Phonology Yearbook 3 (1986) 255–309
Printed in Great Britain

Intonational structure in Japanese and English

Mary E. Beckman
Ohio State University & AT&T Bell Laboratories

Janet B. Pierrehumbert
AT&T Bell Laboratories

1 Introduction

Comparisons between Japanese and English prosodics have usually either focused on the strikingly apparent phonetic differences between the stress patterns of English and the tonal accent patterns of Japanese or concentrated upon formal similarities between the abstract arrangements of the stresses and tones. A recent investigation of tone structure in Japanese (Pierrehumbert & Beckman forthcoming), however, has convinced us that if the proper prosodic phenomena are compared, far more pervasive similarities can be discovered and of a much more concrete sort than hitherto suspected. In particular, there is now extensive evidence that Japanese tonal patterns are very sparsely specified, which suggests that they are much more similar to English intonational structures than earlier descriptions would have allowed. Our investigation of Japanese intonational structure has also revealed a highly structured hierarchy of prosodic phrase types that are defined by various aspects of tone placement and scaling. The existence of these levels of phrasing in Japanese suggests that it would be fruitful to look again at English to see whether similar phrase levels exist in English intonation patterns.

This paper reviews our findings about Japanese intonational structure and re-examines English intonation in the light of these findings. It begins at the most basic level, that of pitch accents and any prosodic units that might be immediately related to them, and then proceeds through higher organisational levels recently discovered in Japanese or already known to exist in English. At each level it reviews the relevant facts about intonational structure in the two languages or outlines what further data are necessary to ascertain how similar or different the two languages are.

2 Accents and the accentual phrase

2.1 The inventory of pitch accents

One of the more salient similarities between English and Japanese is that both have tonal phenomena that can be described in terms of the notion 'accent'. The precise nature of accent is not identical in the two languages, but there is a fundamental likeness in that it involves an association between some well-defined pitch shape in the melody (the 'pitch accent') and some syllable in the text that, by virtue of the association, is 'accented'.

In Japanese, pitch accents are the most straightforward component of an intonation contour. They have a fixed shape consisting of a sharp decline around the accented syllable, a decline that is usually analysed as a drop from a H tone to a L.[1] The precise alignment between this HL shape and the text is also very simple. The place of the accent is lexically contrastive, as in *ka'mi* 'god' *vs. kami'* 'paper', and therefore must be specified in the lexicon. Following Pulleyblank (1983) and Poser (1984), we propose to accomplish this specification by treating the accent as a lexically linked H tone.

The paucity of possible shapes in Japanese is in marked contrast to the rich inventory of shapes that constitute the pitch accents of English. Pierrehumbert (1980) identified seven possible tonal configurations for pitch accents in English, and although recently the H*+H pitch accent has been eliminated as a possible pattern,[2] there still remain the six pitch accent shapes H*, L*, H*+L, H+L*, L*+H and L+H*. When some portion of an utterance is to be accented in English, it is associated to one of these six shapes by linking the starred (or metrically strong) tone of the pitch accent to the accented syllable in the text,[3] as illustrated in Fig. 1. This figure shows several contrasting pitch accent types associated to the word *orange* in the phrase *an orange ballgown.* The first rendition in this figure is a rather neutral way to say the phrase; it would be a perfectly ordinary answer to the question 'What's that?' The second rendition used as an answer to the same question might convey a real or sarcastically feigned judiciousness. And the third might be an expression of astonishment or an impatient reminder. Note that the inventory of pitch accents (and the analyses of the different pitch accent shapes associated to *orange* in Fig. 1) includes some accents that consist of a single tone and others that have two tones. Since the difference between single-tone accents and two-tone accents will be crucial in the subsequent discussion, we motivate it further at this point by comparing it to two alternative treatments of the same phenomena that have been suggested to us.

The first alternative decomposes any bitonal accent into a single-tone pitch accent plus a separate phrasal tone of some sort, either a phrase boundary tone or a 'phrase accent'. (The phrase accent in English is a tone that fills the space after the last pitch accent in a phrase.) Since an accurate description of the English intonation system independently re-

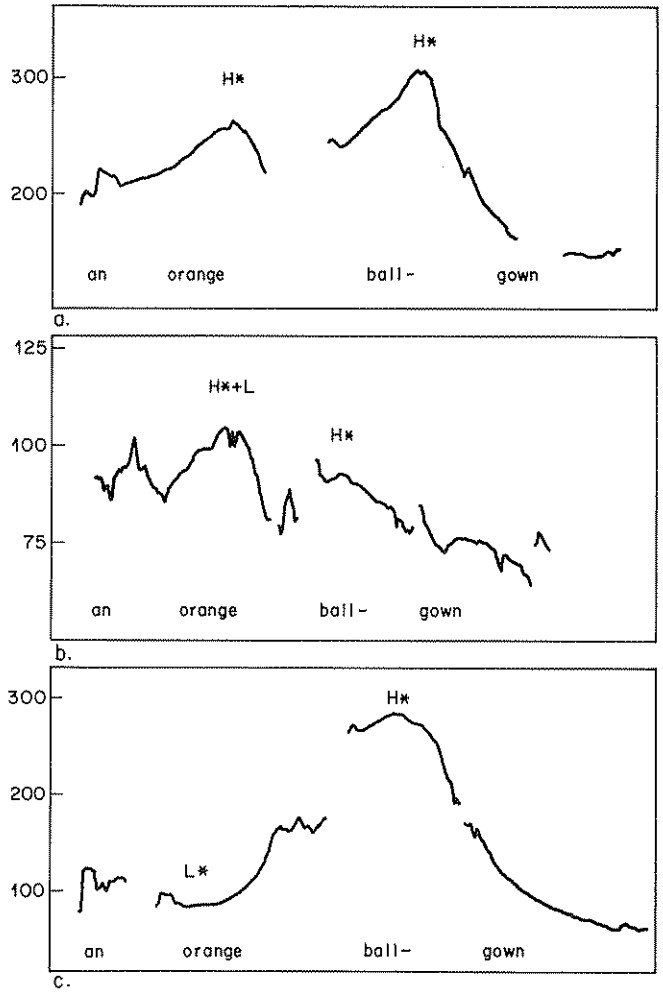

Figure 1

Utterance *an orange ballgown* with (a) H* H* L L% – standard declarative intonation; (b) H*+L H* L L% – a downstepping accent on *orange*; (c) L* H* L L% – surprise–redundancy contour.

quires such phrasal tones, this alternative treatment claims to rid the description of an added abstract category that only complicates the analysis of any intonation contour. Such an alternative analysis, however, makes it difficult to account for the temporal relationship between the two tones of the bitonal pitch accents. A defining characteristic of a pitch accent in English is that it is produced at a rhythmically strong syllable. This is true both of a single-tone accent and of a two-tone accent; the starred tone is phonologically linked to the strong syllable, and the unstarred tone of a two-tone accent precedes or follows it at some given space in time. The phrasal

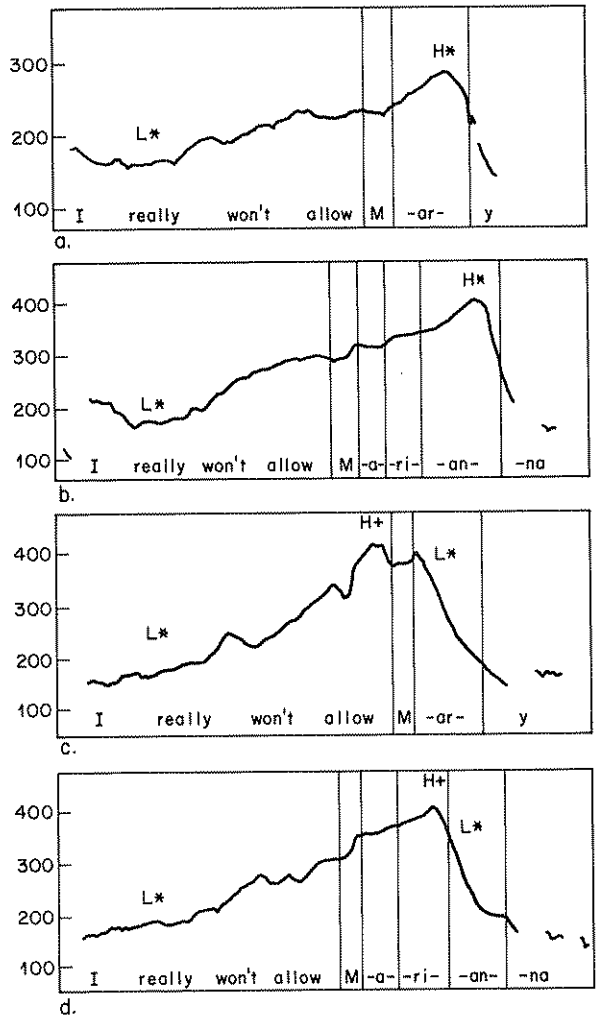

Figure 2

Utterance *I really won't allow Mary/Marianna* with a L* pitch accent on *really* and either (a)–(b) a single tone H* accent or (c)–(d) a bitonal H + L* accent on the name *Mary* or *Marianna*.

tones, on the other hand, are defined by their positions relative to the phrase edge; a boundary tone stays at the phrasal boundary regardless of the rhythmic pattern of the phrase, and the phrase accent fills the space between the last accent and the phrasal boundary. It is not very difficult, therefore, to find intonation contours containing bitonal pitch accents that cannot plausibly be accounted for by analysing the unstarred tone as a phrasal tone, as illustrated in Fig. 2. The two contours in Figs. 2a and 2b first show how the peak that realises a single-tone H* accent is aligned to

the text. In both cases the Fo peak falls near the end of the associated stressed syllable, the first syllable of the word *Mary* in Fig. 2a and the third syllable of the word *Marianna* in Fig. 2b. This is typical for nuclear H* accents when the nuclear stress is not on the last syllable in the phrase. Figs. 2c and 2d illustrate the alignment of a peak when it realises the unstarred H of a bitonal H+L* accent. In this case, the peak occurs just before the onset of the stressed syllable. The different alignment of the peak relative to the stressed syllable is very salient perceptually and also corresponds to a clear difference in interpretation. The pattern in 2a and 2b could be used as an impatient reassertion, whereas that in 2c and 2d has a tone of peevishness or disgruntlement without any implication that the idea should already be known to the listener. The intonation contours in the two sets of utterances must be qualitatively different. How then are we to interpret the peak in 2c and 2d? In the utterance shown in Fig. 2c, there is no audible phrase break before *Mary*, but the occurrence of the peak just before the beginning of the stressed word could perhaps be a boundary tone for some very low-level phrase boundary. In Fig. 2d, however, the peak occurs in the middle of the word, just before the stressed third syllable. This is a very unlikely location for any sort of phrase boundary, making the analysis of the peak as a boundary tone or phrase accent extremely implausible, if not impossible. Such regular patterns of peak alignment relative to the accented syllable are compelling evidence against this first alternative to bitonal accents.

The second alternative to our inventory of two singleton and four bitonal pitch accents is proposed by Ladd (1983). Ladd admits some bitonal accents, but proposes to handle many of the contrasts that we would attribute to pitch accent type with a single feature of peak alignment. Specifically, he rejects the star notation to distinguish between different alignments of the pitch accent with the stressed syllable and he questions the existence of a L+H accent anywhere other than in nuclear position (where he uses it to denote rising configurations involving L* pitch accents and H phrase accents). He does away with our two different types of L+H accents in different ways. He reanalyses the contrast between H* and L*+H accents in prenuclear position as a contrast in the use of an added stylistic feature of peak alignment. That is, he claims that our L*+H accent is a singleton H accent with a positive specification for the feature [delayed peak]. And he does away with L+H* accents apparently by reinterpreting the unstarred L tone as properly belonging to a preceding H*+L accent. That is, sequences which Pierrehumbert (1980) would transcribe as H* L+H*, Ladd transcribes as HL H.[4] We see problems with both of these proposed reinterpretations. The reanalysis of L*+H entails a loss in phonetic explicitness, because the difference between L*+H and H* involves a contrast not only in the timing of the peak but also in the Fo level immediately preceding the peak. A L*+H accent has a valley on the stressed syllable which is as low as that for any L* accent, whereas the H* accent has no such valley. The rejection of L+H* similarly entails a loss in descriptive adequacy. As examples in Pierre-

humbert (1980) clearly show, L + H* contrasts with H* not only after H*, but also utterance-initially and after L*. The utterance-initial cases might be dismissed as involving a L initial boundary tone, but it is difficult to see how Ladd's system would handle the contrast after L*. A final criticism of Ladd's system is that it also fails to describe the full range of contrasts available with H + L accents, such as the contrast between H + L* and H* illustrated in Fig. 2. In the H + L* accent, the peak precedes the stress rather than following it, so the feature [delayed peak] cannot be responsible. In a system involving bitonal accents with starred and unstarred tones, however, this difference is exactly one of the contrasts predicted, and is one of the many reasons for proposing the large inventory of pitch accent types that contrasts the English intonation system to the Japanese intonation system.

The difference between Japanese and English in the number of possible pitch accent shapes is related to another important difference between the two languages. It was noted above that Japanese pitch accents can be treated as lexically linked tones. It is conceivable that a language having more than one possible pitch accent shape also could link the pitch accent in the lexicon and thus use the contrasting shapes as part of the phonological specifications of individual lexical items. The phonemic contrast between accent 1 and accent 2 in Stockholm Swedish, for example, could be analysed as a contrast between a HL pitch accent with the L linked to the accented syllable and a HL with the H linked to the accented syllable.

English, however, does not use pitch accents in this way. The pitch shape for the accent in English is not specific to the accented lexical item. The choice of pitch shape can never contrast different lexical items, but instead contrasts different intonational meanings, as illustrated by the different renditions of *an orange ballgown* presented above in Fig. 1. The meanings of these three answers are quite distinct, and the distinction must be carried by the difference in the shape of the pitch accent on the word *orange*. Yet the difference in meanings is certainly not a difference of lexical choice for the word carrying the accent. The different possible pitch accent shapes thus must be treated as something provided by the intonational system rather than by the lexicon. The only aspect of the pitch accent that is lexically specified is the accent locus, the syllable or syllables in the word to which the starred tone of the accent can be associated.

In Japanese, by contrast, the single possible pitch accent type can be treated as a lexically specified tone shape, as in Swedish. The fact that in Japanese, different melodies at the same accent locus do not contrast lexical items would then be a part of the morphotactics of the language, analogous to those Swedish dialects in which there is only accent 1. Of course the pitch accent in Japanese could also be treated as part of the inventory of intonational structures by specifying only starred syllables in the lexicon and inserting a H* + L along with other intonationally specified tones. This is essentially the technique proposed by Haraguchi (1977), although the spirit of this treatment is somewhat different, since Hara-

guchi's rule associating the starred H tone of the melody to the starred tone bearing unit of the text is part of the derivational phonology of word formation rather than a postlexical phrasal rule. This treatment would make the pitch accent in Japanese rather more similar to the pitch accent in English, but it would not eliminate the crucial difference in the function of the pitch accent. The fact that no other configuration is ever associated to the starred tone bearing unit in Japanese means that the pitch accent cannot partake in the contrast between different intonational meanings in Japanese. The differences produced by varying the pitch accent shape in English cannot be signalled by the choice of pitch accent in Japanese. Thus, whether the pitch accent in Japanese is identified with the lexically linked accent shapes of Swedish or with the intonationally inserted accent shapes of English, Japanese presents a degenerate type that does not allow the lexical or intonational contrast.

2.2 Unaccented phrases and phrase boundaries

Although Japanese tone structure does not allow paradigmatic contrasts in pitch accent shape, it does have a different paradigmatic contrast involving the accent. We said above that the location of the pitch accent is lexically distinctive. There are some lexical items, however, that have no inherent accent locus. A few of these items are unaccented clitics, exactly like the stressless clitics of languages like English. The vast majority of unaccented lexical items, however, are perfectly ordinary words. In the derivation of larger forms such as compound nouns, accents may be inserted into these items by phonological rule, but the insertion of pitch accents is not an inevitable result of producing a well-formed intonation contour. In other words, even at the surface phonological level, there are phrases with no pitch accent, and the presence or absence of accents within an utterance is not determined by some sentential property such as intonational focus, but rather is a lexical property of the components of phrases within the utterance.

In order to understand better this paradigmatic contrast between accent and lack of accent in Japanese, we must introduce the lowest level of phrasing that is well defined by the intonation pattern, a unit that we call the ACCENTUAL PHRASE. The defining mark of the accentual phrase is the presence of two delimitative tones whose occurrence is determined solely by the prosodic phrase structure of the utterance. One of these delimitative tones is a high tone, the PHRASAL H. The phrasal H is phonologically associated to the second sonorant mora of the accentual phrase unless this would conflict with the lexical association of an accent H to the first mora. The other delimitative tone is a low boundary tone that occurs at the beginning of every utterance and at the end of every accentual phrase. When this L tone is not absolutely utterance-initial, we interpret it as properly belonging to the preceding accentual phrase, although a phonological rule associates it to the first syllable of the following phrase if that syllable does not have an associated H tone. The boundary L and the

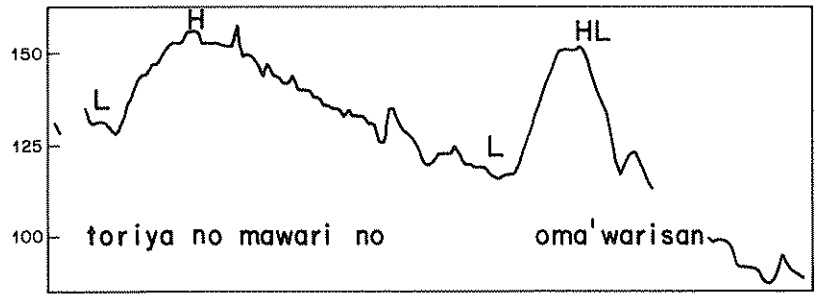

Figure 3

Portion of utterance *Kore-wa toriya-no mawari-no oma'warisan-no hanasi desu* with *toriya-no mawari-no* produced as a single accentual phrase. Speaker ST.

phrasal H together produce a rising pitch shape which marks the beginning of every new accentual phrase.

Note that our description of the delimitative rise that characterises accentual phrases in Japanese differs in several important ways from that of earlier accounts. For example, where we posit a single H tone that can be linked by the phrasal phonology only to a single tone bearing unit toward the beginning of the phrase, earlier autosegmental accounts have associated this H tone to every following tone bearing unit in the phrase until the L of an accent is encountered. Evidence against such H-tone spreading is shown in Fig. 3. In this fundamental frequency contour, the long unaccented accentual phrase *toriya-no mawari-no* has no marked inflections after the initial rise from the L tone on the first syllable. Instead of falling sharply around the L tone on the first syllable of the following phrase, the Fo merely glides smoothly downward in a simple phonetic interpolation between the initial phrasal H and the following boundary L. Pierrehumbert & Beckman (forthcoming) describe an experiment in which many utterances of such unaccented phrases were examined. In this experiment, the number of intervening syllables between the phrasal H on the *toriya*... phrase and the boundary L on the following phrase varied between one and six. It was found that the slope of a regression line fitted to the points between the Fo maximum and minimum corresponding to these linked tones varied inversely with the number of intervening syllables, as illustrated in Fig. 4. Whereas this and other results from the experiment are consistent with the idea that a simple phonetic interpolation connects the two linked tones, they are difficult to reconcile with any account that posits a H tone associated to every syllable in between. Our account therefore differs from earlier accounts in that we posit a much sparser distribution of tones relative to tone bearing units. In this respect, Japanese tone structure is much more similar to English intonation structure than hitherto supposed.

Another difference between ours and earlier accounts has already been

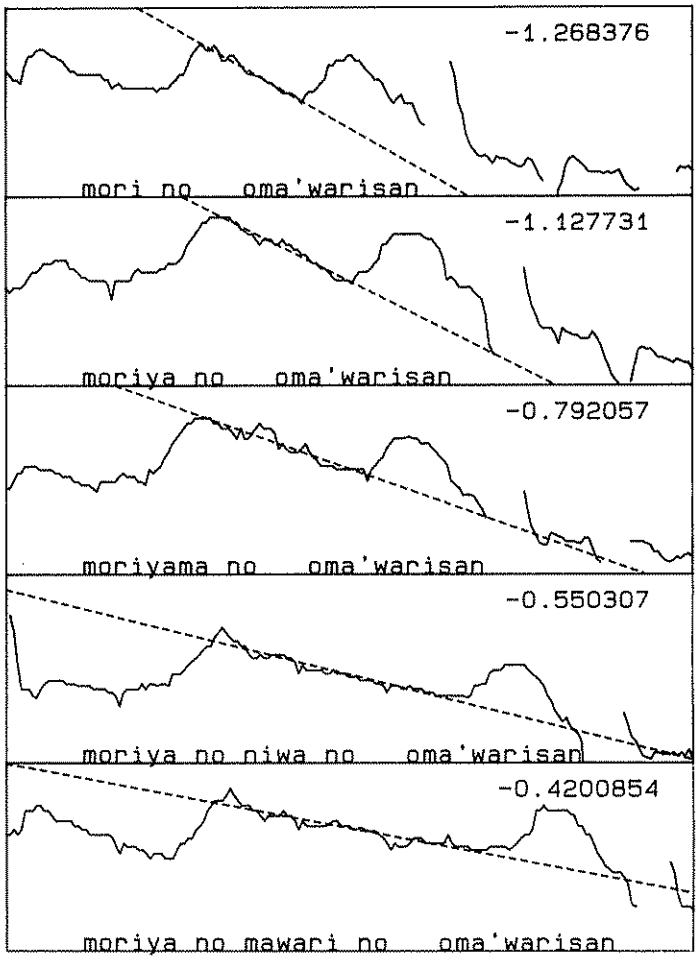

Figure 4

Portion of five utterances of the form *Kore-wa X-no oma'warisan no hanasi desu* with five different lengths for the *X-no* accentual phrase, ranging from three-syllable *mori-no* in top panel to eight-syllable *moriya-no mawari-no* in bottom panel. Speaker OF. Numbers in upper right corners are slopes of the plotted regression lines, which are fit to the points between the peak in *mori* and the minimum for the L at the beginning of *oma'warisan*.

suggested above. Earlier treatments consider the boundary L tone to belong properly to the syllable with which it is associated in the following accentual phrase. Indeed, many of these treatments consider it to be present at all only in the cases where we have it associated by a late rule to that phrase-initial syllable. In our treatment, by contrast, the L tone is always present and, except at the absolute beginning of the utterance, it belongs properly to the preceding accentual phrase. The reasons for

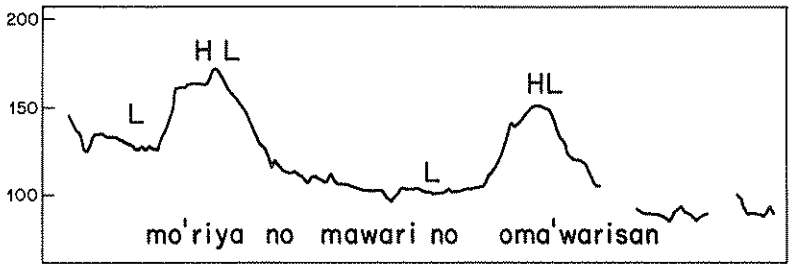

Figure 5

Portion of utterance *Kore-wa mo'riya-no mawari-no oma'warisan-no hanasi desu* produced by speaker ST.

considering the L to be a property of the preceding phrase are discussed in detail in Pierrehumbert & Beckman (forthcoming). They have to do with the realisation of this boundary tone within the local pitch range that is determined by the prominence value of the preceding H tone and by the application of catathesis, a rule of tonal implementation that reduces all tones subsequent to an accent.[5] This determination of the local pitch range holds even in those cases where the L has been associated by rule to the following phrase-initial syllable, as illustrated by Fig. 5, which shows the Fo contour of an accented phrase similar to the one given in Fig. 3. Here the L tone linked to the initial syllable in the following *oma'warisan* is much lower because it has undergone catathesis triggered by the preceding pitch accent. The following phrasal H, on the other hand, has its own local pitch range, and so is as high as the phrasal H after the unaccented accentual phrase. Thus the phrasal affinity of the tone when it is associated is analogous to the syllabic affinity of an ambisyllabic consonant that is linked essentially to the preceding syllable but which becomes associated also to a following syllable by a late resyllabification rule. When it is not linked, however, the L tone still exists and is realised in the fundamental frequency contour around the edge of the phrase.[6]

Although our treatment of the phrasal H and the boundary L differ so radically from that of earlier descriptions, it is in one respect completely within the spirit of earlier treatments. All existing descriptions of Japanese are alike in defining the accentual phrase as the domain of the rising pitch pattern that we attribute to the configuration of L boundary tone and phrasal H.

Most existing descriptions also agree on a second rather important point having to do with the distribution of accents within the accentual phrase. It is usually understood that within an accentual phrase, no more than one accent can occur. If two lexically accented words are grouped together prosodically within an accentual phrase, then one of the accents (the second one) must be deleted, as illustrated in Fig. 6. Whether this deletion is to be treated as an actual deletion of tones provided by the lexicon, or as a failure to insert an intonational tone shape at an accent locus specified

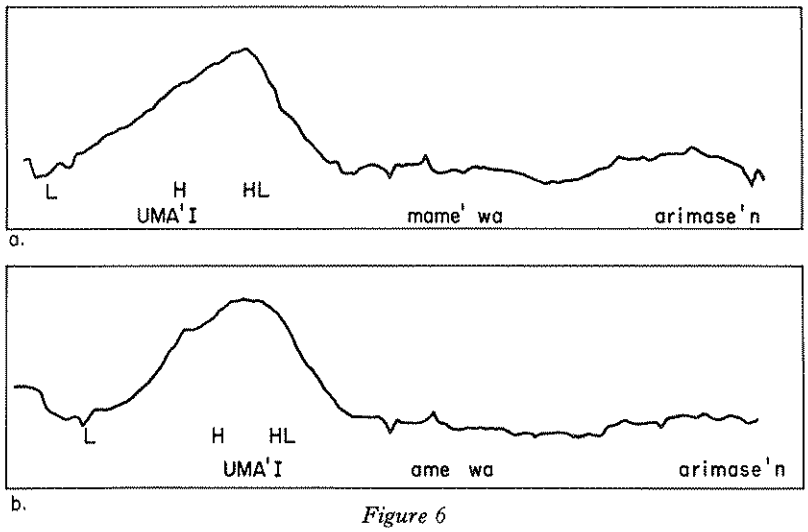

Figure 6

Utterances (a) *uma'i mame'-wa arimasen* and (b) *uma'i ame-wa arimasen*, with the nouns *mame'* or *ame* grouped together in the same accentual phrase with the preceding accented adjective *uma'i*. The tonal contours are identical because the grouping has deleted the lexical accent in *mame'* in (a).

in the lexicon, the end result is the same; an accented accentual phrase can have only one pitch accent. The relationship between the accentual phrase and the accent in Japanese, then, is that accent is distributed among the accentual phrases in an utterance in the same way that potential accent loci are distributed in the lexicon. Although the accent is not necessary to an accentual phrase (since there will be no accent if the phrase contains only unaccented words), its distribution is limited by the phrasal grouping. Just as a form in the lexicon has either only one or no accent, the accentual phrase itself has either only one or no accent (i.e. is accented or unaccented).

2.3 Culminative prominence

The existence of unaccented accentual phrases raises an important question about the nature of the accent in Japanese. In English and other languages like English, we think of accent as being some sort of culmination of tonal prominence, and an intonation contour is not well-defined unless it has at least one accent to associate with the most prominent syllable that is the focal centre of the text. Given the possiblity of utterances without any accents, one might question the propriety of using the term 'accent' to label the linked HL in Japanese. There are several reasons for doing so, having to do with the distribution of pitch accents among accentual phrases and the locally culminative tonal prominence usually given to the linked H tone of the accent.

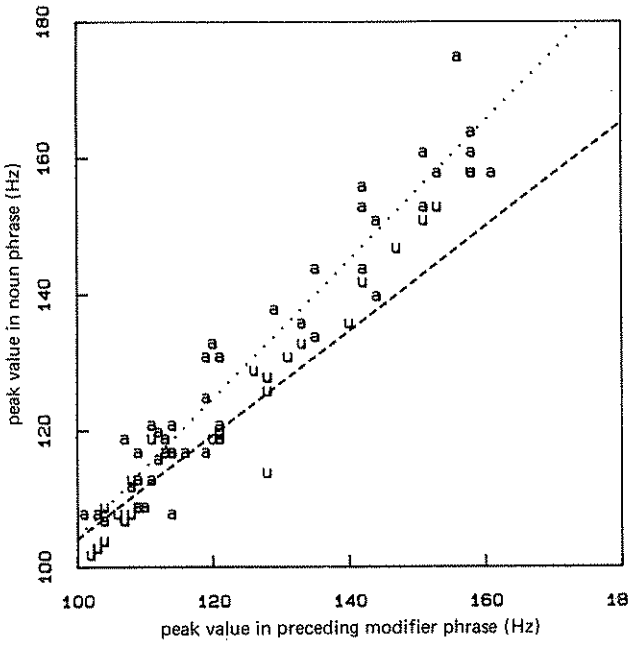

Figure 7

Peak Fo value in noun phrases as a function of the peak value in the preceding modifier. Dotted regression line and plotting point *a* are for accented noun, dashed regression line and plotting point *u* are for contrasting unaccented noun. Utterances produced by speaker SM in variable pitch ranges. At each pitch range the points for the accented nouns are higher.

We have already noted above the fact that only one pitch accent can occur in any given accentual phrase. If the intonational phrasing of an utterance puts two accented words together in the same accentual phrase, one of the accents cannot be realised in the intonation contour. This distributional rule seems to us to be motivated by the same spirit as the rules that govern the distribution of nuclear pitch accents in English. The nuclear pitch accent in English gives a globally culminative tonal prominence to the accented syllable, and there is a rule forbidding any accents after the nuclear accent in an intonation contour. In Japanese, similarly, any accent is culminative to its phrase by the deletion of all subsequent accents. Moreover, its linked H tone normally has greater tonal prominence than that of the phrasal H. This fact was noted by Poser (1984) and confirmed independently in our own experiments. For example, when accented phrases are compared to unaccented phrases, other things being equal, their peak fundamental frequency values are higher, as illustrated in Fig. 7. This suggests that in ordinary intonational circumstances, the accent H is the culmination of tonal prominence for its phrase.

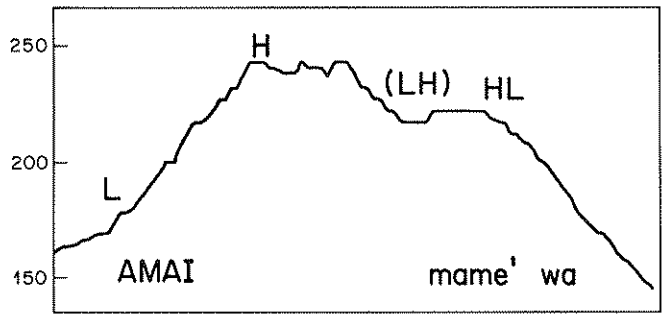

Figure 8

Portion of utterance *Amai mame'-wa arimasen*, with contrastive emphasis on *amai*.

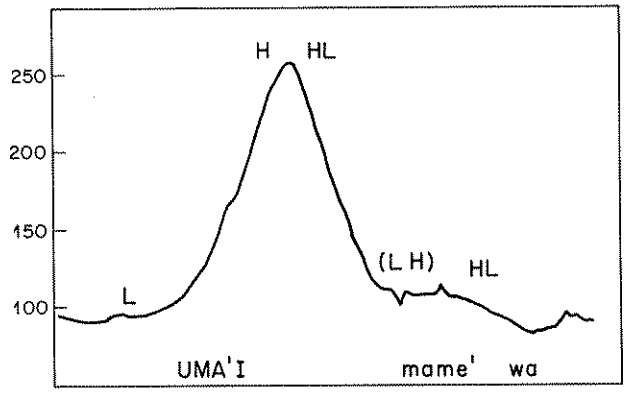

Figure 9

Portion of utterance *Uma' i mame'-wa arimasen*, with contrastive emphasis on *uma'i*.

The situation is somewhat more complicated when we consider the interaction of prominence due to textual focus with the intrinsic prominence of the different tones. In an experiment reported in Pierrehumbert & Beckman (forthcoming) we varied the accentual patterns and location of textual focus in target adjective–noun sequences in a corpus of dialogues. Fig. 8 shows a pattern that we noted in many of the utterances that had initial focus on a sequence of unaccented adjective followed by accented noun. One interpretation of this pattern is that it arises from the dephrasing of the noun and the reduction of the tonal prominence for its accent H in subordination to the focus on the preceding adjective. If this interpretation is correct, then the phrasal H in the adjective bears the culminative tonal prominence for this particular accentual phrase, and phrases such as these would constitute the rare exception to the general principle that the accent H is the culminative tone.

An alternative interpretation, however, is that the accented noun is not

actually dephrased, but merely greatly subordinated to the preceding unaccented adjective. In Pierrehumbert & Beckman (forthcoming) we propose that L tones are scaled downward within the prominence spaces defined by the relevant adjacent H tones. In the pattern shown in Fig. 8, for example, any L tone at the accentual phrase boundary between the adjective and noun would be realised in the very large pitch range of its focused accentual phrase. In other words, it would be scaled downward from the very prominent phrasal H tone that is the culminative tone for the unaccented adjective, and could therefore easily be as high as or higher than the culminative accent H of the following accented accentual phrase, thereby eliminating the characteristic delimitative rise for the second accentual phrase. Thus the consequences of extreme subordination to an accentual phrase to the left would be phonetically similar to the effects of dephrasing. We have used this fact to explain the apparent occurrence of two accents within a single accentual phrase when the first accented word has an extreme emphatic focus, as illustrated in Fig. 9, and there is nothing in theory to block a similar interpretation of the contour in Fig. 8. If this interpretation is correct, then the pitch accent H would consistently be the culminative prominence for every accented accentual phrase.

2.4 The accentual phrase in English

The relationship between accent and the accentual phrase in Japanese makes us wonder whether some similar phrasal unit might not exist in English as well. Does English have an accentual phrase which is the culminative domain of the pitch accent? Such a definition of the accentual phrase would not be unprecedented. Nolan (1984), for example, says that an initially accented 'accent unit' is implicit in analyses such as those of O'Connor & Arnold (1961) or Crystal (1969). Indeed, under some recent versions of metrical theory, such as Prince (1983) and Halle & Vergnaud (1985), two arguments for the existence of an accentual phrase in English can be constructed. First, each pitch accent constitutes a local tonal prominence that gives a textual prominence to the accented item. In these versions of metrical theory, such a locally strong element can arise in only two ways. Either the element is the designated terminal element (DTE) of some prosodic domain, or else it arises through application of a rhythmic alternation rule. However, the generic characteristics of rhythmic alternation rules are incompatible with the way pitch accent placement works. Rhythmic alternation refers to an already established strong element, and iterates away from it until the relevant prosodic domain has been covered. It follows that rhythmic alternation cannot be responsible for the one and only strong element in a domain. Since it is entirely possible to have a phrase with just one pitch accent – for example, a monosyllabic phrase – the rhythmic strength associated with the accent must arise from a rule establishing the DTE of a prosodic domain.

A second argument for the accentual phrase can be based on the fact that it is easier to produce long strings of unaccented syllables utterance-finally

than it is to produce such strings utterance-initially. For example, an utterance of the phrase *constitutional amendment* can easily be produced with a single (nuclear) pitch accent on the third syllable of *constitutional*,[7] but it would be much less natural to say with the nuclear accent on *amendment* and no prenuclear accent anywhere in the first word. This fact would receive a straightforward interpretation under these theories, since the DTE at each level of grouping must be peripheral to the group after discounting one possible extrametrical element. Here the syllables preceding a nuclear accent in *constitutional* could be analysed as a single extrametrical stress foot at the level of the accentual phrase, whereas the longer string preceding a nuclear pitch accent on the second syllable of *amendment* would be an anomalous string of two stress feet unless it were broken up by a pitch accent that grouped the two stress feet of *constitutional* into a separate accentual phrase.

These two arguments for an accentual phrase as a left-dominant grouping of stress feet, however, are rather theory-internal. One reason that the existence of the accentual phrase is less firmly supported in English than in Japanese is that it is not delimited by any boundary tones. The only tones which are properties of the accentual phrase occur at the location of the accent, whereas in Japanese there are tones both at the accent location and at the edge of the accentual phrase. English does have boundary tones, which will be discussed below. However, these belong to levels of phrasing which are hierarchically superior to the accentual phrase. A further source of unclearness arises from the question of whether the accentual phrase should be interpreted as a grouping of stress feet, or as a grouping of prosodic words.

Properly speaking, the accentual phrase should be a grouping of the immediately inferior elements in the hierarchy, and in Japanese, it is clear that the accentual phrase is a grouping of words. The tones that delimit the accentual phrase are aligned to the text with reference to the edges of words. Furthermore, these words are also the domain of certain phonetic rules, such as aspiration of initial voiceless stops or the weakening of medial voiced stops through fricativisation or (in the case of [g]) nasalisation. In English, the prosodic word played an important role in the formulation of stress rules in Liberman & Prince (1977), and it also functions in the lexical phonology. For example, unstressed mid and high vowels are phonetically tensed word-finally, as in *ditto* or *muddy*, where the final syllables are not reduced to schwa even though the flapping of the medial /t/ or /d/ in most American dialects shows that the final vowels in these words contrast in stress with the final vowels of *veto* or *chickadee*. There is also some evidence that the word is a prosodic unit of phonetically interpreted surface representations. In experiments by Nakatani & Schaffer (1978), for example, reiterant renditions of three-syllable phrases such as *noisy dog vs. bold design* were disambiguated by their durational patterns, suggesting some phonetic process such as word-final lengthening.

Unlike in Japanese, however, in English there is no clear relationship between the prosodic word and anything that could be called an accentual

phrase. A long word in English such as *constitution* can have two pitch accents, and in citation form it almost surely will, as shown below:

(1) constitution
| |
H* H* L L%

One might argue from this fact that long words such as these comprise more than one prosodic word. But that argument would be at best rather circular, leaving us with no independent motivation for the accentual phrase as a separate phrasal unit in a directly hierarchical relationship with the word. And at worst it is actually contravened by the fact that it is possible to say *constitution* with two pitch accents without tensing the second vowel; if word-final tensing were to be a postlexical phenomenon associated with word boundaries, and not a lexical phonological rule, this possibility would be clear evidence against the pitch accent being in a direct hierarchical relationship with the prosodic word. One could also argue that the attested durational effects of word boundaries might actually instead be accentual phrase correlates, since experiments demonstrating word-final lengthening have not controlled for intonation pattern. The Nakatani & Schaffer study cited above, for example, does not distinguish between *noisy dog* produced with one or with two pitch accents. Clearly more data are needed about the phonetic correlates of prosodic words in various intonational contexts before the definition of the accentual phrase as a grouping of one or more prosodic words in English can be properly evaluated.

To summarise, then, it is clearly possible to define the accentual phrase as a unit of English prosody. However, the evidence for the accentual phrase as a necessary unit in the prosodic hierarchy in English is much less definitive than for this phrase type in Japanese.

2.5 Accent and the lexicon

The last few subsections have discussed pitch accents in connection with larger tone patterns and phrasal prominence relationships. Japanese and English also show important similarities and differences in how the accent relates to prominence within the lexicon. The characteristic of accentual prominence in the lexicon that is perhaps the most obvious to the general observer is the often noted fact that the possible locations of accents are lexically distinctive. Accent is usually introduced to the student by citing pairs of lexical items that contrast minimally in possible accent loci. This is true of both Japanese and English. The two languages have often been cited as exemplars of this characteristic and have been contrasted to languages such as Czech, in which possible accent loci within lexical items are not distinctive (see, for example, Trubetzkoy 1939).

Another characteristic of the relationship between accent and the lexicon that is also fairly obvious to the casual observer is the fact that these possible accent loci are distributed in the lexicon in a way that is rather

different from the distribution of certain other syllabic features. In Yoruba, every syllable in a lexical item carries a distinctive tone, and there are no restrictions on the distributions of the different tones among the syllables. In Finnish, similarly, every syllable has a distinctively long or short vowel, and there are no restrictions on the distribution of the different vowel lengths. In Japanese, by contrast, only one syllable in any lexical item can carry the linked H tone of the accent. In English, there must be at least one syllable in any lexical item that has a stressed vowel that can be associated to a starred tone. McCawley (1968) and others have singled out such facts about the distribution of possible pitch accent locations within the lexicon as indicative of a fundamental similarity between the two languages, a judgement in which we concur.

On the other hand, this similarity should not be allowed to obscure a striking difference between the two languages in another closely related aspect of the relationship between accent and the lexicon. In both languages, the accentually prominent syllables are marked in the lexicon with only a subset of the phonetic features that are used to mark prominent items in an utterance. The full complement of these features includes the presence of tones (pitch accents and, in Japanese, the phrasal H), which in both languages have more extreme target values under emphasis than otherwise. It includes lengthened segment durations and higher amplitudes, which mark all accents in English and particularly emphatic phrasal prominences in Japanese.[8] And it includes the blocking of phonetic reduction phenomena such as vowel devoicing or deletion, either absolutely in all accented syllables (English), or to a greater extent in accented syllables than otherwise (Japanese). While both English and Japanese use all of these features at least to some extent to mark 'accented' or prominent items in utterances, the two languages have phonologised a different set of these features to regularly mark the prominent syllables that can take accent in the lexicon. As was noted above (§2.1), the H of the accent in Japanese can be considered to be linked in the lexicon. Another way of saying this is that Japanese has phonologised the tonal features of prominence to mark prominent syllables in the lexicon. English, on the other hand, has phonologised the durational/amplitude effects and the accompanying lack of vowel reduction.

The fact that lexically accented syllables in English do not have reduced vowels has been documented so often as not to need further comment here. The connection between duration/amplitude and lexical accent, on the other hand, is more controversial. The early experimental literature on stress perception has sometimes been interpreted as evidence that duration and amplitude are not reliable correlates of accent. The most extreme statement of this view is Bolinger's position that the durational effects are 'ancillary' to pitch accents, that accented monosyllables might have increased durations only 'in order to make room' for an associated pitch obtrusion (Bolinger 1965). We cannot agree with this position. Experiments such as those of Huss (1978) show that even when unaccompanied by pitch accents, lexically stressed syllables in English are marked by relatively

longer durations and higher amplitudes. In stress detection tests reported by Lieberman (1960), Lea (1977) and Beckman (1986), the intensity integral consistently outperforms fundamental frequency as a predictor of stressed syllables. This suggests to us that the duration/amplitude features of accent are associated to accentable syllables in the lexicon and are maintained in production even when these syllables are not intonationally accented. Although these correlates of accent may not be perceived as reliably as are pitch accents, we interpret the fact that they are so reliably produced as indicating that they are phonologised as features of stressed syllables in the lexicon. We are encouraged in this interpretation by more recent perception experiments using stimuli that mimic natural stress patterns more closely than was possible with the methods available to Fry (1958). These experiments show that duration can be used to perceive stress patterns in non-nuclear position (Nakatani & Aston 1978), and that even in nuclear accent position the fundamental frequency pattern may not be such an overriding cue as formerly supposed (Beckman 1986). English differs from Japanese, then, not in having no characteristic prosodic feature associated to accent loci in the lexicon (as Bolinger would have), but in choosing a different feature – namely stress rather than tone.

3 Catathesis

Another point of similarity between English and Japanese intonational structure is that in both languages accents figure in a phenomenon of tonal implementation that has been called 'downstep' or 'catathesis'.[9] Catathesis was first proposed for English by Pierrehumbert (1980), who argued that the inventory of English intonation patterns could be described using only two tones, L and H, provided that a rule of catathesis was posited. This rule, which compresses and lowers the pitch range in certain circumstances, accounts for the apparent 'mid' tones in contours like that shown in Fig. 10. In addition, when applied iteratively, it generates the descending staircase configurations shown in Fig. 11. Examples of such staircases can be found involving up to six or seven step levels; a theory lacking catathesis would accordingly need to posit six or seven phonemically distinct tone levels for English. This is more than has been found in any other language, and furthermore the theory would be hard-pressed to explain why so very few of the many possible combinations of tones are used in practice.

3.1 Catathesis in Japanese

Earlier accounts of tone structure in Japanese have often described phenomena that seem vaguely similar to catathesis. For example, McCawley (1968) describes a 'subordination rule' whereby the H tones in an accentual phrase following another accented accentual phrase are reduced to 'mid' tones. The first to suggest that these reduction phenomena

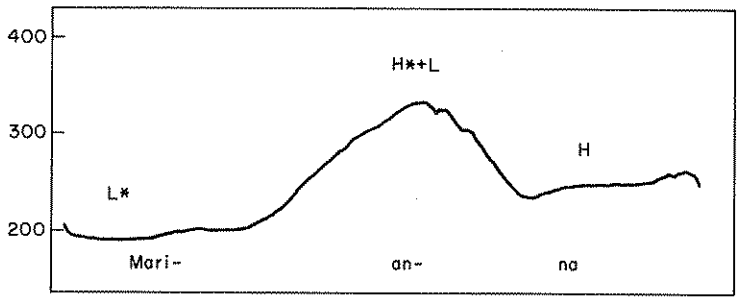

Figure 10

A typical 'calling' contour, ending at an Fo level which has led some authors to posit an English mid tone.

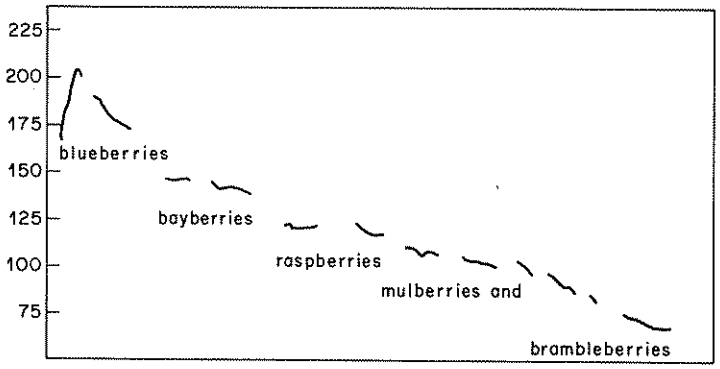

Figure 11

An Fo contour in which many applications of catathesis have produced a descending staircase.

might be formally identical to the processes that produce apparent mid tones in English was Poser (1984), who demonstrated rigorously for one subject that accents trigger catathesis. Comparing H tone values in minimally contrasting sequences of accentual phrases such as *uma'i mirin vs. amai mirin*, he showed that the Fo values in the following phrase were substantially lower when the preceding phrase was accented. We have since demonstrated catathesis for several more subjects, as illustrated in Fig. 12, which plots following accentual phrase peak to preceding peak in a large number of modifier–noun sequences in which the first accentual phrase (the modifier) is accented (plotting point *a*) or unaccented (plotting point *u*). The spread of points along the diagonal in the plot is due to pitch range differences, which were elicited by asking the subject to speak softly or speak up to variable extents in different tokens of the utterances. At every pitch range (i.e. at every region along the diagonal), the second phrase peaks for the *a* data points are lower, having been reduced by the catathesis triggered by the accent in the first phrase.

We have also demonstrated that catathesis chains, just as in English.

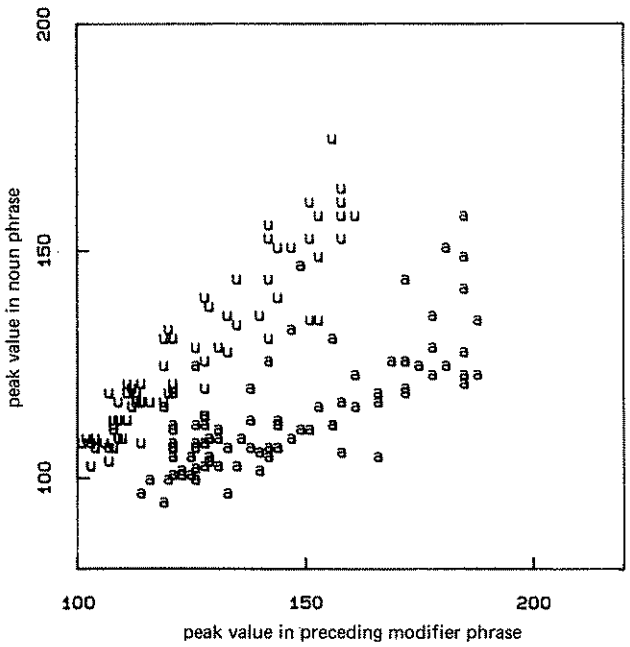

Figure 12

Peak Fo value in noun phrases as a function of the peak value in the preceding modifying phrase. Points labelled 'a' are for utterances in which the preceding phrase is accented, points labelled 'u' for those in which it is unaccented.

Fig. 13 demonstrates this fact by comparing peak values in the last accentual phrase in a series of modifier–modifier–noun. The points labelled '0' are for sequences in which neither the first nor the second modifier clause is accented, those labelled '1' are for sequences in which either the first or the second is accented, and the points labelled '2' are for ones in which both preceding modifier phrases are accented. As the spread of points and the regression lines for the three types show, the final noun phrase in sequences where it has undergone catathesis twice is lower than in those where it has undergone catathesis only once, which in turn is lower than in those where it has not undergone catathesis. Although it is difficult to construct sequences in which more than three applications of catathesis can occur,[10] catathesis does chain, producing descending staircases that are difficult to explain by increasing the number of tone levels, as did the older analysis that posited a 'mid' tone for subordinated accents.

There are two further important facts to note about catathesis in Japanese. The first is that catathesis differs from downdrift in African in that it is triggered by the HL sequence of the pitch accent and not by any other dissimilar sequence of tones. A sequence of H followed by L

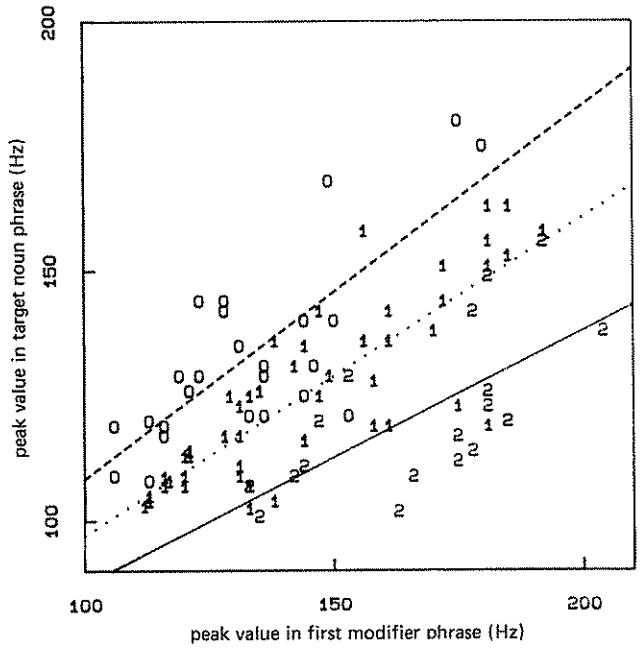

Figure 13

Peak Fo value in the final clause in modifier–modifier–noun sequences as a function of the peak in the first modifier. Dashed regression line and points labelled 'o' are for sequences in which both modifying clauses are unaccented (no applications of catathesis); dotted line and points labelled '1' for sequences in which one modifying clause is accented and the other unaccented (one application of catathesis); and solid line and points labelled '2' for sequences in which both modifying clauses are accented (two applications of catathesis).

can arise in Japanese by means other than having an accented accentual phrase. Indeed, every unaccented accentual phrase also gives rise to such a sequence by having a phrasal H tone followed by the boundary L. But catathesis is triggered only by the HL of the accent.

The second important fact is that the catathesis seems to occur during the triggering tone sequence. In other words, the compression of pitch range seems to be implemented within the accent that triggers the compression, affecting the L tone of the accent itself. We think that this must be so because the accent L is lower than the boundary L before the accent. In fact, in utterances with long post-accent stretches where the accent L can be distinguished from the following boundary L, it is nearly as low as that boundary L, which in addition to undergoing catathesis has also been affected somewhat by declination. (See, for example, the utterance in Fig. 5 above.) These two facts provide interesting points of comparison to the analogous phenomenon in English.

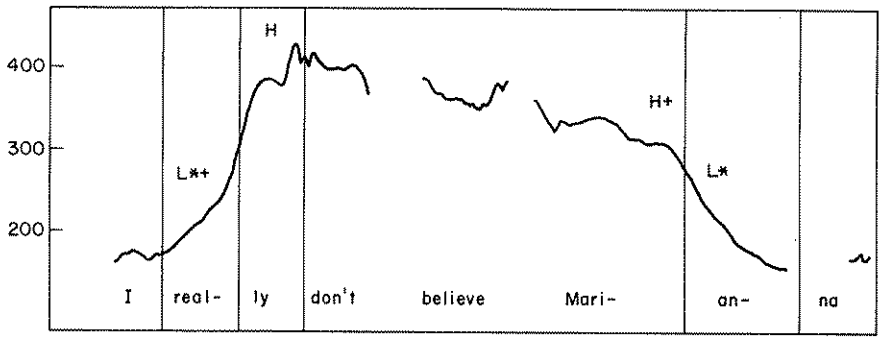

Figure 14
Utterance *I really don't believe Marianna*, with a scooped rise (L* + H) on *really* and a stepped-down accent (H + L*) on *Marianna*.

3.2 Catathesis in English

As noted above, catathesis was first proposed for English in Pierrehumbert (1980). There catathesis was taken to be a phonetic realisation rule triggered by a H L H tonal sequence including a bitonal pitch accent. This treatment of catathesis was motivated by the existence of similar phenomena in African tone languages rather than by comparison to another language with pitch accents. In these tone languages, downdrift is triggered by a sequence of tones H L H with no apparent internal organisation to the sequence. Pierrehumbert (1980) pointed out that catathesis in English differs from downdrift in that its application depends crucially on the tonal organisation of the triggering H L H sequence, but she assumed that it is otherwise similar to downdrift. Her catathesis rule thus applied only within sequences of the form H + L H and H L + H, but aside from the necessity of a ' + ' somewhere in the sequence, the H L H is identical to the downdrift trigger. Cases where it was blocked included contours such as H* L H %, where there is no bitonal accent, but they also included, for example, H* + L L + H*, where the tonal sequence includes bitonal accents but is not strictly alternating.

The comparison of English with Japanese has prompted us to re-examine this earlier account of catathesis in English. We have found two sorts of counterexamples to the earlier account that applied catathesis only where there is a strict alternation of H L H. The first is illustrated in Fig. 14. The lateness of the peak around *really* in this utterance indicates that the accent here is L* + H. The fall from an unstressed syllable to the stressed third syllable in *Marianna* indicates that the accent here is H + L*. Yet the H tone on the second accent has apparently undergone catathesis.

The second counterexample is a fairly unusual form of the 'calling' pattern. This general class of patterns is characterised by ending at a mid pitch level, and is described in Pierrehumbert (1980) as involving a H

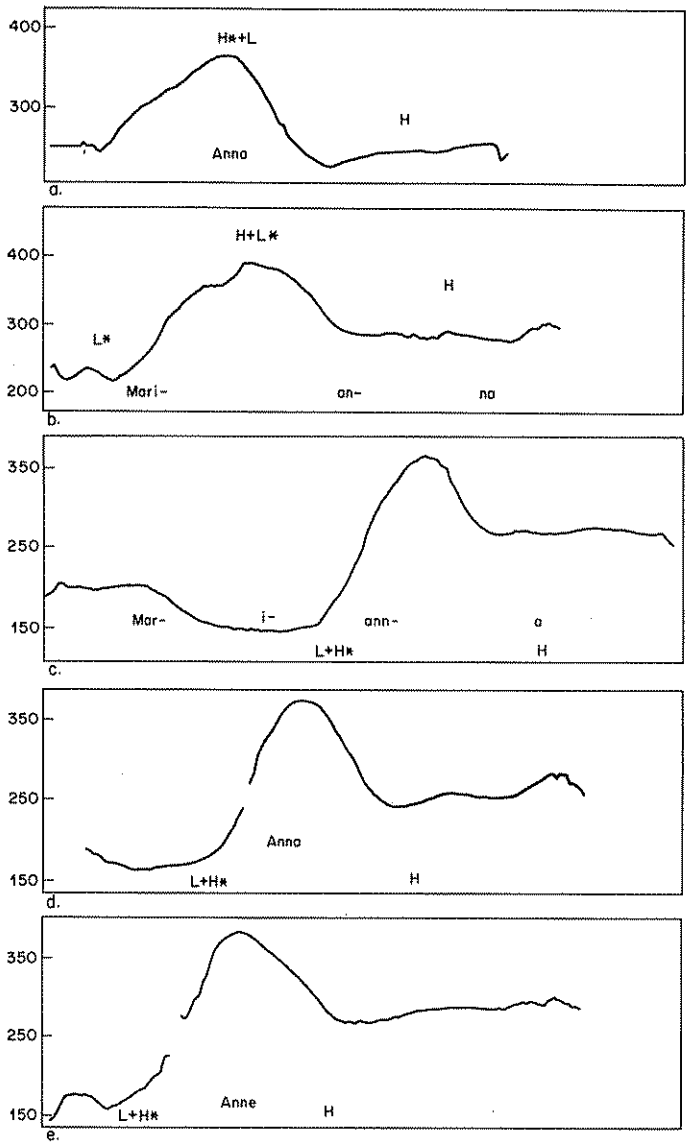

Figure 15

Three types of calling contours – analysed as (a) H*+L H L%, (b) H+L* H L% and (c)–(e) L+H* H L%.

phrase accent which has been subject to catathesis. (The phrase accent, which may be either L or H, spreads over the region from the last, or nuclear, pitch accent to the end of its phrase. We discuss its status at greater length in §4.) When the nuclear pitch accent is H*+L, as in Fig. 15a, a peak on the stressed syllable precedes the mid level. When the

accent is H+L*, the peak precedes the stressed syllable, as in Fig. 15b. This pattern tends to contrast with that in Fig. 15a by having a wheedling quality. The problematic member of this set is shown in Figs. 15c, 15d and 15e. Here, the nuclear accent is evidently L+H (either L*+H or L+H*), since a decided dip in Fo precedes the peak. However, the phrase accent is still at the mid level, indicating that it is a H tone that has undergone catathesis.

The calling contours can all be accounted for by assuming that all bitonal accents trigger catathesis of material after the accent. Utterances such as the one shown in Fig. 14 also can be accounted for if this assumption is made. That is, if any L+H and not just a L+H after a H tone triggers catathesis, then the reduction of the H in the second pitch accent follows automatically from the choice of a bitonal accent on *really*. This new analysis of the catathesis trigger in English makes it exactly analogous to the catathesis trigger in Japanese, where the HL accent cannot occur in a strictly alternating sequence, being necessarily preceded by the phrasal H and followed by the boundary L.[11]

Permitting the two-tone accent *per se* to trigger catathesis also leads to a cleaner analysis of downtrends in Danish. The phonetic characteristics of Danish downtrends are described in detail by Thorsen (1980). Pierrehumbert (1980) suggested a reanalysis of Thorsen's results, based on the idea that the pitch accent in Danish is L*+H, and that sequences of such accents contain the alternating tonal sequence H L H, which triggers catathesis. However, this proposal meets with a difficulty; Thorsen's data show conclusively that in a L*+H L*+H sequence, the second L has already undergone lowering. Under Pierrehumbert's original formulation, catathesis would not affect any tone before the second H tone. A technical solution to this problem is provided, but as Ladd (1983) notes, it is inelegant. Our new observations about English, however, suggest that catathesis would apply in Danish to compress the pitch range immediately after the first L*+H accent. This has exactly the correct consequence of lowering both tones of the second accent.

The reanalysis of English catathesis that we have proposed retains Pierrehumbert's original idea that one particular tonal configuration triggers catathesis to the right. In retaining this important aspect of Pierrehumbert's earlier treatment, we anticipate a criticism that has also been made of the earlier account – namely, that the L of the H*+L accent is too abstract; it must be there to trigger the downstep, but does not usually appear as a well-defined valley in the Fo contour. We do not feel that this criticism is adequate grounds for abandoning our account of the bitonal accent as catathesis trigger. To justify our position we compare our account to the only plausible alternative, recently offered by Ladd (1983). Ladd suggests that catathesis arises as the expression of a distinctive phonological feature on the lowered tone itself. That is, he follows some scholars of African tone languages in positing an underlying distinction between H and ꜜH, where the diacritic 'ꜜ' indicates the presence of a

feature [+downstep]. We do not feel, however, that a [±downstep] feature adequately captures the facts about catathesis.

A major reason for our position is the defective distribution of the catathesised tone. In English, there is no distinction between catathesised ꜜH and plain H at the first accent in a phrase, after the phrase accent, or after a L* accent. These are just the places where our account would predict a neutralisation because the left-hand context for catathesis is necessarily lacking in these cases. Under the downstep-as-feature account, on the other hand, these gaps must be viewed as coincidental. A similar problem occurs in the analysis of Danish. Ladd suggests that Thorsen's data can be explained by positing a downstepped ꜜL in Danish – that is, he transcribes the Danish pitch accent as LH phrase initially and ꜜLH otherwise. The feature account gives no explanation for why the LH and ꜜLH forms are in complementary distribution in this way.

Ladd's suggested treatment of Danish raises a second problem with making [downstep] a phonological feature of tones. If 'ꜜ' can function independently on L and H tones, it is curious that English does not make heavier use of this feature. For example, limiting ourselves to the two accent types H and HL, we would wonder why English does not make a four-way distinction among H, ꜜH, HꜜL and ꜜHꜜL. More generally, the tonal features [high] (for H *vs.* L) and [downstep] should in principle permit a total of twenty different single and two-tone accents. Of these, only a small number appear to be used – only H, HL, L, LH, ꜜH, ꜜHL and HꜜH are mentioned in Ladd's discussion. Furthermore, LH and HꜜH are apparently restricted to nuclear position and ꜜH and ꜜHL have the defective distribution already noted. Eight of the thirteen omissions follow from the language particular restriction of [+downstep] to H tones which was already noted above. The other five gaps appear to be unexplained. In contrast, our present system would in principle permit 10 different accents. Six are actually used, and they can occur in all positions. The lack of the four accents L*+L, L+L*, H*+H and H+H* is systematic; the phonetic realisation rules imply that they would be neutralised everywhere with single-tone accents.

Further advantages of our present position follow directly from the fact that we treat catathesis as a phenomenon affecting the parameters controlling phrasal pitch range rather than parameters of individual tones. This treatment affords two correct predictions not possible in the alternative treatment. First, when catathesis applies, all tones to the right are affected, up to the end of the phrase. The effects stop only when a sufficiently strong boundary is reached to permit a fresh selection of pitch range parameters. Second, in English as well as in Japanese and Danish, catathesis lowers both H and L tones (except that L tones already at the bottom of the range cannot be further lowered). These characteristics of catathesis seem peculiar if it is viewed as a phonological feature of individual tones. We can contrast them to the characteristics of individual accentual prominence, a phenomenon that Ladd treats with a comparable

tonal feature [raised peak].[12] Ladd emphasises that [+raised peak] does not affect the overall pitch range, but only the peak value for the specified tone. If [+downstep] behaved similarly, it too should lower only the specified tone.

To summarise, then, we do not agree with Ladd's claim that positing a feature [±downstep] permits an overall simplification of the English intonational description. On the contrary, there is a trade-off between simplicity and abstractness in describing the intonation system as in many other matters. We would assert that our present theory is considerably simpler than Ladd's, at a moderate cost in abstractness.

Our reanalysis of catathesis in English raises a question that would not have occurred earlier. Since the catathesis is triggered by the bitonal pitch accents themselves and not by the alternation H L H, we now must ask where the catathesis occurs in relationship to the accent-internal tones. In Japanese, as was noted earlier, catathesis seems to apply within the accent itself, affecting the trailing L tone. In English, by contrast, catathesis seems not to apply until after the second tone of the triggering pitch accent. This is especially obvious in sequences of H*+L or H+L* accents, which tend to look like a series of downward steps, as illustrated in Fig. 11 above. Here each subsequent H tone is as low as the L tone of the preceding accent, having undergone catathesis triggered by the accent that contains the L. Similarly, in the calling contour shown above in Fig. 15c, catathesis cannot be taken to apply within the L+H accent, since the H trailing tone of the pitch accent would then be on the same level as the following phrase accent. We therefore conclude that along with all the striking similarities in the composition of the catathesis trigger and so on, there is an interesting difference between the two languages in the timing of the catathesis relative to the triggering tones.

3.3 A short digression on timing

This difference suggests that the internal structure of the accent HL in Japanese might be somehow different from that of the analogous bitonal pitch accent in English. There is another difference between the two languages that at first glance might seem to be related to this difference in the timing of catathesis relative to the triggering accent. In English, the gross differences that exist in syllable durations can be utilised to show that the unstarred tones are realised at some fixed time in relation to the starred tones regardless of the number of potential tone bearing units covered by that time. The peak for the trailing H of a L*+H pitch accent, for example, can occur on the same syllable as the starred tone, as in Fig. 16a, or it can occur two syllables later, as in Fig. 16b. The temporal alignment of the H tone with the different syllables in the text in these two utterances is determined by the difference in phonetic duration between the diphthong followed by consonant cluster in *Stein's* as opposed to the monophthong in an open syllable in *rigamarole*.

In Japanese, by contrast, the phonotactics of the language guarantee

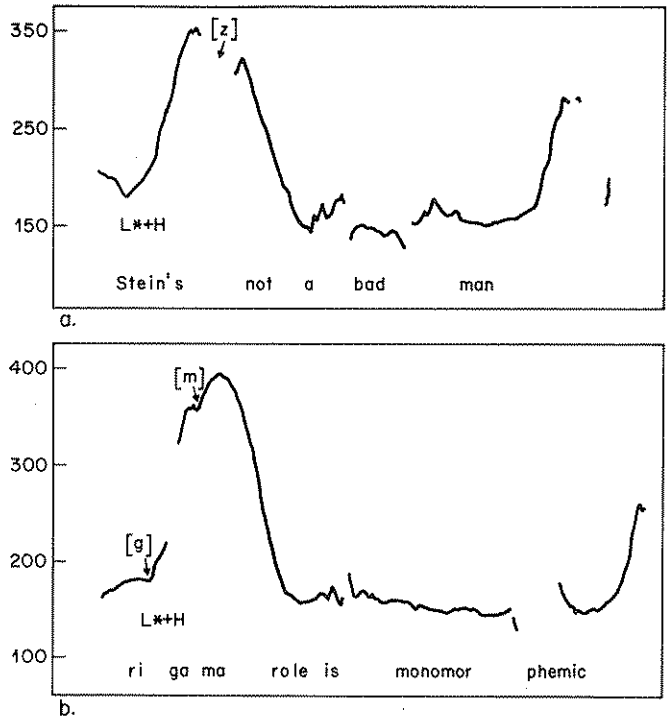

Figure 16

A scooped fall–rise contour (L*+H L H%) on (a) *Stein's not a bad man* and (b) *rigamarole is monomorphemic.*

that there will not be such gross differences in phonetic length among the potential tone bearing units. Therefore, there is no compelling evidence as to whether the L of the accent is a trailing unlinked tone, or is linked to the mora following the H. Here we treat the L of the accent as an unassociated tone, but any predicted timing differences between this and an alternative treatment that associates the L to the mora following the H would not be large. A carefully constructed experiment that contrasted different intrinsic lengths for an intervening consonant might support one treatment over the other, but there are no presently existing data that we know of. The fact that this L in Japanese is affected by catathesis, whereas the trailing unstarred L tone in the analogous English bitonal pitch accent apparently is not, might be interpreted as evidence that the Japanese pitch accent is a somewhat less consolidated unit than is an English bitonal accent. This in turn might constitute an argument for linking the L in Japanese.

The connection between these arguments, however, is rather tenuous. The similarity in what constitutes a trigger for catathesis, moreover, shows that any lesser degree of consolidation for the two tones of the accent in Japanese is a very relative matter. These tones are still much more

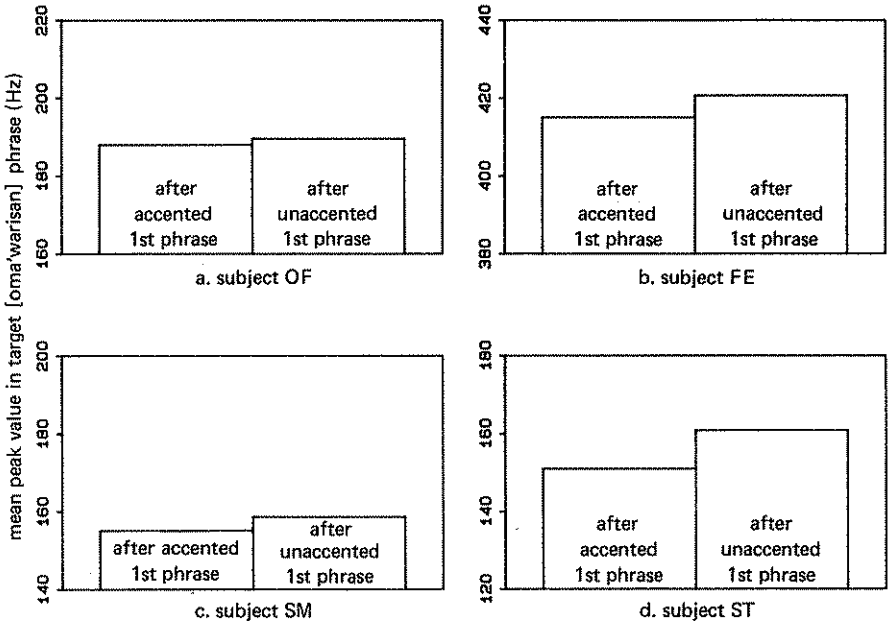

Figure 17

Mean peak Fo values in *oma'warisan* following accented and unaccented phrases for four speakers. Means are averaged over all lengths of preceding phrase from shortest (e.g. *mo'ri-no*) to longest (*mo 'riya-no mawari-no*).

integrated than just any sequence of H and L in the language, since they are the only ones that do trigger catathesis.

The similarities in catathesis and its relationship to accents figure importantly in comparisons of higher levels of prosodic phrasing in the two languages, a topic taken up in the next section.

4 The intermediate phrase

4.1 The intermediate phrase in Japanese

In Japanese, accentual phrases are organised into a larger unit, which we call the intermediate phrase. The intermediate phrase can be as short as a single accentual phrase, and it seldom contains more than three. Its boundary is often marked by a pause or by glottalisation. When these are lacking, we believe that the amount of phrase-final lengthening and the realisation of the L boundary tone still provide evidence for the disjuncture.

The intermediate phrase is also the domain for catathesis, as illustrated in Fig. 17. The data for this figure are drawn from the set of utterances that includes those shown in Figs. 3, 4 and 5 above. In these utterances,

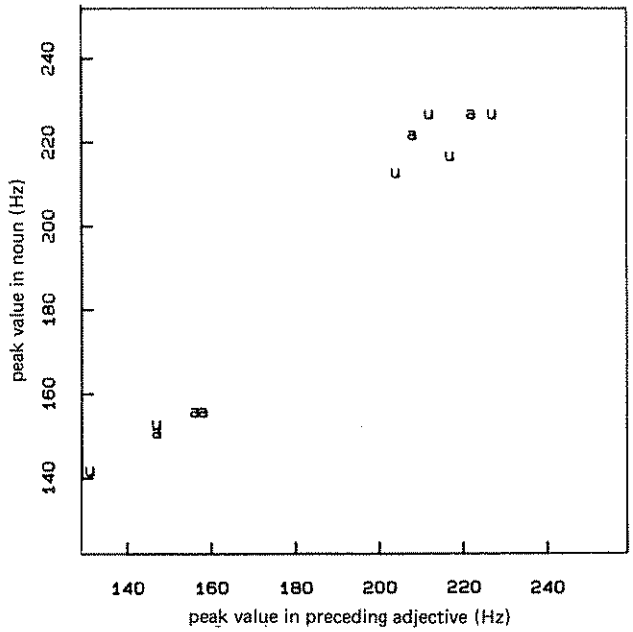

Figure 18

Peak Fo value in noun as a function of the peak Fo value in the preceding modifying adjective in utterances of *uma'i mame'* (plotting points *a*) and *amai ame* (*u*) produced with focus on the noun phrase. Speaker OF.

there usually seemed to be an intermediate phrase break before the *oma'warisan*. The bars in the various panels of Fig. 17 show means for the peak values in *oma'warisan* following an unaccented or an accented first phrase. For each speaker, these two means are practically the same. This relationship is very different from the substantial separation between the means for the peak values which were plotted above in Figs. 12 and 13, even though the modifier–noun sequences in those utterances are otherwise comparable. We interpret this difference as indicating that the intermediate phrase boundary blocked catathesis in the set of utterances that provided the data for Fig. 17. In the utterances that provided the data for Figs. 12 and 13, by contrast, only accentual phrase boundaries occurred between the plotted phrase peaks, and the following phrase peak values (plotted on the y-axis), therefore, showed the lowering of the catathesis that was triggered when the preceding phrase was accented.

This result that catathesis is blocked by an intermediate phrase boundary provides a probe for investigating the influence of focus on phrasing. In another set of utterances, both the location of the focus and the accentuation of the words were varied in an adjective–noun sequence. In listening to the recordings, we suspected that the focused word was ordinarily preceded

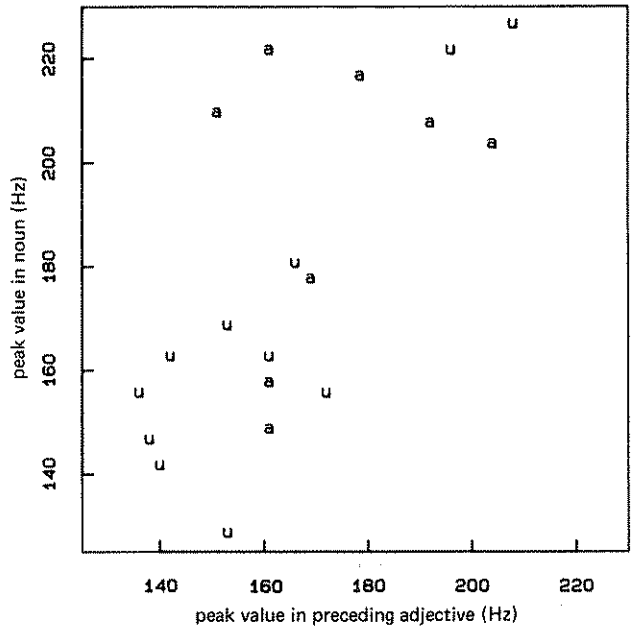

Figure 19

Peak Fo value in noun as a function of the peak Fo value in the preceding modifying adjective in utterances of *uma'i mame'* (plotting points *a*) and *amai ame* (*u*) produced with focus on the noun phrase. Speaker TS.

by an intermediate phrase boundary. This impression was confirmed when we looked for catathesis in the peak relations for this utterance set. Figs. 18 and 19 plot the peak values in sequences with focus on the noun (the second word in the sequence). The plotting character *u* is used for sequences in which both the adjective and the noun are unaccented, and *a* is used for ones in which both are accented. If catathesis applies, the *a* points should lie below the *u* points, as the comparable data do in Figs. 12 and 13 above. If catathesis is blocked, they should lie in the same region of the plot. For subject OF's data in Fig. 18, the points obviously occupy very similar regions. For subject TS's data in Fig. 19, there is more scatter, but a comparison of this figure to Figs. 12 or 13 above still argues that catathesis has not applied here. Thus in these cases, the blocking of catathesis demonstrates the presence of the suspected intermediate phrase break preceding the focused noun.

We can contrast these pre-focus breaks to the same word boundary in post-focus position. In sequences where the focus was on the preceding adjective, catathesis was not blocked. Compare the distribution of points in Figs. 20 and 21 to that in Figs. 18 and 19. The obvious application of catathesis here rules out the possibility that catathesis is blocked whenever the relation between two peaks is affected by emphasis. Instead, for

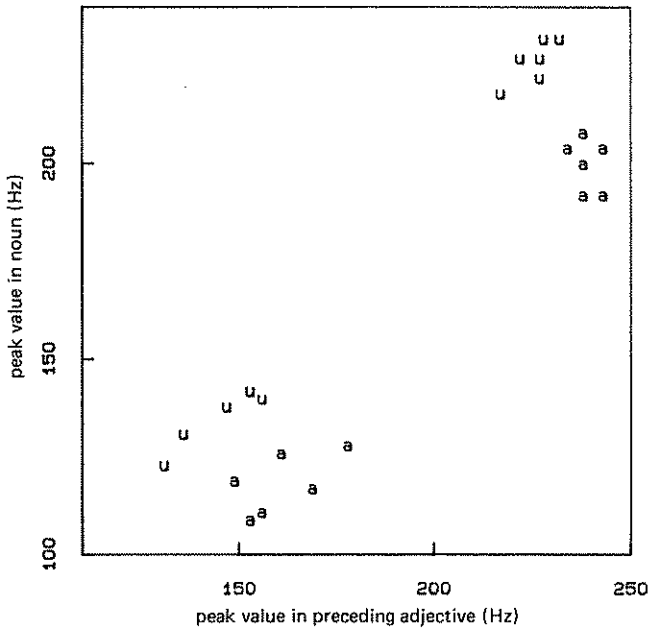

Figure 20

Peak Fo value in noun as a function of the peak Fo value in the preceding modifying adjective in utterances of *uma'i mame'* (plotting points *a*) and *amai ame* (*u*) produced with focus on the adjective phrase. Speaker OF.

materials following a focus, relative prominence and catathesis interact to determine H tone levels.

The metrical theory of phonology provides some motivation for the observed relation of focus and phrasing. Under current versions of metrical theory, such as Selkirk (1984), the strongest element of a higher level of structure is intrinsically more prominent than the strongest element of any lower level of structure. For example, the nuclear stress of an intonation phrase is stronger than any main word stress which does not carry nuclear stress. A good strategy for making a word more prominent, then, is to select a phrasing pattern which makes the word the strongest element of a higher level of structure. We believe that the tendency to introduce a phrase boundary right before the focused item in Japanese, rather than right after, is related to the fact that the intermediate phrase in Japanese is left-dominant. That is, the strongest element is at the beginning of the phrase. This explanation of the relationship between focus and phrasing is not really rigorous evidence for a left-dominant intermediate phrase, however, since intonation phrase breaks are sometimes introduced before a focused item in English, even though the English intonation phrase is generally believed to be right-dominant. In particular, when the focused word is in a position to carry nuclear stress, it will seem

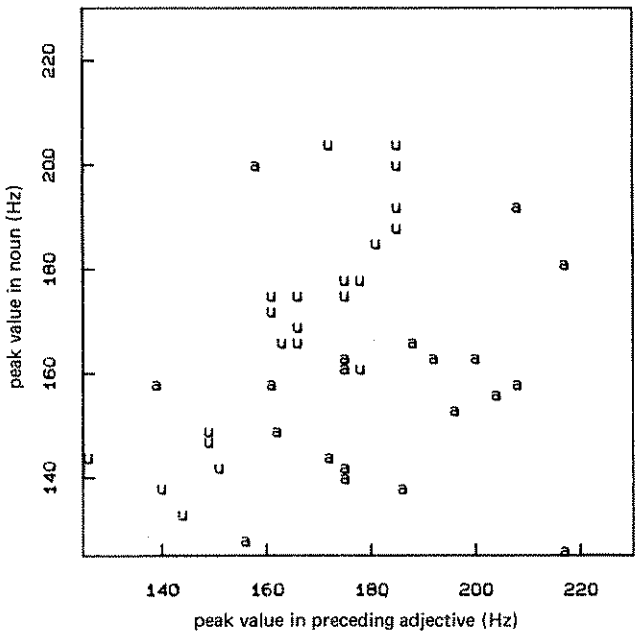

Figure 21

Peak Fo value in noun as a function of the peak Fo value in the preceding modifying adjective in utterances of *uma'i mame'* (plotting points *a*) and *amai ame* (*u*) produced with focus on the adjective phrase. Speaker TS.

even more prominent if it is set off as a phrase by itself, rather than being the main stress of a complex phrase. (We will return to such examples in §4.4.)

4.2 The intermediate phrase and the intonation phrase

Having made these observations, we can compare the Japanese intermediate phrase to the English intonation phrase or tone group. In the treatment of Pierrehumbert (1980), the English intonation phrase has two phrasal tones following the last pitch accent, the phrase accent and boundary tone. Each can be either L or H, so that altogether four terminal configurations are possible.[13] Note that in this framework, the inventory of pitch accents proper is taken to be the same in nuclear and prenuclear position, and the sequence of nuclear pitch accent, phrase accent and boundary tone is a phonological decomposition of what would be referred to as the 'nuclear tone' in the British tradition. When the nuclear stress is at some distance from the end of the phrase, the phrase accent and boundary tone characterise the Fo configuration over what is termed the 'tail' in the British school. If the nuclear stress is on the last syllable of the phrase, the

extra tones of the intonation phrase still occur; they are merely crowded onto a single syllable together with the nuclear pitch accent.

In terms of its phonological level, then, the Japanese intermediate phrase contrasts with the English intonation phrase just described in having only a single terminal tone. Also unlike in English, this extra tone is always L. Moreover, since it is found at boundaries separating accentual phrases within an intermediate phrase as well as at intermediate phrase boundaries, it must actually be terminal to the accentual phrase rather than being an intermediate phrase property.

On the other hand, there is one situation in which a sequence of two terminal tones does occur in Japanese. At the end of a yes/no question, a H boundary tone follows the L boundary tone of the accentual phrase. This H tone resembles the terminal tones of the English intonation phrase in its interpretation; its presence assigns a particular pragmatic force to the utterance taken as a whole. One possible hypothesis is that the boundary H is optional at an intonation phrase boundary in Japanese, where an intonation phrase is taken to consist of one or more intermediate phrases. However, since the preceding L is actually terminal to the accentual phrase, it is also possible to view the boundary H as an optional intermediate phrase property. The fact that we did not observe it utterance-medially would then be attributed to its pragmatic interpretation rather than to its phonological affiliation.

In addition to these differences in phonological structure, there are also differences in usage between the Japanese intermediate phrase and the English intonation phrase. For one thing, the intermediate phrase is typically smaller, including fewer syntactic constituents. For example, most of our subjects divided sequences of the following form into two intermediate phrases, with the break after the first modifying clause, as indicated by '|':

(2)	Kono arai	ayaori-no	obizi ga.
	This rough	twill	obi-cloth.

In English, it is possible to place an intonation break at the comparable location in the sequence, but this would only be done if the speaker was making an effort to be exceptionally clear or emphatic. A second example of this difference is that an intermediate phrase break can apparently occur in the middle of compounds in Japanese. We interpret Kubozono's (1985) descriptions of the Fo contours in compounds like the following:

(3)	be'ika	neage	hantai	u'ndoo
	rice-price	price-raise	protest	movement

as indicating that catathesis was blocked after the word *neage*. We interpret the failure of catathesis as evidence for an intermediate phrase boundary. In English, by contrast, even long compounds such as *49th Street Subway Stop* are ordinarily pronounced as a single intonation phrase.

4.3 The intermediate phrase in English

Since the English intonation phrase is not comparable to the Japanese intermediate phrase, one might ask if English does have a level of phrasing which is comparable. We believe that it does. Specifically, we believe that the phrase-accent plus boundary-tone configuration in Pierrehumbert (1980) should be reanalysed as involving correlates of two levels of phrasing. The phrase accent would then be a terminal tone for the intermediate phrase, while only the boundary tone is terminal to the intonation phrase. Since an intonation phrase is made up of one or more intermediate phrases, both the terminal tone for the intermediate phrase and the one for the intonation phrase are seen in sequence at the end of the intonation phrase. The situation is thus similar to that seen in Japanese yes/no questions, where the L boundary tone terminal to the accentual phrase is followed by a H boundary tone belonging to the larger phrase. The chief difference at the level of phonological or phonetic description is that the English phrase accent can spread; there is a relatively abrupt transition from the last target level specified by the pitch accent to the target level for the phrase accent, which is then maintained over the remainder of the phrase. This is particularly true for the L phrase accent. A synthesised H* L configuration will sound peculiar if the transition from the H* pitch accent to the L phrase accent is too gradual. It is less true for the H phrase accent, whose temporal placement in the tail of the contour is extremely variable. However, the H phrase accent is still very different from the situation in Japanese, where we have not encountered any cases in which the level for the accentual phrase L boundary tone was sustained. Even in the longest unaccented accentual phrases, linear interpolation from the phrasal H to the boundary tone appears to yield a better description of the Fo contour than does any account involving tone spreading (see §2.2 above).

This reanalysis of English suggests that phrase accents should be able to occur medially without being followed by an intonation phrase boundary tone. We have recently found a number of intonational types which are strong candidates for being examples of such medial phrase accents. Some of these examples have emerged from attempts to apply the Anderson *et al.* (1984) intonation synthesis program in speech synthesis, while others are phenomena we have noted in natural discourse. In presenting these examples, we would like to stress that the indicated phrasing is in most cases not the only possible phrasing. Phrasing in English is highly facultative. In extremely slow or emphatic speech, the intermediate phrase boundaries shown could easily be replaced by full intonation phrase boundaries. On the other hand, the phrasing would be less articulated if the speech was rapid, if shorter words were used in the indicated constructions, or if some of the words did not convey important information. We would also like to stress that the examples here have been picked because they are cases in which a level of intermediate phrasing seems clearly motivated. It is our suspicion that the use of this level of phrasing

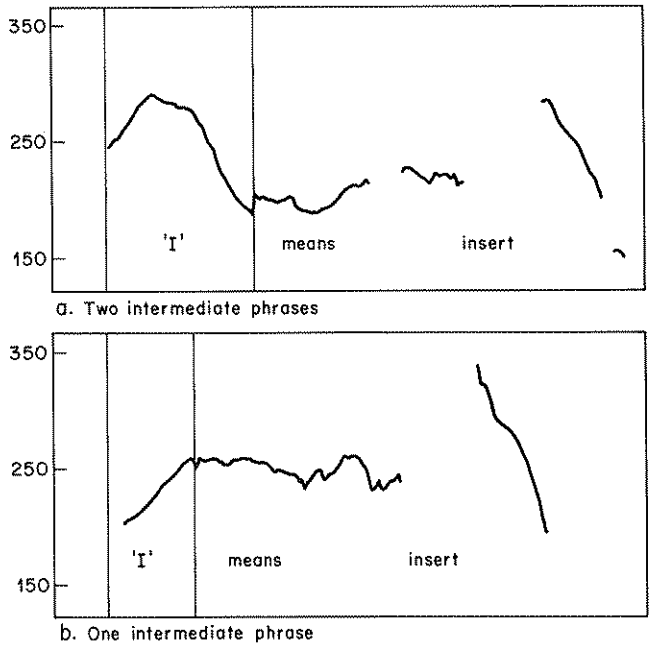

Figure 22

Utterance '*I*' *means insert* said with (a) an intermediate phrase break after '*I*', and (b) no phrase break.

is extremely pervasive, and that many cases which we would previously have taken to involve an intonation phrase boundary actually have only an intermediate phrase boundary.

One class of cases was identified in the course of Hirschberg & Pierrehumbert's (1986) investigation of intonation in a computer aided instruction system. The system uses synthetic speech to teach people how to use a text-editor, and a number of sentences define the meanings of special keys. For example, the 'i' key must be used when one wishes to insert some text, and the computer explains:

(4) 'I' means insert.

Fig. 22a shows a natural Fo contour for a rhetorically explicit rendition of this sentence. Compare this contour to the one shown in Fig. 22b, where the sentence is produced as a single phrase and the word *I* comes across as much less salient. In Fig. 22a, the word *I* is longer than in Fig. 22b. Also, the Fo peak is followed by a fall to a rather low value in Fig. 22a, whereas in Fig. 22b the Fo peak occurs at the end of the syllable and is not followed by a fall. We would like to say that Fig. 22a has two intermediate phrases, with the first marked by a L phrase accent, whereas Fig. 22b has only one intermediate phrase. We do not believe that Fig. 22a has a full intonation phrase boundary after *I*. A full intonation break would probably have a

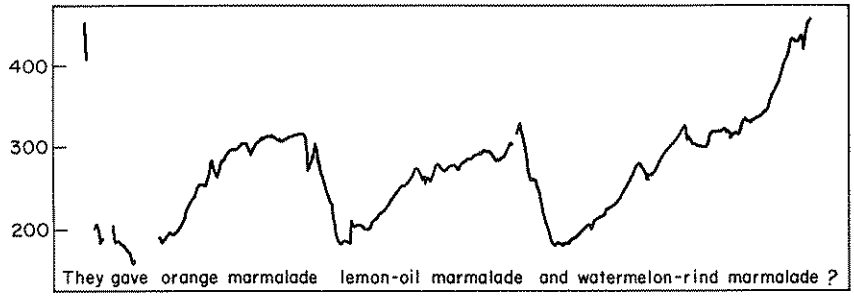

Figure 23

Fo contour for utterance *They gave orange marmalade, lemon-oil marmalade, and watermelon-rind marmalade?* produced as a simple list with question intonation.

continuation rise in this construction (a sequence of L phrase accent and H boundary tone). Furthermore, synthesising the sentence with a full intonation break gives the impression of an excessively strong boundary, at least for normal rates of speech.

The pattern in Fig. 22a seems to be fairly typical. It appears that a definiendum is in general set off as a separate intermediate phrase in rhetorically clear intonation. This observation appears to be true even in cases where the effect of the sentence is definitional, but its form is less stereotyped. For example:

(5) Use 'h' | for reverse.
(6) Use 'hint' | if you need help.

(Here '|' is used to indicate the boundary location.)

A second class of case where the intermediate phrase boundary seems motivated is a frequently observed form for multi-phrasal yes/no questions. Fig. 23 displays an Fo contour for a particularly simple case, the questioned list in (7):

(7) They gave orange marmalade, lemon-oil marmalade, and watermelon-rind marmalade?

Each noun phrase in this list has a rise, but the final rise is by far the largest. Also, it is the only one that shows an abrupt upturn at the end. The two others have a fast rise just after the stress followed by a gradual rise up to the end of the phrase, but the gradual rise is very similar to the medial portion of the contour on the last list item and there is no further upward inflection at the end. To account for this Fo contour in the framework of Pierrehumbert (1980), it would have been necessary to suppose that there are three intonation phrases. The last intonation phrase has a H phrase accent and a H boundary tone, but the first two end in a H phrase accent and a L boundary tone. A problem with this analysis

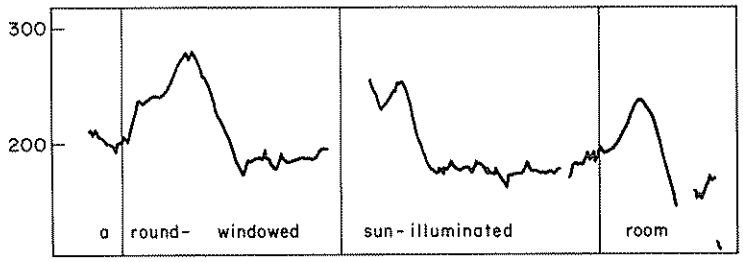

Figure 24
Fo contour for utterance *a round-windowed, sun-illuminated room.*

is that it is unclear why the boundary tones would differ in this way. Indeed, one's intuition is that the three phrases should be treated in exactly the same way. However, if the same Fo contour with a final upturn for the boundary H is applied to all three list items, the sense of disjuncture is increased to the point where one might wonder if three different people were being queried about the three different items. Our new account is much cleaner. Under the new account, the three items are treated in parallel as intermediate phrases with a H phrase accent. The final H boundary tone then becomes a property of the entire construct. For each item to be assigned a final H boundary tone, each would have to be a separate intonation phrase. This captures directly the feeling that the level of disjuncture is increased by the change of melody.

A third class of cases where intermediate phrasing is commonly used is in sequences of modifiers which are to be interpreted in parallel. A closely related use is to disambiguate the scope of conjunction. Fig. 24 shows an Fo contour illustrating the first of these uses, for the sentence:

(8) A round-windowed, sun-illuminated room.

Note the way that the L phrase accent has spread over the deaccented regions occupied by the words *windowed* and *illuminated.* In this sentence, the existence of disjuncture is conveyed by a comma in writing. However, the authors of the Olive/Liberman text-to-speech system report to us that interpreting commas in modifier strings as full intonation breaks results in a marked impression of disfluency. Except in the case of very long and complex modifiers, an intermediate phrase boundary appears to be called for.

Figs. 25 and 26 are examples where we believe that the intermediate phrasing disambiguates the scope of conjunction. The sentence:

(9) A pale orange and yellow ballgown.

is ambiguous between an interpretation in which only the orange is pale, and one in which both the orange and the yellow are pale. In the utterance of this sentence shown in Fig. 25, *pale* is set off as an intermediate phrase, and *orange and yellow* are put together in the following intermediate phrase;

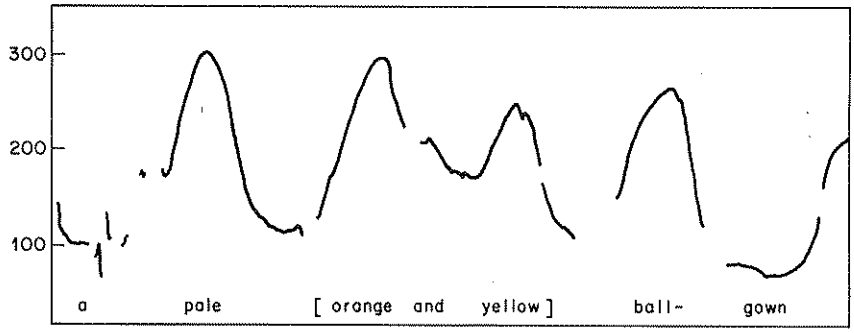

Figure 25

Fo contour for utterance *a pale orange and yellow ballgown* with *pale* modifying the conjunct *orange and yellow.*

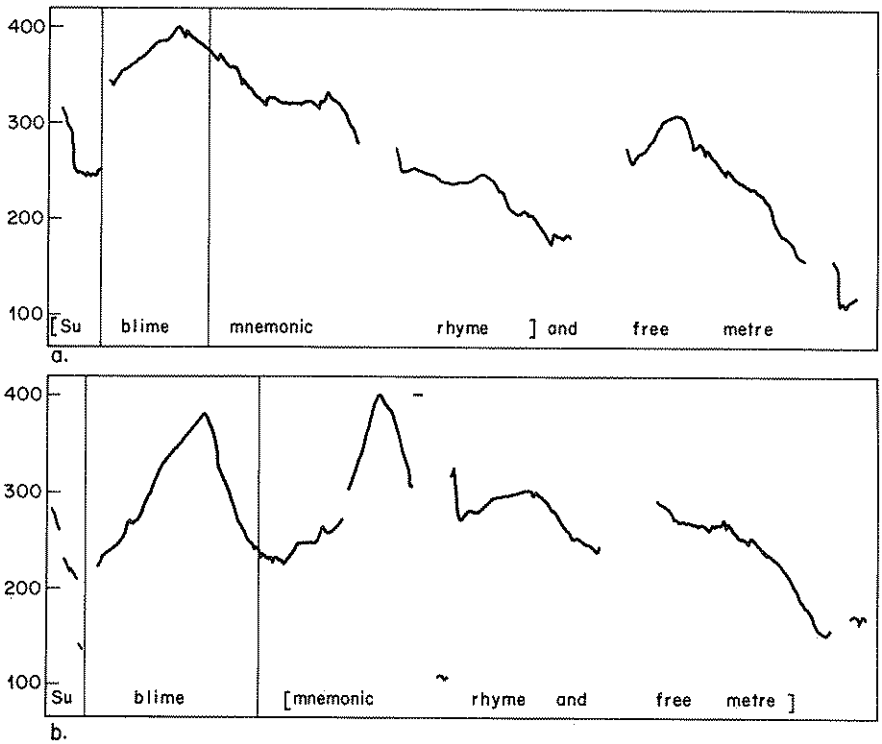

Figure 26

Fo contours for two versions of *Sublime mnemonic rhyme and free metre.* In version (a) *sublime* modifies only *mnemonic rhyme.* In version (b) it modifies both noun phrases.

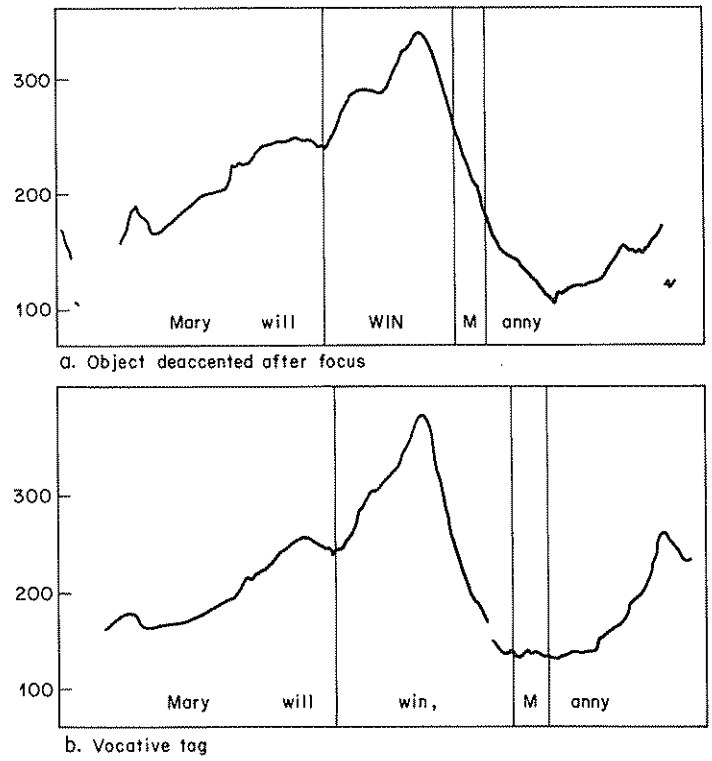

Figure 27

Fo contours for two segmentally matched utterances *Mary will win Manny*. In version (a) *Manny* is the object of the transitive verb *win*. In version (b) *Manny* is a vocative tag following the intransitive verb.

this phrasing unambiguously conveys the second interpretation. Fig. 26 shows two utterances which illustrate the contrast between the two possible interpretations of:

(10) Sublime mnemonic rhyme and free metre.

In the first utterance (Fig. 26a), a boundary after *rhyme* indicates that the mnemonic rhyme is sublime, but not necessarily the free metre. In the second utterance (Fig. 26b), the break occurs after *sublime*, and here both the rhyme and the metre are taken to be sublime. Note that in this second utterance, the stressed syllable of *sublime* is considerably longer than in the first utterance, and it carries the fall due to the presence of the L phrase accent.

As a fourth example of intermediate phrase breaks, we consider the treatment of tags. Many authors (Bing 1979; Gussenhoven 1984) have felt that tags are prosodically in closer construction with the main clause than a separate intonation phrase would be. On the other hand, there are obstacles to treating them as part of the same intonation phrase as the main

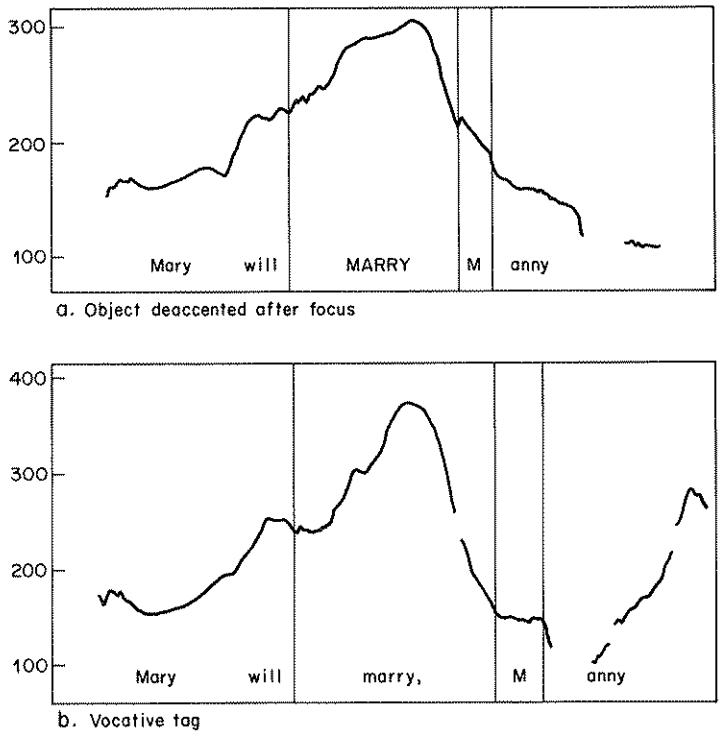

Figure 28

Fo contours for two segmentally matched utterances *Mary will marry Manny*. In version (a) *Manny* is the object of the transitive verb *marry*. In version (b) *Manny* is a vocative tag following the intransitive verb.

clause. Figs. 27–29 illustrate their special properties. We believe that these special properties can be accounted for by positing an intermediate phrase boundary between the main clause and the tag.

The two utterances in Fig. 27 exemplify a contrast which was first pointed out to us by M. Liberman. In the first utterance (Fig. 27a), the verb *win* is under focus and its object *Manny* is deaccented because it is in postnuclear position in the same phrase. In the second utterance (Fig. 27b), *win* is intransitive and *Manny* is interpreted as a vocative tag. Note that in the second utterance, the Fo fall is complete by the end of *win*, so that the Fo level on the [m] of *Manny* is low and level. In the first utterance, on the other hand, the Fo fall only begins on *win*, and it continues through the [m] of *Manny* and into the following vowel. A related observation is that the syllable *win* is much longer in the utterance with the vocative tag. These differences may be summarised by saying that in the utterance with the vocative tag, *win* has a duration and Fo pattern which is typical of a phrase-final nuclear stressed syllable. In the utterance with a deaccented object, *win* has the typical phonetics of a syllable which has nuclear stress but is not phrase-final. When the verb is bisyllabic with

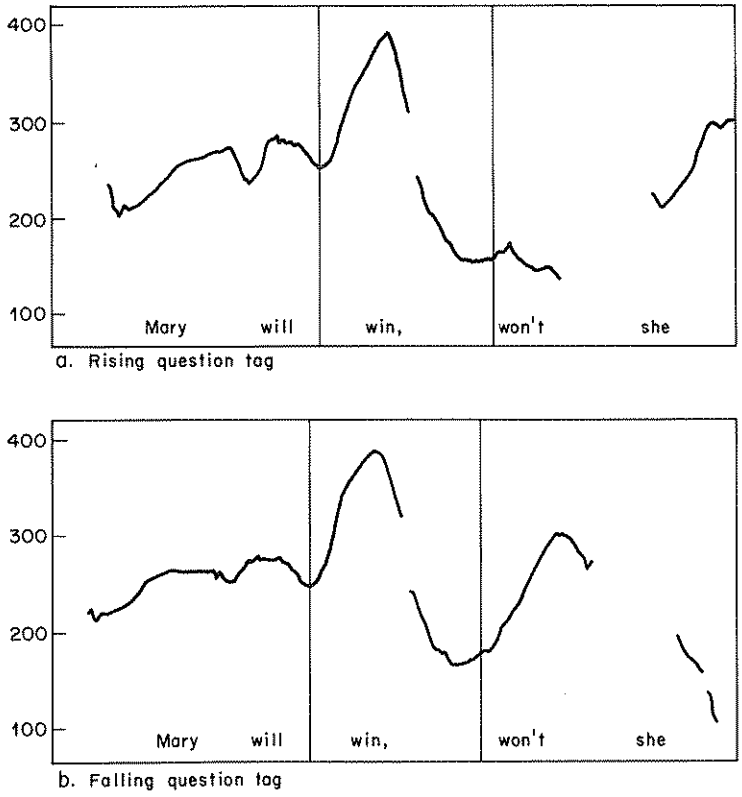

Figure 29

Fo contour for two versions of *Mary will win, won't she.* Version (a) has a rising tag contour, version (b) a falling tag contour.

initial stress, as in the utterances in Fig. 28, a similar contrast is observed. Here, however, the contrast is somewhat less striking, because even in the utterance with the vocative tag, the stressed syllable is not absolutely phrase-final.

These special properties of vocative tags are also shared by question tags. The two utterances in Fig. 29 illustrate two possible treatments of the question tag. The first has the same phonetics as the vocative tags just discussed. In the second, the treatment of *win* is the same, but the tag is handled differently; it has an Fo peak and fall. If the contour in Fig. 29b is treated as a single phrase, then its nuclear accent must be on the tag; the tag has the last accent, and many different theories agree that the last accent in a phrase is the most prominent one. This conclusion seems very counterintuitive; most listeners feel that the main stress of this construction is on *win*, with a subordinated stress on the tag. Positing an intermediate phrase boundary before the tag provides a way to handle the subordination. Since each intermediate phrase can have its own pitch range, we can take

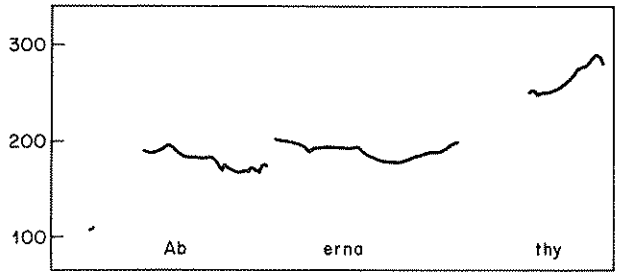

Figure 30

A low-rising contour produced on a vocative in isolation. This might be used to gently attract the attention of someone in the same room.

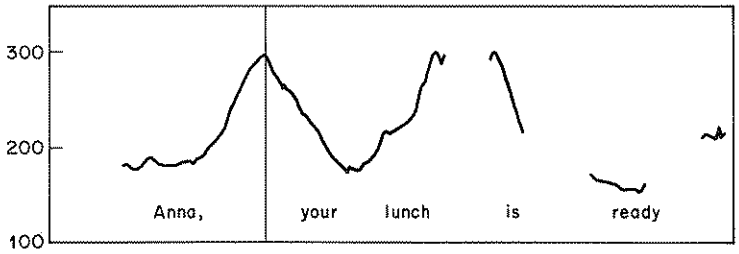

Figure 31

A low-rising contour on a preposed vocative.

the tag as a whole to have a pitch range which is subordinated to the pitch range of the main clause. The consequence is that its strongest element is not as intonationally prominent as the strongest element of the main clause.

This treatment of tags immediately raises the question of whether every tag has a pitch accent. Some authors have suggested that vocative tags, in particular, are deaccented. We believe that the intuitions of these authors are based on the subordination of the tag phrase as a whole, and that even vocative tags do have an accent. The transcription for the tags in Figs. 27b, 28b and 29a would in this account be L* L H%. We see two arguments for this position, though neither is really conclusive. One is based on the comparison to tag questions. The falling contour for tag questions, shown in Fig. 29b, cannot be generated without positing a H* accent on the tag. The rising contour, shown in Fig. 29a, evidently differs in melody, but we see no reason to believe that it differs in rhythmic structure. Accordingly, we would suggest that it has a L* rather than a H* accent. But then, there seems to be no compelling reason to view the vocative tags as phonologically distinct from the rising tag questions, and so they too must have a L* accent.

A second argument is based on the comparison of vocative tags to vocative expressions in other positions. A low-rising pattern is a common

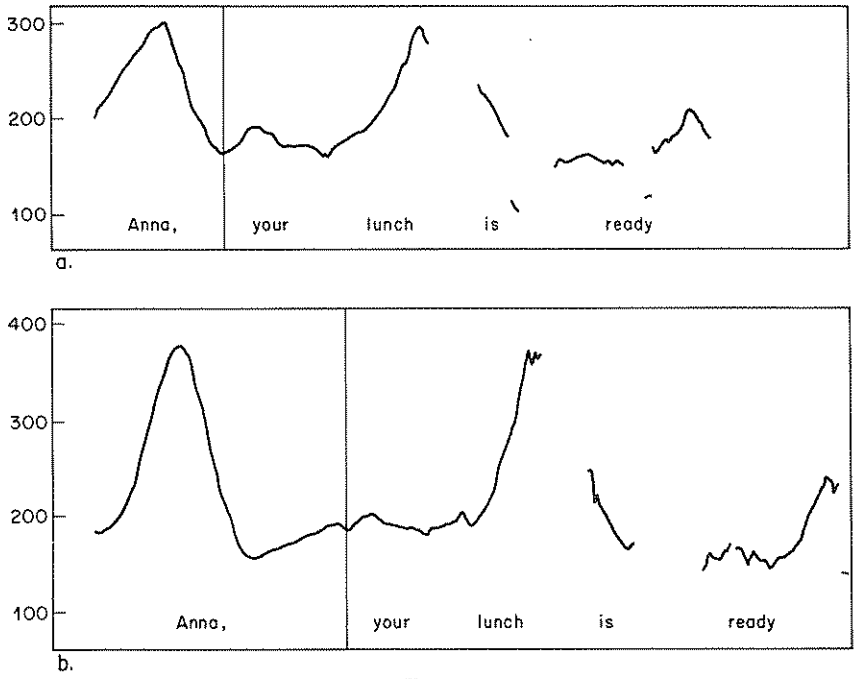

Figure 32

Two other possible patterns for a preposed vocative. Pattern (a) is probably H* L, and is separated by an intermediate phrase boundary from the main clause. Version (b) is L + H* L H %, with an intonational phrase boundary separating it from the main clause.

one for a vocative produced in isolation, as shown in Fig. 30. Under all theories of English intonation that we are familiar with, a word produced in isolation must have an accent. In Fig. 30, the accent is evidently a L*. The low-rising pattern can also be found in preposed position, as in Fig. 31. Here, it alternates with other patterns which have accents, for example the ones shown in Fig. 32. We see no reason to single out the low-rising pattern in this position as lacking a pitch accent.

It is of course true that the vocative tags differ from these other cases in occurring only with a low-rising melody. This fact is both striking and surprising, since choice of melody is not in general determined by syntax. Consider, for comparison, syntactic yes/no questions, which can be produced with either a falling or a rising melody, depending on their pragmatic force. However, this striking fact is not in principle amenable to a phonological explanation. No matter what phonological representation is chosen for the vocative tag contour, independent principles must be brought in to explain why the vocative tags are assigned that contour and not some other one. The interesting question is how general these principles are. At one extreme, one might propose that the vocative tag contour is simply an intonational idiom. But it would be more interesting

if the restricted choice of contour fell out from more general considerations. Our suspicion is that the vocative tags are not actually syntactically special, but rather occupy the same syntactic position as other postponed noun phrases. Their special status has to do with their conditions for felicitous use, and when these conditions are made precise, it seems likely that they will explain the choice of intonation pattern.

The transcription proposed here for vocative tags differs from that in Pierrehumbert (1980). Pierrehumbert (1980) did not posit a level of intermediate phrasing, but did suggest that intonation phrases ending in a vocative tag have two phrase accents. One phrase accent ended the main clause, and the other described the Fo value up to the boundary tone on the tag. We see two weaknesses in this proposal. First, without positing a phrase boundary between the two phrase accents, the characteristic pattern of duration and tonal alignment is not accounted for. Second, the proposed transcription has no use other than for tags. We feel sceptical of a solution which relies on such a specialised addition to the phonological system. To make an analogy, we would feel similarly sceptical of a theory under which a syllable could have four consonants in the onset just in words for expressing affection.

4.4 Catathesis and the intermediate phrase in English

We have already shown that the intermediate phrase is the domain of catathesis in Japanese. For English, Pierrehumbert (1980) took the intonation phrase to be the domain for catathesis. However, some examples involving use of the L + H accents call this assumption into question. We believe that these examples can be handled by taking the intermediate phrase, rather than the intonation phrase, to be the domain for English catathesis.

The structural description for catathesis given in §3 predicts that all chains of L + H accents should show successively lower H tones. Intonation patterns which have this property certainly occur, as shown in Fig. 33. However, it is easy to find cases which do not, for example Fig. 34. A common feature of these examples is that they have focus or emphasis on the word whose accent has failed to undergo lowering. Recalling that focus in Japanese induces an intermediate phrase boundary, we would suggest that in case like Fig. 34, too, an intermediate phrase boundary has blocked catathesis. Note that cases occur in English in which a full intonation break has evidently been introduced before a focused item in order to make it more salient. For example, the answer in (11) can be produced with a continuation rise and a pause at the % juncture:

(11) But it doesn't make any sense? Why did she do it?
– Well I think she did it % for JOHN'S sake.

In the utterance in Fig. 34, the sense of disjuncture is not as striking as in (11), but the emphasised element still seems to have been set off. This impression is accounted for by supposing that a phrase boundary has been

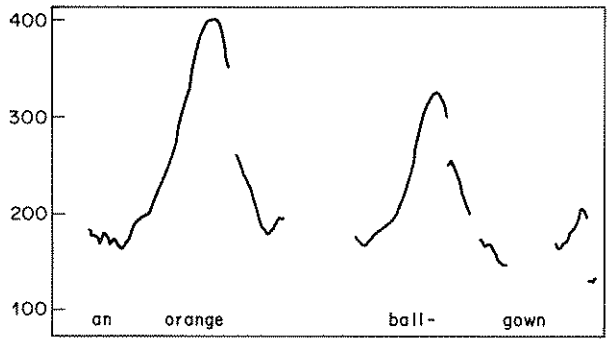

Figure 33

Fo contour for utterance *an orange ballgown* in which the second of two L+H accents has been lowered by catathesis.

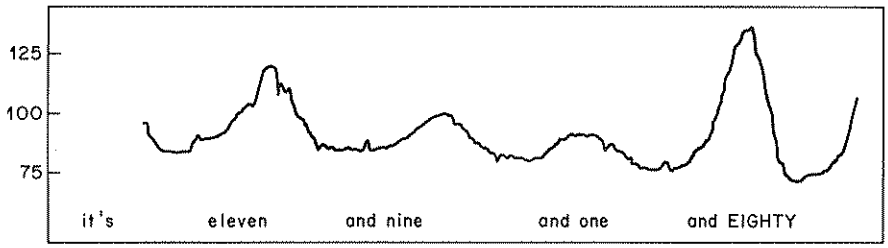

Figure 34

Fo contour for utterance *eleven and nine and one and EIGHTY* with downstepping pitch accents on the first three numbers, but catathesis blocked on the focused *eighty*.

introduced before the focused element in Fig. 34, but a weaker one than in (11).

Further support for this analysis comes from considering the Fo contours already presented in Figs. 25 and 26. In these, words which are grouped together as a single intermediate phrase participate in catathesis. At the intermediate phrase boundaries, the pitch range appears to be reset. It is not merely the case that the first pitch accent of the new phrase fails to be lowered relative to the immediately preceding one; this could happen because the immediately preceding accent was not bitonal. It is also the case that the pitch range compression due to earlier applications of catathesis fails to propagate to the first accent of the new phrase.

Rigorous testing of this proposal is made difficult by the way that phrasal pitch range can be manipulated. Fig. 35 shows an Fo contour with a decided downtrend, in which catathesis might be taken to have applied. However, the intonation pattern evidently has three intonation phrases, as evidenced by the continuation rises on the first two. Moreover, the relative pitch range of these phrases can be accounted for independently by principles proposed in Brown *et al.* (1980). They suggest that the pitch

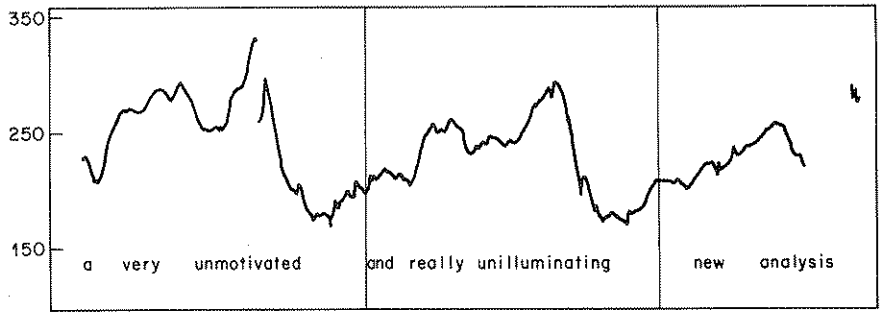

Figure 35

Fo contour of three intonation phrases, in which phrasal manipulations of overall pitch range mimic catathesis.

range is raised when initiating a new topic, and lowered when concluding one. For discourse segments consisting of only one topic, a downtrend is accordingly predicted. Fig. 35 is therefore not a counterexample to the claim that the intermediate phrase is the domain of catathesis. A more promising avenue for testing is the interpretation of cases in which catathesis has been blocked and the pitch range has been reset. The analysis predicts that failure of catathesis should strongly bias the interpretation of some ambiguous constructions. For example, scope of modification could be signalled in this way. Perceptual experiments would be necessary to test this claim.

5 Even larger units of phrasing?

In Liberman & Pierrehumbert's work on English, phonetic effects were identified whose domain was evidently larger than the phrasal units just discussed. Specifically, Liberman & Pierrehumbert (1984) found that the ends of declarative sentences were subject to a process of final lowering. Final lowering is a gradual compression and shift of the pitch range which occurs in anticipation of the end of a declarative utterance. It affects the scaling of accents as well as the realisation of the postnuclear tones.[14] Subsequent experiments reported in Pierrehumbert & Liberman (1983) suggested that the time interval for final lowering is about half a second. This means that in multi-phrase utterances, only the last intonation phrase is typically affected. Unless the last intonation phrase is quite short, only the later portion of it will be affected.

In our experiments on Japanese, we found that final lowering is also characteristic of Japanese declarative intonation. Fig. 36 demonstrates the nature of final lowering by comparing average values at successive measurement points in a declarative sentence to the same data in the corresponding yes/no question. The mean values are virtually identical in

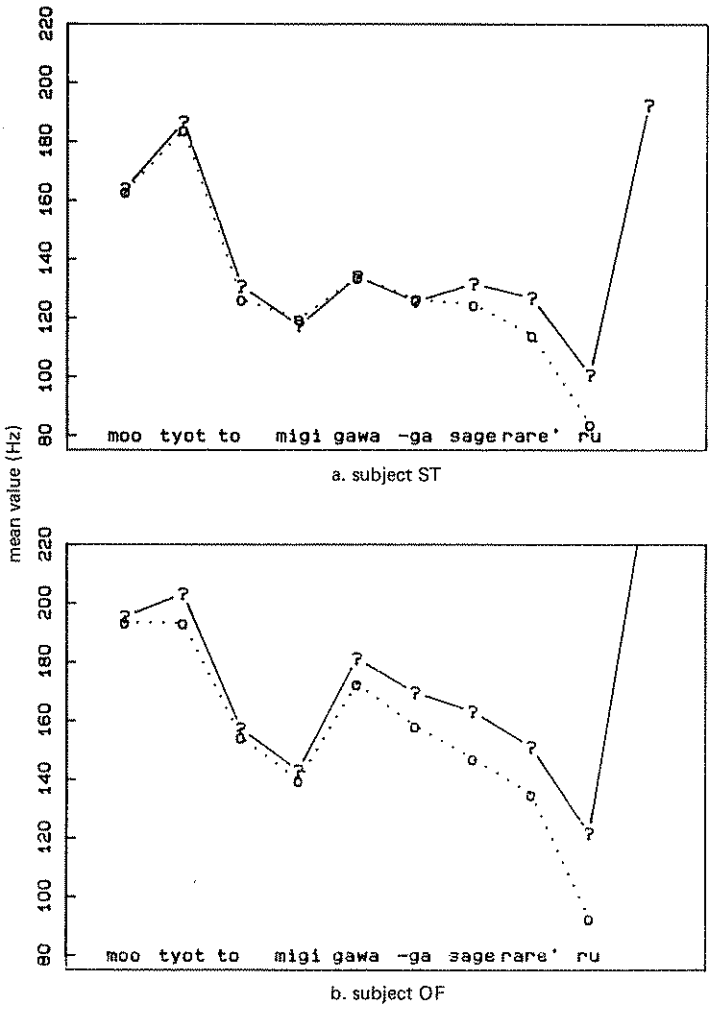

Figure 36

Averaged values at successive measurement points in two versions of *moo tyo'to migigawa-ga sagerare'ru*. Dotted lines connecting points labelled 'o' are for declarative renditions, solid lines connecting '?' for interrogative renditions. Utterances by subject ST (a) and by OF (b).

the initial portion of the sentence, and then they diverge increasingly during the final portion.

Our experiments on Japanese also identified a small but significant declination effect. (We define declination as a lowering of the pitch range which operates in time from the beginning of the utterance, without regard to the tonal description.) This conclusion was based on an examination of the peak relations in the utterance set illustrated above in Figs. 3–5. In this

set, the tonal analysis of all sentences was the same; only the length of the first accentual phrase was varied. Without declination, the value of the second phrase peak relative to the first phrase peak should be identical in all cases. In fact, however, the value was slightly smaller in the longer sentences; declination affected the peak in *oma'warisan* more the further away it was from the peak in the preceding clause. The magnitude of the effect was about 10 Hz per second for the subject with the greatest effect; one subject showed no effect at all. This observation was true for the unaccented phrases of different lengths illustrated in Fig. 4, and also for the contrasting sentences with accented first phrase.

In the utterances in the declination experiment, the boundary between the first and second phrases was an intermediate phrase boundary. We know this because the second phrase did not show the effects of catathesis when the first phrase was accented (see Fig. 17 in §4.1). Since declination affected the relation of the two phrases, it follows that declination is not a process whose domain is the intermediate phrase. Instead, it must be operating at some larger level of structure.

In English, the existence of declination is controversial. Many studies reporting declination have not controlled the materials phonologically in any way that would allow them to separate declination from catathesis and final lowering. For example, Pierrehumbert (1979) reported that listeners compensated for declination in judging the relative height of two intonational peaks. However, since only two peaks were involved and the second was near the end of the utterance, these results can equally be interpreted to mean that listeners compensated for final lowering. The detailed model of fundamental frequency scaling in Liberman & Pierrehumbert (1984) does not include declination, and still achieves a very close fit to their data. However, the Japanese results suggest that the declination effect would be very small if it exists in English. Furthermore, the materials in Liberman & Pierrehumbert (1984) were not designed to separate declination from catathesis and final lowering. It is therefore likely that a model with a small declination component would provide an equally good fit to the data, with only slight adjustments of values for other parameters. Thus declination is not inconsistent with the data but is not actively supported either.

To summarise, then, final lowering and declination in Japanese appear to operate over the largest units examined, namely sentences with complex phrasing patterns. In English, final lowering characterises the end of some unit larger than the intonation phrase. We might expect declination to share this large domain if it actually exists in English. In our earlier work, such as Beckman & Pierrehumbert (1985), these considerations led us to posit a very large unit of phonological structure, which we called the 'utterance'. This unit was taken to provide the domain for assignment of declination and final lowering.

However, recent work by Hirschberg & Pierrehumbert (1986) calls this assumption into question. They investigated how final lowering can be used to signal discourse structure in synthetic speech. The specific domain

was a computer-aided instruction system designed to teach naive users the vi text editor (Nakatani *et al.* 1985). The intonation of the synthetic speech was controlled using the Anderson *et al.* intonation synthesis program. The main finding was that rhetorically effective intonation can use final lowering hierarchically to convey the hierarchical structure of the discourse; final lowering applies only at the ends of sentences which conclude a topic or subtopic. There were some indications that the amount of final lowering actually varies continuously, depending on how important a unit is being concluded. Consider, for example, the beginning of the introduction to the tutoring system, which we put in outline form to indicate the discourse structure:

(I) Hello.
  (A) Welcome to word processing.
    (1) That's using a computer to write letters and reports.
  (B) Word processing makes typing easy.
    (1) Make a typo?
      (a) No problem.
      (b) Just back up, type over the mistake, and it's gone.
    (2) And, it eliminates retyping.
      (a) Need a second draft?
      (b) No problem.
      (c) Just change the first, and you've got the second.
(II) Today, the computer will teach you word processing...

In this text, final lowering was applied in only three places: at the end of IA1 (*That's using a computer to write letters and reports*); at the end of IB1b (*Just back up, type over the mistake, and it's gone*); and at the end of IB2c (*Just change the first, and you've got the second*). More final lowering was applied in IB2c than in IB1b and IA1, because IB2c ends a top-level discourse unit, whereas the other two end subordinated units.

Note that the stretch of discourse from the beginning of the text to the first application of final lowering is not, by this algorithm, a constituent in the discourse structure. Our feeling is that it is not plausibly viewed as a phonological constituent either. This feeling was reinforced by our recent attendance at a particularly verbose and rambling lecture, where the speaker made no use of final lowering for as much as five minutes at a time. We also note that final lowering may be omitted at the end of utterances where the speaker is setting a topic which he expects the other conversational participant to elaborate. For example, Wh-questions frequently seem to lack final lowering. Obviously, the Wh-question and the other party's reply cannot be treated as a single phonological phrase, of any level.

In the intonation synthesiser, the amount of final lowering affects the value for a L boundary tone. The manipulations of final lowering just described varied the terminal value by up to 10%, or 7 Hz for a low-pitched male voice. Thus Hirschberg & Pierrehumbert's work would question the claim made by many authors that the lowest point of a

declarative terminal fall is an invariant property of a given speaker's voice. There are two possible reasons for this contradiction between the synthesis experiment and the earlier reports. First, studies such as Liberman & Pierrehumbert (1984) have involved highly simplified discourse structures in which all declarative sentences ended a topic. In such utterances, the final lowering would not, by hypothesis, be varied. Second, the end of the utterance is a very difficult place to measure fundamental frequency, because of disturbances of voice quality typically found there. Even when the pitch is clear to the ear, it may not be measurable because of pitch-tracking errors. This means that the variance for values measured at this point is large relative to the magnitude of the effect we are describing. This would especially be the case for spontaneous speech recorded in natural conditions, such as Boyce & Menn (1979) describe. In synthetic speech, on the other hand, the terminal fundamental frequency is well-defined because the disturbances in voice quality are not reproduced. Accordingly, the relation of terminal value to perceptual effect is readily apparent.

There is a strong relationship between Hirschberg & Pierrehumbert's results on final lowering and observations by other authors about choice of pitch range. The *beginning* of a new topic is commonly observed to be marked by an expansion of the pitch range (see Brown *et al.* 1980, and other references cited there). In their synthesis of the script for the vi tutor, Hirschberg & Pierrehumbert implemented these ideas rigorously. They adjusted the overall pitch range for the first phrase of each sentence so as to reflect depth of embedding in the hierarchical discourse structure. The resulting synthesis was extremely satisfactory, and far superior to versions in which the pitch range for initial phrases was not varied or was varied depending on the number of phrases within the same sentence.

Note that the stretch of discourse from one pitch range expansion to the next is not necessarily a constituent in the discourse structure, by this algorithm. This is the case because the pitch range can be expanded to initiate a subtopic when the larger topic marked by a previous expansion is not yet complete. The stretch from a pitch range expansion to the matched application of final lowering is a constituent. But it may be a very extended one, containing a great deal of internal structure.

We wonder whether a connection can be made between the use of pitch range expansion to initiate a topic, and the use of final lowering to terminate one. If this analogy were a strict one, we would expect the pitch range expansion, or initial raising, to be time-dependent. That is, it should show its greatest effect at the very beginning of an utterance, and gradually dwindle in strength as the utterance proceeded. Furthermore, it should operate without regard to the type and location of tonal elements. Note that these are exactly the characteristics of declination, as we have defined it. The empirical predictions of this view are that declination should be regularly found in experimental tasks where each utterance stands alone as a topic unit. In more complex materials, it would only be observed in some utterances, those which initiated a new topic. Furthermore, it seems possible that the amount of declination would vary

depending on how major a topic change was being initiated. An experiment which addresses some of these points is reported in Umeda (1982). Her results are somewhat difficult to interpret, because her materials were not in any way phonologically analysed. However, her conclusions are consistent with the predictions we have just laid out. Specifically, she found reliable downtrends only in sentences produced in isolation, and not in medial sentences in a paragraph.

Returning to issues of phonological structure, we can now summarise our beliefs about levels of phrasing above the intonation phrase. Initial results on final lowering and declination suggested that they were characteristic of a phonological unit on the scale of a sentence or utterance. More detailed investigation undermines this view. Final lowering appears to be controlled by the discourse structure in a way which makes it seem implausible that it defines a unit of phonological phrasing. It seems possible that declination fits the same pattern. These conclusions make the phonological status of final lowering and declination somewhat problematic. There is a strong possibility that they contrast with the internal characteristics of the intonation phrase in being truly paralinguistic. We suspect that this will prove true of declination and final lowering in Japanese as well as in English.

6 Conclusion

Japanese and English, then, seem to share intonational features and rules of many different sorts. There is a very general similarity between the two intonation systems in that both organise the tone features into a hierarchy of prosodic structures, from the grouping of tones into pitch accents at a local level to the choice of phrase-terminal tones and the manipulation of pitch range over larger domains. In addition, there are more detailed similarities at many specific levels of the prosodic hierarchy. Thus in both languages, pitch accents or accentual phrase tones are limited to being one or two-tone units. In both languages, the two-tone accents trigger catathesis. In both languages, catathesis seems to occur only within some intermediate level of phrasing, which is closely connected to such organisational effects as focus domain or scope of conjunction. In both languages, larger prosodic units are marked by boundary tones which are aligned to the edges of the unit and are not phonologically associated to any particular tone bearing unit. In both languages, the boundary tone at some level can be H, and the choice of a H tone at this level has similar pragmatic consequences in the two languages.

Of course, a detailed comparison also reveals a few differences between the two languages. For example, Japanese accents have a fixed tonal shape, and their potential loci in lexical items are not marked by having longer durations or a different set of vowel types as are English accent loci. Japanese also has lexically unaccented words and consequently can have well-formed utterances without any pitch accents, which would be

impossible in the English intonation system. Because of the lexical origin of these accentual-phrase-internal features in Japanese, the range of possible intonational variation is considerably smaller than in the English intonation system; aside from the single choice between having or not having a H boundary tone, the only sources of variation seem to be different choices of phrasing and of pitch range.

However, these differences seem rather minor next to the many major similarities between the two languages. The fact that tonal specification in both languages is similarly sparse and the fact that these sparsely specified tones are organised into hierarchical structures which have similar effects on their alignment and scaling seem especially indicative of a more fundamental similarity between the two languages. We wonder whether these similarities might be characteristic of intonational structure in general. It will be interesting to see whether detailed examination of prosodic systems in other languages reveals similar uses of tone.

NOTES

[1] See Pierrehumbert & Beckman (forthcoming) for a critique of analyses such as those of Kawakami (1956, 1961) or Clark (1978), which involve dynamic tones (tone changes as primitives) instead of tone levels.

[2] Pierrehumbert (1980) used the H*+H to explain certain 'hat patterns' that we would now analyse as involving ordinary H* accents produced in an elevated but compressed pitch range. This reanalysis was a natural outcome of the new treatment of pitch range introduced by Liberman & Pierrehumbert (1984).

[3] Note that by 'accented' here, we mean 'intonationally prominent'. When applied to English, the term 'accent' could also refer to a rhythmic or durational prominence that is not necessarily accompanied by a pitch accent. In referring to such non-tonal prominences, we will use the term 'stress' instead.

[4] The HL shape figures also in another claimed advantage for Ladd's analysis. Ladd uses his HL accent not only in this reanalysis of the prenuclear sequence H* L+H*, but also to denote any phrase-final configuration in which a tone is followed by a L phrase accent. Both L*+H and H* in this configuration result in a following fall, and Ladd considers it to be a strength of his system that they would both be analysed then as HL, allowing the functional generalisation that 'all falls are HL' (1983:730). However, this generalisation would hold only for such phrase-final configurations. When prenuclear tonal sequences are considered, his system does not rule out other sequences containing falls of some sort, such as H L and H L+H.

[5] This rule of tonal implementation is discussed further in §3 below.

[6] It is in this respect very much like the H and L boundary tones that align to the edges of intonational phrases in English without being associated to any syllable.

[7] This would be the usual way to produce a contrastive emphasis on the first word.

[8] The marking of strong emphasis by duration in Japanese has been phonologised to some extent, as in the many adverbs that have alternative neutral and emphatic forms – for example *amari* ~ *ammari*, *bakari* ~ *bakkari* and *yahari* ~ *yappari*.

[9] The term 'downstep' was first used to refer to certain well-known phenomena in African tone languages that vaguely resemble the tonal implementation phenomenon we describe here. Since we are not in a position now to evaluate the precise extent of the formal similarity between the two classes of phenomena, we will use the term 'catathesis' only, withdrawing the term 'downstep' as it was used in Pierrehumbert (1980) and Beckman & Pierrehumbert (1985) out of deference to the Africanist usage.

[10] The reason for this difficulty has to do with higher levels of phrasing that will be discussed in §4.1.

[11] This proposal requires two significant revisions in the transcriptional conventions proposed in Pierrehumbert (1980). First, L*+H H was formerly taken to differ from L* H by the abruptness of the Fo rise at the nuclear syllable. We now believe that such contours are really L* H, and that the varying abruptness of the rise is merely a stylistic dimension. Further investigation of low-rising questions in conjunction with work on the Anderson *et al.* (1984) intonation synthesiser has suggested that the timing of the H phrase accent varies considerably, with no obvious difference in semantic interpretation. Second, the transcription H*+L L+H was previously held to describe a class of two-peaked contours in which the second H did not undergo catathesis. Under the present proposal, catathesis would apply in this case, but not in the case of H* L+H. Also, some such contours obviously may involve an intermediate phrase boundary (see §§4.3 and 4.4).

[12] Ladd's major motivation for the feature [raised peak] appears to be a misinterpretation of results in Liberman & Pierrehumbert (1984). In these experiments, subjects were prompted to produce a particular relative prominence relation between two H* accents. The peak Fo values for these accents then turned out to be reliably separated in data on their production. Ladd takes these data to indicate that the peak Fo values were being controlled by a binary feature. However, Liberman & Pierrehumbert do not share this view. They believe that the linguistic system in general treats this dimension of variation as gradient, and that the data took the form they did because only two values along this gradient dimension were elicited in the experimental conditions.

[13] In this framework, the boundary tone is raised, or upstepped, when it follows a H phrase accent. If the boundary tone is H, this rule generates the characteristic rise–plateau–rise shape of questions having a L* nuclear accent in English. If it is L, the result is a high level tail. Clearly, the description is empirically equivalent to one in which the phrase accent is either L or H and is followed by an optional H boundary tone.

[14] Because final lowering can affect a nuclear accent peak as well as the postnuclear region, it cannot be identified with the declarative terminal fall of Lieberman (1967). We would describe the terminal fall itself as the reflex of the H* L L% tonal transcription.

REFERENCES

Anderson, Mark D., Janet B. Pierrehumbert & Mark Y. Liberman (1984). Synthesis by rule of English intonation patterns. In *Proceedings of the IEEE International conference on Acoustics, Speech, and Signal Processing*. 2.8.2–2.8.4.

Beckman, Mary E. (1986). *Stress and non-stress accent*. Dordrecht: Foris.

Beckman, Mary E. & Janet B. Pierrehumbert (1985). Synthesizing Japanese using a downstep model. *JASA* 77. S38.

Bing, Janet D. (1979). *Aspects of English prosody*. PhD dissertation, University of Massachusetts, Amherst. Distributed by Indiana University Linguistics Club.

Bolinger, Dwight (1965). Pitch accent and sentence rhythm. In I. Abe & T. Kanekiyo (eds.) *Forms of English : accent, morpheme, order*. Harvard: Harvard University Press. 139–180.

Boyce, S. & L. Menn (1979). Peaks vary, endpoints don't: implications for linguistic theory. In *Proceedings of the 5th Annual Meeting of the Berkeley Linguistics Society*. 373–384.

Brown, G., K. Currie & J. Kenworthy (1980). *Questions of intonation*. London: Croom Helm.

Clark, Mary (1978). *A dynamic treatment of tone with special attention to the tonal system of Igbo*. PhD dissertation, University of Massachusetts, Amherst.

Crystal, David (1969). *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.

Fry, Dennis B. (1958). Experiments in the perception of stress. *Language and Speech* **1**. 126–152.

Gussenhoven, Carlos (1984). *On the grammar and semantics of sentence accents*. Dordrecht: Foris.

Halle, Morris & Jean-Roger Vergnaud (1985). Stress and the cycle. Paper presented at the colloquium *Phonologie Pluri-linéaire*, Lyon.

Haraguchi, Shosuke (1977). *The tone pattern of Japanese: an autosegmental theory of tonology*. Tokyo: Kaitakusha.

Hirschberg, Julia & Janet Pierrehumbert (1986). Intonational structuring of discourse. *Proceedings of the 24th Meeting of the Association for Computational Linguistics, New York*. 136–144.

Huss, Volker (1978). English word stress in post-nuclear position. *Phonetica* **35**. 86–105.

Kawakami, Sin (1956). Buntoo no intoneesyon. *Kokugogaku* **25**. 21–30.

Kawakami, Sin (1961). On the relationship between word-tone and phrase-tone in Japanese language. *Onsei no kenkyuu* **9**. 169–177.

Kubozono, Haruo (1985). On the syntax and prosody of Japanese compounds. *Work in Progress, Department of Linguistics, University of Edinburgh* **18**. 60–87.

Ladd, D. Robert (1983). Phonological features of intonational peaks. *Lg* **59**. 721–759.

Lea, Wayne A. (1977). Acoustic correlates of stress and juncture. In L. M. Hyman (ed.) *Studies in stress and accent. Southern California Occasional Papers in Linguistics* **4**. 83–119.

Liberman, Mark Y. (1975). *The intonational system of English*. PhD dissertation, MIT. Distributed by Indiana University Linguistics Club.

Liberman, Mark & Janet Pierrehumbert (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. T. Oehrle (eds.) *Language sound structure*. Cambridge, Mass.: MIT Press. 157–233.

Liberman, Mark & Allan Prince (1977). On stress and linguistic rhythm. *LI* **8**. 249–336.

Lieberman, Philip (1960). Some acoustic correlates of word stress in American English. *JASA* **32**. 451–454.

Lieberman, Philip (1967). *Intonation, perception and language*. Cambridge, Mass.: MIT Press.

McCawley, James D. (1968). *The phonological component of a grammar of Japanese*. The Hague: Mouton.

Nakatani, Lloyd H. & Carletta H. Aston (1978). Acoustic and linguistic factors in stress perception. Ms, AT&T Bell Laboratories.

Nakatani, Lloyd H., D. Egan, L. Ruedisueli & P. Hawley (1986). TNT: a talking tutor 'n' trainer for teaching the use of interactive computer systems. Paper presented at the Conference on Human Factors in Computing Systems.

Nakatani, Lloyd H. & Judith A. Schaffer (1978). Hearing 'words' without words: prosodic cues for word perception. *JASA* **63**. 234–245.

Nolan, Francis (1984). Auditory and instrumental analysis of intonation. *Cambridge Papers in Phonetics and Experimental Linguistics* **3**.

O'Connor, J. D. & G. F. Arnold (1961). *Intonation of colloquial English*. London: Longman.

Pierrehumbert, Janet B. (1979). The perception of fundamental frequency declination. JASA **66**. 363–369.

Pierrehumbert, Janet B. (1980). *The phonology and phonetics of English intonation*. PhD dissertation, MIT.

Pierrehumbert, Janet B. & Mark Y. Liberman (1983). Intonational invariance under changes in pitch range and length. Paper presented at the symposium *Prosody: Normal and Abnormal*, Zürich.

Pierrehumbert, Janet B. & Mary E. Beckman (forthcoming). Japanese tone structure. Paper submitted to *Linguistic Inquiry*.

Poser, William J. (1984). *The phonetics and phonology of tone and intonation in Japanese*. PhD dissertation, MIT.

Prince, Allan S. (1983). Relating to the grid. *LI* **14**. 19–100.

Pulleyblank, Douglas (1983). *Tone in Lexical Phonology*. PhD dissertation, MIT.

Selkirk, Elisabeth O. (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, Mass.: MIT Press.

Thorsen, Nina (1980). A study of the perception of sentence intonation: evidence from Danish. *JASA* **67**. 1014–1030.

Trubetzkoy, Nikolai S. (1939). *Grundzüge der Phonologie. Travaux du Cercle Linguistique de Prague*.

Umeda, Noriko (1982). Fo declination is situation dependent. *JPh* **10**. 279–291.