschberg, Julia, and Gregory Ward. 1992. "The Influence of Pitch Range, Duration, Amplitude and Spectral Features on the Interpretation of the Rise-Fall-Rise Intonation Contour in English." *Journal of Phonetics* 20:241–251.

Hualde, José, and Pilar Prieto. 2016. "Towards an International Prosodic Alphabet (IPrA)." *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7 (1): 5.

Jun, Sun-Ah, ed. 2005. *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

Kingdon, Roger. 1958. *The Groundwork of English Intonation*. London: Longman.

Kiparsky, Paul. 2014. "New Perspectives in Historical Linguistics." In *The Routledge Handbook of Historical Linguistics*, edited by Claire Bowern and Bethwyn Evans, 64–102. London: Routledge.

Ladd, D. Robert. 1978. "Stylized Intonation." *Language* 54:517–539.

Ladd, D. Robert. 1983. "Phonological Features of Intonational Peaks." *Language* 59:721–759.

Ladd, D. Robert. 2008a. *Intonational Phonology*. 2nd ed. Cambridge: Cambridge University Press.

Ladd, D. Robert. 2008b. Review of *Prosodic Typology: The Phonology of Intonation and Phrasing*, by Sun-Ah Jun. *Phonology* 25:372–376.

Ladd, D. Robert. 2017. Review of *Uptalk: The Phenomenon of Rising Intonation*, by Paul Warren. *Journal of the International Phonetic Association* 47:369–374.

Leben, William. 1976. "The Tones in English Intonation." *Linguistic Analysis* 2:69–107.

Liberman, Mark. 2006. "Uptalk Is Not HRT." *Language Log*. http://itre.cis.upenn.edu/~myl/languagelog/archives/002967.html.

Mayo, Catherine, Matthew Aylett, and D. Robert Ladd. 1997. "Prosodic Transcription of Glasgow English: An Evaluation Study of GlaToBI." In *Intonation: Theory, Models and Applications: Proceedings of an ESCA Workshop, Athens*, edited by A. Botinis, G. Kouroupetroglou, and G. Carayiannis, 231–234. Athens: ESCA and University of Athens.

Moulton, William G. 1947. "Juncture in Modern Standard German." *Language* 23:212–226.

Palmer, Harold. 1922. *English Intonation, with Systematic Exercises*. Cambridge: Heffer.

Pierrehumbert, Janet B. 1980. "The Phonology and Phonetics of English Intonation." PhD diss., MIT.

Pierrehumbert, Janet B., and Julia Hirschberg. 1990. "The Meaning of Intonational Contours in the Interpretation of Discourse." In *Intentions in Communication*, edited by P. R. Cohen, J. Morgan, and M. E. Pollack, 271–311. Cambridge, MA: MIT Press.

Pike, Kenneth L. 1943. *Phonetics: A Critical Analysis of Phonetic Theory and a Technique for the Practical Description of Sounds*. Ann Arbor: University of Michigan Press.

Pitrelli, John, Mary E. Beckman, and Julia Hirschberg. 1994. "Evaluation of Prosodic Transcription Labeling Reliability in the ToBI Framework." Paper presented at the Third International Conference on Spoken Language Processing. https://www.isca-speech.org/archive/icslp_1994/i94_0123.html.

Price, Patti J., Mari Ostendorf, Stefanie Shattuck-Hufnagel, and Cynthia Fong. 1991. "The Use of Prosody in Syntactic Disambiguation." *Journal of the Acoustical Society of America* 90:2956–2970.

Silverman, Kim, Mary Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert, and Julia Hirschberg. 1992. "ToBI: A Standard for Labeling English

Prosody." Paper presented at the Second International Conference on Spoken Language Processing, Banff, Alberta, Canada.

Steedman, Mark. 2014. "The Surface-Compositional Semantics of English Intonation." *Language* 90:2–57.

Venditti, Jennifer. 1997. "Japanese ToBI Labelling Guidelines." *Working Papers in Linguistics, Ohio State University Department of Linguistics* 50:127–162.

Warren, Paul. 2016. *Uptalk: The Phenomenon of Rising Intonation*. Cambridge: Cambridge University Press.

Wightman, Colin, Stefanie Shattuck-Hufnagel, Mari Ostendorf, and Patti Price. 1992. "Segmental Durations in the Vicinity of Prosodic Phrase Boundaries." *Journal of the Acoustical Society of America* 91:1707–1717.