

Prominence marking in the Japanese intonation system

Jennifer J. Venditti, Kikuo Maekawa, and Mary E. Beckman

Abstract: Japanese uses a variety of prosodic mechanisms to mark focal prominence, including local pitch range expansion, prosodic restructuring to set off the focal constituent, post-focal subordination, and prominence-lending boundary pitch movements, but (notably) *not* manipulation of accent. In this paper, we describe the Japanese intonation system within the Autosegmental-Metrical model of intonational phonology, and review these prosodic mechanisms that have been shown to mark focal prominence. We point out potentials for ambiguity in the prosodic parse, and discuss their larger implications for the development of a tenable general theory of prosody and its role in the marking of discourse structure.

1 Introduction

Beginning at least as early as Trubetskôï (1939), Arisaka (1941), and Hattori (1961), Japanese has played an important role in developing our understanding of prosodic structure and its function in the grammar of all human languages. In much of this literature, the primary consideration has been to account for apparent similarities between the culminative distribution of lexical stress in the lexicon of languages such as English and the distribution of pitch accents in simple and derived words of Japanese. For example, McCawley (1970) and Hyman (2001) both cite Japanese in rejecting an older typology of word-level prosodic systems in which languages are categorized by a simple dichotomy into “stress languages” such as English or Catalan and “tone languages” such as Yoruba or Cantonese. More recently, the analysis of Japanese has played a critical role in the development of computationally explicit compositional theories of the elements of intonation contours and their relationship to the prosodic organization of utterances. For example, Fujisaki and Sudo (1971) and Pierrehumbert and Beckman (1988) provide two different accounts of Japanese intonation patterns in which phrase-level *pitch range* is specified by continuous phonological control parameters that are independent of the categorical specification of more local tone shapes in lexicon. This incorporation of continuous (or “paralinguistic”) specifications of phrasal pitch range directly into the phonological description was a critical element in the development of what Ladd (1996) has named the “Autosegmental-Metrical” framework (hereafter, the “AM framework”). Beginning with Pierrehumbert’s (1980) model of the grammar of English intonation contours, the AM framework has been widely adopted in descriptions of intonation systems and in comparisons of prosodic organization across languages. (See Jun (2005) for a recent collection of such descriptions and Gussenhoven (2004) for a review of work in this framework since Ladd (1996).)

Such cross-language comparisons, in turn, are critical for our understanding of *prominence marking*, and of the ways in which prosody is used in the articulation of discourse-level categories such as *topic* and *focus*. Japanese is widely cited as a language that has a morphological marker for TOPIC. Work on other languages, including English, Catalan, and Hungarian (e.g., Valduví (1992), Roberts (1996)), strongly suggests that the morphosyntactic articulation of topic and focus is related, at least in part, to phonological constraints on the prosodic marking of prominence contrasts. For Japanese, too, we see reference to contrasting degrees of prosodic prominence as a necessary condition on some interpretations of phrases marked for TOPIC (see, e.g., Kuno (1973), Nakanishi (2004), Heycock (2007)). To fully understand how such categories as topic and focus are realized in Japanese, therefore, it is critical to understand how the prosodic marking of contrasts between prominent versus non-prominent elements works in the language.

In this chapter, we will review the currently standard AM framework account of how the prosodic marking of prominence works in the Japanese system (in Section 3). We will show that, just as in the English prosodic system, there are several different prosodic mechanisms in Japanese for marking some constituents as having focal prominence and others as being relatively reduced and out of focus. We will then describe four phenomena that are the locus of lively discussion and controversy in the further development of this AM framework account (in Section 4), before discussing the larger implications that these phenomena have for the development of a tenable general theory of the role of prosody in the marking of discourse prominence (in Section 5). Before we begin this description, however, we need to make explicit some assumptions about what prosody is and about the nature of phonetic representations that we use to study it. An understanding of these assumptions is essential background for understanding the AM framework description of Japanese. Section 2 lays out this background.

2 Elements of an AM description of Japanese intonation

In describing the prominence-marking mechanisms of Japanese, we will use the account of the Japanese intonation system that is encoded in the X-JToBI labeling conventions (Maekawa et al., 2002). These conventions build on the earlier J_ToBI conventions described by Venditti (1997, 2005), adding tags for elements of intonation contours that have not been noted in the previous literature in English, although they are observed even in relatively formal spontaneous speech such as the conference presentations in the Corpus of Spontaneous Japanese (henceforth “CSJ”; Maekawa 2003, Kokuritsu Kokugo Kenkyuujo [National Institute for Japanese Language] 2006). Both the original and the expanded tag set are intended to describe intonation contours for standard (Tokyo) Japanese. The intonation systems of other dialects are known to differ, some rather dramatically. However, they have either not been studied at all (most regional dialects) or have been studied less extensively (e.g., the Osaka system, as described by Kori (1987) and others). Both tag sets also assume a description couched in the AM framework.

2.1 The AM framework

As noted above, the term “Autosegmental-Metrical” was coined by Ladd (1996) to refer to a class of models of phonological structure that emerged in work on intonation in the 1980s, beginning with Pierrehumbert’s (1980) incorporation of Bruce’s (1977) seminal insights about the composition of Swedish tone patterns into a grammar for English intonation contours. All AM models are deeply compositional, analyzing sound patterns into different types of elements at several different levels. The most fundamental level is the separation of autosegmental content features from the metrical positions that license them. This is roughly comparable to the separation of lexis from phrase structure in a description of the rest of the language’s grammar.

More specifically, the “A” (Autosegmental) part of an AM description refers to the specification of content features that are autonomously segmented — i.e., that project as strings of discrete categories specified on independent tiers rather than being bundled together at different positions in the word-forms that they contrast. For example, in Japanese, changing specifications of stricture degree for different articulators can define as many as six different manner autosegments within a disyllabic word-form, as in the word *sanpun* [sambuɴ] ‘3 minutes’, where there is a sequence of stricture gestures for sibilant airflow, for vowel resonances, for nasal airflow, for a complete stopping of air, and then a vowel and a nasal again. However, there are constraints on which oral articulators can be involved in making these different airflow conditions, so that only one place feature can be specified for the middle two stricture specifications in a word of this shape.

Therefore, in the grammar of Japanese (as in many other languages), place features should be projected onto a different autosegmental tier from the manner features. This projection allows for a maximally general description of the alternation among labial nasal [m] in *sanpun*, dental nasal [n] in *sandan* ‘3 steps’, velar nasal [ŋ] in *sangai* ‘3rd floor’, and uvular nasal [ɴ] in the citation form *san* ‘3’. That is, the projection of consonant place features onto a separate autosegmental tier from the manner features accounts for the derivational alternation in these Sino-Japanese forms in terms of the same general phonological constraints which govern how postvocalic nasals are adapted into the language in monomorphemic loanwords such as *konpyuuta* ‘computer’ or *hamu* ‘ham’ (where the [m] corresponds to a post-vocalic [m] in the English source word), *handoru* ‘steering wheel’ (where the [n] corresponds to a post-vocalic [n]), *hankachi* ‘handkerchief’ (where [ŋ] < [n]), and *koon* ‘corn’ (where [ɴ] < [n]). A key insight of Firth (1957), which was translated into Autosegmental Phonology by North American interpreters such as Langedon (1964), Haraguchi (1977), and Goldsmith (1979), is that the traditional consonant and vowel “segments” that are named by symbols such as the [s], [a], [m], [p], [u], and [ɴ] in *sanpun* are not primitive elements in the phonological grammar. Rather, they are complex categories defined by the intersection of place and manner specifications for the word-form when the array of parallel streams of autosegments on different tiers is collapsed into a one-dimensional string.

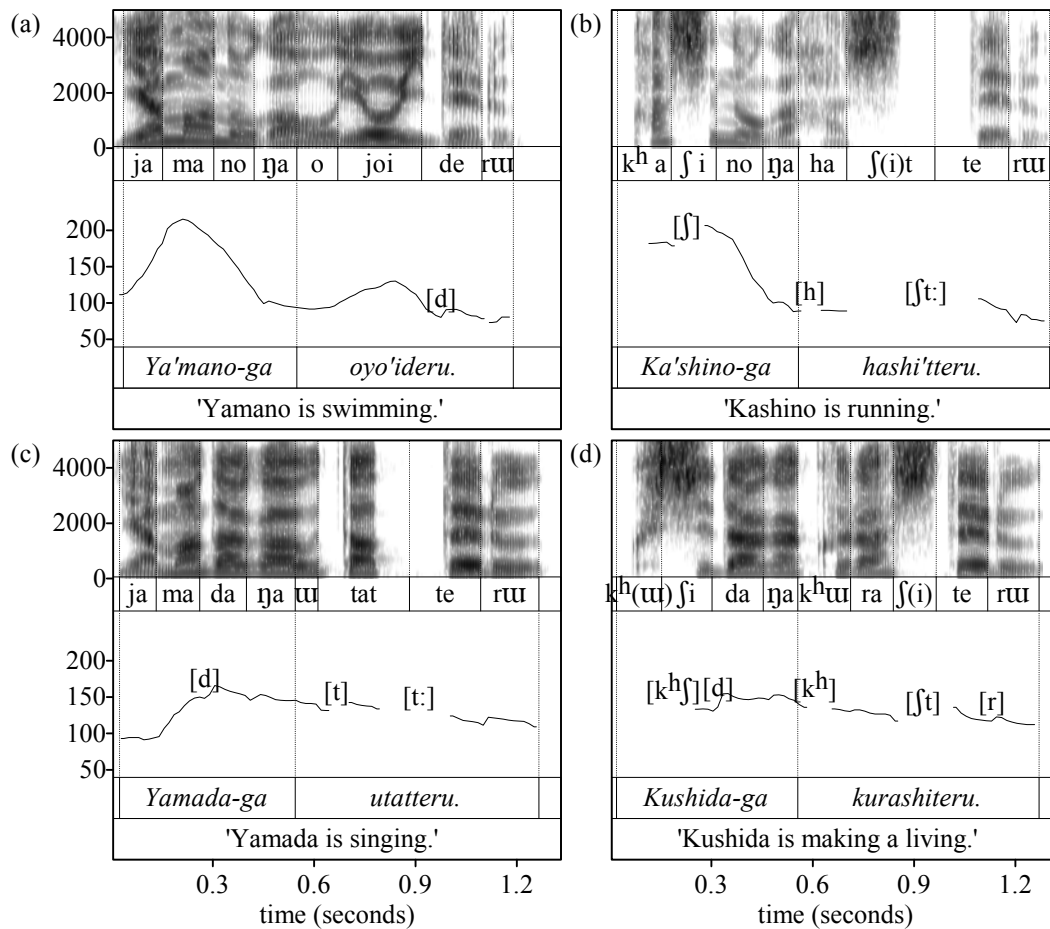
These traditional “segments” are also the names for the C or V leaf nodes in the other part of an AM description of a spoken utterance. More generally, the “M” (Metrical) part specifies the hierarchy of prosodic constituents that defines the meter or rhythm of an utterance containing the word-form. For example, in Japanese, there is a low-level prosodic constituent *syllable* (typically abbreviated as σ) that coordinates the place and manner features of words to alternate between more sonorant manners licensed to occur at the head and less sonorant manners licensed to occur only at the edges. In an utterance of *sanpun*, this syllable-level meter is CVC.CVC. In many other words, there is a more regular rhythmic alternation between C and V leaf-node types, as illustrated by the utterances in (1).

All but five of the 33 syllables in these utterances have a perfectly alternating CV structure. The exceptional cases are of two types. The first is the bare V at the beginning of the verbs *oyoideru* and *utatteru* in (1a) and (1c). These 2 V-shape syllables, like the 28 CV-shape syllables, are *short*. That is, each contains just one *mora* (abbreviated as μ), a constituent between the syllable and the CV leaf nodes which licenses the specification of vowel place and height features as the head of the syllable.

The other three exceptional syllables are *long*. A long syllable contains at least one extra mora after the head V that is the sole obligatory leaf constituent. This following mora can be either a V or a C. If it is a V, it can be the second part of a geminate vowel or it can be a less sonorous vowel than the head vowel, as in (1a), where the [i] in the second syllable of *oyoideru* has lesser sonority than the preceding head vowel [o]. If the following mora is a C, it can be a moraic nasal, as in each of the two syllables of *sanpun*, or it can be the first part of a geminate obstruent, as in the second syllables of the verbs *hasitteru* and *utatteru* in (1b) and (1c).

As *sanpun* shows, a long syllable that is closed by the moraic nasal can be word-final. By contrast, a long syllable-that is closed by an obstruent must have a following syllable to provide the second C position for the necessarily geminate consonant. This constraint is one of the motivations for positing prosodic constituents above the syllable as well as below. That is, the place and manner autosegments for a geminate consonant necessarily associate to a sequence of C nodes that is medial to a *prosodic word* (abbreviated as ω).

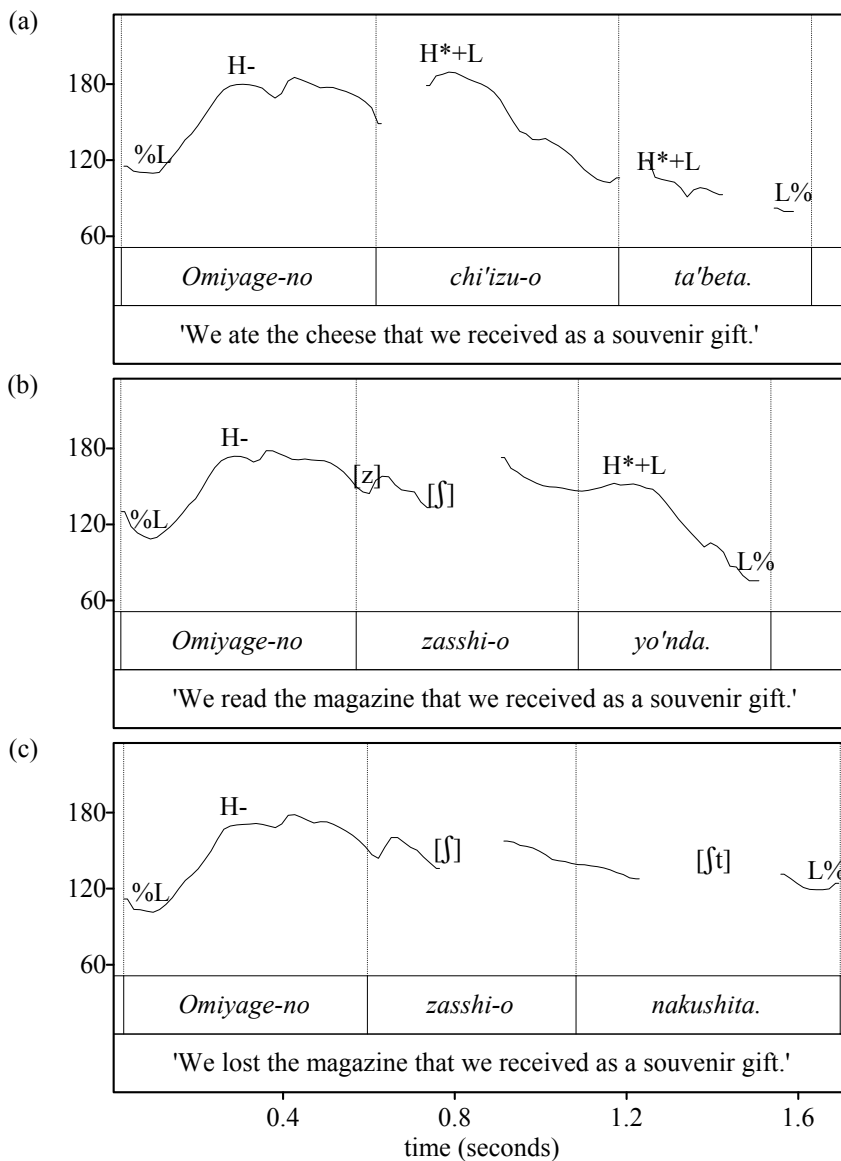
- (1) Spectrograms and fundamental frequency (F0) contours for utterances of four sentences with the structure [proper noun-NOM]_{NP} [V (cont.)]_{VP}: (a) and (b) contain two accented words, and (c) and (d) contain two unaccented words.



To recap, then, the hierarchy of prosodic constituents in the phonological description of an utterance is analogous to the hierarchy of syntactic constituents in its syntactic description. By this analogy, the prosodic constraint that each syllable must have a V (vowel) category autosegment positioned to be its head is comparable to the syntactic constraint that a verb phrase must have a V (verb) category word positioned to be its head. It is in this sense that we say that the inventory of segments for a language is analogous to the lexis of the language. One argument in favor of this analogy is that the prosodic category of a segment such as [i] is inherently ambiguous in the same way that the syntactic category of a word such as *happyoo* is inherently ambiguous. That is, *happyoo* can be parsed either as the verb ‘to announce, present’ or as the noun ‘announcement, presentation’ in different morphosyntactic contexts. Similarly, the palatal constriction that gives rise to the low first formant and high second formant on either side of the medial [o] in the word *oyoideru* in (1) can be parsed either as a C (written as ‘y’) in the syllable onset or as a V (written as ‘i’) in the second mora of the syllable. A second argument in favor of this analogy is the way that empty structural positions are interpreted. That is, the implicit but unrealized subject noun phrase in sentences such as (2a) or (2b) can be recovered from the syntactic parse of the rest of the sentence. Similarly, the implicit head vowel in the second syllable of *hashitteru* in (1b) and third syllable of *kurashiteru* in (1d) can be recovered from the prosodic parse of the syllable structure

even though there is no interval containing the usual acoustic traces of the vowel manner specification in the signal. These two tokens of the segment are “devoiced” and so cannot show the vowel formant values that cue the explicit [i] in *oyoideru*. Nonetheless, the listener parses the presence of some implicit head V for the second syllable of *hashitteru* and third syllable of *kurashiteru* from the fact that [ʃt:] and [ʃt] are not grammatical sequences of segments for any position in a prosodic word of the language. A good grasp of how listeners resolve prosodic ambiguities and of how they recover the intended autosegmental features of an implicit prosodic node in phonological ellipsis is critical for understanding the ways in which phonological structures above the word are manipulated to mark discourse prominence.

(2) Annotated fundamental frequency contours for utterances of three sentences with the structure [[____]NP [[Adjective Noun-ACC]NP Verb-past]VP] containing (a) two accented words, (b) one accented word, and (c) no accented words.



In introducing the prosodic constituents mora, syllable, and prosodic word in this section, we defined them primarily in terms of the distribution of place and manner autosegments for consonants and vowels. In describing what listeners do to parse the prosodic categories of ambiguous segments or to recover elided segments, we made passing reference to their phonetic interpretation, using the spectrograms in (1) as a convenient representation of the place and manner autosegments. We will use the fundamental frequency (F0) contours that are displayed below the spectrograms as a convenient phonetic representation of another type of autosegment that is critical for understanding prosodic structures above the word.

2.2 Tone features and prosodic constituents above the word

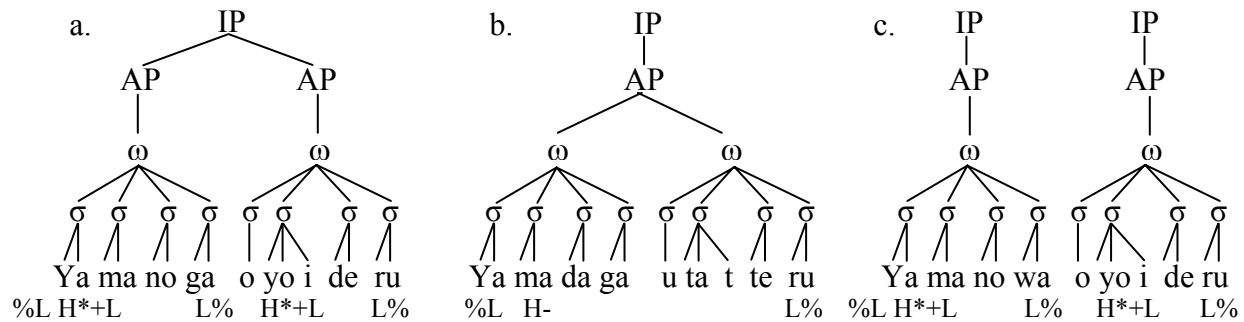
In addition to word-internal constituents such as the syllable, Japanese also has larger constituents above the word: various types of *prosodic phrase*. The X-JToBI model describes two levels of prosodic phrasing: the *accentual phrase* (abbreviated as AP) and the *intonation phrase* (abbreviated as IP). When our speaker produced the fluent utterances of the sentences in (1a) and (1b), he grouped the prosodic words in each utterance into two APs, which in turn were grouped together to form a single IP for the utterance as a whole. In order to define these two larger groupings, we must refer to the distribution of another type of autosegment besides the place and manner autosegments that defined the C and V nodes of the prosodic tree. This third autosegmental tier is the string of *low (L)* and *high (H) tones* that correspond to the target pitch values that the speaker produces at different prosodic positions in the utterance.

The fundamental frequency (F0) contours below the spectrograms in (1) are a convenient phonetic representation of these target pitch values, and we have annotated each of them for some of the C and V autosegmental targets that also affect the pitch pattern in ways that can make F0 contours difficult for the uninitiated to read. In particular, there is a discontinuity in the F0 representation at each interval where the [-voice] autosegment is in effect. For example, there is a break in the F0 contour during the [ʃ] of *Kashino* and during the [h] and the [ʃt:] sequence in the first three syllables of *hashitteru* in (1b). Also, [+voice] consonants, such as the [d] and [r] in *oyoideiru* in (1a), can cause sharp dips in the pitch, because the speed of opening and closing of the glottis depends in part on the difference in air pressure below and above the glottis, which in turn depends on a steady venting of air through the mouth. Experienced readers will automatically parse these effects as cues to the consonant or vowel segments rather than cues to the tones, and see that the tone sequences in utterances (1a) and (1b) are essentially identical. There is a high tone target near the beginning of the subject NP, then a sudden fall to a low tone target, then a rise to a second high target on the second syllable of the VP, followed by another fall to a low target. The pitch rise at the beginning of the VP, which is especially apparent in (1a), marks the boundary between the first and the second AP.

In the previous section, we described how constriction targets can be called by their metrical functions as well as by their actual content, so that a close palatal constriction is symbolized differently depending on whether it marks a syllable boundary (C=[y]) or a syllable head (V=[i]). All tone targets, too, can be called by their metrical functions as well as by their pitch levels. For example, in the X-JToBI model, the low target at the AP boundary between the NP and the VP in (1a) and (1b) is analyzed as a *boundary tone*, which is aligned to the phrase edge. Following the usual AM conventions for indicating the demarcative prosodic function directly in the symbol for the category, this boundary tone is symbolized as [L%] in an X-JToBI tagging of the utterance. Since the start of the utterance as a whole is also necessarily the start of a new AP, there is another instance of this low boundary tone marking the beginning of these two utterances as well as the

beginning of the utterances in (1c) and (1d). Following common practice in the AM literature, the X-JToBI conventions tag this boundary tone as [%L], using the function marker as a prefix to indicate that this boundary tone marks an initial boundary after a silent pause. Of course, this utterance-initial [%L] target is explicitly realized only when the text of the phrase begins with a sequence of [+voice] consonant and vowel autosegments on which the F0 can be realized, as in (1a) and (1c). However, even when the phrase begins with a [-voice] segment, as in (1b) and (1d), the [%L] can be recovered from the prosodic parse in the same way that the implicit [i] in the second syllable of *hashitteru* is recovered in (1b). In both (1a) and (1c), the explicit low tone defines the start of a rise in F0. The high tone target that defines the end of this rise is also explicitly realized in (1b) and (1d), providing part of the context for positing an elided [%L]. In (1a) and (1b), this high tone target is the [H*+L] accent, while in (1c) and (1d) it is an AP-level *phrasal tone*, symbolized with [H-]. These tone sequences and their metrical affiliations are shown in (3a-b) for the utterances in (1a) and (1c).¹

(3) Metrical structure and associated tone targets for (a) *Ya'mano-ga | oyo'ideru* (see (1a)), (b) *Yamada-ga | uttateru* (see (1c)), and (c) *Ya'mano-wa || oyo'ideru* (see (4a), solid line).

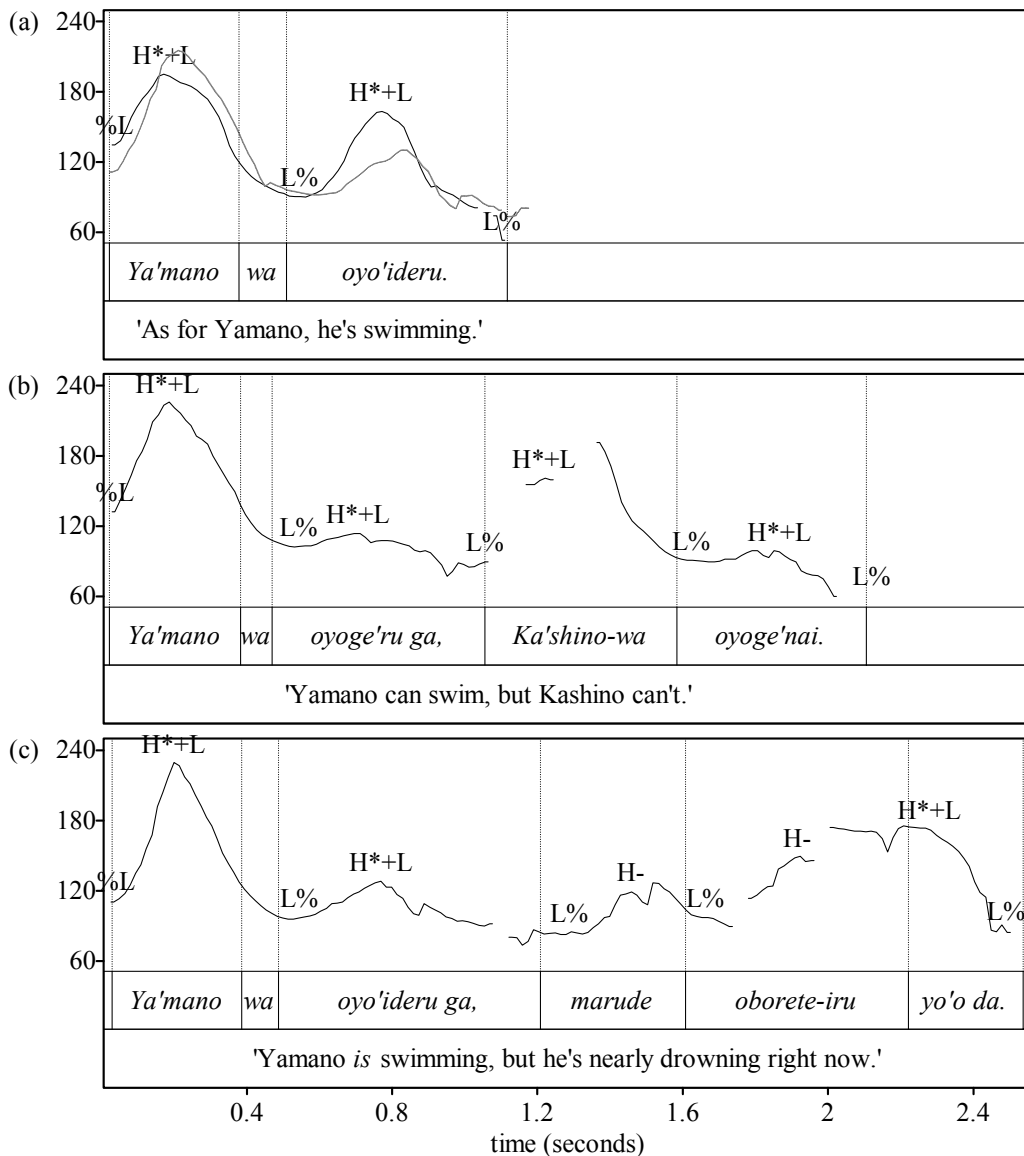


Each of the utterances in (1) constitutes a single IP, and the second [L%] boundary tone at the end of the utterance marks the final boundary of this IP as well as the final boundary of the second AP. By contrast, the utterance of *Ya'mano-wa oyo'ideru* in (4a, solid line) is two IPs, so that the medial [L%] boundary tone here marks an IP boundary as well as an AP boundary. (The phonetic markers to the strength of this medial disjuncture include the noticeable phrase-final lengthening on the *-wa* as well as other less well-documented hiatus markers, such as a possible change in voice quality or glottal stop. However, they do not include a silent pause, so there is no separate initial [%L] boundary tone marking the beginning of the second IP.) There are many other tone sequences besides simple [L%] that can mark the end of a prosodic phrase, giving rise to a variety of boundary pitch movements (BPMs) with contrasting pragmatic functions. We will discuss the inventory of BPMs that can mark the end of phrase in the next section, but first we must describe one more element that can occur in the tone sequence for some APs. This other AP-level element is the *akusento-kaku*, a sequence of tones that is tagged in Japanese ToBI as [H*+L]. This tone

¹ Following standard practice in the literature on Japanese phonology, we will from this point on use an apostrophe following the vowel to mark the location of the *akusento-kaku* in the roman transliteration of any word that is lexically specified as containing a syllable that associates to this [H*+L] tone sequence. We also will use | to indicate an AP boundary and || to indicate an IP boundary, when we need to highlight the distinction.

sequence is what results in the steep fall in each AP in (1a) and the first AP in (1b). The APs *Ya'mano-ga* and *Ka'shino-ga* contrast with *Yamada-ga* and *Kushida-ga* in that each has an *akusento-kaku* on its first syllable. The APs *oyo'ideru* and *hashi'tteru* similarly contrast with *utatteru* and *kurashiteru* in that each has an *akusento-kaku* on its second syllable, and there would be a steep fall after the second syllable in *hashi'tteru* in (1b) if the vowel in this syllable were not effectively deleted. That is, the listener parses the implicit high target for a [H*+L] from the surrounding pitch dynamics and from the relationship between the low tone pitch targets on the first and the third syllables, even though the syllable bearing the high tone of the *akusento-kaku* is completely voiceless.

(4) F0 contours for utterances of four sentences illustrating different pitch range relationships between *Ya'mano-wa...* (or *Ya'mano-ga...*) and the following verb phrase. The gray dashed line in (a) is a copy of the F0 contour in (1a).



This term *akusento-kaku* is usually rendered in English as ‘pitch accent’ or ‘accent’, but we introduce it first by the Japanese term, to try to decouple this use in translation from the way these two English terms are used in AM models of several other well-studied languages. In descriptions of English, for example, *accent* refers to any of a number of prominence-lending pragmatic morphemes that can be associated with the metrically strongest syllables in the most prominent content words of an utterance. In English, the location of these prominence-lending morphemes is determined by prosodic constraints such as OBLIGATORYHEAD (the principle that the intonation contour for each intonational phrase must contain at least one pitch accent associated to the metrically strongest prosodic position, see Hyman (2005)) in interaction with the information structure of the utterance in its discourse context. For example, a native speaker of English may choose to place a particular rising-falling configuration of accents on each of the nouns and verbs in the sentences in (5) to invoke the relevant contrast sets for each person and each activity. When speakers of English begin to learn Japanese, they may interpret the rises and falls in examples (1a) and (1b) in terms of such head-marking tonal morphemes in their native language. But this is a misparsing of the prosodic function of the tone shapes in these two Japanese utterances. While the steep fall in *Ya'mano* in example (1a) is because this noun has a pitch accent on the first syllable, that fact is a property of the word itself, regardless of the context in which it is uttered. The specification of a [H*+L] fall at this point is part of the phonological contrast between accented *Ya'mano* and unaccented *Yamada*. That is, Japanese is like conservative varieties of Basque in having a lexical contrast between words specified for an accentual fall and words without this specification. (See Hualde (1991) and Hualde et al. (2002) for descriptions of the conservative Basque system and of the influence of contact with Spanish in other varieties.)

(5) [context: Who's doing what?]

YaMAno is SWIMming. KASHIma is RUNning. YaMAda's just standing there SINGing to 'em.

Tokyo Japanese is crucially different from Basque, however, in that this lexical specification completely constrains the distribution of [H*+L] sequences in the surface intonation pattern. That is, in Basque, an unaccented word must be grouped together with a following accented word, and if a sentence contains only lexically unaccented words, an accent will be inserted somewhere, either in the last argument before the sentence-final verb or earlier if there is narrow focus on an earlier constituent (see Ito (2002a)). By contrast, there is no post-lexical insertion of accent in Japanese. If a Japanese sentence contains no lexically accented words, then the intonation contour of an utterance of that sentence will not contain any accents, whatever the focus pattern. Instead, it will be composed only of the same tones as the contours for (1c); there will be the rises for the [%L H-] or [L% H-] sequences that mark the beginning of every AP, and there may be additional boundary tones marking the end of some phrases, but there will be no other tones. Any fall in pitch within an AP will not be the sharp localized fall of the accent [H*+L] but some more or less gradual interpolation between the high pitch target for the phrasal high and some following low tone. If these words have all been grouped together into one long AP, as in (2c), this gradually falling F0 can describe a fairly long stretch of the contour from the high target at the end of the rise for the initial [%L H-] to the IP-level boundary [L%] at the end.

Note that this account of the long gradual fall in (2c) assumes underspecification of tones relative to potential tone-bearing units. That is, none of the morae between the demarcative rise in *omiage* and the boundary tone at the end of the utterance is described as having an associated tone target of any kind. This differs from older accounts within the Autosegmental Phonology

framework, such as Haraguchi (1977), who specifies a [H] tone target for each of the ten sonorant morae after the demarcative rise from the first to the second V in *omiage*. In such older accounts, the description of the fall that is actually observed over this purported sequence of ten [H] tone targets is relegated to *declination* — an “extragrammatical” or “paralinguistic” manipulation of global pitch trends. Pierrehumbert and Beckman (1988) argued for their underspecified AM account (which X-JToBI follows in this regard) on the basis of slope measurements for unaccented APs containing varying numbers of syllables. In their data, the slope of the fall was inversely correlated with its length, just as predicted in an account that models the fall as an interpolation from the [H-] phrase tone to some later low target. In this AM account, then, the fall in (2c) is attributed to the transition between different target values assigned to the [H-] and the [L%], just as the slightly rising plateau-like interval in the first AP in (2a) is attributed to a transition between only slightly different target values for the phrasal [H-] at the beginning and the higher tone target of the [H*+L] accentual fall at the syllable that is lexically associated to the *akusento-kaku*.

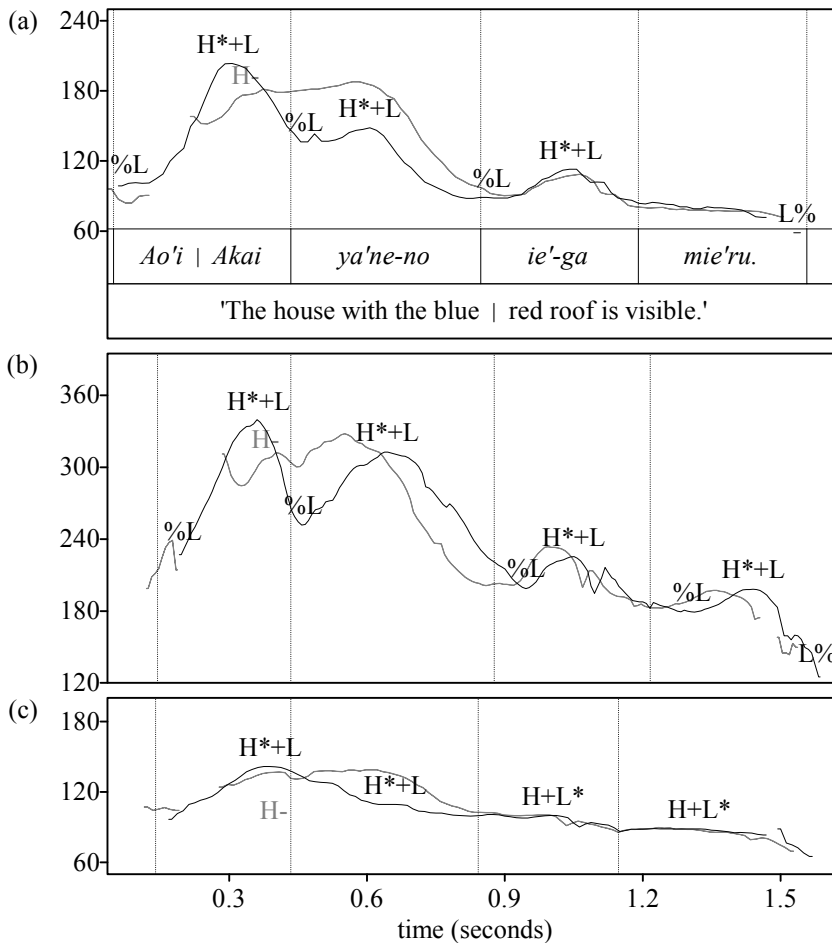
The F0 contour for (2a) also illustrates another phenomenon that was treated differently in at least one older account in the Autosegmental Phonology framework. The [H*] target at the accent in *ta'beta* is considerably lower than that in *chi'izu*. The same relationship holds between the accent on the VP and the accent on the preceding NP in each of (1a) and (1b). McCawley (1968) accounted for this relationship by positing a *mid (M) tone* for Japanese — i.e., a tone target intermediate between [H] and [L] like the [M] tone necessary to describe the lexical tone contrasts on monosyllabic words in Yoruba and Cantonese. In the AM account, by contrast, the lowering of the tone target on the verbs in (1a), (1b), and (2a) is attributed instead to the component of the grammar that specifies parameters of the backdrop *pitch range* for each AP and for each IP.

Pierrehumbert and Beckman (1988) argue for the AM treatment on the grounds that there are many more intermediate levels than just one. For example, the medial AP-level [L%] boundary tone in (1a) is certainly lower than either the preceding or the following accent peak. Yet it is higher than an IP-level [L%] boundary tone at the end of the utterance. This can be seen by contrasting the AP-level [L%] to the utterance-final [L%] in that example or by contrasting the lower IP-level [L%] after *oyoge'ru-ga* in utterance (4b) to the merely AP-level [L%] after the following *Ka'shino-wa*. This difference between the truly low target for the [L] at an IP boundary versus the intermediate [L] at a mere AP boundary is a fairly reliable cue to the level of the disjuncture. It seems to be the same kind of boundary strengthening effect noted for consonant features in many languages by researchers such as Pierrehumbert and Talkin (1992), Fougeron and Keating (1997), and Jun (1998). Other types of systematic difference among tones are illustrated in the F0 contours of the two example sentences that are overlaid in (6).

One difference in (6) is contained just within the *Akai ya'ne-no* phrase. Here the high target for the phrasal [H-] is lower than the high for the following accent peak, and this is true for all three speakers. The same relationship also accounts for the rise in pitch over the plateau-like interval in the *Omiyage-no chi'izu-o* phrase in (2a). We have also seen flat or slightly falling shapes for longer accented phrases, such as the one in (2b). In read speech, however, a small rise seems typical for short phrases such as the *Akai ya'ne-no* phrase. That is, if [H-] is typically lower than the high target of [H*+L], then we can explain the difference in peak heights between unaccented *Yamada-no* in (1c) and accented *Ya'mano-no* in (1a) and between comparable minimal pairs in prior experiments by many researchers, including Pierrehumbert and Beckman (1988). This difference may be related to the same dissimilative *pre-L raising* seen in many other languages, including Yoruba (Laniran and Clements (2003)), Thai (Gandour et al. (1994)), and Taiwanese (Peng (1997)).

Another systematic difference seen in (6) is between the *ya'ne-no* phrase in the two overlaid utterances. The accent peak on *ya'ne* is lower after accented *ao'i* than it is after unaccented *akai*, as predicted by McCawley's rule lowering [H] to [M] in this context. However, each of the subsequent accent peaks in both sentences also is lower than its immediately preceding accent peak. Moreover, in the female speaker's utterances in (6b), where the medial [L%] AP-boundary tones are all clearly visible as local F0 minima, there is the same successive lowering of F0 values across the two or three utterance-medial AP boundaries.

(6) Overlaid F0 contours for utterances of two sentences produced (a) by the same male speaker who produced the utterances in all utterances shown in earlier examples, (b) by a female speaker, and (c) by a different male speaker. In all three panels, the thin black line is for the *Ao'i* ... sentence and the thick grey line is for the *Akai*... sentence.



Which of these intermediate levels should be singled out as the target level [M]? The X-JToBI model follows Pierrehumbert and Beckman (1988) in calling none of them [M]. Instead, all tone levels in Japanese are either [H] or [L], with the intermediate F0 values attributed to another type of phonological representation than the categorical contrast between these two tones — namely, a representation of backdrop reference pitch levels. For example, beginning with Poser (1984), all accounts in the AM framework model the systematically lower value for the [H] tone in *yane* after the accented adjective in (6) in terms of *downstep*, defined as a compression of the pitch range that

begins at the *akusento-kaku* and affects all tone targets up to the following IP boundary. Other current formal models of Japanese intonation patterns account for some of the systematic F0 target differences between and within utterances in subtly different ways from the account in the X-JToBI model. However, despite their differences, all current formal models of the Japanese intonation system share one property that distinguishes them from older Autosegmental Phonology accounts such as McCawley's (1968) or Haraguchi's (1977). All provide some way of representing the kind of continuous variation discussed in these examples directly in the grammar. Understanding this aspect of the AM framework model and other compatible models is critical for understanding one of the basic mechanisms for focus marking in the language. We expand on this aspect in the next section.

2.3 The relationship between tone features and pitch range features

Pierrehumbert and Beckman (1988) describe an intonation synthesis program that models the differences in F0 value among the various [H] or [L] in terms of *prominence* relationships at two different levels of specification. One level involves *tone scaling* within the local pitch range. For example, in utterance (4b) the more prominent IP-level [L%] target after the *oyoge'ru-ga* is scaled lower in the pitch range than the merely AP-level [L%] targets at the ends of the preceding and following phrases. Similarly, the more prominent [H] target of the accent is scaled higher in the pitch range than the phrasal [H-]. In the synthesis program, these two local prominence relationships are specified in terms of two constants indicating tone-scaling ratios. The first constant positions an IP-final [L%] proportionally closer to the *bottom-line* value for the local pitch range, and the second one positions any accent [H*] target proportionally closer to the *top-line* for the local pitch range. Such variation in tone-scaling changes the target position within the pitch range, but leaves the pitch range intact.

By contrast, other prominence relationships are modeled by varying the pitch range. The synthesis program provides for a paradigmatic choice of hertz values for the bottom-line and for the top-line of the pitch range of each IP. It also provides for the paradigmatic choice of an “AP prominence” factor — a ratio value that determines the top-line for each AP relative to the pitch range of the IP that dominates it. For example, the accent peak on *Ya'mano-wa* in (4a) is lower than the accent peak on *Ya'mano-ga* in (1a), because the topic IP (4a) is not in focus relative to the following predicate. In the synthesis program, this relationship is generated by specifying a lower top-line value for the first IP in (4a). In a similar way, the accent peak on the *Akai ya'ne-no* phrase is very much higher than the peaks on *ie'* and *mie'ru* in all three speakers renditions of this sentence in (6) because the initial AP is produced in a pitch range that is somewhat expanded relative to a more neutral reading. This reflects the focus structure of the sentence in the kinds of discourse contexts that the speakers could imagine for it. The distinctive roof color identifies which house it is that is visible. In the synthesis program, this is specified by an AP prominence ratio greater than 1.0 so that the top-line for the first AP is above the top-line for the IP as a whole.

The discourse-level considerations just described are not the only ones at play in determining the top-line value for an AP. There is also the general phonological constraint that AP-level high tone targets are lower in APs that occur after an accented AP within the same IP. This constraint is reflected in the lower value for the accent peak on *oyo'ideru* in (1a), where the subject and predicate are grouped together into one IP, relative to the accent peak on *oyo'ideru* in (4a), where the predicate stands alone in an independent IP. The same relationship can be seen also on *ya'ne* in *Ao'i ya'ne-no* as compared to the accent peak in the *Akai ya'ne-no* phrase in (6). In the synthesis program, these differences are modeled by a *downstep* ratio that proportionally reduces the distance

of the IP top-line above the bottom-line immediately after each accent [H*] target. The very much lower values of the peaks on *ie'-ga* and *mi'eru* in (6) then reflect the combined effects of focus subordination and of downstep.

The types of prominence relationships that are modeled in Pierrehumbert and Beckman's synthesis program are also modeled in a comparably direct way in Hiroya Fujisaki's model that was first described in Fujisaki and Sudo 1971 and which has been adopted in an impressive body of later studies that became the basis for the intonation synthesis modules in several leading Japanese text-to-speech and concept-to-speech systems (e.g., Fujisaki and Hirose 1993, Fujisaki et al. 1994, Hirai et al. 1996, Kiriya et al. 2002). In the Fujisaki framework, the prominence relationship among different intonation phrases is modeled by specifying different amplitudes for their *phrase commands*. The phrase command is a local pulse aligned near the beginning of the phrase and then convolved with a decaying declination function. The prominence relationship among different APs, on the other hand, is modeled by the specification of different amplitudes for the *accent commands*. The accent command is a square wave that generates the rise at the beginning of each AP and the subsequent steep fall after the accent peak in an accented AP. The differences between the X-JToBI AM framework model and Fujisaki's model primarily involve the treatment of unaccented phrases. The accent command in Fujisaki's model implies a [H] target at the end of an unaccented AP. In other respects the two frameworks provide comparable accounts of the many intermediate F0 target levels observed within and across utterances. The different frameworks are comparable enough that observations couched in one set of synthesis parameters can be translated readily into the parameters of the other model. That is, this kind of formal analysis-by-synthesis yields data that can be compared across frameworks in a way that is not possible with the informal pre-theoretic observations reported in studies such as Selkirk and Tateishi (1991), who assume a direct and deterministic mapping between syntactic structure and prosodic structure instead of building an independent formal model of the prosodic structure on its own terms. Both synthesis models also provide an explicit representation of continuous variation in the alignment of tone targets relative to the consonant and vowel segments of the associated text. This explicit representation of tone-text alignment is critical for describing the contrasting boundary pitch movements at the ends of intonational phrases, which we describe in the next section.

2.4 Boundary Pitch Movements

The preceding two sections have discussed the relevant metrical structure at both the AP and the IP level, as well as the tonal autosegments that mark the lower level of the AP and the pitch range compression that is triggered by the accent tones and affects all following tones until a new pitch range is chosen for the following IP. The literature in English focuses almost exclusively on these aspects of the Japanese intonation system, and all were encoded even in the earlier J_ToBI conventions (Venditti, 1997). By contrast, the last major element of Japanese intonation, the inventory of rises and more complicated tonal events that are licensed to occur after a phrase-final [L%] boundary tone, is rarely given more than passing mention.

In the X-JToBI conventions, these events are called *boundary pitch movements* (BPM). A BPM occurs when the AP-level boundary [L%] is followed by one or more other tone targets, including at least one [H], to create a simple rise, a concave rise, a rise-fall, or some even more complicated F0 movement. There is a long-standing consensus that BPM are pragmatic morphemes that are licensed to occur at the end of a phrase to contribute to the pragmatic interpretation of the phrase (see Kindaichi 1951, Ohishi 1959, Kawakami 1963, Miyaji 1963, Kori 1989b, Fujisaki and Hirose 1993, Muranaka and Hara 1994, Venditti et al. 1998, Katagiri 1999,

among many others). Despite this primordial characterization of their general function, however, we do not yet have either an exhaustive inventory of BPM or an exact description of the contrasting discourse meanings that they convey. In this section we discuss the inventory of BPM categories in the X-JToBI scheme, which were used to tag all 178,000+ phrase boundaries in the 45-hour ‘Core’ portion of the CSJ (henceforth CSJCore), yielding the counts in (7). We compare this list of BPM to the inventory proposed by Kawakami in his insightful 1963 paper on phrase-final rises.

(7) Distribution of phrase-final tones in CSJCore. APS (Academic Presentation Speech) refers to CSJ monologues which were collected from live academic presentations, SPS (Simulated Public Speaking) refers to monologues collected from lay subjects who were asked to give short presentations about everyday topics, D refers to dialog recordings, and R refers to read speech of (i) passages from popular science books or (ii) ‘reproduced’ transcriptions of the APS/SPS spontaneous speech (see Maekawa 2003 for details).

	L% only	L%+H%	L%+HL%	L%+LH%	L%+HLH%	Total
APS	56736	19647	3041	125	1	79550
SPS	64546	11148	6997	259	9	82959
D	7152	1705	565	34	4	9460
R	5118	1636	33	1	0	6788
Total	133552	34136	10636	419	14	178757
%	74.7%	19.1%	5.9%	0.2%	0%	(100%)

As the table in (7) shows, many prosodic phrases in the CSJCore end simply in [L%], without any additional boundary pitch events. This is true also of all of the prosodic phrases in our read example utterances (2-4), (6), and (8a). However, phrases ending in other boundary configurations are not rare even in fairly scripted spontaneous speech. For example, nearly a quarter of the phrases in the academic presentations (APS) in the CSJCore are tagged with some kind of BPM after the [L%].

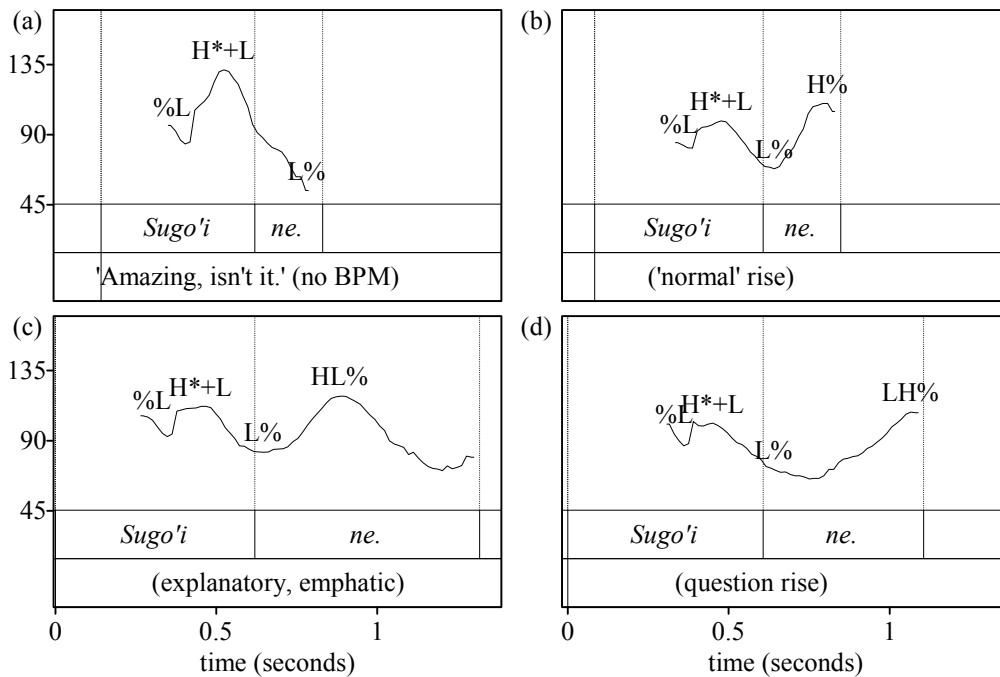
The most common BPM tag in the CSJCore is [H%]. This corresponds to the shape that Kawakami (1963) termed the *futsuu no jooshoochoo* ‘normal rise’. The pitch rises sharply from the onset of the final syllable of the phrase, as shown in (8b). The [H%] BPM should not be confused with the superficially similar ‘question rise’ illustrated in (8d). Its function is instead more closely related to that of [HL%], another BPM with which it also shares an alignment pattern for the [H] tone target. In the [HL%] BPM, as in [H%], the F0 rises sharply at the onset of the final syllable of the phrase, before it falls again. The final syllable is usually lengthened as well, as in the example utterance in (8c).

In each of the examples in (8), the BPM occurs on a sentence-final pragmatic particle *ne*. Katagiri (1999) describes *ne* as occurring only in dialog, where it indicates lack of commitment to the proposition just uttered (in contrast to *yo*, which indicates the speaker’s acceptance of the proposition into the mutual belief space). Katagiri also describes how this core meaning of *ne* interacts with the meaning of a rising versus a falling phrase-final intonation pattern to induce a variety of conversational implicatures, such as confirmation question, polite assertion, and the invoking of external authority. It is important to note, however, that Katagiri does not distinguish between the two types of rising BPM illustrated in (8b) versus (8d). Thus, it is possible that some of the meaning contrasts he shows are associated with tonal differences.

Katagiri is not alone in his failure to distinguish more than one rising BPM. Indeed, in the literature in English, this is the rule rather than the exception. For example, Pierrehumbert and

Beckman (1988) also recognized only one rising BPM, which they analyzed as a [H%] boundary tone. Moreover, they elicited this rise only in interrogative uses of the sentence-final particles *ka* and *no*. The occurrence of a rising BPM in association with an utterance-final interrogative *ka* or *no* marks a very salient easily-grasped speech act, so it is not surprising that all researchers have recognized at least one rising BPM in at least this one discourse context.

(8) F0 contours for utterances of the sentence *Sugo'i ne*. 'Amazing, isn't it.' pronounced with four different final boundary-tone configurations.

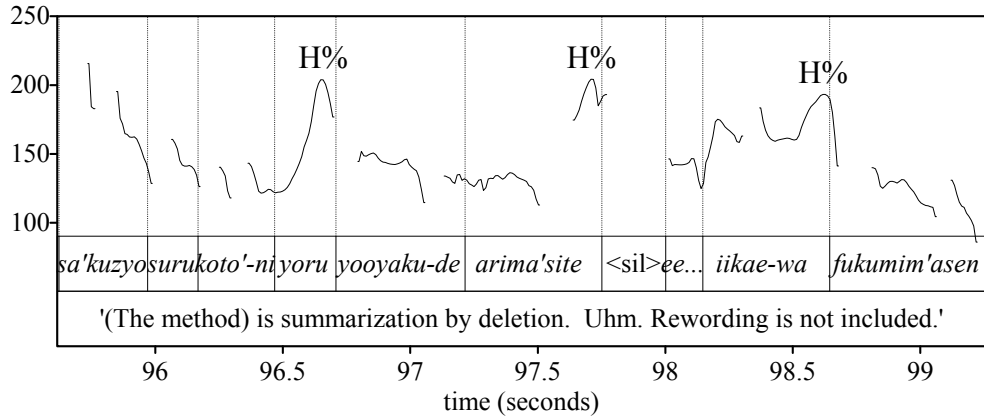


Pierrehumbert and Beckman also followed McCawley (1968), Haraguchi (1977), and Poser (1984) in assuming that this rise occurs only in sentence-final position. Based on this (mistaken) assumption, they posited a prosodic constituent above the IP (termed the *utterance*) as the domain that licenses the [H%] target. However, as (9a) shows, the BPM that is tagged as [H%] in X-JToBI readily occurs sentence-medially in spontaneous speech (see also Nagahara (1994) and Nagahara and Iwasaki (1994) for a similar insight). Often the phrase-final syllable on which [H%] is realized is a postposition or case particle, although this need not be the case. For example, there are three instances of [H%] in (9a), but only one of these is associated to a case particle. Moreover, as the counts in (7) show, this BPM occurs frequently in all genres represented in CSJCore. That is, unlike the particles *ne* and *yo*, the simple rise BPM is by no means limited to dialog.

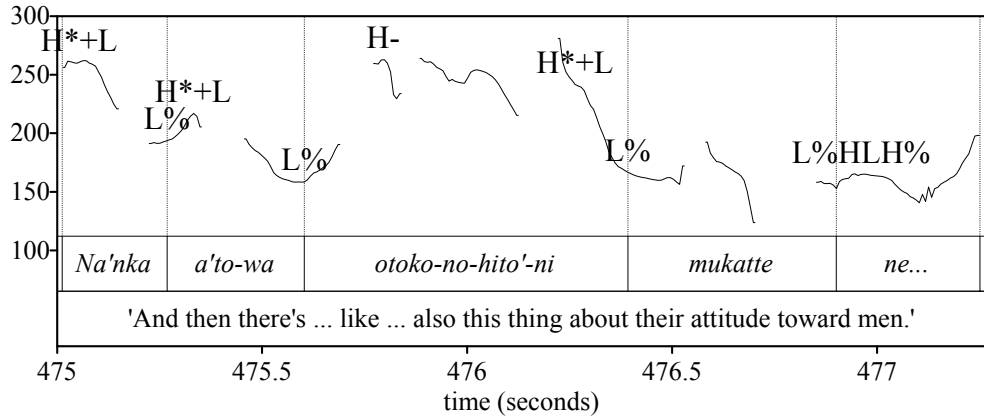
In the (mostly non-English) literature that does distinguish at least two different rising BPM, the function of the rise that is tagged as [H%] in X-JToBI is described as one of imparting prominence to the constituent to which it associates (see Ohishi (1959), Kori (1989b,1997), Venditti et al. (1998), Taniguchi and Maruyama (2001), Oshima (in press), among many others). This function contrasts with the function of the 'question rise' in (8d), and relates [H%] instead to [HL%], the rise-fall BPM illustrated in (8c). The rise-fall BPM in (8c) is not included in Kawakami's (1963) inventory, but it was among the shapes examined by Venditti and her colleagues in a perception experiment (Venditti, Maeda, and van Santen, 1998) matched with an extensive set of acoustic measurements (Maeda and Venditti, 1998). In the perception study

(9) F0 contours for three utterances from the CSJCore:

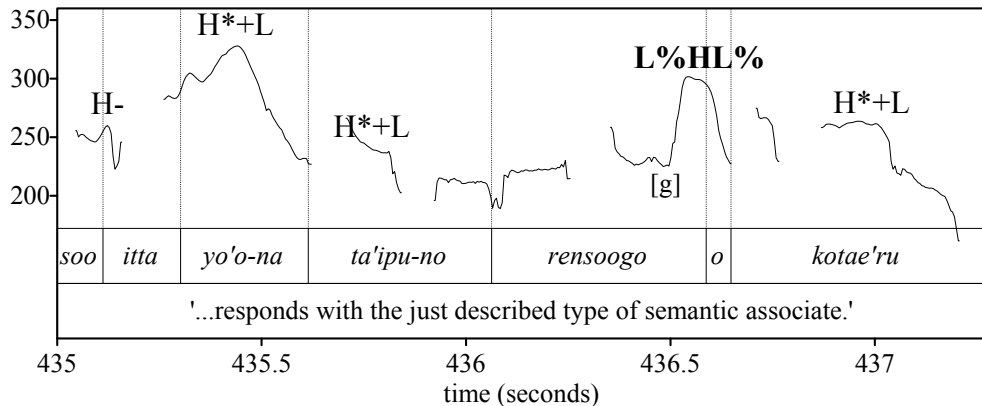
a. An example of H% marking sequential phrases. (Tags for the accent tones and AP-initial rise tones are suppressed in order to highlight the [H%] BPM.)



b. An example of the HLH% boundary pitch movement (from a conversation between two young women about how Chinese women are different).



c. An example of the HL% PNLP: underlined extract of <AME> to ittara <AMAI> toka soo itta yo'o-na ta'ipu-no rensoogo^-o kotae'ru ... 'When presented with <CANDY>, the subject responds with <SWEET> or some such associate of this type ...' from a talk on processes of auditory lexical access of words that have near homophones differing minimally in lexical accentuation, such as *ame* 'candy' and *a'me* 'rain'.



Venditti et al. (1998) asked listeners to rate each stimulus on several semantic scales. They observed that listeners perceive [HL%] to be highly ‘explanatory’ and ‘emphatic’. That is, Japanese listeners expect speakers to use [HL%] in contexts where they are explaining some point to their interlocutor, and want to focus attention on a particular phrase in this explanation. Although it is not as frequent as [H%], and is markedly rare in read speech, [HL%] occurs at 4% to 8% of the phrase boundaries in the different genres of spontaneous speech in the CSJCore.

The next most frequent BPM in the CSJCore is much rarer, occurring at only 0.2% of phrase boundaries. This is the [LH%] rise illustrated in (8d). This BPM may occur sentence-finally, and is observed most often at the ends of utterances expressing questions. As noted already, this BPM is superficially similar in form to [H%] in that the F₀ rises on the last syllable of the phrase. However, unlike with [H%], where the onset of the rise aligns with the onset of the final syllable, the rise of [LH%] occurs later within that syllable, making the overall shape of [L% LH%] more “scooped” or “concave” than that of [L% H%]. In some cases, especially in cases expressing incredulity, the final syllable is lengthened as well. In the Venditti et al. 1998 perception study, stimuli carrying [LH%] were perceived as ‘seeking confirmation’ from the interlocutor, and those which also had a lengthened final syllable yielded a high ‘disbelief’ rating (i.e. incredulity questions).

The X-JToBI [LH%] category corresponds to two different rise shapes in Kawakami (1957), who describes simple information-seeking questions such as *ka'eru?* or *ka'eru no?* ‘Are you returning?’ as having a *futsuu no jooshoochoo* ‘normal rise’, and distinguishes this shape from interrogatives indicating ‘deep’ questioning, which he describes as having a *hanmon no jooshoochoo* ‘return rise’. In their acoustic analysis of BPM contrasts, Venditti et al. (1998) showed that both the alignment and overall F₀ curve shape categorically distinguished all questions from prominence-lending rises: in both information-seeking and incredulity questions ([LH%]), the rise onset occurs later in the syllable and the rise shape is more concave than in the prominence-lending rise ([H%]). By contrast, the phonetic distinction between the two question types was fuzzier. The duration of the final syllable varied continuously from the shortest most clearly information-seeking question (e.g. *ka'eru no?*) to the most drawn-out clearest examples of an incredulity question (e.g. *ka'eru no?!!*). In addition, the location of the rise onset (and other points during the rise) was correlated with the varying durations of the C and V autosegments on which the rise was realized. Given these results, we now understand Kawakami’s *futsuu no jooshoochoo* and *hanmon no jooshoochoo* to be descriptions of the extreme endpoints of a continuum that includes many intermediate degrees of emphatic lengthening. That is, the gradient nature of the relationship between the phonetic dimensions and the continuum of contrasting degrees of incredulity suggests an analysis akin to the one that Hirschberg and Ward (1992) propose for the uncertainty versus incredulity interpretations of the English fall-rise contour.

The rarest BPM distinguished by X-JToBI is [HLH%], a very complex shape in which the F₀ rises, then falls, then rises again. This BPM also is not included in Kawakami’s inventory of rises. It seems characteristic particularly of child-directed speech, where it can give a wheedling or cajoling quality to the utterance to which it attaches. The counts in (7) show that [HLH%] occurs a total of only 14 times on the 178,000+ boundaries of the CSJCore, one of which is shown in (9b). This is likely because the speaking genres that CSJ includes are for the most part not the types in which [HLH%] readily occurs; we would expect to count more instances of [HLH%] in interactions between mothers and their young children, or in conversations between young female friends.

In addition to the BPM tags in (7), the X-JToBI conventions provide a number of diacritics which mark distinctions within categories. Several of these could well be phonologically distinct tonal sequences. They correspond to categories that Kawakami discusses (e.g. ‘floating rise’, ‘hooked rise’, etc.) which do not seem to have any kind of gradient relationship between the phonetic dimensions of contrast and a continuum of pragmatic interpretation. However, research on these distinctions is even less well developed than research on the differences between the two types of [LH%]. Moreover, the functions that have been suggested for them are not directly relevant to the theme of this paper. Therefore, the only X-JToBI diacritic that we will mention here is for an F0 rise that Kawakami does not mention as a distinct BPM type.

The shape of this rise is convex, like that of the [H%] BPM, with the [L%] target that begins the rise aligned early in the syllable to which it associates, and it has been described by Ohishi (1959), Kori (1989b, 1997), and others as having a similar prominence-lending function. This rise contrasts with the [H%] BPM, however, in that it occurs on the penultimate syllable of the phrase rather than the last. In the X-JToBI conventions, this earlier rise is termed the PNLP (for *penultimate non-lexical prominence*). Because the F0 typically also falls immediately after the PNLP to a [L] target at the following phrase boundary, the X-JToBI convention is to mark the PNLP using the [HL%] rise-fall BPM label on the tones tier, with an additional tag (‘PNLP’) on the comments tier. A total of 1,162 cases of PNLP were tagged in CSJCore (62.7% in APS, 33.7% in SPS, 1.9% in dialog, and 1.6% in read speech). In the example in (9c), the PNLP tagged rise-fall is indicated in bold.

3 The prosodic marking of prominence

Having provided an overview description of the Japanese intonation system, we can now turn to the various ways in which intonation patterns are manipulated in the marking of focal prominence in the language. Since this is an area of research where the literature has been muddled considerably by drawing analogies to pre-theoretical intuitions about English focus-marking mechanisms, it is useful to begin by describing one way in which focal prominence is *not* marked in Japanese.

3.1 Distinguishing *accent* from *akusento*

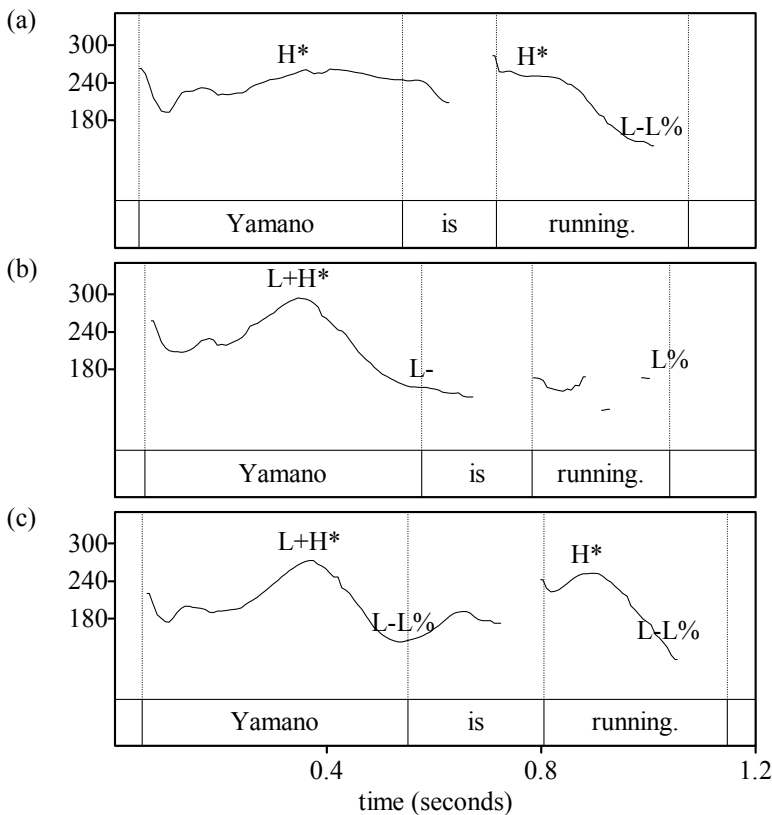
Much of the theoretical literature on focus-marking in English, German, and Dutch invokes a structural property that is variously called *accent*, *nuclear pitch accent*, *sentence stress*, or *nuclear stress*. As this jumble of terms suggests, focus-marking in these languages involves a complex interplay of prominence markers at several levels of the prosodic hierarchy. We cannot do justice here to the richness of these systems, however, and instead will discuss just one example from English that we hope will let us state as clearly as possible what we mean when we say that standard Japanese has no analog to the notion “accent” when it is used as a synonym for “nuclear stress” in these Germanic languages. Especially, we must emphasize that the notion “accent” in that sense is neither a formal nor a functional analog to ‘accent’ as a translation of *akusento-kaku*.

In Section 2.2, we alluded to a common cross-language misparsing of the *akusento-kaku* in words such as *Ya'mano*, *oyo'ideru*, and *hashi'tteru* in (1a) and (1b). The utterances in (10) illustrate how the English intonation system contributes to this misparsing. In (10a) the first syllable of *running* is associated to the nuclear pitch accent in the intonation contour. By *nuclear pitch accent*, we mean a configuration of tones that marks the word (or some larger constituent) as the focus of the utterance and simultaneously invites the listener to make some inference about the pragmatic relationship between the focused constituent and the shared model of the discourse context. In (10a), this configuration is the fall in pitch defined by the transition from the [H*] *pitch*

accent on the first syllable of *running* to the following [L-] *phrase accent*. In (10b), the second syllable of the (English) word *Yamano* is associated to the same kind of fall that makes *running* focally prominent in (10a).

The tone labels in (10) follow the AME_ToBI tagging conventions described in Beckman et al. (2005). Other AM models of English offer different analyses of the falling pitch that is aligned earlier in (10b) as compared to (10a). For example, Gussenhoven (2004) labels the fall as [H*+L], because he disputes the relevance of the notion “phrase accent” for English. Despite these differences across frameworks, however, there is a broad consensus about the focus-marking function of the fall, and about its culminative role. That is, any complete well-formed intonation contour of English has exactly one nuclear pitch accent per intonation phrase. In (10c), then, the sentence is divided into two intonation phrases so that each of *Yamano* and *running* can be associated to a nuclear pitch accent. That is, the restructuring of the sentence into two intonation phrases in (10c) preserves the culminative distribution of nuclear pitch accents, with the result that there is exactly one word in each IP that is marked as bearing the one obligatory pitch accent in the intonation contour for the IP (what Liberman and Prince (1977) termed the “designated terminal element” of the phrase).

(10) F0 contours of the English sentence *Yamano is running.* produced in response (a) *What’s happening in the next leg of the triathlon relay?* (b) *Who’s doing this leg?* and (c) *Who’s doing what?*



A noteworthy property of the English prosodic system is that this culminative distribution of nuclear pitch accents relative to IP boundaries in connected speech mirrors the culminative distribution of potential association sites for the nuclear pitch accent in polysyllabic words in the

lexicon. In the words *Yamáno* and *rúnning* in (10), as in *rélay*, *dóing*, *átom*, *atómic*, *rétail*, *ráttele*, *legislátion*, and many other polysyllabic words of English, this potential association site is the penultimate syllable. In *átomize*, *ráttelesnake*, and *législatiure*, it is the first syllable. The terms *primary stress* and *lexical stress* refer to this potential for association to the nuclear pitch accent. As Hyman (2005) and many others before him have pointed out, the obligatory designation of exactly one primary stress per word imparts a sense of word-hood. Thus, compound words such as *ráttelesnake* and *láv degree* can be distinguished from phrases such as *ráttele it óff* and *láv and órder* because of the *Compound Stress Rule* — a morphological process whereby all but one of the primary stresses in a compound word lose the potential to associate to accent. Also, function words such as *the*, *of*, and *but* (which associate to pitch accents only in unusual circumstances) are commonly understood to be prosodically *cliticized* onto a neighboring content word. This link between the culminative distribution of primary stresses in words in the lexicon and the culminative distribution of nuclear pitch accents in the intonation phrases of a discourse gives rise to the terms *sentence stress* and *nuclear stress* to designate the stressed syllable in the word that associates to the nuclear pitch accent in an IP.

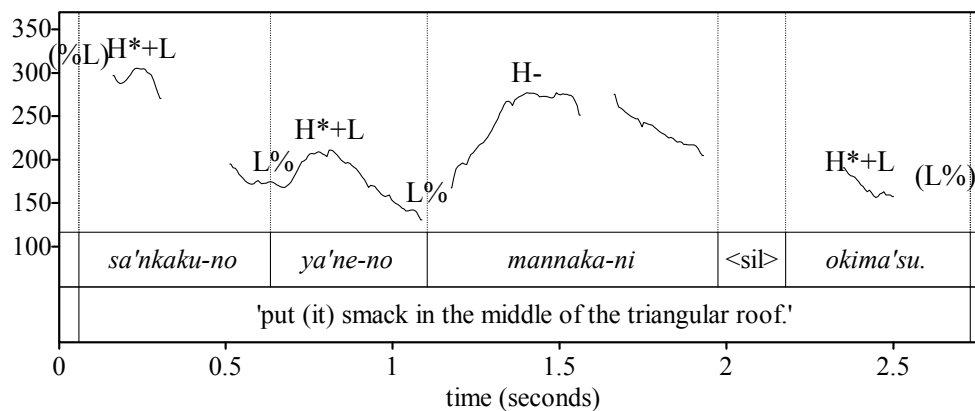
Another noteworthy property of the English intonation system is that association to the nuclear pitch accent is often accompanied by an exaggeration of other phonetic markers of prominence which together constitute the formal properties that define the *stress foot*, a lower-level constituent in the English prosodic hierarchy. These include, for example, (i) the aspiration of the foot-initial /t/ in *atomic* and *retail* as opposed to the lenited pronunciations typical of the foot-medial /t/ in *atom* and *rattler*, and (ii) the constrained distribution of reduced/elided vowels that yields the morphological alternation between a full long /a/ in the foot-initial *stressed* syllable of *atómic* and the reduced short /ə/ or deleted vowel (making the [m] syllabic) in the corresponding *unstressed* syllable in *átom*. Thus, the aspiration of foot-initial voiceless stops can be exaggerated when the word is accented (e.g. Docherty (1992), among many others), whereas foot-medial /t/ can be weakened further and even deleted in unaccented contexts in running speech (Raymond et al. 2006). Similarly, the morphological alternation between full vs. reduced/deleted vowels in related forms such as *atomic* and *atom* is echoed by the optional prosodic alternation between disyllabic pronunciations of words such as *rattler* ([ræʔ.lər]) in fast, unaccented contexts, and trisyllabic pronunciations ([ræ.rə.lər] or even [ræ.tɹ.lər]) when the stressed syllables in these words are associated with accents in the intonation contour (as noted by Beckman (1996a), among many others). Such segmental prominence (“stress proper”) at the foot level is linked to the prosodic marking of focus in English because lexical stress constrains where pitch accents can and cannot go: any pitch accent in the intonation contour must associate to the head of a stress foot. This is the source of the unequivocally syllabic “strong” variants of words such as *of*, *the*, and *but* in contexts where they receive a pitch accent.

It is important to keep in mind that these links between syllable stress (in the lexicon) and pitch accent (in the intonation system) are particular to English and the handful of other languages that are like English. The prosodic organization of Japanese words and intonation phrases is quite different from that of English, and researchers should not be misled by the superficial phonetic resemblances that lead native speakers of Japanese to misparse the distribution of nuclear pitch accents in an utterance such as (10c) in terms of the distribution of *akusento-kaku* in an utterance such as (1a), and vice versa.

The differences can be summarized most succinctly by contrasting the applicability of the notion “stress” in describing the two prosodic systems. In descriptions of English, this term is an indispensable shorthand way of capturing the many profound implicational dependencies among

categorical markers of segmental prominence at the foot level and categorical markers of tonal prominence at the IP level. In Japanese, by contrast, there are no such fundamental links between segmental prominence and tonal prominence, and the term “stress” is at best a somewhat misleading translation of technical terms for pragmatic functions such as *kyoochoo* (better translated as ‘emphasis’ or ‘focal prominence’). For example, where vowel lenition in English is constrained to occur in weak unaccented syllables, vowel lenition readily occurs in accented syllables in Japanese even in very deliberate read speech. The deletion of the [i] in the accented second syllable of *hasi'tteru* in (1b) is by no means anomalous, as can be seen by the counts that Maekawa and Kikuchi (2005) report in all of the different genres in the CSJ. Also, where the typically unaccented (and lexically unstressed) pronunciation of function words in English is often described in terms of the notion “clitic” — i.e., the prosodic grouping of function words onto adjacent content words in running speech — many function words in Japanese, such as *ma'de* ‘until’, *yo'ri* ‘from’, *su'ra* ‘even’, and *no'mi* ‘only’, are lexically accented, and while these words are almost always realized as clitics (i.e., produced without the [L% H-] rise that would demarcate them from the preceding noun phrase), the accented syllable nonetheless is often produced with an associated F0 fall for the *akusento-kaku*, as noted by Sagisaka and Sato (1982), and Kubozono (1993), among others. (See Maekawa and Igarashi, 2006, for counts in the CSJ.) Conversely, content words that are lexically unaccented (i.e., lexically marked as having no possible association to an *akusento-kaku*) do not surface with an associated pitch accent even under focus. This fact is illustrated by the utterance in (11), in which the focally prominent word *mannaka* ‘smack in the middle’ is lexically unaccented and surfaces without any accent, whereas the following verb *okima'su* is lexically accented and surfaces with an accent but with none of the intonational hallmarks that make the preceding locative phrase the focus of the sentence. We turn to these hallmarks in the next section.

(11) F0 contour of an utterance from a spontaneous narrative, illustrating a very prominent but unaccented focal constituent.



3.2 Phrasal pitch range and focal prominence

It has been widely noted in the literature that Japanese uses local pitch range expansion to mark focal prominence (see Kindaichi, 1951; Kawakami, 1957; Fujisaki et al. 1984, 1988, 1994; Poser 1984; Pierrehumbert and Beckman 1988; Kori 1989a, 1989b, 1997; Takeda and Ichikawa 1990; Maekawa 1997; Nagao and Amano 2000; Ito 2002a, among others). Examples (4a) and (11) illustrate this. In (4a), the F0 contour for an utterance of *Ya'mano-wa oyo'ideru* is overlaid on the F0 contour of an utterance of *Ya'mano-ga oyo'ideru*. The rendition of the *Ya'mano-wa...* sentence

(solid line) invokes a garden-variety “thematic” interpretation of *wa* (see Kuno 1973, Heycock 2007), making the following verb the focal constituent in the utterance. By contrast, the rendition of the *Ya’mano-ga...* sentence (dotted line) is an “out-of-the-blue” broad-focus utterance. In keeping with these contrasting focus patterns, the pitch range is expanded on *oyo’ideru* in the *Ya’mano-wa...* sentence relative to that on the same word in the *Ya’mano-ga...* sentence. Conversely, the pitch range on *Ya’mano-ga...* is expanded relative to that on *Ya’mano-wa...*, marking the subject as part of the focus constituent, which is the whole sentence in the broad-focus case. Since both *Ya’mano* and *oyo’ideru* are lexically accented, this pitch range expansion is most readily observed around the [H] tone of the [H*+L] *akusento-kaku* which shows both a more extreme rise from the preceding [%L] or [L%] boundary tone and a more extreme fall to the [L] of the *akusento-kaku*. The utterance in (11), by contrast, shows that a word need not be accented to be a target of pitch range expansion.

This utterance is from a spontaneous monolog narrative that Venditti and Swerts (1996) elicited using a puzzle task in which the speaker is asked to describe the moves chosen to “build a house” from colored pieces of paper. In the just-completed move, the speaker has placed a triangular piece for the roof (*sa’nkaku-no ya’ne*), and in the utterance just before (11) in the monolog, she states that she will next place a pink rectangular piece for a window. In (11), then, she describes where exactly on the roof she will put the window piece. She produces prosodic prominence on the locative phrase *mannaka-ni* ‘smack in the middle’, which is the focal constituent. Note that this focused phrase is unaccented. That is, it does not contain the sharp fall in F0 that is present in APs containing words that are lexically specified as accented. Nonetheless it is clearly marked for focal prominence by the expanded local pitch range. As noted above, this is a key difference from prominence-marking in English, where a pitch accent is inserted on a rhythmically strong syllable post-lexically to cue the pragmatic function of focal prominence. English speakers also employ pitch range expansion to mark focus, but it is localized to the tone targets of the nuclear pitch accent. Japanese is different in that accents cannot be inserted in phrases that do not contain lexically accented words. Therefore, when unaccented phrases such as *mannaka-ni* in (11) receive focal prominence, the local pitch range expansion must target tones other than the [H*+L] of the *akusento-kaku*. Specifically, the pitch range expansion is most easily observed on the [%L H-] boundary-marking sequence at the beginning of the focused unaccented AP.

This kind of dramatic rise at the beginning of the focus constituent is often termed *reset* or *F0 reset*. In both unaccented and accented phrases, the percept of reset can be enhanced in a gradient way by a contrast effect, as noted by Maekawa (1997), who conducted an experimental analysis of F0 heights of all phrases in an utterance containing a focused constituent. He showed that the pitch range of the pre-focal phrase varies in inverse relationship to the degree of focal emphasis; more extreme pitch range expansion on the focused phrase was accompanied by more extreme pitch range compression on the immediately preceding phrase. Ito (2002a) similarly observed that the F0 values of pre-focal peaks were lower than those in a baseline (broad focus) condition.

One topic of lively discussion in the early AM literature on Japanese was the exact extent of the local pitch range expansion after reset. This debate centered on the notion of the “domain” of pitch range specification and the “restructuring” of the prosodic constituent hierarchy that would license a local pitch range expansion such as that observed on *mannaka-ni* in (11). That is, in Fujisaki’s synthesis system, as in other models from that era when the paradigmatic specification of continuous variation in pitch range was first incorporated directly into the phonological formalism, there are two ways in which the extreme rise in F0 at the beginning of *mannaka* in this utterance can be formalized. First, it can be generated by specifying an unexpectedly large amplitude for the

accent command on *mannaka-ni*.² (In the Pierrehumbert and Beckman synthesis model, this corresponds to the specification of an AP prominence ratio value that is much higher than 1.0 for the *mannaka-ni* phrase.) Second, it can be generated by inserting an unexpected new phrase command with an especially large amplitude around the beginning of *mannaka-ni*. (In the Pierrehumbert and Beckman model, this corresponds to an unexpected IP boundary after *ya'ne-no*, with an unusually high top-line value for the new IP that begins at *mannaka-ni*.)

In order to try to decide between these two alternative accounts (i.e. a local AP prominence or an IP break), Pierrehumbert and Beckman (1988) analyzed utterances containing adjective-noun sequences such as *uma'i mame'-wa* ‘tasty beans-TOP’ that differed minimally by which word received a corrective narrow focus, comparing these utterances to baseline utterances with broad focus on the noun phrase as a whole. In one analysis, they had a fluent speaker of Japanese listen to each utterance and tag the perceived degree of disjuncture between the two words. In this test, focused nouns were usually perceived as being set off from the preceding adjective by an IP boundary. Sometimes this percept could be attributed to a visible pause or glottalization. Another analysis exploited the fact that application of downstep is blocked across an IP boundary: Pierrehumbert and Beckman used materials that systematically varied the adjective between accented *uma'i* ‘tasty’ and unaccented *amai* ‘sweet’, and they observed whether downstep had applied by comparing the peak F0 in the following focused noun in the accented context vs. the unaccented context. They found that the F0 height of the focused noun was *not* significantly lower when following an accented word. They interpreted this non-application of downstep as further evidence indicating that an IP boundary is inserted immediately preceding a focused phrase. That is, they concluded that focus marking involves the second of the two mechanisms described above — i.e., the insertion of an unexpected high-amplitude phrase command in the Fujisaki model, or insertion of an IP boundary in the Pierrehumbert and Beckman model.

Kubozono (1993, 2006), by contrast, argues for the first of the two mechanisms, corresponding to the unexpectedly large accent command in the Fujisaki model (or a local AP prominence in the Pierrehumbert and Beckman model). He shows summary statistics suggesting that downstep is not always blocked, and thus concludes that late narrow focus within an adjective-noun sequence (or any other syntactic constituent that normally is produced as one IP) does not involve a restructuring of the prosodic hierarchy. Rather, the focused AP is produced with a “metrical boost” realized on the [H] tone targets. We will return to this point in Section 4.2 after describing other aspects of focus marking that also have been described in terms of restructuring, albeit at a lower level of the prosodic hierarchy.

3.3 Prosodic subordination after focal prominence

Another older observation reconfirmed by Pierrehumbert and Beckman’s study was that focal prominence has measurable effects on following material as well as on the focused constituent itself. That is, putting focal prominence on a word in an utterance induces some kind of *prosodic subordination* of all following words in the intonation phrase. For example, in the utterance in (11), the F0 on the verb *okima'su* is much lower than the [H-] target in the preceding *mannaka*. In (4b), similarly, the accent peak on the verb *oyo'ideru* is very much lower than the accent peak on the preceding *Ya'mano-wa*, in keeping with the “contrastive” interpretation of the *wa* in this utterance.

² Recall that the Fujisaki model uses a square wave “accent command” to model *both* accented and unaccented phrases; in the unaccented case the fall at end of the square wave is obscured either by a following accent command, or by cessation of voicing for a following silent pause.

That is, the degree of pitch range reduction on the verb in this utterance is more than can be accounted for simply by downstep, as can be gauged by comparing the peak F0 on *oyo'ideru* in (4b) to the *oyo'ideru* in the broad-focus utterance of *Ya'mano-ga oyo'ideru* in (4a, dotted line).

This post-focal prosodic subordination is realized in various ways which can be modeled in terms of more or less elaborate manipulations of the prosodic structure. In the simplest case, the only feature that is affected is the pitch range; there is an extreme pitch range reduction on all words in the post-focal region, but all tone targets are realized in the F0 contour. An example of such a realization is the small but observable rise-fall of the [L% H*+L L%] on *oyo'ideru* in (4b). In these cases, the demarcative rise from the AP-level [L%] to a following [H] target (the phrasal [H-] or the [H*] of an early accent) is preserved on the post-focal constituent, although the manifestation of this rise can be very subtle because of the substantially reduced range.

Quite often, however, the AP-initial rise is *not* realized after a focused word. Pierrehumbert and Beckman described this later type of post-focal prosodic subordination as the process of *dephrasing*, which they defined as a total deletion of the [L% H-] demarcative AP-initial rise and (in cases where this deletion would yield two *akusento-kaku* within the resulting merged AP) deletion of all but the first [H*+L] accent fall. That is, they posited a restructuring under focus that not only inserted a new IP boundary to set off the beginning of the focused constituent, but also erased any following AP boundary within the IP. They related this restructuring to the morphological processes that determine the accent patterns of compound words.

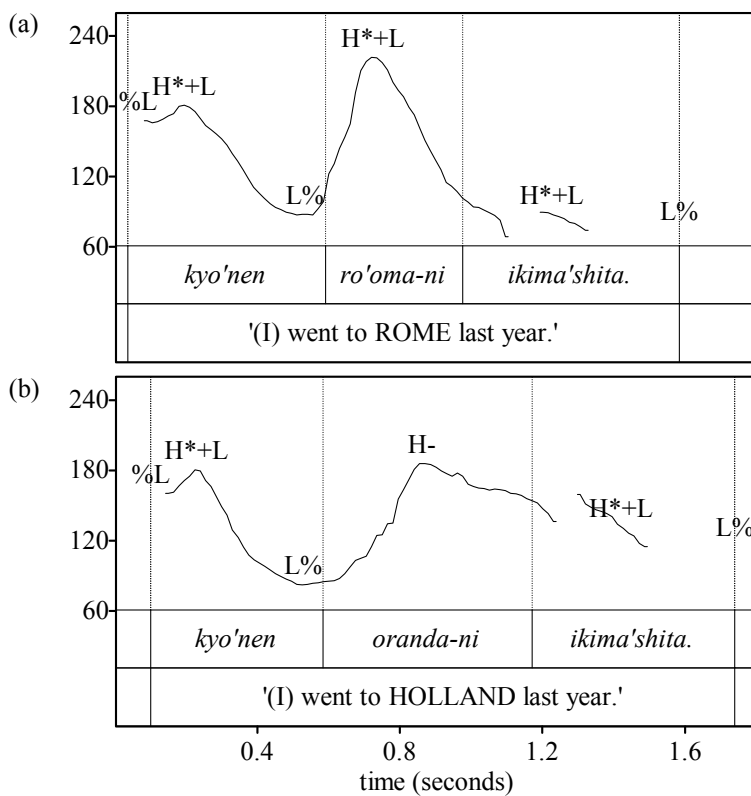
Pierrehumbert and Beckman illustrated dephrasing with figures of utterances that placed narrow focus on the adjective in their target adjective-noun sequences. The figures show an extreme F0 rise marking the beginning of the adjective, but no F0 rise at the beginning of the following noun. The lexical accents of the component words interacted with dephrasing to yield three distinct contours. When the adjective was accented, the purported dephrasing was realized as a very steep fall at the *akusento-kaku*, followed by a gradual fall located in the lower part of the speaker's pitch range. This combination of steep fall followed by gradual fall resembles the F0 contour on the *chi'izu-o ta'beta* portion of example (2a). Notice there is no F0 rise onto *ta'beta* in this case. When the adjective was unaccented, on the other hand, dephrasing resulted in a more or less gradual decline in the upper part of the pitch range. In an unaccented adjective + accented noun sequence (e.g., *AMAI mame'* in Pierrehumbert and Beckman's materials³), such dephrasing resembles the gradually declining interpolation between the phrase-initial [H-] and the [H*+L] on the verb shown in (2b). In the unaccented adjective + unaccented noun sequence (e.g. *AMAI ame*), the shape resembles the interpolation between the initial [H-] and the utterance-final [L%] shown in (2c). Pierrehumbert and Beckman interpreted all three of these patterns as indicating a prosodic restructuring that groups the post-focal noun into the same AP as the focused adjective. They identified this dephrasing with the morphological process of prosodic cliticization that distinguishes auxiliary verbs such as *iru* and *mi'ru* from their full clause counterparts, making the single AP *yo'nde-mi'ru* 'try reading' distinct from the two AP sequence *yo'nde mi'ru* 'read and then see' (as described by Poser (1984), for example).

Later work by Kori (1997) gives further support for this account, and also clarifies the exact nature of the prosodic restructuring. Specifically, Kori's examples show that post-focal prosodic subordination effectively deletes the [L% H-] sequence that normally marks the boundary between two AP, but it need not delete the tone targets of the subsequent *akusento-kaku* when this

³ Following the usual conventions in the literature on focus-marking, we sometimes adopt the shorthand notation of writing the focus constituent in small caps in lieu of explicating the context.

dephrasing occurs within a sequence of two or more accented words. Kori describes cases of prosodic subordination after a focused accented word where post-focal accents are in fact maintained, noting they are prosodically weakened due to the fact that they (i) are realized in a reduced pitch range and (ii) are lacking the phrase-initial rise. This is shown by the “shoulder” of the post-focal accent on *ikima'shita* ‘went’ in (12a). In contrast, Kori notes that prosodic subordination after unaccented focused words is realized differently, although it can be modeled in terms of the very same process of dephrasing at the AP level. That is, after an unaccented focused word, the deletion of the AP-initial [L% H-] rise before any post-focal word results in the F0 remaining relatively high in the speaker’s range, as illustrated in (12b). As Kori points out, despite the striking superficial difference between low F0 after an accented focused word and higher F0 after an unaccented focused word, there is a common underlying prosodic restructuring. As he puts it, “At first glance, the behavior under focus seems to differ depending on the accentuation, but actually you can think of both of these realizations as ‘weakening the autonomy of post-focal words’” (Kori (1997:178)). That is, prosodic subordination can be accomplished by dephrasing (deletion of the demarcative [L% H-] rise) in both accentual contexts.

(12) F0 contours of utterances illustrating post-focal prosodic subordination after an early narrow focus on (a) an accented word versus (b) an unaccented word.



To understand the significance of examples such as (12a), where a focused accented word is followed by another accented word within the same AP, recall that Pierrehumbert and Beckman (1988) followed Fujisaki et al. (1984) and others in assuming that dephrasing deletes the tones of the second *akusento-kaku* as well as the demarcative tones at the AP boundary itself. Under this account, the culminative distribution of accents in compound words (as described by McCawley (1968) and Kubozono (1993), among many others) is mirrored in the culminative distribution of

accents in the post-lexical process of dephrasing for focus marking. Of course, in Fujisaki's model this linkage between dephrasing and accent deletion is unavoidable, given that the square wave of the accent command generates both the phrase-initial rise and the later fall at the accented syllable. In the AM framework, by contrast, there is no necessary coupling between the demarcative rise and the accent fall. This means that dephrasing can delete the [L% H-] sequence that marks the boundary between two accented words *without* deleting [H*+L] in the second word. Thus, as noted above, function words such as *yo'ri* and *su'ra* can be cliticized onto the preceding content word and still surface with the lexically specified fall at the accent.⁴ A related process of subordination of following accents seems to be involved in the contrast that Poser (1984) noted between *yo'nde-mi'ru* 'try reading' (where the auxiliary verb is cliticized onto the main verb) and *yo'nde mi'ru* 'read and see' (where the two verbs form separate APs). Because he assumed a full specification account, Poser described this contrast in terms of a prosodic grouping that necessarily involved a complete deletion of all [H] tone targets in the auxiliary verb *mi'ru*, making the auxiliary reading lose its accent to be *yo'nde-miru*. However, Maekawa's (1994) statistical analysis of F0 slopes in post-focal *yo'nde-mi'ru* versus *yo'nde-iru* 'is reading' shows that this account is not tenable. That is, the contrast between accented *mi'ru* and unaccented *iru* is preserved even for the auxiliary verb readings of these two verbs. And, as the example in (12a) shows, narrow focus on the first word in a sequence of accented words can be realized by a prosodic subordination of post-focal material that drastically lowers the [H*] target without completely eliminating the accent fall in the second word.

Note that by this account (in contrast to Pierrehumbert and Beckman's original formulation), dephrasing after the focus constituent is a "post-lexical" intonational process that is qualitatively different from the morphological processes that determine the distribution of accents in compound words. As described at length by McCawley (1968), Poser (1984), and Kubozono (1993), among many others, when two accented words are conjoined in compound word formation, the output is a word with at most one accent (e.g. *sha'kai* 'society' + *se'ido* 'system' → *shakaise'ido* 'social system').⁵ By contrast, by the account of post-focal prosodic subordination presented here, when two or more accented words are conjoined by the deletion of the boundary-marking rise, the output is a single AP that can contain two or even more *akusento-kaku*. Note, too, that while his account is somewhat different from the one presented here, Kubozono (1993) also differentiates the morphological processes that enforce a culminative distribution of accents in compound word formation from the post-lexical grouping of words into phrases to reflect the syntactic organization of sentences in a discourse. He reserves the term *accentual phrase* for the output of the compound accentuation rules and terms the post-lexical prosodic grouping the *minor phrase*. While Kubozono (1993) does not discuss post-focus subordination in these terms, we can say that his "accentual phrase formation" is roughly analogous to the English Compound Stress Rule in the same way that his "minor phrase formation" is roughly analogous to the application of the Nuclear Stress Rule to larger constituents such as noun phrases in the older literature on the syntax-prosody mapping in English. This recalls the analogy that Pierrehumbert and Beckman (1988) made between post-focal dephrasing in Japanese and *deaccenting* in English — i.e., the constraint in the

⁴ Kubozono follows Poser (1984) in calling this pattern "total downstep"; see pattern B in Kubozono (1993:104).

⁵ The exceptions to this culminative distribution involve a handful of productive prefixes such as *ju'n* 'quasi' and *mo'to* 'former' (the so-called *Aoyagi prefixes* (see Aoyagi (1969)), which retain their accent in addition to the accent on the base noun, as in *mo'toda'ijin* 'former minister'.

English intonation system against the occurrence of any further pitch accents within an IP after a word that bears the nuclear pitch accent. These analogies serve to highlight both the profound formal differences and the deeper functional similarities between the two intonation systems. Where Japanese uses prosodic grouping to mark prosodic subordination, English uses the culminative distribution of stress markers such as association to a nuclear pitch accent. Nonetheless, in both languages, there is a formal relationship between the mechanisms that mark morphological subordination in compound word formation and higher-level subordination of post-focal material within the IP. We will return to this point in Section 4.3.

3.4 Prominence-lending boundary pitch movements

In the two previous sections we have described how speakers can use prosody to cue focal prominence in Japanese (i) by inserting an IP boundary and manipulating the pitch range before and after that boundary to make for a dramatic reset at the beginning of the focus constituent, and (ii) by processes of prosodic subordination after the focus constituent which includes the deletion of the rise at the beginning of following APs as well as an extreme pitch range compression that can reduce subsequent accent peaks to the same level as the [L] target of the first [H*+L] when the focus constituent itself is lexically accented. We now turn to a different kind of restructuring to cue focus whereby the speaker sets off the focus constituent from adjacent material on *both* sides. That is, in addition to the reset that marks the beginning of the focused word, there is some salient marking of the end of the focus constituent, to separate it prosodically also from following material. One way that speakers can mark the end of the focus constituent is by inserting a silent pause. When this pause is inserted after an unaccented word, it simply interrupts the interpolation from the phrasal [H-] to the next tone target in an attention-getting way. In (11), for example, the downward sloping interpolation from the [H-] in *mannaka-ni* toward the lowered [H*+L] in *okima'su* suddenly stops for 200 ms before resuming again at the level that it would have reached if the two words had been separated instead by a filled pause. Another way of marking the end of the focal constituent to set it off from following material is by producing a BPM. As noted above, two of the four BPM that are tagged as tonally distinct categories in the X-JToBI scheme — namely [H%] and [HL%] — have been linked to a prominence-lending function. Both of these are illustrated in (13), and we will describe each of them in turn.

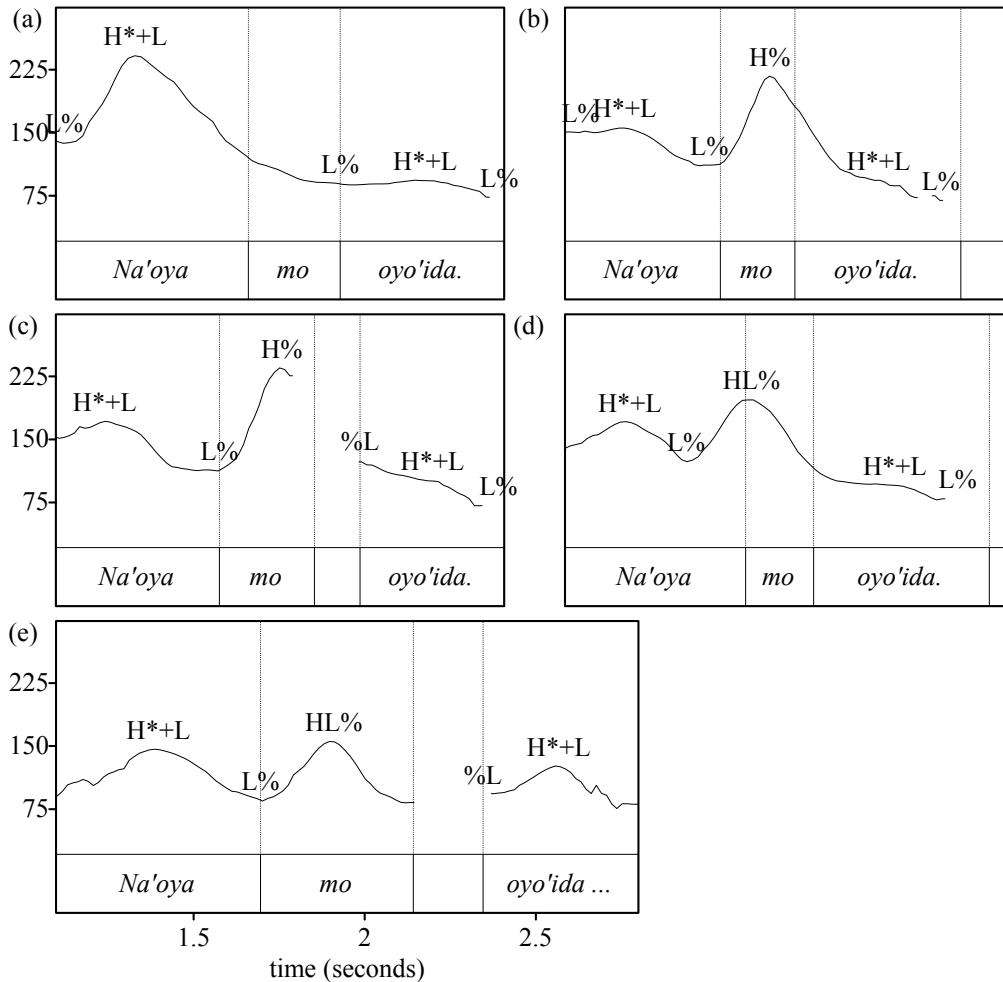
3.4.1 The Prominence-Lending Rise (PLR)

The [H%] prominence-lending rise (henceforth, PLR) is the BPM most commonly associated with a prominence-lending function. For example, focal prominence can be marked on the phrase *Na'oya-mo* 'Naoya-also' in the sentence *Kinoo Na'oya-mo oyo'ida* 'NAOYA also swam yesterday' by inserting an intonation phrase break before *Na'oya* accompanied by a large pitch reset, as described above and shown in (13a), or by putting a PLR at the end of the focused phrase, as shown in (13b-c). This second mechanism was noted by Ohishi (1959), Kawakami (1963), Miyaji (1963), Kori (1989b,1997), Muranaka and Hara (1994), Venditti (1997), among others, and more recently has been investigated experimentally by Venditti et al. (1998), Taniguchi and Maruyama (2001), Maruyama and Taniguchi (2002), Oshima (2005), and Mesbur (2005).

The example in (13a) shows the use of expanded pitch range to cue focal prominence on *Na'oya-mo* 'Naoya-also'. Note the prosodic subordination of the following verb *oyo'ida* 'swam', similar to (12a). Figures (13b-e) illustrate the use of the various BPM to cue this same function: (13b) shows [H%] marking *Na'oya-mo*, with no pause following, (13c) shows [H%] with a pause following, and (13d) gives an example of the 'penultimate non-lexical prominence' (PNLP), which

we will discuss below. These three examples show prosodic subordination of the verb after the BPM, similar to (13a). The final contour (13e) gives an example of the [HL%] ‘explanatory’ BPM, which was also illustrated in (8c).

(13) F0 contours of the underlined portion of *Kinoo*, *NA'oya-mo oyo'ida.* ‘NAOYA also swam yesterday.’, illustrating five different ways of marking the focal prominence on *Na'oya-mo*.



Since the PLR often occurs on phrases ending with a particle/postposition, many accounts of its form and function use terms such as “prominent particles” — which might be misinterpreted as suggesting that the particle itself is the focus constituent. The [H%] does make the final syllable in a phrase stand out in the intonation contour, and its occurrence on a phrase-final particle can mark narrow focus on the particle itself, as in Ohishi’s (1959) example of meta-linguistic contrast in (14a).⁶ More typically, however, the focus constituent is the entire phrase ending with the particle that is aligned with the PLR, as in Ohishi’s example in (14b). This fact has been pointed out by a

⁶ In text-only examples, we will indicate a prominence-lending BPM by underlining the phrase-final syllable of the focus constituent, followed by the symbol \uparrow (to represent the F0 rise that is realized on that syllable for [H%]) or \wedge (to represent the rise-fall shape for [HL%]). The symbol \wedge also will be inserted after the penultimate syllable to represent a PNL rise.

number of scholars, including Ohishi himself, who says: “When prominence is placed on a particle, the thing that is focused is not just the particle, but it is the word that that particle is affiliated or packaged with” (Ohishi, 1959:95). Maruyama and Taniguchi (2002:18) also highlight this point, saying: “Prominence on particles prosodically marks the right edge of the phrase which includes the focused element, and has the function of prosodically indicating the focus structure of an utterance which may not be represented by its syntax”.

- (14) a. A. *Gakkoo-no*[↑] *tooro* 'nkai? (Ohishi, 1959)
 ‘Did you say it’s a debate ABOUT the school?’
 B. *Iya, gakkoo-de*[↑] *yatta n de* 'su.
 ‘No, it’s that they held it AT the school.’
- b. *Watashi-no a* 'ni-wa[↑] *otonashi* 'i.
 ‘MY OLDER BROTHER is shy.’
- c. *Watashi-no A* 'NI-wa *otonashi* 'i. [with pitch range reset on *a* 'ni-wa]
 ‘My older BROTHER is shy.’
- d. *San-senchi-gu* 'rai akete[↑] *shita-ni okima* 'su. (Venditti, 1997)
 ‘I will OPEN UP ABOUT 3 CM and put it below there.’
- e. [Context: ‘Is it the Mr. Yamada who lives in Miyajima?]
Iie. Ao 'mori-ni *sunde-iru*[↑] *Yamada-san de* 'su. (Maruyama and Taniguchi, 2002)
 ‘No. (It’s) the Mr. Yamada who LIVES IN AOMORI’
- f. *Ka* 'zuya-wa *bi* 'iru-o *sa* 'nba[↑] *chuumon shima* 'shita. (Oshima, 2005)
 ‘Kazuya ordered THREE BOTTLES of beer.’

A further point to make is that while [H%] can occur on a particle, it need not do so, as is shown by the examples from various authors in (14d-f), and by the contour in (9a). Indeed, it appears that the only constraint on prominence-lending BPM placement (besides the need to be at or near the constituent edge) is that a prominence-lending BPM cannot occur sentence-finally.⁷ In this position, focal prominence can only be cued by reset — i.e., by having a phrase boundary before the focus constituent where there can be a pronounced pitch range expansion relative to prior material.

This constraint against marking focus on utterance-final phrases with PLR complements the constraints on marking focus by the mechanisms identified in the previous section. That is, on utterance-initial words, it is difficult to mark focus by inserting an IP boundary and expanding the pitch range, because the beginning of an utterance must be the beginning of a new IP in any case, and because there is no preceding material to produce in a relatively reduced pitch range to contrastively enhance the phrase-initial rise.⁸ This functional motivation for the constraint is reflected in Ohishi’s (1959) discussion of two of his examples, given here as (14b) versus (14c).

⁷ To clarify, [H%] can occur sentence-finally, on sentence-final particles such as the *ne* in (8b), but this usage does not have a prominence-lending function and is not an example of PLR.

⁸ Compare the asymmetry in English between marking narrow focus on non-final constituents with an early nuclear accent, versus the difficulty in marking narrow focus by nuclear pitch accent placement in utterance-final position (although it is possible to set off the final word with a preceding silence and by choosing L+H* as the nuclear accent so as to produce a steep rise onto the peak that is associated with the accented syllable).

Specifically, he suggests that while local pitch range expansion can indicate narrow focus on *a'ni* 'older brother' in (14c), it is impossible to use pitch-range expansion to focus the entire noun phrase *watashi-no a'ni* 'my older brother'. However, the speaker can still mark that phrase as the focus constituent by producing a [H%] PLR on the final syllable, as in (14b).

Another related point is that some speakers find it difficult to cue focus by expanding the range of an unaccented phrase, without setting off the focus constituent somehow from following material. That is, it is generally acknowledged among Japanese researchers (with suggestive instrumental support in an experimental study by Ito (2002a)) that it is difficult to expand the range of an unaccented phrase to mark focus. However, other experimental studies by Kori (1989a) and Pierrehumbert and Beckman (1988) have observed pitch range expansion on unaccented APs, and the example in (11) shows that it can occur in spontaneous speech, too. The feeling of awkwardness probably originates in the lesser salience of the AP-initial boundary rise when there is no downstep to reduce the F0 values of the [L%] and any following [H*+L]. In such cases, the use of the [H%] PLR is a likely alternative to inserting a pause in the middle of the AP, as in (11).

Using PLR to mark focal prominence is optional, just as pitch range expansion on the focal constituent is optional. In fact, Kori (1997) states that the only obligatory prosodic marker of prominence is the post-focal prosodic subordination discussed in Section 3.3. In a naturalness rating perception study, Maruyama and Taniguchi (2002) found that [H%] did indeed cue focus, but [H%] alone was not sufficient when the post-focal region was not prosodically subordinated as well. That is, while their intonation contour with PLR followed by a compressed pitch range (their pattern B) was judged by listeners as a highly natural way of focusing that phrase, the contour with PLR and no compression of the following pitch range (their pattern D) was judged as unnatural. They concluded, as Kori did, that post-focal prosodic subordination is crucial for cueing focus naturally.⁹

That said, post-focal pitch range compression does not occur in all cases where apparently focal material is set off from following phrases by the insertion of [H%]. The utterance in (9a) illustrates this point, representing a very deliberate speaking style that is common in monolog narratives such as the academic talks and prepared presentations in the CSJ; the speaker produces a string of phrases each deliberately set off (and given a kind of focal prominence) by [H%], with no obvious post-focal subordination. This use of [H%] to set off each phrase as a separate focal constituent is reminiscent of the short intonation phrases and frequent use of [L+H*] nuclear pitch accents in American radio and television news broadcast style, where prosody is used to break up the densely

⁹ Maekawa (1991) offers further evidence for this conclusion. He compared the prosodic patterns observed on the WH-question in (i) to that of the minimally different polarity question in (ii).

- (i) *Na'ni-ga mie'ru?* 'What do you see?'
 (ii) *Na'ni-ka mie'ru?* 'Do you see something?'

Utterances of (i) typically have an expanded pitch range on *Na'ni*, with the predicate *mie'ru* being realized in a very reduced pitch range. By contrast, utterances of (ii) typically show a lower peak on *na'ni* and an prominent AP boundary or even a reset on the predicate. In a perception experiment using synthetic stimuli (where the /k/ vs. /g/ consonants were crucially replaced with white noise), Maekawa showed that listeners' judgments were influenced more by the degree of subordination of the predicate than by the height of the question word itself. That is, the percent of WH-question judgments increased as the F0 height of the predicate decreased, regardless of the height of the focused WH-word. This finding emphasizes the crucial importance of the post-focal region in the interpretation of focal prominence.

information-packed paragraphs typical of news stories. That is, examples such as (9a) may point to another function of [H%] in discourse, a use that shares the foregrounding function of PLR in (14), but without the concomitant backgrounding of all material outside of the focus constituent. In the information-packed paragraphs of this kind narrative monolog, the [H%] helps speakers and listeners chunk information in the speech stream. This “Here, parse this!” function is alluded to in Muranaka and Hara’s (1994:396) descriptions of the PLR on particles as “making clear the syntactic structure”, or “uttered with higher F0 when the preceding noun phrase is long, [indicating] the end of the phrase”.

3.4.2 The rise-fall BPM and PNL

The [HL%] BPM is also said to have a prominence-lending function, and some of the remarks just made about when and why a speaker chooses to mark focus with PLR also apply to the rise-fall. In the discourse in (15), for example, speaker B uses [HL%] to put prominence on *ano 'hito* ‘that person’. At the same time, however, [HL%] also expresses the speaker’s irritation at B’s misunderstanding. In their perception study using ratings on various semantic scales, Venditti et al. (1998) found that listeners perceive [HL%] as highly ‘emphatic’ (*kyoochoo shite iru*) and ‘explanatory’ (*joohoo-o setsume shite iru*) — i.e., the rise-fall BPM imparts a very strong sense of the speaker forcefully explaining some point to the listener. In fact, [HL%] was rated higher on these two scales than [H%] was. However, in this study listeners did not rate [HL%] as indicating the speaker’s irritation, as it seems to do in the context in (15).

- | | |
|--|--------------------------------|
| (15) A: (talking to B) <i>ana'ta-ga Yamada-san de'su ka?</i> | Are you Mr. Yamada? |
| B: <i>Iie, watashi-wa Yoshida de'su.</i> | No, I'm Yoshida. |
| C: <i>Soshite watashi-ga[↑] Yamada de'su. (H%)</i> | And I am Yamada. |
| | |
| A: (talking to B) <i>wakarima'shita.</i> | I understand. |
| <i>ana'ta-ga Yamada-san de'su.</i> | You are Mr. Yamada. |
| B: <i>Iie, ano'hito-ga[^] yamada-san na'n de'su.(HL%)</i> | No, THAT PERSON is Mr. Yamada. |

The chunking function described for [H%] and illustrated in (9a) is also very characteristic of the rise-fall BPM; in many utterances in the CSJ presentation speech narratives, [HL%] sets off nearly every phrase in an especially deliberate way. This use of the rise-fall BPM seems especially common in dialogs among young speakers, especially young girls. In fact, some speakers use [HL%] so frequently that it seems to lose any local focalizing function and becomes more of a marker of the speech style, and of the speaker’s social identity in choosing the style. That is, the rise-fall BPM in these speakers and this style may have be like the frequent use of the English high-rise contour by the younger speakers of American English that is described by McLemore (1991), where the frequency of use effectively weakens the meaning of this contour that is described by Hirschberg and Ward (1995), among others, and makes it instead a marker of solidarity (with the speaker’s fellow sorority members, in the case of McLemore’s study).

As noted earlier, X-JToBI uses the [HL%] sequence of the rise-fall BPM to mark another distinct prominence-lending shape — namely, the penultimate non-lexical prominence, or PNL, illustrated earlier in (9c). In that utterance, the rise is localized on the penultimate syllable of the phrase *soo itta yo'o-na ta'ipu no rensoogo-o* ‘the aforementioned type of semantic associate’ (indicated here by underlining). The PNL puts narrow focus on the newly introduced term *rensoogo* ‘semantic associate’ which is being defined in this appositional use of *no*.

We describe the PNLP together with the rise-fall BPM in this subsection because it is tagged in the CSJCore using the [HL%] tone label (in conjunction with the PNLP comment label). However, this choice of the rise-fall BPM label in the tagging conventions should not be interpreted as indicating a commitment to the implied phonological analysis on the part of any of the researchers who developed the X-JToBI tagging conventions. Rather, the [HL%] tag was chosen as a quick-and-dirty first approximation to the superficially similar rise-fall shape that PNLP typically shows. For example, in (9c), the F0 falls over the duration of the accusative case particle *o* onto the first syllable of the following phrase, the verb *kotae'ru* ‘respond’. However, there is an alternative account available for this fall. Since the beginning of the phrase immediately following a medial PLR [H%] (like the beginning of a phrase after a silent pause) is marked by an initial [%L] boundary tone, the F0 fall after a PNLP could be the transition to this independently motivated [%L] target. The PNLP rise, then, could be a leftward-displaced instance of the PLR [H%].

This alternative account is appealing to us, because the PNLP is often cited in the literature (e.g., by Ohishi (1959) and Kori (1989b, 1997)) as a variant of the “prominent particle” phenomenon we attribute to the PLR [H%]. Moreover, our impression from the instances that we have seen is that the PNLP is more similar to [H%] than it is to [HL%] in both form and function. Compare, for example, the slope of the [L% H%] rise in (13b) to that of the PNLP in (13d). In addition, the typical lengthening of the syllable that is aligned with the [HL%] (as seen in (13e)) is not present in cases of either PNLP or [H%]. However, differentiating between these two (and several other more controversial) alternative accounts is a research question in its own right. Moreover, it is a question that requires more extensive corpus work than would have been possible before the development of resources such as the CSJ. We will return to this point again briefly in Section 4.4.

4 Ambiguities of the prosodic parse

At various points in our description of focus-marking mechanisms in Section 3, we alluded to points of disagreement or potential disagreement among linguists currently working on Japanese intonational phonology. Many of these are points where the phonetic evidence that might differentiate between competing prosodic parses is inherently lacking or ambiguous. In this section, we will briefly describe four of these points.

4.1 Post-focal dephrasing

In Section 3.3, we described the prosodic subordination of post-focal material in terms of the notion “dephrasing” which we defined as the deletion (or non-realization) of the [L% H-] sequence that marks the beginning of every well-formed accentual phrase. By this definition, the junctural cues are definitive. When no [L% H-] sequence is realized between two accents and there are no segmental cues to support an AP boundary in the absence of this tonal cue, dephrasing has occurred. When two accented words are phrased together into a single AP, both accent falls can be realized, although the second will typically be reduced in prominence, so that its [H*] target is at the same F0 level or a somewhat lower level than the [L] target of the first accent, as in *ta'beta* in (2a) and *ikima'shita* in (12a).

Note that by this account, for sequences of accented words, the line between dephrasing and extreme pitch range compression is a fuzzy one, particularly when there is only one or two syllables separating the two accents. The F0 contour over the sequence of words *yo'o-na ta'ipu* in (9c) illustrates this source of ambiguity in the prosodic parse. The inflection point at the beginning of the small rise to the accent peak on *ta'ipu* could be parsed as the [L] of the accent on *yo'o-na* or

as a [L%] boundary tone at the word boundary. The X-JToBI transcriber must pay close attention to subtle cues such as the length of the word-final vowel in *yo'o-na* to distinguish between these two parses. This means that there are many different degrees of clarity to the phrasal disjuncture between two accented APs. Moreover, this continuum of degrees of perceived disjuncture is in a gradient relationship to the many different degrees of subordination that are also possible. The post-focal dephrasing evident in (12a), then, is toward one end of a continuum that includes (at the other end) the much less compressed pitch range of the following verb phrases relative to the *Ya'mano-wa* and *Ka'shino-wa* phrases in example (4b).

By our account, in other words, sequences of two accented words are vulnerable to the kind of prosodic parsing ambiguity described by Beckman (1996b) because the F0 pattern in the middle of a sequence of two accents within a single AP is phonetically very similar to the rise that marks the boundary between two APs. By contrast, sequences of two unaccented words are not phonetically susceptible to this type of misparsing of the intended prosodic structure, because a rise at the boundary between two unaccented words can only be parsed as [L% H-]. Of course, sequences of unaccented words are vulnerable to the opposite misparsing, because the [L%] is not subject to downstep and because (other things being equal) the phrasal [H-] is scaled lower in the pitch range than the peak in the accent [H*+L]. These two facts together mean that the demarcative rise will be much smaller for sequences of two unaccented words as compared to sequences of two accented words for the same degree of prosodic subordination of the second word, so that listeners may misinterpret an intended sequence of two unaccented APs as having undergone post-lexical dephrasing.

Note that this account of dephrasing as one endpoint of a continuum of subordination differs from the more categorical distinction assumed in some other accounts of post-focal subordination. That is, there is a broad consensus in the AM literature that dephrasing does occur. However, many other phonologists have assumed a slightly different account of dephrasing which does not go as far as our definition in differentiating the post-lexical intonational process from the morphological processes that account for the accent patterns of compound words. Specifically, in most previous accounts other than Kori (1997), the output of post-lexical dephrasing is assumed to preserve the (almost perfectly) culminative distribution of *akusento-kaku* in the lexicon. When two accented words are phrased together into a single AP, these accounts say, the second *akusento-kaku* must be deleted along with the demarcative rise.¹⁰ In applying these models in analyses of actual utterances, then, the presence of a second [H*+L] fall in a sequence of two accented words is interpreted as unequivocal evidence that there is an intervening AP boundary even when the second peak is reduced to a target level at or even below the level of the inflection point for the [L] of the first accent. By these other accounts, then, the fall on *ta'beta* in (2a) must indicate an AP break after the object NP, making the prosodic structure of this utterance different from that in the syntactically identical sentences in (2b) and (2c). Moreover, the different shapes in the post-focal region in (12a) versus (12b) are interpreted as indicating different focus-marking mechanisms depending on the lexical accentuation of the word in focus.

For example, Sugahara (2002) observes what she takes to be evidence of an effect of lexical accentuation in determining prosodic phrasing in the post-focal region. That is, like Kori (1997),

¹⁰ That is, even though AM accounts such as Pierrehumbert and Beckman's (1988) differ from Fujisaki's account in modeling the phrasal [H-] as an independent target from the [H*] of the accent fall, most previous AM accounts effectively model dephrasing as the formal equivalent of deletion of the whole accent command in the Fujisaki model.

she describes a pattern of no AP-initial rise and continuous gradual fall on post-focal (contextually “given”) words in unaccented sequences and a contrasting pattern with a “bumpier” transition when the focal word and the following word are accented. In account for this difference, she posits a dependency between accentual phrasing and lexical accentuation status. That is, she notes that post-focal material is not subject to dephrasing when it is contextually “new” information or at the left edge of a major syntactic constituent (XP), as might be expected if accentual phrasing tends to reflect both the discourse context and the syntactic structure of an utterance, but she also concludes that post-focal material is not subject to dephrasing when it is lexically accented. This is a somewhat surprising effect of the lexical tone pattern if we assume the strictly modular division of classical Generative Phonology between the phonological grammar that generates phonological structures and the phonetic processes that implement them.

Other researchers also have observed apparent interactions of dephrasing with lexical accentuation. Most notably, Ito and her colleagues (Ito et al. (2003), Ito and Speer (to appear)) document such an interaction in spontaneous narratives, where the contrast between “given” and “new” information arises naturally from the non-linguistic task that naive subjects were asked to perform. We can be confident, therefore, that it is not an artifact of eliciting the contrasting information structures in the laboratory using contrived dialogue scripts. More extensive corpus work, in conjunction with analysis-by-synthesis methods, needs to be done before we can say whether the apparent interaction can be explained away by the different, complementary phonetic susceptibilities to ambiguity in the parsing of tonal cues to disjuncture in accented-accented sequences as compared to unaccented-unaccented sequences.

4.2 Prosodic restructuring or local boost?

In Section 3.2, we described pitch range expansion under focal prominence in terms of a prosodic restructuring, such that an IP boundary is inserted before the focal constituent. As Kubozono (1989, 1993, 2006) and others have pointed out, however, there is an alternative account available, because the reset at the beginning of the focus constituent is potentially ambiguous: it could signal the start of a new IP, or it may be a more local boost that raises the topline of the current AP only, without affecting the prosodic organization of the utterance at the IP level.

To determine locations of IP boundaries, researchers sometimes compare the height of the target peak to that of the previous one: if the target is the same height or higher than the previous peak, they assume that there is an IP boundary between the two phrases. Kubozono (2006) refers to this as the “syntagmatic approach.” Beckman and Pierrehumbert (1992) suggest that, as a research methodology, this approach is circular; typical peak height relationships can vary to contrast two categorically different levels of prosodic disjuncture (as in the contrast between IP and AP in the Japanese prosodic hierarchy), but peak height relationships can also vary in a continuous, iconic way to reflect discourse structure, including many different degrees of relative foregrounding versus backgrounding of constituents. They therefore suggest that the only way to definitively determine whether or not an IP boundary exists between two phrases is to systematically vary the accentuation of the first phrase and observe the downstep effects (if any) on the second phrase. Kubozono (2006) refers to this as the “paradigmatic approach” and he, too, suggests that it is a better method for deciding whether an IP boundary has been inserted before a focus constituent.

Using this paradigmatic approach, Pierrehumbert and Beckman (1988) found no significant effect of previous accentuation on focused peaks in their data. Specifically, while they found a small (albeit insignificant) mean difference for the peak height value in a focused word depending

on the accent of a preceding word, a scatterplot showing the peak height relationship for each token in the data set suggested that this was due to variability in the phrasing, with a few tokens showing a dramatic syntagmatic reduction after accent, and all the others showing reset. This pattern led them to propose that focal prominence typically introduces an IP break before the focused element. This proposal has been widely adopted in subsequent analyses of focus-related effects on phrasing in Japanese (Nagahara, 1994; Ito, 2002a, among others).

In contrast to Pierrehumbert and Beckman’s findings, Kubozono (2006) found a significant difference in mean values for F0 peaks in focused WH-phrases such as the *na’ni-o* ‘what-ACC’ in (16). Here, the accentuation of the two phrases preceding the target word *na’ni* was systematically varied, yielding a somewhat lower average peak value on *na’ni* in (16a) relative to the mean in (16b) for 5 of 7 speakers. This small difference in mean values appeared even though the pitch range of the WH-word appeared to be expanded under focus in every token. Following his earlier work on local pitch range expansion at major syntactic boundaries (see Kubozono 1989, 1993), Kubozono posits a local F0 “boost” which occurs on focused phrases without undoing the effects of downstep from preceding accents, as would be predicted if reset was associated with an unexpected IP boundary just before the target word. Kubozono concludes that pitch range expansion for focus marking can occur without the prosodic restructuring posited by Pierrehumbert and Beckman.

- (16) a. ACCENTED ACCENTED ACCENTED (FOCUS)
Ana’ta-wa *Ao’mori-de* *Na’oko-to* *na’ni-o* *mima’shita ka?*
 you-TOP Aomori-LOC Naoko-with what-ACC saw-Q
 ‘What did you see with Naoko in Aomori?’
- b. ACCENTED UNACCENTED UNACCENTED (FOCUS)
Ana’ta-wa *Oomori-de* *Naomi-to* *na’ni-o* *mima’shita ka?*
 you-TOP Oomori-LOC Naomi-with what-ACC saw-Q
 ‘What did you see with Naomi in Oomori?’

How can we account for the discrepancies between these two studies? Kubozono points to one possible source of the difference in results — namely, a difference between his materials and the materials used in most other experiments, including Pierrehumbert and Beckman’s. Kubozono examined WH-phrases, which are (under most accounts of focus in questions) natural targets of focal prominence.¹¹ By contrast, most other laboratory studies of focus have used scripted dialogs that are designed to prompt the readers to perform a contrastive or even corrective focal prominence. For example, the materials in Pierrehumbert and Beckman (1988) used the “contrastive *wa*” construction illustrated in (4b). As Prince (1981) and many others have pointed out, the discourse phenomena that have been termed “focus” are no more homogeneous than the prominence-lending mechanisms that languages use to mark focus. Perhaps the differentiation of the contrastive use of *wa* from the ordinary thematic use of *wa* requires a more dramatic setting off of the focus constituent than does the “inherent” focus of the WH-word.

Another important point that emerges from comparing these studies is that the potential for misparsing the degree of prosodic disjuncture before a focus constituent interacts with the

¹¹ See, for example, the contrast between WH-questions and yes-no questions studied in Maekawa (1991).

interpretation of focal prominence, and that this interaction depends on the lexical accentuation in a way that is reminiscent of the ambiguities regarding dephrasing after the focus constituent discussed in Section 4.1. That is, when preceding material is unaccented, the listener can directly compare the F0 before and after the [L% H-] rise to gauge the degree of relative pitch range expansion on the focal constituent — as in Kubozono’s syntagmatic approach — and the speaker can induce the percept of focal prominence even with a fairly small reset at the beginning of the focused element. When the preceding material is accented, on the other hand, the listener must gauge the intended pitch range relationship against the backdrop of Bayesian expectations regarding downstep — taking a necessarily paradigmatic approach by evaluating the reset relative to an internal representation of the typical effects of being after an accented word within the same IP versus being after an accented word across an IP boundary.

We suspect that this difference between the two cases is what gives rise to another interaction with lexical accentuation that Kawakami (1957) observed. He points out that after an accented word, pitch range expansion for focus marking must at least mimic the effect of blocking the downstep, creating syntagmatic cues that suggest an IP boundary, as in (17a). After an unaccented word, on the other hand, the speaker can put focal prominence on a word simply by producing the AP-initial rise within a syntactic constituent that is typically produced as one AP, as in (17b).

- (17) a. *Ka're-no* || *ANE desu.* (Kawakami, 1957)
 ‘(She)’s his OLDER SISTER.’
 b. *Watashi-no* | *ane desu.*
 ‘(She)’s my OLDER SISTER.’

Thus, just as the more extreme degrees of subordination of post-focal material leads to an ambiguity of parsing at the AP level in a sequence of two accented phrases, more extreme degrees of emphasis on the focal material leads to an ambiguity of parsing at the IP level. Moreover, this phonetic ambiguity at the IP level depends on the lexical tone pattern in a way that suggests that the components of the grammar that determine the prosodic structure of an utterance are not rigidly encapsulated away from the phonetic processes that implement that structure.

4.3 Scaling of high tones within the AP

In Section 4.1, we defined dephrasing as a post-lexical process that differs from the morphological processes involved in compound word formation by targeting the cues to the AP boundary alone, without deleting one of the accents when two accented words are conjoined. We suggested that this difference between the two processes may help to explain an otherwise puzzling dependency between dephrasing and the lexical accentuation of the focused and post-focal constituents. If we are correct in saying that the conjoining of two morphemes into a compound word differs prosodically from the conjoining of two words into a single AP, then we might also expect different behavior in cases of meta-linguistic contrast on one of the elements in a compound word, as in (18).

The sentences in (18a-c) are based on Gussenhoven (2004), who cites Mariko Sugahara (personal communication, 2000) as saying that the two elements of a compound word cannot be focused separately. More specifically, he says that a compound word such as *mukashi-ba'nashi* ‘legend’ in (18a-c) cannot be divided into two APs to have different pitch ranges to express the difference between narrow focus on *mukashi* in (18a) and narrow focus on *hanashi* in (18b), and he shows F0 contours of elicited utterances of (18a-c) which “are all basically the same”

(Gussenhoven 2004:205). His figure suggests two further pieces of evidence for the purported prosodic indivisibility of the compound word: (i) the initial consonant in the second morpheme is [b] rather than the underlying [h], showing that the morphological process of *rendaku* (“sequential voicing”) has applied, and (ii) the location of the accent fall in the F0 contour reflects the compound accent pattern rather than the underlying final accent of *hanashi* ‘tale’.

- (18) a. *OTOGI-ba 'nashi jana 'kute, MUKASHI-ba 'nashi da yo.* (Gussenhoven, 2004)
 ‘It’s not a NURSERY tale, but a LEGEND [literally: OLDENTIMES-tale].’
 b. *Mukashi MONOGA 'TARI jana 'kute, mukashi-BA 'NASHI da yo.*
 ‘It’s not an old-time STORY, but a LEGEND [literally: oldentimes-TALE].’
 c. *Watashi-ga mi 'tai no wa E 'IGA jana 'kute, MUKASHI-BA 'NASHI da yo.*
 ‘What I want to see is not a FILM, but a LEGEND [literally: OLDENTIMES-TALE].’
 d. *Ho 'n-DANA de 'wa nakute ho 'n-BAKO de 'su.* (Kawakami, 1957)
 ‘It’s not a bookSHELF, it’s a book-BIN.’
 e. *Koogyoo-DA 'IGAKU da ' ka koogyoo-GA 'KKOO da ' ka...*
 ‘Whether it’s an industrial COLLEGE, or an industrial SCHOOL ...’

By contrast, Kawakami (1957) says that corrective focal prominence on the second element in compound words such as *ho 'n-dana* versus *ho 'n-bako* in (18d) and *koogyoo-da 'igaku* versus *koogyoo-ga 'kkoo* in (18e) can be expressed by having the same kind of AP-initial rise in the middle of the compound word that he observes in sequences of words such as (17b). At the same time, he points out that even when this prominence-lending rise occurs, the forms show the other prosodic hallmarks of being compound words. Specifically, (i) *rendaku* applies to both the underlying *tana* ‘shelf’ and *hako* ‘box’ in (18d), and (ii) the compound accent pattern applies to put an *akusento-kaku* on the first syllable of *daigaku* ‘college’ and *gakkoo* ‘school’ (both of which are unaccented as simplex forms) in (18e).

The conflicting cues to the nature of the word-internal juncture that Kawakami describes for examples (18d-e) is reminiscent of the conflicting cues to the lexical status of the initial morpheme in English words such as *illiterate* and *impolite* when corrective focus associates the nuclear pitch accent to the prefix, as in (19).

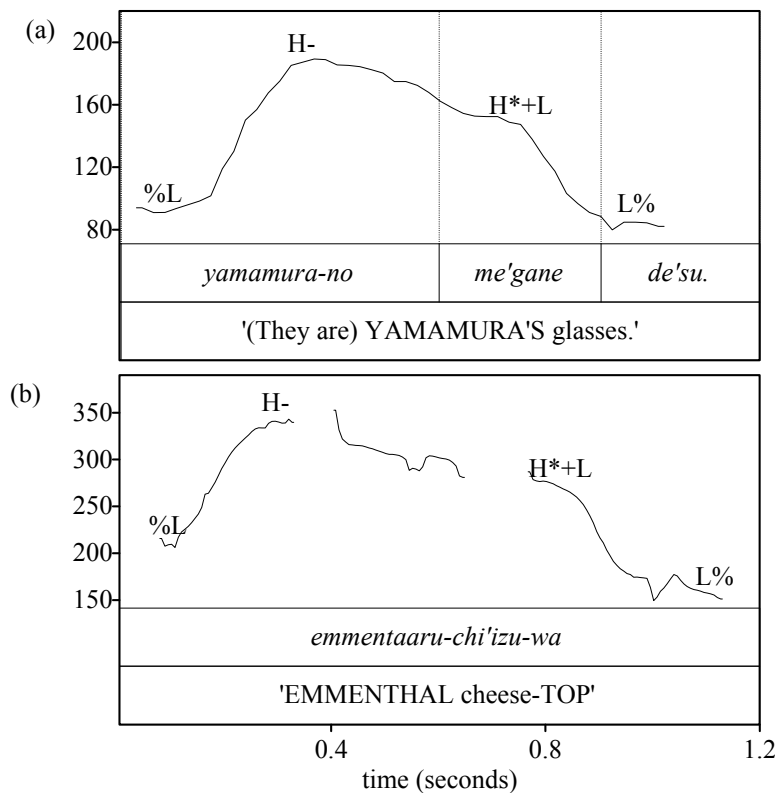
(19) He’s NOT polite. In fact, he’s downright IMPolite.

Hay (2003) shows that native speakers find it easier to accent the prefix in such contexts of corrective focus when the frequency of the derived form has low token frequency relative to the stem, making the prefixed form more decomposable into its component morphemes. We suspect that similar considerations are at play in the apparent discrepancy between Gussenhoven’s description of the impossibility of inserting an AP boundary in order to place focal prominence on the second element of *mukashi-ba 'nashi* in (18b) and Kawakami’s description of the focalizing AP-initial rise at the beginning of the second part of *ho 'n-BAKO* in (18d).

At the same time, it is important to point out that Gussenhoven’s identification of a common source of difficulty for expressing the narrow focus in (18a) and (18b) assumes that focus-marking on the first element of a compound should be the mirror image of focus-marking on the second element. There is no logical necessity to this assumption. As we noted in Section 3, there are complementary constraints on the availability of different focus-marking mechanisms at different positions in an utterance. Prosodic restructuring to begin the focus constituent with an unexpected

IP boundary and reset is not a possible way to express narrow focus on an utterance-initial word, whereas post-focal dephrasing is not a possible way to express narrow focus on an utterance-final word. Analogously, the production of unexpected tonal cues to an AP boundary that Kawakami described for (18d-e) should not be able to express the early narrow focus of (18a) as distinct from the whole-word focus of (18c), and if there is an analog to post-focal dephrasing, we might expect it be constrained to occur on non-final elements of a compound word. The utterance in (20b) suggests a possible analog to post-focal dephrasing within a compound word form, so long as it is accented and shows the late accent placement that is a typical output of the compound accentuation rules.

- (20) F0 contours for two utterances with early narrow focus. The utterance in (a) is in response to *Are those Suzuki's glasses?* The utterance in (b) is from one of the recipe narratives in Venditti (2002). The context describes why Emmenthal cheese (as opposed to other types of cheeses) has qualities that make it perfect for use in fondue.



We mentioned in Section 2.2 that the $[H^*]$ target of the accent is inherently scaled higher in the local pitch range for the AP than is the phrasal $[H-]$. In Section 4.1, we discussed the implications of this difference for the interpretation of rises in the middle of sequences of accented words where the second accent is early in the word. Here we discuss another implication of this relationship of $[H-] < [H^*]$. When an AP groups together an unaccented word with a following accented word, this difference in inherent tone-scaling often results in an upward-sloping interpolation from the initial $[H-]$ to the following accent $[H^*]$. This is the pattern over the initial AP *Akai ya'ne* 'red roof' in all three panels of (6) and in the first two words *Omiyage-no chi'izu* 'souvenir cheese' in (2a). The productions of *mukashi-ba'nashi* that Gussenhoven shows for the three focus patterns in (18a-c) also have this $[H-] < [H^*]$ tone scaling relationship, and he follows most earlier AM

accounts in interpreting this as evidence that focal prominence on a word can only be marked by changing the pitch range specification for the AP containing the word as a whole, without affecting the inherent tone-scaling relationship of [H-] < [H*] within the phrase. As Pierrehumbert and Beckman (1988:108) put it, “focus is a property of the AP as a whole, and is therefore reflected in the realization of the accent H as well as the phrasal H.”

Example (11) already shows that this may be an over-simplification. The [H-] target on the focused *mannaka-ni* is very much higher than the [H*] target at the following accent in *okima'su*. Examples (12a) and (20a) show that this unexpected relationship of [H-] >> [H*] can be produced to mark narrow focus on the first element in an AP even when there is no silent pause interrupting the interpolation from the [H-] to the [H*], as was the case in (11). Example (20b) shows that this unexpected relationship of [H-] >> [H*] can even be produced in a compound word. That is, since the sequence *emmentaaru-chi'izu* is marked as a morphologically conjoined sequence by the compound accentuation pattern (*emmenta'aru* being underlyingly accented), the downsloping pattern in (20b) is comparable to the patterns that Kawakami observed in (18d-e) — an instance where focal prominence is marked on just one of the elements in a compound, by reversing the expected relationship between tones and the prosodic structures that they mark. Examples such as (18d-e) and (20b) thus exemplify another way in which the prosodic parse can be ambiguous; the segmental cues and compound accentuation pattern tell the listener that these forms are very closely conjoined morphologically, but the post-lexical AP-boundary tones in (18d-e) and the independent scaling of the two [H] targets that allows for the [H-] >> [H*] relationship in (20b) tell the listener that the two sub-elements of the compound form are independent enough that the second or the first of them can be marked separately as the focus constituent.

4.4 Prominence-lending edge markers

A final point of potential ambiguity that we will briefly address here concerns the phonological analysis of prominence-lending pitch movements. At first glance, the rise-fall F0 movements located on or near the particle *mo* for the [L%H%] sequence in (13b), the PNL in (13d), and the [L%HL%] in (13e) are all strikingly similar in shape to the [%L H*+L] sequence for the AP-initial rise and lexical accent on the first syllable of *Na'oya* in (13a). This surface similarity has prompted some researchers to wonder whether the movements in (13b-e) are in fact pitch accents (as opposed to boundary tones), and that is, in fact, how some researchers model them. Fujisaki and Hirose (1993), for example, generate all of these rises (including the more scooped question rise) by inserting an extra accent command. Oshima (2005) similarly argues that all particles are underlyingly accented, not just bimoraic ones such as *ma'de* and *su'ra*. This allows him to analyze the rise-fall shape on the particle *mo* in (13b-e) as the resurfacing of the accent tones under narrow focus. That is, Oshima would interpret the PLR and other prominence-lending BPM in these examples as an exaggeration of an underlying [H*+L] fall on *MO*, marking the particle as the focus constituent.

The X-JToBI analysis that we adopt is different. Although the contour on *mo* in (13b-e) does superficially resemble the shape around the lexical accent on the first syllable of *Na'oya* in (3a), we do not consider these BPM to be the realization of an underlying accent on the mono-moraic particle *mo*. This X-JToBI analysis is supported by Mesbur's (2005) experimental observation that multi-moraic particles that have lexically specified accent on some non-final mora (e.g. *gu'rai* ‘about, as much as’) can be uttered with *both* the rise-fall of the lexical accent *and* a final [H%] PLR under focus, as in (21).

- (21) *Omame gu'rai[†] eiyoo no a'ru mono wa na'i n da kara...*
'(You'd better eat them) since there's nothing with as much nutrition as beans.'

That is, examples such as (13a) show that prominence can be marked on a lexically accented word simply by expanding the local pitch range so as to increase the [H*] target. If it were the case that one simply expands the pitch range of an underlying accent to make the particle prominent, as in Oshima's (2005) account, then we would expect speakers to do the same for *GU'RAI*. The fact that speakers can preserve the lexical accent in *gu'rai* while putting an additional [H⁰] rise at the end of the phrase suggests that the PLR is a phenomenon separate from the lexical accent.

At the same time, we can see clearly why Oshima (2005) and Fujisaki and Hirose (1993) were attracted to this account of prominence-lending BPM as accents instead of as boundary tones. And an analysis that equates the prominence-lending properties of these BPM with focalizing mechanisms that expand the accent fall in (13a) is especially attractive for the BPM that X-JToBI calls PNLP ('penultimate non-lexical prominence'). The rise-fall movement of PNLP is transcribed as a [HL⁰] boundary tone, even though it is not realized at the phrase edge, as is the canonical 'explanatory' [HL⁰] (with which it shares its X-JToBI label) and the [H⁰] PLR (with which it is allied in alignment and duration characteristics). Thus, PNLP is inherently ambiguous. It looks and acts like the [H⁰] boundary tone of the PLR, but it violates our expectation that the phonetic realization of a boundary-marking tonal morpheme should be localized at the edge to which it is phonologically affiliated.

However, other languages have similarly ambiguous tones that can be analyzed as boundary-marking events despite their dislocation from the phrase edge. Compare, for example, the behavior of English phrase accents, which are associated with the phrase edge but spread leftward to fill up the space between the nuclear accent and the phrase edge. (The [L-] tones labeled in each panel of (10) are an example of such phrase accents.) Grice et al. (2002) review a number of languages that have prominence-marking tones with such an ambiguous or dual affiliation, including Greek and Hungarian. The prominence-lending function of the [H⁰] and [HL⁰] in (13b-d) already suggests an affinity between these boundary tones and the phrase accents of English, Greek, Hungarian, etc. This prominence-marking function makes them susceptible to a diachronic reanalysis like the one that may have given rise to the possibility of post-lexical insertion of accents in Basque. The PNLP shares with the phrase accents of English, Greek, Hungarian, etc., an even more inherently ambiguous parse because of its displacement from the edge to which it is phonologically associated. It will be interesting to see whether this displacement is the harbinger of a sound change that will make Tokyo Japanese more similar to Basque.

5 Theoretical implications

We began this chapter by reviewing aspects of the Japanese intonation system that are especially relevant for understanding the interplay between information structure and prosodic organization. We then reviewed the literature that describes the ways in which prosody is used by the language to mark focal prominence on constituents in running speech. The literature that we reviewed shows that there is a rich variety of prominence-marking mechanisms even when morphosyntactic mechanisms such as scrambling are ignored.

In reviewing this literature, we have tried to highlight points of consensus that emerge from comparing how a particular phenomenon or particular set of phenomena is treated in models that make different assumptions about such aspects of the intonation system as whether the [H-] tone at the beginning of an accentual phrase is an independent target from the [H*] tone at the accent fall.

When different models make analogous generalizations about older observations leading to similar predictions about future observations, this makes us confident that the generalizations are robust.

One generalization that emerges from comparing treatments across frameworks in this way is that pitch range specification plays an integral role in the intonational phonology. That is, as we saw in Section 4.2, it is possible to disagree about whether the reset that marks the beginning of the focus constituent in examples such as (16) involves the insertion of an unexpected IP boundary or a boost to the pitch range after a mere AP boundary. But there is no disagreement that speakers have a flexibly continuous control of the pitch range on either side of this boundary, and that they can use this control to produce a continuum of degrees of reset to indicate different relationships of foregrounding the focus constituent against the background of reduced pitch range on the preceding constituent. Moreover, while the exact details of the analysis differ across different frameworks, there is strong agreement that this continuous control of pitch range interacts with the control of categorical contrasts in lexical tone specification to produce superficially very different patterns, as shown in (12). Nonetheless, listeners can parse this variation in terms of their expectations about how words and syntactic phrases will be reflected in the prosodic organization of an utterance in order to recover the intended focus pattern.

A closely related (and overarching) generalization concerns the importance of prosodic grouping and the role of constituent edges in the marking of focal prominence. The exaggerated reset in utterances such as (11) targets the beginning of the focus constituent, as does the unexpected AP boundary in examples such as (17b), whereas the prominence-lending boundary pitch movements discussed in Section 3.4 target the end of the focus constituent. This focus on edges contrasts to the importance of head positions in the English intonation system. That is, prominence-marking in Japanese, unlike prominence-marking in English, does not target any such position as the syllable with primary stress in the focused word or the “designated terminal element” of the intonation phrase.

Despite this profound formal difference between the two languages, however, there is still a similarity concerning the associated types of phonetic properties that mark some constituents as more prominent and others as less prominent. For example, just as pitch range is manipulated to produce a more extreme rise to mark the beginning of a focused word in Japanese, pitch range is expanded to produce a more extreme rise or more extreme fall around the culminative pitch accent associated to the syllable that is the obligatory head of a focused word in English. If we understand the phonology-phonetics interface as being a counterpart to the syntax-semantics interface at the other end of the grammar, then this speaks to a deeper similarity in the ways in which the grammatical patterns are interpreted. However, this deeper “semantic” similarity can only be described accurately once the analysis moves beyond the fixation on more superficial similarities that lead to the cross-language misparsings described in Section 3.1. That is, one has to first analyze each language on its terms (and not just look at surface similarities) before one can see how the two languages are fundamentally the same.

Moreover, once each prosodic system is analyzed fully in its own terms, another related deeper similarity emerges. In each language, when there is no narrow focus within an utterance, some of the prominence markers are associated with constituents near the edge of the utterance. Looking at English first, we can say that the generalization about association to the obligatory head means that there must be a nuclear pitch accent somewhere in every well-formed IP. When there is no narrow focus, the nuclear pitch accent goes on the last content word (by the Nuclear Stress Rule). In Japanese, similarly, there has to be at least one IP-initial rise at the beginning of every well-formed utterance, and when there is no narrower focus prompting an IP break and reset later on, the rise

from the utterance initial [%L] makes the immediately next [H] target (whether a phrasal [H-] or the [H] of a [H*+L]) the highest (most prominent) peak in the utterance. Thus, it is difficult to distinguish broad focus on the utterance as a whole from narrow focus on the relevant edge constituent in both languages, even though the edge is different — namely, last in English, first in Japanese.

This generalization across the two languages predicts that there should also be complementary patterns of *focus projection* within the VP in transitive clauses. That is, since objects follow verbs in English, placing the nuclear pitch accent on the object NP should be ambiguous between narrow focus on the NP and broader focus on the VP, as indeed it is, as was shown experimentally by Gussenhoven (1983) and others. Conversely, since objects come before verbs in Japanese, producing a prominent reset before the object and dephrasing the following verb should be ambiguous between narrow focus on the NP and broader focus on the VP, as indeed it is, as was shown experimentally by Ito (2002a, 2002b).

Comparing across languages in this way also highlights important lacunae in our understanding of one or the other prosodic systems. For example, Ito (2002a, 2002b) tested her predictions about focus projection only with sequences of lexically accented object followed by lexically accented verb. Much previous work on English focus projection, comparably, had not examined the effects of varying aspects of the tune other than the location of the nuclear pitch accent, but more recent work has shown, for example, that the broad focus reading of late nuclear accent placement is less readily available when the pitch accent on the object is [L+H*] rather than [H*] (e.g., Welby 2002). Are there comparable interactions in Japanese, such that focus projection is easier or more difficult in other types of sequences besides accented object followed by accented verb?

Comparing across models also highlights other substantial gaps in our knowledge. For example, one lesson to draw from the discrepancy between Kubozono's results and Pierrehumbert and Beckman's results described in Section 4.2, is that we need more work on the prosody-pragmatics interface, work that applies more different methods, drawn from other fields such as sentence processing.

Another, equally important lesson that we glean from this comparison across frameworks is the reminder that phonological relationships such as the contrast between an IP boundary and a mere AP boundary are grammatical abstractions that native speakers acquire in the course of extensive exposure to a rich variety of language-specific cues. Some of these cues come from parsing the F0 contour in terms of tone targets and pitch range specifications. Others come from parsing the spectral patterns in terms of the dynamic flow of consonant constrictions and vowel resonance patterns. A frequently encountered congruence of cues from multiple sources in the signal induces a stronger abstraction, allowing adult native speakers to produce conflicting cues when necessary to convey the intended information structure, as in the *mannaka-ni okima'su* sequence in (11), where the F0 contour suggests that the adverbial phrase and the verb are grouped together into one AP, but the long pause suggests that an IP boundary separates them. The native listener, conversely, can accommodate to conflicting cues to recover both the intended focus pattern and the syntactic grouping that the prosodic phrasing also cues. In a similar way, although *rendaku* and the compound accent pattern of *ho'n-dana* typically will occur on sequences of morphemes that are grouped together into a single AP, the speaker can produce the tonal cues to an AP boundary to convey narrow focus on the *tana*, and can do so without “undoing” the other prosodic cues to the morphological grouping. Moreover, the speaker can produce these conflicting cues in full expectation that the listener will recover both the intended information structure and the intended morphological grouping. The disagreements described in Sections 4.1 and 4.2 come about because

the same kind of statistical dependencies drive the various interactions between the preferred focus-marking mechanism and lexical accent pattern, making for inherent ambiguities. What this lesson means for the theoretical linguist, then, is that the next generation of models of intonational phonology needs to do better justice to these complex interactions among discrete contrasts (e.g., between accented and unaccented words) and continuous variation (e.g., more versus less extreme degrees of pitch range expansion or compression at the edges of focus constituents), and the ways that native speakers and listeners take advantage of the statistical dependencies among different patterns.

References

- Aoyagi, Seizô. 1969. A demarcative pitch of some prefix-stem sequences in Japanese. *Onsei no Kenkyuu* 14: 241-247.
- Arisaka, Hideyo. 1941. Akusento no kata no honshitsu ni tsuite. *Gengo Kenkyuu* 7: 83-92.
- Arvaniti, Amalia, D. Robert Ladd, and Ineke Mennen. 1998. Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics* 26: 3-25.
- Beckman, Mary E. 1996a. When is a syllable not a syllable? In *Phonological structure and language processing: Cross-linguistic studies*, ed. Takashi Otake and Anne Cutler, 95-123. Berlin: Mouton de Gruyter.
- Beckman, Mary E. 1996b. The parsing of prosody. *Language and Cognitive Processes* 11: 17-67.
- Beckman, Mary E., and Atsuko Shoji. 1984. Spectral and perceptual evidence for CV coarticulation in devoiced /si/ and /syu/ in Japanese. *Phonetica* 41: 61-71.
- Beckman, Mary E., and Janet Pierrehumbert. 1992. Comments on chapters 13 and 14. [original title: Strategies and tactics for thinking about F0 variation]. In *Papers in Laboratory Phonology II: Segment, Gesture, Prosody*, ed. Gerard J. Docherty and D. Robert Ladd, 387-397. Cambridge, UK: Cambridge University Press.
- Beckman, Mary E., Julia Hirschberg and Stefanie Shattuck-Hufnagel. 2005. The original ToBI system and the evolution of the ToBI framework. In *Prosodic typology: The phonology of intonation and phrasing*, ed. Sun-Ah Jun, 9-54. New York: Oxford University Press.
- Bolinger, Dwight. 1951. Intonation: levels versus configurations. *Word* 7: 199-210.
- Bruce, Gösta. 1977. *Swedish word accents in sentence perspective*. Lund: Gleerup.
- de Jong, Kenneth. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America* 97: 491-504.
- D'Imperio, Mariapaola. 2000. *Perceptual considerations on tone targets and their alignment*. Doctoral dissertation, Ohio State University, Columbus, OH.
- Docherty, Gerard J. 1992. *The timing of voicing in British English obstruents*. Dordrecht: Foris Publications.
- Firth, J. R. (John Rupert). 1957. *Papers in linguistics, 1934-1951*. London: Oxford University Press.
- Fougeron, Cécile, and Patricia A. Keating. 1997. Articulatory strengthening at the edges of prosodic domains. *Journal of the Acoustical Society of America* 101: 3728-3740
- Fujisaki, Hiroya, and H. Sudo. 1971. Synthesis by rule of prosodic features of connected Japanese. In *Proceedings of the 7th International Congress on Acoustics (ICA)*, 133-136. Budapest.
- Fujisaki, Hiroya, and Keikichi Hirose. 1984. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustical Society of Japan (English)* 5(4): 233-242.
- Fujisaki, Hiroya, and Keikichi Hirose. 1993. Analysis and perception of intonation expressing paralinguistic information in spoken Japanese. *Lund Linguistics Working Papers* 41: 254-257.

- Fujisaki, Hiroya, Sumio Ohno, Masafumi Osame, Mayumi Sakata and Keikichi Hirose. 1994. Prosodic characteristics of a spoken dialogue for information query. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 1103-1106. Yokohama, Japan.
- Gandour, Jack, S. Potisuk and S. Dechongkit. 1994. Tonal coarticulation in Thai. *Journal of Phonetics* 22: 477-492.
- Goldsmith, John A. 1979. *Autosegmental phonology*. New York: Garland.
- Grice, Martin, D. Robert Ladd, and Amalia Arvaniti. 2000. On the place of phrase accents in intonational phonology. *Phonology*, 17: 143-85.
- Gussenhoven, Carlos. 1983. Testing the reality of focus domains. *Language and Speech*, 26: 61-80.
- Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.
- Haraguchi, Shōsuke. 1977. *The tone pattern of Japanese: An autosegmental theory of tonology*. Tokyo: Kaitakusha.
- Hattori, Shiro. 1961. Prosodeme, syllable structure and laryngeal phonemes. *Bulletin of the Summer Institute of Linguistics: Studies in Descriptive and Applied Linguistics* 1: 1-27. Tokyo: International Christian University.
- Hay, Jennifer. 2003. *Causes and consequences of word structure*. New York and London: Routledge.
- Heycock, Caroline. 2007. Japanese *-wa*, *-ga*, and information structure. This volume.
- Hirai, Toshio, Norio Higuchi and Yoshinori Sagisaka. 1997. Comparison of F0 control rules derived from multiple speech databases. In *Computing prosody: Computational models for processing spontaneous speech*, ed. Yoshinori Sagisaka, Nick Campbell and Norio Higuchi, 211-223. Berlin: Springer-Verlag.
- Hirschberg, Julia, and Gregory Ward. 1992. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics* 20: 241-251.
- Hirschberg, Julia, and Gregory Ward. 1995. The interpretation of the high-rise question contour in English. *Journal of Pragmatics* 24: 407-412.
- Hualde, José Ignacio. 1991. *Basque phonology*. London: Routledge.
- Hualde, José Ignacio, Gorka Elordieta, Iñaki Gaminde and Rajka Smiljanić. 2002. From pitch-accent to stress-accent in Basque. In *Laboratory phonology 7*, ed. Carlos Gussenhoven and Natasha Warner, 547-584. Berlin: Mouton de Gruyter.
- Hyman, Larry. 2001. Tone systems. In *Language typology and language universals: An international handbook (vol. 2)*, ed. Martin Haspelmath, Ekkehard König, Wulf Oesterreicher and Wolfgang Raible, 1367-1380. Berlin and New York: Walter de Gruyter.
- Hyman, Larry. 2005. Word-prosodic typology. Paper presented at the “Between Stress and Tone” conference, Leiden, June 2005.
- Igarashi, Yosuke, Hideaki Kikuchi and Kikuo Maekawa. 2006. Inritsu jooohoo [Prosodic information]. In *Nihongo hanashi kotoba koopasu no koochikuhooh [Construction of The Corpus of Spontaneous Japanese]*, Kokuritsu Kokugo Kenkyuujo [National Institute for Japanese Language (NIJL, formerly NRI)] Report 124, 347-453.
- Ito, Kiwako. 2002a. *The interaction of focus and lexical pitch accent in speech production and dialogue comprehension: Evidence from Japanese and Basque*. Doctoral dissertation, University of Illinois at Urbana-Champaign.

- Ito, Kiwako. 2002b. Ambiguity in broad focus and narrow focus interpretation in Japanese. In *Proceedings of Speech Prosody*, 411-414. Aix-en-Provence.
- Ito, Kiwako, and Shari R. Speer. To appear. Using interactive tasks to elicit natural dialogue. In *Methods in empirical prosody research*, ed. P. Augurzky and D. Lenertova. Mouton de Gruyter.
- Ito, Kiwako, Shari R. Speer and Mary Beckman. 2003. The influence of given-new status and lexical accent on intonation in Japanese spontaneous speech. Poster presented at the 16th Annual CUNY Conference on Human Sentence Processing, 34. Boston, Massachusetts, March 2003.
- Jun, Sun-Ah. 1998. The accentual phrase in the Korean prosodic hierarchy. *Phonology* 15: 189-226.
- Jun, Sun-Ah. 2005. *Prosodic typology: The phonology of intonation and phrasing*. New York: Oxford University Press.
- Katagiri, Yasuhiro. 1999. Dialogue functions of Japanese sentence-final particles. In *Proceedings of LP'98 (Linguistics and Phonetics '98)*, ed. Osamu Fujimura, Brian D. Joseph, and Bohumil Palek, 77-90. Charles University in Prague: The Karolinum Press.
- Kawakami, Shin. 1963. Bunmatsu nado no jooshoochoo ni tsuite [On phrase-final rises]. Reprinted in 1995 in *Nihongo akusento ronshuu [A collection of papers on Japanese accent]*, 274-298. Tokyo: Kyûko Shoin.
- Kawakami, Shin. 1957. Tookyoogo no takuritsu kyoochoo no onchoo. [Tonal prominence in Tokyo Japanese]. Reprinted in 1995 in *Nihongo akusento ronshuu [A collection of papers on Japanese accent]*, 76-91. Tokyo: Kyûko Shoin.
- Kindaichi, Haruhiko. 1951. Kotoba no senritsu. *Kokugogaku* 5(2): 37-59.
- Kiriyama, Shinya, Keikichi Hirose and Nobuaki Minematsu. 2002. Control of prosodic focuses for reply speech generation in a spoken dialogue system for information retrieval on academic documents. *Proceedings of Speech Prosody*, pp. 431-434.
- Kokuritsu Kokugo Kenkyuujo [National Institute for Japanese Language (NIJL, formerly NRLI)]. 2006. *Nihongo hanashi kotoba koopasu no koochikuhoo [Construction of The Corpus of Spontaneous Japanese]*. NIJL report No. 124. Tokyo. Available online at <http://www2.kokken.go.jp/~csj/public/>
- Kori, Shiro. 1987. The tonal behavior of Osaka Japanese: An interim report. *Ohio State Working Papers in Linguistics (OSUWPL): Papers from the Linguistics Laboratory* 36: 31-61.
- Kori, Shiro. 1989a. Fookasu jitsugen ni okeru onsei no tsuyosa, jizokujikan, F0 no eikyoo [Acoustic manifestation of focus in Tokyo Japanese: The role of intensity, duration and F0]. *Onsei Gengo* 3: 29-38.
- Kori, Shiro. 1989b. Kyooshoo to intoneeshon [Emphasis and intonation]. In *Nihongo no onsei/on'in (joo) [Japanese phonetics/phonology (1)]*, vol. 2 of *Koza Nihongo to Nihongo Kyooiku*, ed. Miyoko Sugito, 316-342. Tokyo: Meiji Shoin.
- Kori, Shiro. 1997. Nihongo no intoneeshon: kata to kinoo [Japanese intonation: Form and function]. In *Akusento, Intoneeshon, rizumu to poozu [Accent, Intonation, Rhythm, and Pause]*, vol. 2 of *Nihongo Onsei*, ed. Tetsuya Kunihiro, 169-202. Tokyo: Sanseido.
- Kubozono, Haruo. 1989. Syntactic and rhythmic effects on downstep in Japanese. *Phonology* 6: 39-67.
- Kubozono, Haruo. 1993. *The organization of Japanese prosody*. Tokyo: Kuroshio.
- Kubozono, Haruo. 2006. Focus and intonation in Japanese: Does focus trigger pitch reset? In *Potsdam Working Papers: Interdisciplinary Studies on Information Structure*, ed. Shinichiro Ishihara, Michaela Schmitz and Anne Schwarz. Potsdam, Germany.
- Kuno, Susumu. 1973. *The structure of the Japanese language*. Cambridge, Mass: MIT Press.

- Ladd, D. Robert. 1996. *Intonational phonology*. Cambridge, UK: Cambridge University Press.
- Langendoen, D. Terence. 1968. *The London school of linguistics; a study of the linguistic theories of B. Malinowski and J. R. Firth*. Cambridge, Mass: MIT Press.
- Laniran, Yetunde, and G. N. Clements. 2003. Downstep and High tone raising: Interacting factors in Yoruba tone production. *Journal of Phonetics* 31(2): 203-250.
- Lieberman, Mark, and Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8(2): 249-336.
- McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.
- McCawley, James D. 1970. Some tonal systems that come close to being pitch accent systems but don't quite make it. *Papers from the 6th Regional Meeting of the Chicago Linguistic Society (CLS)* 6, 526-532.
- McLemore, Cynthia Ann. 1991. *The pragmatic interpretation of English intonation: Sorority speech*. Doctoral dissertation, University of Texas, Austin.
- Maeda, Kazuaki, and Jennifer J. Venditti. 1998. Phonetic investigation of boundary pitch movements in Japanese. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 631-634. Sydney, Australia.
- Maekawa, Kikuo. 1991. Perception of intonational characteristics of WH and non-WH questions in Tokyo Japanese. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS)*, vol. 4/5, 202-205. Aix-en-Provence.
- Maekawa, Kikuo. 1994a. Intonational structure of Kumamoto Japanese: A perceptual validation. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 119-122. Yokohama, Japan.
- Maekawa, Kikuo. 1994b. Is there 'dephrasing' of the accentual phrase in Japanese? *Ohio State University Working Papers in Linguistics* 44: 146-165.
- Maekawa, Kikuo. 1997. Effects of focus on duration and vowel formant frequency in Japanese. In *Computing prosody: Computational models for processing spontaneous speech*, ed. Yoshinori Sagisaka, Nick Campbell and Norio Higuchi, 129-153.
- Maekawa, Kikuo. 1998. Phonetic and phonological characteristics of paralinguistic information in spoken Japanese. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 635-638. Sydney, Australia.
- Maekawa, Kikuo. 2003. Corpus of Spontaneous Japanese: Its design and evaluation. In *Proceedings of the ISCA and IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR2003)*, 7-12. Tokyo, Japan.
- Maekawa, Kikuo, and Yosuke Igarashi. 2006. 2-moora yuukaku joshi no inritsujoo no dokuritsusei: Nihongo hanashi kotoba koopasu no bunseki [Prosodic independence of bimoraic accented particles: Analysis of the Corpus of Spontaneous Japanese]. *Journal of the Phonetic Society of Japan* 10(2).
- Maekawa, Kikuo, Hideaki Kikuchi, Yosuke Igarashi and Jennifer J. Venditti. 2002. X-JToBI: An extended J_ToBI for spontaneous speech. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 1545-1548. Denver, Colorado.
- Maekawa, Kikuo, and Hideaki Kikuchi. 2005. Corpus-based analysis of vowel devoicing in spontaneous Japanese: An interim report. In *Voicing in Japanese*, ed. Jeroen van de Weijer, Kensuke Nanjo, and Tetsuo Nishihara, 205-228. Mouton de Gruyter.
- Maruyama, Takehiko, and Miki Taniguchi. 2002. Bun no shooten koozoo to kyokushoteki takuritsu [Focus structure of sentences and local prominence]. *Kansai Linguistics Society* 22: 18-28.

- Mesbur, James. 2005. Evidence for prominence-lending rise in focus phrases in Japanese. Unpublished manuscript, University of Pennsylvania.
- Miyaji, Yutaka. 1963. IV Intoneeshon [Intonation]. *Kokuritsu Kokugo Kenkyuujo [National Institute for Japanese Language (NIJL, formerly NRLI)] Report 23*, 178-208.
- Muranaka, Toshiko, and Noriyo Hara. 1994. Features of prominent particles in Japanese discourse. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 395-398. Yokohama, Japan.
- Nagahara, Hiroyuki. 1994. *Phonological phrasing in Japanese*. Doctoral dissertation, University of California at Los Angeles.
- Nagahara, Hiroyuki, and Shoichi Iwasaki. 1994. Tail pitch movement and the intermediate phrase in Japanese. Paper presented at the Linguistic Society of America (LSA) annual meeting, January 1994. Washington, DC.
- Nagao, Kyoko and Shieaki Amano. 2000. A role of fundamental frequencies in the perception of emphasized words. Conference presentation abstract. *Journal of the Acoustical Society of America* 108(5): 2465.
- Nakanishi, Kimiko. 2004. Prosody and scope interpretations of the topic marker *wa* in Japanese. In *Topic and focus: Cross-linguistic perspectives on intonation and meaning*, ed. Chung-Min Lee, Matthew Gordon and Daniel Büring. Dordrecht: Kluwer.
- Ohishi, Shotaro. 1959. Purominensu ni tsuite: Tookyoogo no kansatsu ni motozuku oboegaki [Prominence: Observation of the Tokyo Dialect]. *Kotoba no Kenkyuu (Papers of the Kokuritsu Kokugo Kenkyuujo [National Institute for Japanese Language (NIJL, formerly NRLI)])*, 87-102.
- Oshima, David Y. 2005. Boundary tones or prominent particles?: Variation in Japanese focus-marking contours. *Berkeley Linguistics Society (BLS)* 31.
- Peng, Shu-hui. 1997. Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics* 25: 371-400.
- Pierrehumbert, Janet B. 1980. *The phonology and phonetics of English intonation*. Doctoral dissertation, Massachusetts Institute of Technology.
- Pierrehumbert, Janet B., and Mary E. Beckman. 1988. *Japanese Tone Structure*. Cambridge, Mass: MIT Press.
- Pierrehumbert, Janet, and David Talkin. 1992. Lenition of /h/ and glottal stop. In *Papers in Laboratory Phonology II: Segment, Gesture, Prosody*, ed. Gerard J. Docherty and D. Robert Ladd, 90-116.
- Poser, William J. 1984. *The phonetics and phonology of tone and intonation in Japanese*. Doctoral dissertation, Massachusetts Institute of Technology.
- Prince, Ellen F. 1981. Toward a taxonomy of given/new information. In *Radical Pragmatics*, ed. Peter Cole, 223-255. New York: Academic Press.
- Raymond, William D., Robin Dautricourt, and Elizabeth Hume. 2006. Word-internal /t, d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change* 18: 55-97.
- Roberts, Craige. 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics. *OSU Working Papers in Linguistics* 49: 91-136.
- Sadanobu, Toshiyuki. 2004. A natural history of Japanese pressed voice. *Onsei kenkyuu* 8: 29-44.
- Selkirk, Elisabeth O., and Koichi Tateishi. 1991. Syntax and downstep in Japanese. In *Interdisciplinary Approaches to Language: Essays in Honor of S.-Y. Kuroda*, ed. Carol Georgopoulos and Roberta Ishihara, 519-543. Dordrecht: Kluwer.

- Sluifjter, Agaath M.C., and Vincent van Heuven. 1996. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100: 2471-2485.
- Sugahara, Mariko. 2002. Conditions on post-FOCUS dephrasing in Tokyo Japanese. In *Proceedings of Speech Prosody*, 655-658.
- Taniguchi, Miki, and Takehiko Maruyama. 2001. Shooten koozoo to zyosi no takuritu [Focus structure and prominence of particles]. *Kansai Linguistics Society* 21: 56-66.
- 't Hart, Johan, and René Collier. 1980. *Cursus Nederlandse Intonatie*. Louvain: Acco.
- Trubestkôï, Nikolâi Sergeevich. 1939. *Grundzüge der phonologie*. (*Travaux du Cercle linguistique de Prague No. 7.*) Prague: Cercle linguistique de Prague. [Translated 1969, by Christiane A. M. Baltaxe as *Principles of Phonology*. University of California Press.]
- Vallduví, Enric. 1992. *The informational component*. New York: Garland.
- Venditti, Jennifer J. 1997. Japanese ToBI Labelling Guidelines. *Ohio State University Working Papers in Linguistics* 50: 127-162. [First distributed in 1995.]
- Venditti, Jennifer J. 2000. *Discourse structure and attentional salience effects on Japanese intonation*. Doctoral dissertation, Ohio State University.
- Venditti, Jennifer J. 2005. The J_ToBI model of Japanese intonation. In *Prosodic typology: The phonology of intonation and phrasing*, ed. Sun-Ah Jun, 172-200. New York: Oxford University Press.
- Venditti, Jennifer J., and Marc Swerts. 1996. Intonational cues to discourse structure in Japanese. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 725-728. Philadelphia, Pennsylvania.
- Venditti, Jennifer J., Kazuaki Maeda and Jan P. H. van Santen. 1998. Modeling Japanese boundary pitch movements for speech synthesis. In *Proceedings of the Third ESCA workshop on speech synthesis*, 317-322. Jenolan Caves, Australia.
- Welby, Pauline. 2003. Effects of pitch accent position, type, and status on focus projection. *Language and Speech*, 46: 53-82.
- Welby, Pauline. 2003. *The slaying of Lady Mondagreen, being a study of French tonal association and alignment and their role in speech segmentation*. Doctoral dissertation, Ohio State University.
- Zwicky, Arnold. 1972. Note on a phonological hierarchy in English. In *Linguistic change and generative phonology*, ed. Robert Stockwell and Ronald Macauley, 275-301. Bloomington, IN: Indiana University Press.