

Notes on the Phonology of Prominence

D. Robert Ladd
 Department of Linguistics
 University of Edinburgh, Scotland

ABSTRACT

An explicitly phonological approach to prosody, of the sort provided by autosegmental and metrical theory, is necessary for understanding the relationships among different prosodic phenomena. Prominence is not a simple phonetic property of syllables but a complex phenomenon reflecting the metrical structure of a phrase or utterance, the location of (intonational) pitch accents, and the utterance's paralinguistic aspects.

INTRODUCTION

It is difficult to give a coherent brief overview of the topic "prominence", because the topic itself is not one but several. A quick glance at the six workshop papers grouped together under this rubric shows that we are dealing with a many-headed beast: one (Cruttenden) is about the pragmatics of sentence-stress, one (Gronnum) deals with the relation between stress and rhythm, one (Grabe et al.) studies the use of certain English-specific prosodic cues in human sentence-processing, and one (Hermes & Rump) explores the contribution of pitch range to perceived prominence. Only two - Campbell's and Fant & Kruckenberg's - deal with something that is clearly the same topic, namely the role of a variety of acoustic cues, especially duration cues, in signalling both prominence and phrase boundaries. Given this variety, it is pointless to try to discuss each of the six papers and relate them to a single core of theoretical issues and ideas. Instead I would like use them as points of reference in an argument for approaching the study of prominence - and indeed, prosody in general - in explicitly phonological terms.

Specifically, I wish to argue that the theoretical framework provided by autosegmental and metrical phonology (Liberman & Prince 1977, Pierrehumbert 1980, and much work since then; for a review relevant to intonation see Ladd 1992) is essential to reconciling the diversity of methodologies and points of view of the papers in this section. I hasten to add that I am more or less agnostic about many specific issues within autosegmental and metrical phonology, and those are not my topic here. My point in invoking the general autosegmental/metrical approach is to suggest that it provides an appropriate way of thinking about the *relationships* among prosodic phenomena, and between prosodic form and function.

By this I mean two things. First, observable acoustic properties like F0, duration, and intensity are not direct correlates of functional categories like focus, nor the direct realisation of morphosyntactic structures. They are rather the correlates of phonological categories and phonological structures, and as such may only indirectly reflect focus, phrase boundaries, and so on. I shall return to this point below. Second and more specifically, what I find important about the autosegmental/metrical point of view is the idea that stress, duration, rhythm, and prosodic grouping form one coherent cluster of phenomena, and intonation forms another. This is what is implied by a representation of the sort in Fig. 1. The prominence relationships and surface constituency - stress and phrasing, in other words - are represented in the metrical tree, and the intonation is represented in the tonal string. At the risk of being accused of Procrustean misrepresentation, I think that many of the observations in Gronnum's paper are based on just such an implicit distinction between duration, stress, and foot structure on the one hand, and pitch features and rules for their alignment with prominent syllables on the other.

of segments or syllables. It would admittedly be stretching things to claim that Campbell's notion of "duration contour" is inspired by metrical phonology, but metrical phonology is at least consistent with Campbell's leading idea that duration has a complex structure, and that by discovering that structure we make it possible to identify prominence acoustically without reference to F0.

PHONOLOGY AND FUNCTION

A more general consequence of a phonological approach to prosody is that it makes us less likely to look for direct links between phonetic form and communicative function. Everybody knows that prosodic features have a bewildering variety of "functions" - signalling focus, emphasis, phrasing distinctions, lexical distinctions, speaker attitude, and many more. To my mind, this variety of functions is bewildering only if we assume that the link between sound and meaning is direct.

By way of illustration, consider the "delimitative" cues to dividing up the stream of speech into chunks corresponding to words and phrases. There are many delimitative cues in segmental phonology, but these are essentially accidental. For example, in American English one can clearly distinguish the utterances *can't race* and *can trace* by the allophonic variation in the three segments /n/, /t/, and /r/, but it is not very revealing to say that the "function" of segmental allophonic variation is delimitative. Segmental allophony exists (for whatever reason), and it may be exploited by listeners when it happens to provide "delimitative" information.

In the same way, I believe that if we find prosodic features playing a delimitative or otherwise disambiguating role in sentence processing, we should always assume that this role is a useful accident, and only later consider the possibility that delimitation is the central function of a given prosodic feature. This is relevant to the paper by Grabe et al. I find it unremarkable that listeners can distinguish the unshifted stress pattern of *Chinese teacher* 'teacher of Chinese' from the shifted pattern of *Chinese teacher* 'teacher who is Chinese', even when they have heard only the first syllable, or that listeners should make syntactic use of that information as soon as they get it. These findings tell us about human sentence processing, not about the function of prominence. That is, Grabe et al.'s results tell us how quickly listeners can make use of information from a variety of sources, not that the "function" of stress shift is syntactic disambiguation. Like segmental allophony, stress shift exists - apparently for reasons of maintaining as regular an alternating speech rhythm as possible. In the specific case of *Chinese teacher*, stress shift happens to have an effect that we can put to use in syntactic processing, just like the allophony of /n/, /t/, and /r/. As far as I am aware, Grabe et al. interpret their findings in more or less the way suggested here, but I have discussed this example at length because it is precisely the sort of finding that might be taken as evidence for the essentially delimitative function of prominence, or for a "functional" explanation of stress shift. I do not think that such interpretations are worth pursuing.

If the linguistic function of prominence is not delimitative, what is it? In my opinion the phenomena dealt with in Cruttenden's paper are closer to the essence of what prominence is for. Prominence is like a grammatical category, similar to number or case. As with number or case, there are broad similarities across languages in the way prominence is distributed, but also language-specific pragmatic, syntactic, and phonological principles. Cruttenden's paper discusses a couple of specific points on which languages differ. I find this work very useful as a corrective to the idea that, as a universal of intonation systems, pitch prominences go on focused or otherwise emphasised words. Focus may help to govern the distribution of pitch accents, but focus can apply to whole constituents, and accents, by definition, must be associated with individual words. We must therefore assume that phonological and other rules come into play here in building the prominence structure of an utterance on the basis of the intended focus (cf. e.g. Gussenhoven 1983, von Stechow & Uhmman 1986). The simple idea of a direct correspondence between pitch accent and some intuitive notion of "focus" on individual words does

explain a substantial percentage of cases, but it ignores the language-specific differences, and gets in the way of understanding the remaining percentage of cases that don't fit the simple pattern.

PHONOLOGY AND PARALINGUISTICS

So far I have suggested that a phonological perspective on prominence will keep us from looking for too direct a link between simple phonetic properties and communicative functions. In the remainder of the paper I wish to explore a third puzzling area of prosody where I believe we can make progress by considering the phonological implications of experimental findings. This is the problem of "gradience".

Prosody appears to differ from the rest of phonology in the way it treats continuous acoustic dimensions such as pitch range and duration. Instead of such acoustic continua being divided up into discrete categories, as is usual in phonological organisation, at least in some cases gradual acoustic changes can result in steadily perceptible non-categorical shades of meaning. This is clearly the way a lot of paralinguistic signalling works - the broader the smile, the happier the smiler - but it is not very common in language proper, except in prosody. Perhaps the most common example is one directly relevant to prominence, and well-documented in the papers by Fant & Kruckenberg and by Hermes & Rump: the gradual increase in emphasis or contrastiveness associated with a gradual increase in overall pitch range or with gradual increases in intensity or duration. Such gradient prominence is allowed for even in rigorously phonological descriptions. Within the autosegmental/metrical approach to prosody, both the degree of emphasis of individual accents within phrases and the overall pitch range of individual phrases within utterances are normally said to be freely variable.

I have argued elsewhere (Ladd forthcoming) that the concept of free gradient variability of prominence poses serious theoretical and empirical problems. Here I wish to go beyond my earlier arguments and propose that much of what is taken as evidence for gradient prominence of individual words actually reflects paralinguistic differences of *overall* prominence, emphasis, interest, etc., which affect the perception of the most prominent word. That is, when listeners are presented with a short test utterance under experimental conditions and asked to judge the degree of prominence of the accented word, what they are really doing is judging the degree of emphasis *of the utterance as a whole* and applying that judgement to the accented word. This is exactly analogous to something that Fant & Kruckenberg report for word prominence and syllable prominence: they found that judgements of a word's degree of prominence correlate highly with judgements of the prominence of the word's stressed syllable. I therefore claim that for the most part prominence really is a largely categorical, phonological matter: this is what is implied by the presence or absence of pitch accent in the tonal string. Gradient prominence on an accented word is simply gradient overall prominence or emphasis, which the speaker, under experimental conditions, interprets as affecting the accented word. I disagree with the notion that the prominence of every individual accent can vary freely; I take quite literally the categorical nature of relative prominence implied by the autosegmental/metrical representation.

Of course, in some cases individual words really *do* have extra emphasis or prominence within a given utterance, like the word *do* in a colourful reading of this sentence. In this case the presence of extra emphasis is signalled categorically by the presence of extra intensity, extreme pitch range, and so on. This claim is also supported by Fant & Kruckenberg: they found that intensity comes into play as a correlate of perceived degree of prominence only for "emphatic" accents. Implicitly, that is, they draw a categorical rather than a gradient distinction between those accents for which intensity is relevant and those for which it is not. I believe this is correct: I think there is a fairly discrete boundary between a neutral reading like *He's Ukrainian* (interpreted as providing new information about the subject), and a paralinguistically marked emphatic reading *He's UKRAINIAN* (interpreted as em-

phasing the adjective, e.g. to contrast with erroneously presupposed *Russian*). More generally, I think that many paralinguistic dimensions - certainly including local emphasis - have a "neutral" range, and it is only beyond this neutral range that classic gradience sets in.

If this general interpretation of gradient prominence is to hold up, we need to distinguish paralinguistic effects on an utterance as a whole from paralinguistic effects on individual words - and to distinguish both from phonological distinctions of relative prominence. I believe this can be done. To begin with, we have the results of several studies on pitch range in at least three different languages (Lieberman & Pierrehumbert 1984 on English, Bruce 1982 on Swedish, Pierrehumbert & Beckman 1988 on Japanese, and others). These studies all find that the relative pitch range of individual accents in a phrase and of individual phrases in an utterance remains *virtually constant* when the overall pitch range of the utterance is experimentally varied for paralinguistic reasons (e.g. by getting subjects to "speak up" or to talk as if they were "more involved"). That is, gradient variability generally seems to affect the pitch range of utterances or larger chunks of discourse as a whole without affecting the pitch range relations within the utterance. This is very difficult to explain if the pitch range of each accent and each phrase is freely and independently variable, but makes sense if the relative prominence of accents within an utterance is tightly constrained by the phonology and not controlled paralinguistically.

More concretely, Ladd, Verhoeven and Jacobs (forthcoming; henceforth LVJ) have provided experimental evidence that accentual prominence is perceived in a way consistent with the proposal just made. Their study replicates and extends an earlier discovery by Gussenhoven & Rietveld (1988). Gussenhoven & Rietveld found that, in an utterance with two accent peaks (e.g. a sentence like *Her mother's a lawyer*), a decrease in pitch range on the first accent causes a *decrease* in the perceived prominence of the *second* accent. If the degree of emphasis on accents were independently variable, one would expect the opposite effect - i.e. one would expect a decrease on one accent to enhance the prominence of the other. LVJ explain Gussenhoven & Rietveld's finding by suggesting that, for moderate peak heights, gradient variability applies to the pitch range *of the utterance as a whole*, not to each accent individually. This means that lowering one accent lowers the perceived degree of overall emphasis of the utterance, and hence the degree of emphasis on all the accents of the utterance. This is again the utterance level analogue of the word/syllable effect found by Fant & Kruckenberg.

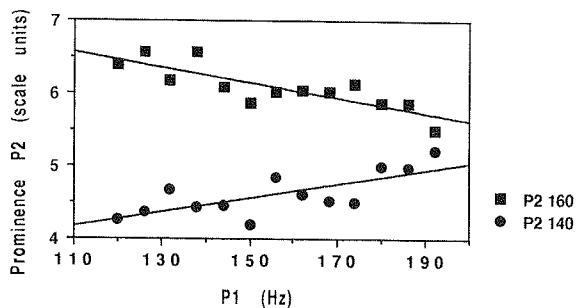


Fig. 2 Perceived prominence of the second accent peak of a two-accent utterance, as a function of the F0 of the first accent peak. The two F0 levels of the second peak (140 and 160 Hz) show different effects. From Ladd, Verhoeven and Jacobs (forthcoming).

However, LVJ also found that Gussenhoven & Rietveld's effect is reversed (and the more expected effect obtained) when the second accent peak is *relatively high*. This is shown in Fig. 2. LVJ suggest that this reversal reflects a categorical difference between "normal" or "neutral" accent peak height - with which pitch range is evaluated globally - and "emphatic" peak height - which overrides global pitch range and treats every accent in its own right. If LVJ are correct, gradient variability applies across-by-accent only beyond the threshold of the emphatic. This, as noted earlier, is quite consistent with Fant & Kruckenberg's implicitly categorical distinction between emphatic and non-emphatic accents.

Obviously, the most rigorous test of the idea that there is a categorical distinction between neutral and emphatic accents will come from something akin to a categorical perception experiment. A promising pilot study along these lines has just been done under my supervision by Rachel Morton, as the basis of her undergraduate honours dissertation in Edinburgh. Under certain conditions Morton's results seem to show a stepwise increase in perceived prominence as overall pitch range increases, suggesting a categorical shift from neutral to emphatic. But further work needs to be done before this result can be regarded as established.

REFERENCES

- G. Bruce (1982), "Developing the Swedish intonation model", *Working Papers* (Lund University Dept. of Linguistics) no. 22, pp. 51-116.
- J. Edwards & M. E. Beckman (1988), "Articulatory timing and the prosodic interpretation of syllable duration", *Phonetica* 45, pp. 156-174.
- D. B. Fry (1955), "Duration and intensity as physical correlates of linguistic stress", *Journal of the Acoustical Society of America* 27, pp. 765-768.
- C. Gussenhoven (1983), "Focus, mode, and the nucleus", *Journal of Linguistics* 19, pp. 377-417.
- C. Gussenhoven & T. Rietveld (1988), "Fundamental frequency declination in Dutch: testing three hypotheses", *Journal of Phonetics* 16, pp. 355-369.
- D. R. Ladd (1992), "An introduction to intonational phonology", in *Papers in Laboratory Phonology II: Segment, Gesture, Prosody*, ed. by G. J. Docherty & D. R. Ladd (Cambridge University Press, Cambridge), pp. 321-334.
- D. R. Ladd (1993), "In defense of a metrical theory of intonational downstep", in *The Phonology of Tone: The Representation of Tonal Register*, ed. by H. v.d.Hulst and K. Snider (Mouton DeGruyter, Berlin), pp. 109-132.
- D. R. Ladd (forthcoming), "Constraints on the gradient variability of pitch range (or) Pitch Level 4 Lives!", to appear in *Papers in Laboratory Phonology III*, ed. by P. Keating (Cambridge University Press, Cambridge).
- D. R. Ladd, J. Verhoeven, & K. Jacobs (forthcoming), "Influence of adjacent pitch accents on each other's perceived prominence: two contradictory effects", to appear in *Journal of Phonetics*.
- M. Liberman & J. Pierrehumbert (1984), "Intonational invariance under changes in pitch range and length", in *Language Sound Structure*, ed. by M. Aronoff and R. Oerhle (MIT Press, Cambridge MA), pp. 157-233.
- M. Liberman & A. Prince (1977), "On stress and linguistic rhythm", *Linguistic Inquiry* 8, pp. 249-336.
- J. Pierrehumbert (1980), *The Phonology and Phonetics of English Intonation*, PhD Dissertation, MIT.
- J. Pierrehumbert & M. Beckman (1988), *Japanese Tone Structure* (MIT Press, Cambridge MA).
- A. von Stechow & S. Uhmman (1986), "Some remarks on focus projection", in *Topic, Focus, and Configurationality*, ed. by W. Abraham & S. de Meij (Benjamins, Amsterdam), pp. 295-320.