

Jennifer J. Venditti

Introduction

The prosodic structure and accompanying intonational “melody” of an utterance can provide a rich source of information to the sentence-processing mechanism. In spoken language, speakers mark prominences and group words into prosodic constituents by varying the pitch, timing, and other aspects of the voice. The intonational contour resulting from these variations can provide listeners with cues to the syntactic structure, information structure, and a variety of other linguistic structures. Thus, a crucial part of our understanding of spoken-language processing is our understanding of prosody, and how this structure maps onto the linguistic structures that we parse.

This chapter introduces Standard (Tokyo) Japanese prosodic structure and intonation, and outlines one of the most recent descriptions of Japanese intonation, the Japanese ToBI system. Then, it summarizes the literature on the mapping between prosodic structure and syntactic structure including recent studies examining the role of prosody in Japanese sentence processing, even in silent reading.

Japanese prosodic structure and intonation

Japanese intonation has been described from a variety of theoretical perspectives, ranging from mathematical models employing a composite of superimposed curves (e.g. Fujisaki & Sudo, 1971; Venditti & van Santen, 2000) to phonological models employing abstract tonal targets or patterns (e.g. McCawley, 1968; Poser, 1984; Pierrehumbert & Beckman, 1988; Venditti, 1995).¹ Here we will focus on the phonological model of Japanese intonation described by the *Japanese ToBI* labeling scheme, which is currently the most widely used approach in psycholinguistics and phonology. This labeling scheme, originally

¹ The terms *prosody* and *intonation* are commonly used interchangeably in the literature on suprasegmental sound structure. Often, the term *prosody* will refer to the underlying prominences and constituent structure of speech, while *intonation* will refer to the realization of this structure by acoustic means, primarily by pitch variation (see Beckman, 1996 for details).

formulated as J-ToBI in Venditti (1995) and Venditti (2005), and more recently extended as X-JToBI by Maekawa et al. (2002), grew out of Beckman and Pierrehumbert’s phonological and experimental analysis of Japanese intonation (Beckman & Pierrehumbert, 1986; Pierrehumbert & Beckman, 1988). Here an overview of the core notions of this model is provided.

Accent

One major feature of Japanese intonation is the contrast between *accented* and *unaccented* words at the lexical level. In English, an *accent* refers to an intonational prominence associated with a metrically strong syllable or word at the sentence level. The location of such pitch accents in English is determined not only by structural considerations, but also by the pragmatics and communicative intent. For example, a speaker may choose to place a focal accent on *PAT* in *She made PAT do it* to mark an opposition between *Pat* and other salient people in the discourse. In contrast to this discourse-based function of accent, in Japanese accentuation is a property of the word itself, regardless of the context in which it is uttered. A word is either accented or unaccented, and this specification can be found in the dictionary. Figure 28.1 shows an example of this contrast.

The left panel shows the fundamental frequency (F0) contour (aka *pitch track* or *intonation contour*) of the accented phrase *ue'rumono* “those who are starved.” The pitch rises at the start of the phrase, then falls sharply near the end of the accented mora /e/ (marked by an apostrophe in the transcription). The right panel shows the unaccented phrase *uerumono* “something to plant,” which has a similar (albeit smaller) phrase-initial rise, but exhibits a gradual F0 decline toward the end of the phrase. The presence or absence of a sharp fall in F0 such as that shown in the left panel, denoted by the tone H*+L, is the hallmark of the accented vs. unaccented distinction in Japanese: accented words display the sharp fall near the end of the accented mora, while unaccented words lack such a fall. The source of the initial rise and gradual fall displayed in the unaccented case is due to tones associated not with the accent but with the entire phrase, as we will describe next.

Prosodic phrasing

The other main feature of Japanese intonation is prosodic phrasing. When a speaker utters a string of words in connected discourse, he/she groups them into constituents in the sound structure. These units, termed *prosodic phrases* or *intonation phrases*, are demarcated by delimitative tones, variation in segmental durations and pitch range, and optional pauses. The X-JToBI model describes two levels of prosodic phrasing: the *accentual phrase* and the *intonation phrase*, each of which is marked by different types of tones. Words are grouped together

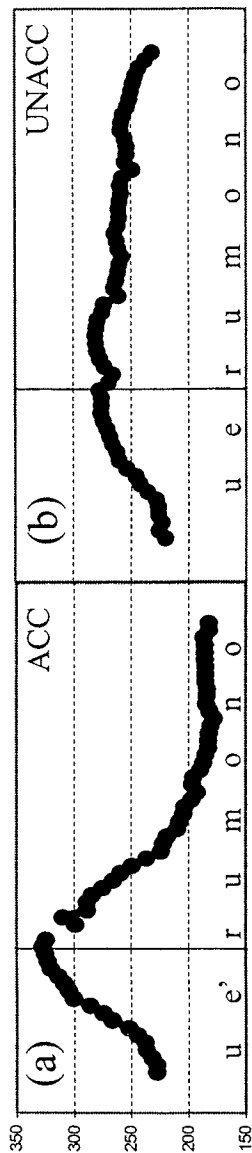


Figure 28.1 Waveform and F0 contours of the accented phrase *ue'rumono* "those who are starved" (left panel) versus the unaccented phrase *urumono* "something to plant" (right panel). The x-axis plots time, and the y-axis plots F0 (which is linearly related to perceived pitch) in hertz. Vertical lines mark the onset of the flap /r/. From Venditti (2005: figure 7.1).

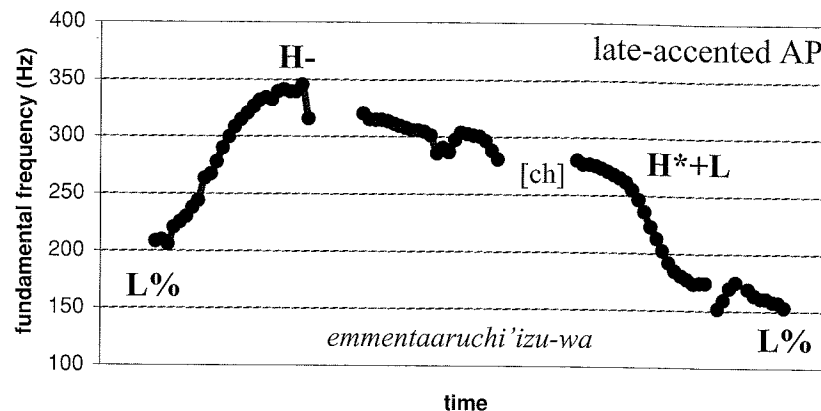


Figure 28.2 F0 contour of the AP *emmentaaruichi'izu-wa* "Emmenthal cheese-TOP." Gaps and fine jitter in the contour are due to segmental perturbations (e.g. from voiceless segments), and are not relevant in assessing the overall shape of the curve.

to form accentual phrases, which are in turn grouped together to form higher-level intonation phrases.

The accentual phrase

The *accentual phrase* (AP) is a grouping of words delimited by three tones: a low *boundary tone* at the start of the phrase (labeled L% in X-JToBI), a high *phrase tone* near the second mora of the phrase (labeled H-), and another L% boundary tone at the end of the phrase. In the X-JToBI approach, these three tones act as tonal targets, and the observed intonation contour is a result of linear interpolation between these targets. In figure 28.1 (right panel), since the phrase *urumono* is unaccented, the H*+L accent tone is absent. The only tones that are present here are the phrasal and boundary tones of the AP; the observed intonation contour is a result of the L% H- L% tonal sequence. In figure 28.1 (left panel), the sharp F0 fall is due to the H*+L falling pitch accent on the accented mora /e/. The rise at the start of the phrase is due to interpolation between the AP-initial L% boundary tone and this accent, while the low plateau near the end of the phrase is due to interpolation between the L of the accent and the L% AP-final boundary tone. The H- phrase tone (which normally occurs near the second mora of the AP) is obscured by the H*+L accent occurring on this same mora. When the lexical accent is late in the phrase, the H- and H*(+L) tones are distinct, as shown in figure 28.2.

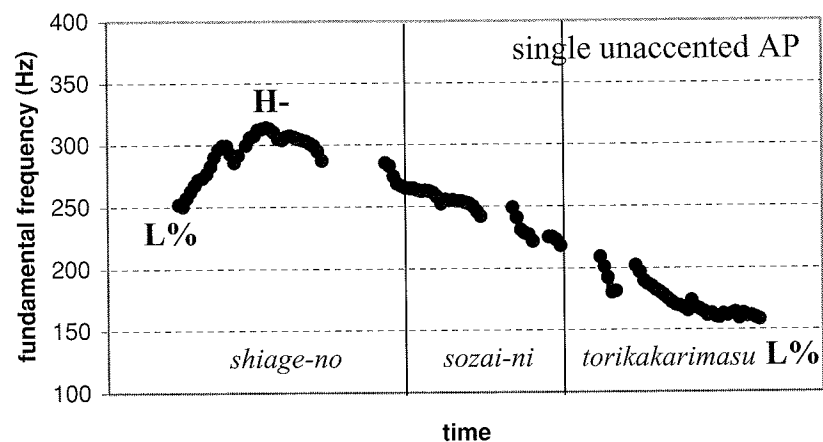


Figure 28.3 F0 contour of the sentence *shiage-no sozai-ni torikakarimasu* “Now we start on the final ingredient,” in which all words have been dephrased into a single AP. Vertical lines mark *bunsetsu* boundaries.

In continuous speech, a sequence of words may combine to form a single AP marked by the L% H- L% delimitative tonal pattern. Figure 28.3 shows an entire sentence forming a single AP. In general, unaccented words tend to phrase together with adjacent unaccented (U) or accented (A) words, while sequences of accented words tend to form their own APs (see example contours of UU, UA, AU, and AA combinations in Venditti, 1994; Beckman, 1996).

The intonation phrase

The *intonation phrase* (IP) is the next and highest level up in the prosodic hierarchy of Japanese, and represents a grouping of accentual phrases. An IP can have an optional *boundary pitch movement* (BPM) at its right edge: a sharp short rise (H%), a scooped extended rise (LH%), or a small “hump” (HL%), each having its own pragmatic meaning/function. In addition to carrying these optional boundary tones, the IP is the phonological domain within which pitch range is specified and *downstep* applies.

Downstep is the phonological process whereby the pitch range of an AP (be it accented or unaccented) is compressed immediately following an accented AP. Only accented APs “trigger” downstep; unaccented APs do not. Figure 28.4 shows schematic contours illustrating this phonological phenomenon.

Each trace in figure 28.4 consists of two sequential APs. In the AA condition the first AP is accented, and thus downstepping results in a compression of pitch range on the second AP, *relative to* the range of the same phrase in the

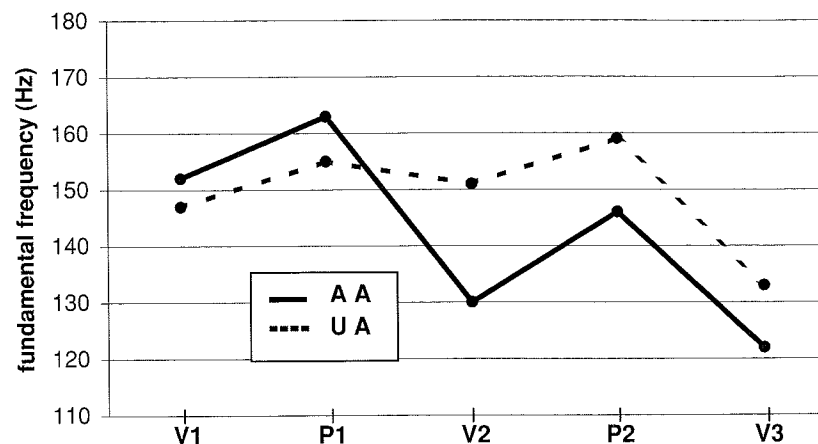


Figure 28.4 Schematic F0 contours depicting the application (solid) versus non-application (dotted) of downstep. Modeled after Kubozono (1993: figure 15.2).

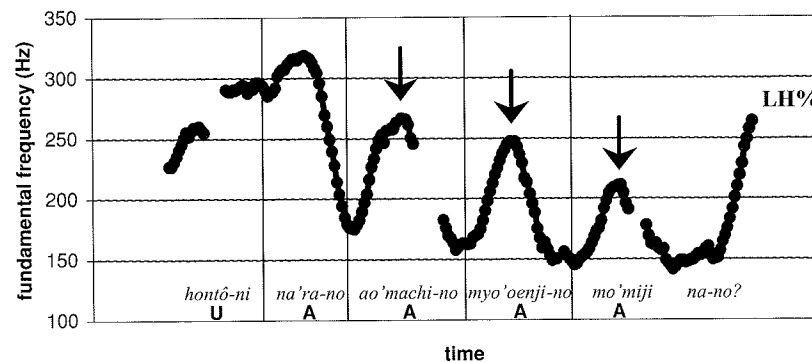


Figure 28.5 F0 contour showing sequential downstep within a single IP: “Are those really the autumn leaves from the Myōenji temple in the Aomachi district of Nara?” Vertical lines mark AP boundaries.

baseline condition (i.e. the dotted UA contour).² As long as a trigger is present, downstep will apply sequentially within the bounds of an intonation phrase, often resulting in a staircase-like descending pattern of APs. Figure 28.5 shows an example of downstep chaining. (This IP also contains a LH% rising BPM at its right edge marking the interrogative.)

² The difference in F0 height of the peak on the first AP (P1) is due to inherent height differences of H- vs. H* (e.g. Pierrehumbert & Beckman, 1988).

It has been experimentally observed that downstep does not apply across an IP boundary. Instead, after the boundary the pitch range is “reset” to a level independent of the (initial) range of the previous IP. A variety of linguistic factors, including pragmatic focus, information status, discourse structure, etc. will determine the pitch range of each separate intonation phrase.

Prosodic cues to syntactic structure

In order for prosody to inform parsing, there must be some concrete and reliable mapping between the prosodic structure of the sound stream and the syntactic structure of the utterance being parsed. Determining the nature of this mapping has been the topic of much experimentation and theorizing. While there is general agreement that these structures do relate, the mapping is certainly not one-to-one, and is often quite illusive. Probably the two most studied structures in the Japanese prosody–syntax mapping literature are left-branching (LB) vs. right-branching (RB) noun phrases, and relative clause constructions.

Ambiguous NPs have been a favorite topic of study in the Japanese psycholinguistics literature. Selkirk and Tateishi (1991) examined the intonation of left-branching [[XY]Z] vs. right-branching [X[YZ]] structures, specifically with regard to downstepping patterns. They found that while downstep occurred between elements X and Y in LB structures, downstep was blocked in RB structures, suggesting that an IP (aka their *major phrase*) boundary intervened. To account for this, they propose a mapping parameter whereby a prosodic phrase boundary coincides with the left edge of a syntactic maximal projection (X^{\max}). Kubozono (1989b, 1993) conducted extensive experimental analyses on these types of constructions. He observed downstep in *both* LB and RB structures, contrary to Selkirk and Tateishi’s findings. In addition, Kubozono found that the peak F0 on Y was higher in RB than in LB structures. He explains this difference by positing a process of *metrical boost*, whereby the F0 is boosted on an element which lies at the left edge of a right-branching syntactic structure, independent of whether downstep has occurred or not. In a subsequent study, Venditti (1994) found that certain speakers may choose to place an IP break in this position in RB NPs (hence blocking downstep), while other speakers may choose to produce downstep with additional boost. However, regardless of the means a given speaker chooses, the goal is the same: right-branching structures are prosodically marked in Japanese by an increase in F0 (and possibly other acoustic features). Below we discuss the similar marking of relative clause boundaries as well.

Prosody in processing

Given that prosody can provide cues to the syntactic structure of an utterance, are listeners able to use these cues in online sentence comprehension to avoid

potential ambiguity and garden-paths? One commonly cited potential for temporary ambiguity in Japanese is the fact that a clause such as *Ma’ri-ga yo’nda* “Mari read” can be a simplex sentence by itself, or can also be a subordinate clause of a complex NP, as in *Ma’ri-ga yo’nda ho’n-wa . . .* “the book that Mari read . . .”. From text alone, this ambiguity is resolved when the head noun is encountered. However, Venditti and Yamashita (1994b) showed that, in spoken language, listeners could resolve the ambiguity sooner, even before the verb had ended. Kondo and Mazuka (1996) replicated this finding in their eye–voice span data: in first-pass online reading, subjects uttered sentence-final verbs differently from subordinate verbs. They note that this was most likely because subjects could see if the sentence would end or not in their parafoveal vision.

Another place for potential confusion in online processing is parsing the argument structure of relative clauses (e.g. Mazuka & Itoh, 1995). For example, in reading text alone, the string in (1a) might on first-pass be interpreted as forming a simplex sentence. Once the head noun *child* is encountered (see 1b), the parser would then have to reanalyze to a RC construction. But even this reanalysis is still ambiguous, since the argument structure is unclear until the final verb. Here the dative NP *Nobu’yuki-ni* must be part of the matrix clause, hence a syntactic clause boundary occurs after this NP.

- (1) a. Ma’yuko-ga Nobu’yuki-ni me’ron-o na’geta . . .
 Mayuko-NOM Nobuyuki-DAT melon-ACC threw
 b. Ma’yuko-ga Nobu’yuki-ni [me’ron-o na’geta] e’nji-o azu’keta.
 Mayuko-NOM Nobuyuki-DAT melon-ACC threw child-ACC left in care
 “Mayuko left the child who threw the melon in Nobuyuki’s care.”

How might the processing be different for listeners hearing spoken language? Since listeners are able to distinguish simplex sentences from embedded clauses even before the head noun is encountered, this potential pitfall may be avoided. In addition, perception studies have shown that major syntactic boundaries such as RC boundaries are marked prosodically in Japanese, most often by an IP break and downstep reset. For instance, Uyeno et al. (1980) examined the perception of globally ambiguous RCs as in (2) by manipulating the F0 height of element X, keeping all other acoustic features such as pausing, duration, amplitude, etc. the same.

- (2) W X Y Z
 a’rutoki ni’geta e’nji-ga ka’ita
 at one time ran away child drew

They found that differences in the intonation contour determined RC interpretation: when the F0 of X was the same or higher than W, listeners

interpreted the structure as center-embedded [W[X]YZ] *The child who ran away drew (it) at one time*. Azuma and Tsukuma (1991) found that F0 was perceptually more important than pausing in determining interpretation in such ambiguous RCs. A production experiment by Venditti (1994) found no down-step between elements W and X in center-embedded structures, for all speakers, suggesting that an IP break does indeed occur at this syntactic boundary. In addition to such globally ambiguous RCs, Kondo and Mazuka (1996) and Hirose (1999) found that speakers can use IP breaks to mark the start of the relative clause in temporarily ambiguous structures as well. Thus, prosody can be a useful tool for avoiding potential ambiguity in spoken-language processing.

Might prosody also have a role in written-language processing, i.e. in reading? In Kondo and Mazuka's (1996) eye-tracking experiment, they found that in first-pass reading, speakers reading aloud construct an intonation contour that is based on a very local syntactic analysis; their calculated eye-voice span suggests that readers have access to lexical information only in the current *bunsetsu* plus the next one. If the disambiguating information is outside this span, Kondo and Mazuka suggest that speakers have no reason to produce a prosodic boundary (but they may insert one upon re-reading once the intended structure is clear). However, their data show that, if the disambiguating information is in the immediately following *bunsetsu*, readers *do* produce a prosodic phrase break. This suggests that prosody can be meaningful even in first-pass reading aloud.

Data from recent experiments by Hirose (1999, 2003) suggest that prosody may be even more useful to readers, even silent readers, than was previously thought. Hirose presents carefully controlled data showing that in first-pass reading, speakers reading aloud produce an IP break after the first two accented *bunsetsu* (i.e. APs), without having any information about what the structure of the utterance will end up being. For example, this observation predicts that an IP break will occur after *Nobu'yuki-ni* in (1b), even in first-pass reading. This finding seems to contradict Kondo and Mazuka's (1996) claim that readers have no reason to place a break there, since they do not yet have access to disambiguating information. However, Kondo and Mazuka's data was not controlled for important prosodic factors such as accentuation and accentual phrasing, thus making it impossible to compare their data with Hirose's. However, Hirose's findings are consistent with Kubozono's (1989b, 1993) controlled experiments showing a phenomenon called *rhythmic boost*, whereby the third accentual phrase in a (4-*bunsetsu* long) left-branching structure exhibits an increased F0 (a type of rhythmic balancing). The effect of rhythmic boost is similar to that of metrical boost in right-branching structures, as described above. Hirose claims that such prosodic breaks produced in first-pass reading may in essence be "recycled" in subsequent reanalysis. Her proposal is motivated by a curious effect in her data, whereby a purely arbitrary phonological characteristic, namely the

accentuation and number of APs of a subject NP, affects the ease/difficulty of processing the RC in the reanalysis stage – this effect occurs *even* in silent reading. She proposes that the IP break that is generated by rhythmic principles in first-pass (silent) reading can then be "recycled" in later stages to encourage certain reanalysis preferences. In terms of our example (2b), an IP break would be generated after *Nobu'yuki-ni* in the first pass, and this would facilitate reanalysis of the RC structure in which there indeed should be a break occurring there. Hirose's proposal is based on the Implicit Prosody Hypothesis put forth by Fodor (1998) to describe other similar prosodic effects in silent reading.

Discussion

The goal of this chapter was to illustrate ways in which prosody has been shown to influence Japanese sentence processing. We began by providing details of the prosodic structure of Japanese – a thorough understanding of these fundamentals is essential to any work concerning the psycholinguistics of prosody. We then provided a brief overview of studies of the prosody-syntax mapping in Japanese. We also reviewed recent studies showing how prosody can be relevant, and indeed is often quite useful, in Japanese sentence processing, even in silent reading. However, there are still a number of open research questions. For example, it will be important to extend Hirose's work on implicit prosody by examining the processing of utterances with varied accentuation and phrasing patterns. In addition, much work is still needed to determine the contexts in which prosodic cues are reliable, versus those in which these cues may be optional. Finally, we need to step beyond syntactic structure to examine the ways in which prosody can affect the online processing of other linguistic structures as well, such as focus or discourse structures.