

# Focus and Intonation in Japanese: Does Focus Trigger Pitch Reset?\*

*Haruo Kubozono*

Kobe University

This paper discusses how focus changes prosodic structure in Tokyo Japanese. It is generally believed that focus blocks the intonational process of downstep and causes a pitch reset. This paper presents experimental evidence against this traditional view by looking at the prosodic behavior of Wh words, which receive focus lexically in Japanese as in other languages. It is demonstrated, specifically, that the focused Wh element does not block downstep although it receives a much higher pitch than its preceding element. This suggests that presence of lexical focus does not trigger pitch reset in Japanese.

*Keywords: Focus, Intonation, Japanese, Pitch Reset, Downstep*

## 1 Introduction

The relationship between prosody and information structure is attracting serious attention in the literature (e.g. Zubizarreta 1998). This topic has been discussed in the literature of Japanese prosody, too (see, for example, Truckenbrodt 1995 and Ishihara 2003), but there is not much work specifically dealing with the interface between intonation and focus. As far as I know, Poser (1984) presented the first experimental work that looked at this issue in terms of the prosodic

---

\* An earlier version of this paper was read at the Second Workshop on Prosody, Syntax and Information Structure (WPSI 2) held at Potsdam University in March 2005. I would like to thank the participants in this workshop for various valuable comments. I am also grateful to Donna Erickson for checking and commenting on this manuscript, to Shingo Sugiyama and Shinji Ogawa for their help in the experiment and to all those who participated in the experiment as subjects. All remaining errors are of course my own. The work reported in this paper has been supported by Grant-in-Aid for Scientific Research (A) (Japan Society for the Promotion of Science, Grant no. 17202010).

organization of utterances in (Tokyo) Japanese.<sup>1</sup> By carrying out an experiment in which a certain element bears contrastive focus (or emphasis, in his terminology), Poser reported that focus does not block the intonational process of catathesis (or downstep) and, hence, that focus does not introduce a major intonational phrase boundary. On the other hand, Pierrehumbert and Beckman (1988) (henceforth P&B) claimed that focus blocks catathesis/downstep and, hence, starts a new major intonational phrase which they called ‘intermediate phrase’.<sup>2</sup> Since this intonational phrase is defined as the domain of pitch reset, as we will see shortly below, P&B argued that focus resets the pitch range. Their view regarding the interaction between focus and intonation structure seems popularly accepted in the studies of Japanese prosody in general.

As in the general literature dealing with focus, both Poser and P&B mean contrastive or corrective focus by the term ‘focus’ and they are not different in this respect. However, they reached entirely different conclusions as regards the effect of focus on intonation structure. One thing that can be said with some certainty is that it is generally difficult to obtain sufficiently natural speech when we examine contrastive focus in a controlled experiment. When we use corrective or contrastive focus in the material, subjects tend to exaggerate the contrast and hence to put extra boost on the focused element. There are two ways to avoid this potential problem: one is to analyze spontaneous speech in a non-controlled experiment, while the other is to look at the kind of focus in a controlled experiment that does not involve contrast. We chose the second option in this study and decided to analyze sentences with a Wh element which

---

<sup>1</sup> In this paper, ‘Japanese’ refers to the standard Tokyo Japanese unless otherwise stated.

<sup>2</sup> “Focus blocked propagation of catathesis...focus caused the introduction of a prosodic boundary, at which the pitch range was reset. We will call this level of phrasing the intermediate phrase” (P&B 1988: 19–20).

---

is known to receive focus lexically, independent of the pragmatics of the sentence.

Before going into the details of the experiment, let us define some basic notions and shared ideas about the intonation structure of Japanese (section 2). In section 3, we will review past works on the interface between focus and intonation by specifically comparing Poser's (1984) experimental results with P&B's (1988). In section 4, we will sketch our experiment and present its major results. This will be followed by a discussion of the data in section 5, where we consider the implications of the data for the modeling of Japanese intonation, especially as regards the hierarchical organization of prosody.

## 2 Background

We need to define four basic notions here: lexical accent, downstep, minor phrase and major phrase.

### 2.1 Lexical accent

'Lexical accent' is ambiguous in Japanese phonetics and phonology. First of all, lexical items in Tokyo Japanese fall into two types: 'accented' and 'unaccented'. 'Accented' words involve a sudden pitch fall at the phonetic output, whereas 'unaccented' words do not show any such fall even when they are followed by a particle (like the nominative particle *ga*). This distinction is generally preserved in phrases and sentences, where accented and unaccented words give rise to phrases with or without a sudden pitch fall, respectively. Some examples of the two lexical types are given in (1): accent is denoted by an apostrophe placed on the vowel immediately before the pitch fall, whereas unaccented words are transcribed with no such mark.

## (1) a. Accented words

kyóoto ‘Kyoto’, kóobe ‘Kobe’, nagásaki ‘Nagasaki’, náoko ‘Naoko’,  
dóitu ‘Germany’, pótudamu ‘Potsdam’, umái ‘tasty’, haréru ‘to clear’

## b. Unaccented words

tookyoo ‘Tokyo’, oosaka ‘Osaka’, hirosima ‘Hiroshima’, naomi ‘Naomi’,  
berurin ‘Berlin’, amerika ‘America’, amai ‘sweet’, hareru ‘to swell up’

The term ‘lexical accent’ sometimes refers to accent patterns specified at the word-level, including the unaccented one in (1b). This convention of referring to both accented and unaccented words as having lexical accent is popular in the traditional literature of phonetics and phonology in Japan (Akinaga 1985). Quite often, however, the term ‘lexical accent’ is used to refer to the sudden pitch fall observed in accented words; words in (1a) bear a lexical accent on the vowel marked by the accent mark, whereas those in (1b) lack such a phonological feature. In this paper, we will follow this second convention, thus characterizing (1b) as lexical items with no ‘lexical accent’.

The distinction between accented and unaccented words is supposed to be lexical in the sense that all morphologically simplex words are specified in the lexicon with respect to their accentedness. While this is largely true in the native vocabulary, it is not always the case in the foreign and Sino-Japanese vocabulary. Recent studies on Japanese accent have revealed that accentedness is predictable to a considerable extent in these two types of words, especially on the basis of their syllable and mora structure (Kubozono 1996, Kubozono & Fukui 1996, Ogawa 2004).

## 2.2 Downstep

The lexical distinction between accented and unaccented words exerts profound effects on Japanese intonation. In addition to the presence or absence of an

abrupt pitch fall, accented and unaccented words/phrases also differ in pitch height (see Figure 1 below). Namely, the former exhibits a higher pitch peak than the latter (Poser 1984); very often, it also starts with a higher pitch than the latter (Kubozono 1988/1993). The extra pitch boost in accented words, or ‘accentual boost’ (Kubozono 1988), can be interpreted as a concomitant feature of the sudden pitch fall that follows. In other words, accented words receive a higher pitch in order to display the following pitch fall more clearly.

Lexical accent exerts a considerable effect on the following material, too, by lowering the pitch level of the following phrase. This can be seen very clearly by comparing two sentences or expressions, one with an accented first phrase and the other with an unaccented one. In Figure 1, for example, the second phrase *nomímono* ‘drink’ is realized at a considerably lower level when it follows an accented phrase *umái* ‘tasty’ than when it follows an unaccented one *amai* ‘sweet’.

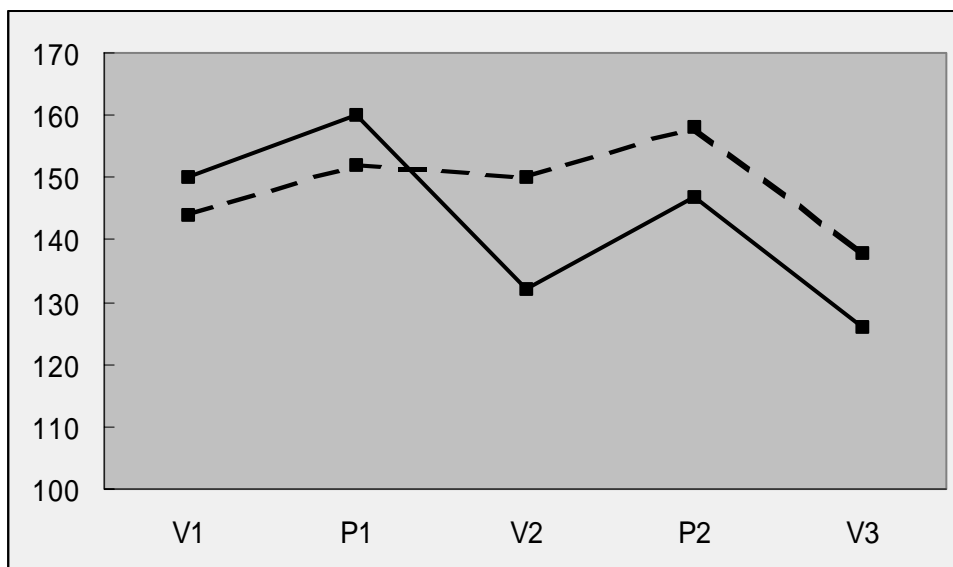


Figure 1 Schematic comparison of two contours: a sequence of two accented phrases (solid line) vs. a sequence of an unaccented phrase plus an accented one (dashed line).

Poser (1984) and P&B (1988) called this lowering process ‘catathesis’, for which Kubozono (1988), Selkirk and Tateishi (1991) and others used the term ‘downstep’. We will use the latter term throughout this paper.

Figure 1 illustrates the two effects of lexical accent on Japanese intonation. Accented words are realized at a higher pitch level than their unaccented counterparts on the one hand (*accentual boost*), but they realize the following material in a lower region on the other (*downstep*). Because of these effects, the two contours in Figure 1 show a clear separation when the pitch peak in the second phrase is plotted as against the pitch peak in the first phrase: the solid line has a greater pitch value than the dashed line for the first phrase but a smaller pitch value for the second phrase (P&B 1988).

In addition to these, it was also demonstrated that downstep occurs iteratively (Poser 1984, Kubozono 1988); the pitch level goes further down if a third phrase follows in Figure 1. In this sense, the downstep effect is cumulative.

Whether downstep has occurred between two adjacent phrases can be tested in two independent ways. In a syntagmatic dimension, it can be tested by comparing the pitch level of two adjacent phrases within a single utterance. In Figure 1, for example, the second phrase is downstepped in the solid line since it is lower in pitch than its preceding phrase. This high-low relation of two adjacent phrases within a single utterance is the syntagmatic clue to downstep. In a paradigmatic dimension, on the other hand, downstep can be defined by comparing the pitch contours of two utterances, like the two contours in Figure 1, one beginning with an accented phrase and the other beginning with an unaccented one. Downstep can be identified if the second phrase is realized at different pitch levels depending on the accentedness of the first phrase. Unlike the syntagmatic definition, this definition of downstep does not require us to compare two peaks within a single utterance, but rather the peaks of corresponding elements on two different utterances.

Of the two definitions of downstep, it is the paradigmatic definition that was originally proposed (Poser 1984), and was generally adopted in 1980s (e.g., P&B 1988, Kubozono 1988). In contrast, the syntagmatic definition was adopted popularly in 1990s and later years (e.g., Selkirk & Tateishi 1991, Nagahara 1994). This approach to downstep, although more simplistic than the paradigmatic one, poses some serious problems in actual prosodic analysis. The most serious problem is that of ambiguity concerning the identification of downstep. Namely, it is not clear how much lower a given phrase has to be before it is recognized as being ‘downstepped’ as relative to its preceding phrase. It often happens that a phrase is only slightly lower in pitch than its preceding accented phrase. In such a case, it is difficult to tell in any objective sense whether or not downstep has occurred between the two phrases. Equally serious is the existence of cases where the two definitions of downstep do not agree with each other. These cases will be discussed in detail in sections 3 through 5 below.

The process of downstep thus defined has both phonological and phonetic aspects. It is a phonological process in the sense that it is triggered by a phonological factor, i.e., lexical accent. It is a phonetic process, on the other hand, if one looks at its effect. The second phrase in Figure 1 shows a bigger pitch rise in the utterance with downstep (solid line) than in the utterance without downstep (dashed line) to compensate for the greater pitch fall triggered by the lexical accent. This compensatory pitch rise is obviously insufficient, with the result that the downstepped phrase nevertheless attains a lower peak (P2 in Figure 1) than the non-downstepped phrase. Thus, the phrase following an accented phrase is realized at a lower pitch level than the one following an unaccented phrase. This whole process is phonetic in nature since the magnitude of the lowering process is not binary but variable depending on several factors such as the phonological length of the phrases, the location of accent, etc. (Kubozono 1988). That the effect of downstep is basically phonetic rather than

phonological can also be supported by perceptual evidence. Shinya (2005), for example, showed that generally speaking, native speakers are not conscious of the downstep effect in either a syntagmatic or paradigmatic sense. In particular, they are not conscious of the pitch difference between the first and second phrases in the utterance with downstep (solid line in Figure 1) or between the downstepped second phrase (solid line) and the non-downstepped one (dashed line).

### 2.3 Major and Minor Phrases

Most theoretical studies of Japanese intonation today assume two prosodic levels under the topmost ‘Utterance’, which is only vaguely defined in the literature. The level just under the utterance is the level of ‘intermediate phrase’ (P&B 1988) or ‘major phrase’ (Poser 1984, Kubozono 1988, Selkirk & Tateishi 1991). Thus, each utterance is supposed to consist of one or more major phrases. The prosodic level that is further lower is that of ‘accentual phrase’ or ‘minor phrase’: each major phrase consists of one or more minor phrases.

The minor phrase is generally defined as the domain of initial pitch rise as well as the domain within which at most one lexical accent can be phonetically realized. A sequence of accented words/phrases usually display a staircase like the solid line in Figure 1, where each syntactic phrase (often referred to in Japanese as *bunsetsu*) constitutes one minor phrase with an independent initial pitch rise and accentual fall (if it contains a lexically accented word). On the other hand, a sequence of two syntactic phrases is often amalgamated into one minor phrase if the first phrase does not have a lexical accent (Poser 1984, Kubozono 1988). In this case, the whole sequence shows only one initial pitch rise and at most one accentual fall.

The Major Phrase is generally defined as the domain of downstep, the pitch lowering process described above. If downstep is identified between two

adjacent minor phrases as in the solid contour in Figure 1, those phrases belong to one and the same Major Phrase. If, on the other hand, it is blocked between two minor phrases, there is a Major Phrase boundary between them.

In P&B's work, the Major Phrase (or 'intermediate phrase' in their terminology) is also defined as the domain of pitch reset. So if downstep is blocked between two minor phrases, there is a Major Phrase boundary between them, with the pitch being 'reset' at the beginning of the second phrase. We follow P&B to assume that the Major Phrase is the domain of pitch reset as well as downstep.

Selkirk & Tateishi (1991) do adopt the general definition of the Major Phrase as the domain of downstep, but they also take a syntax-driven approach in identifying Major Phrase boundaries. According to their top-down approach, a Major Phrase boundary is inserted at the left edge of every XP. We will see evidence against this approach later (see footnote 3 and section 4.2).

### 3 Review of past work on focus prosody

With a view to examining the effect of focus on intonation structure, Poser (1984) used the set of four phrases in (2) and compared the peaks of the second adjective *aóí* 'blue'. In (2a) and (2b), no word is emphasized, whereas the adjective *aóí* is emphasized in (2c) and (2d). Emphasized words are capitalized in (2) and the rest of this paper.

- (2) a. amai aóí kudámono 'sweet, blue fruit'  
b. umái aóí kudámono 'tasty, blue fruit'  
c. amai AÓI kudámono 'sweet, BLUE fruit'  
d. umái AÓI kudámono 'tasty, BLUE fruit'

Poser obtained the following average peak values (Hz) from a single speaker for the three phrases/words comprising the sentences in (2): the first adjective (*umái* or *amai*), the second adjective (*aói*) and the noun (*kudámono*).

(3)	<i>amai</i> or <i>umái</i>	<i>aói</i>	<i>kudámono</i>
a.	171.3	168.9	159.8
b.	176.7	163.2	135.7
c.	168.4	187.7	151.8
d.	175.9	183.4	133.6

A comparison of these values indicates that emphasized elements are more boosted in pitch than non-emphasized ones. For example, the adjective *aói* is considerably higher in (3c,d) than in (3a,b), respectively. In terms of downstep, a comparison of (3a) and (3b) reveals that the adjective *aói* was significantly lower in (2b) than in (2a) ( $T=3.48$ ,  $p<0.005$ ). This indicates the presence or absence of a lexical accent in the immediately preceding adjective (*umái* vs. *amai*) has exerted an effect on the height of the adjective *aói*; namely, *aói* is downstepped in (2b) due to the presence of a lexical accent in the preceding adjective. Poser then compared the pitch peaks of the same adjective in (2c) and (2d), in both of which the adjective itself is emphasized. He reports that *aói* is lower in (2d) than in (2c) with a difference that is ‘only marginally significant’ ( $T=1.98$ ,  $p=0.03$ ) (Poser 1984: 301).

What Poser’s data suggest is that emphasis does not block downstep although it may weaken the lowering effect as compared with the non-emphasis context. This result is particularly interesting because the downstepped element, *aói*, has a higher peak than its preceding element, *umái*, in (2/3d): 183.4 Hz vs. 175.9 Hz. We will see similar paradoxical cases in the following sections.

While Poser's data suggest that focus on a particular element does not block downstep and, hence, it does not introduce a Major Phrase boundary, P&B (1988) presented quite different data and drew an entirely different conclusion. As mentioned in section 1, they claimed that focus blocks the intonational process of downstep and, hence, starts a new Major Phrase, or what they called 'intermediate phrase'. The sentences they used to make this claim are the following:

- (4) a. amerika-níwa amai KÉEKI-wa arimásu-ga amai AME-wa arimasén.  
 'In America there are sweet CAKES, but there aren't sweet CANDIES.'
- b. amerika-níwa umái NINZIN-wa arimásu-ga umái MAMÉ-wa arimasén.  
 'In America there are tasty CARROTS, but there aren't tasty BEANS.'
- c. amerika-níwa umái KÉEKI-wa arimásu-ga umái AME-wa arimasén.  
 'In America there are tasty CAKES, but there aren't tasty CANDIES.'
- d. amerika-níwa amai NINZIN-wa arimásu-ga amai MAMÉ-wa arimasén.  
 'In America there are sweet CARROTS, but there aren't sweet BEANS.'

In order to see an effect of focus on downstep, P&B compared (4a) and (4b) with respect to the pitch peaks of the adjective-noun sequences, where *wa* is a particle denoting contrast:

- (5) a. (=4a) amai AME-wa 'sweet CANDIES'  
 b. (=4b) umái MAMÉ-wa 'tasty BEANS'

They demonstrated that these two phrases do not exhibit a typical downstep pattern, a pattern whereby they are clearly separated when the peak of their first word is plotted against the peak of the second word. On the basis of this experimental result, P&B claimed that the focused element is realized at the

same pitch level irrespective of any difference in phonological structure of the preceding material; thus, the pitch contour is ‘reset’ by focus.

Given P&B’s data thus described, one may quite naturally wonder why they did not compare (4a) and (4c) or (4b) and (4d), respectively. These suggested comparisons are given in (6) and (7).

- (6) a. (=4a) *amai AMÉ-wa* ‘sweet CANDIES’  
 b. (=4c) *umái AMÉ-wa* ‘tasty CANDIES’

- (7) a. (=4b) *umái MAMÉ-wa* ‘tasty BEANS’  
 b. (=4d) *amai MAMÉ-wa* ‘sweet BEANS’

It is not clear why P&B did not compare the two phrases in these pairs, but this appears to be a drawback in their analysis. The two phrases in (5) differ not only in the accentedness of the first word but in that of the second word as well. A pair of phrases like this cannot be used to examine an effect of downstep since the second word in (5b) differs from the second word in (5a) in more than one way. In terms of accentual boost, the second word in (5b) should bear a higher pitch than its counterpart in (5a) due to the lexical accent it contains, or ‘accentual boost’. If downstep takes place, on the other hand, the second word in (5b) should bear a lower pitch than its counterpart in (5a) because of the lexical accent in its preceding word, *umái*. Given these antagonistic forces that may operate on the second word in (5b), this particular word may well be realized at much the same pitch level as its counterpart in (5a). In other words, the effect of accentual boost may well diminish the effect of downstep in the second word in (5b).

This interpretation is supported by Kubozono’s (1988) experimental data, which contain the following pairs of phrases.

(8) a. umái méron ‘tasty melon’

b. amai oimo ‘sweet potato’

(9) a. umái nomímono ‘tasty drink’

b. amai yamaimo ‘sweet yam (potato)’

When pronounced in a non-contrastive context, the second words in each pair did not exhibit a noticeable pitch difference with respect to their height. However, the phrases in (8a) and (9a) did show a clear effect of downstep when they were compared with the phrases in (10) and (11), respectively: *méron* is realized at a significantly lower pitch level in (8a) than in (10), and *nomímono* is significantly lower in (9a) than in (11).

(10) amai méron ‘sweet melon’

(11) amai nomímono ‘sweet drink’

Thus, the second words in (8a) and (9a) are downstepped due to the presence of an accent in the preceding material, but this downstep effect cannot be seen—i.e., it is masked—when they are compared with the second words in (8b) and (9b), respectively. After all, in order to see whether or not there is a downstep effect, one must compare two phrases that contrast minimally with each other. Comparing the two phrases in (5), as P&B did, is equivalent to comparing the two phrases in (8) or those in (9), and will not answer the question of whether downstep has taken place between the two relevant phrases.

## 4 Experiment

### 4.1 Method

We used the following frame sentence in our experiment: TOP, ACC and Q stand for topic, accusative and question markers, respectively.<sup>3</sup>

(12) *anáta-wa X-de Y-to náni-o mimásita-ka?*

You- TOP X-in Y-with what-ACC see-PAST-Q

‘What did you see with Y in X?’

We put either *aómorí* ‘Aomori’ or *oomorí* ‘Oomori’ in the X slot and *náoko* ‘Naoko’ or *naomi* ‘Naomi’ in the Y slot. This created the four combinations in (13), which differ in the accentedness of the nouns in the two slots. [AA-Wh] stands for a sequence of two accented phrases plus the Wh word (which is always lexically accented in Tokyo Japanese), whereas [UU-Wh] refers to a sequence of unaccented noun phrases plus the Wh word.

(13) a. [AA-Wh] ...*aómorí-de náoko-to náni-o...*

‘What did you see with Naoko in Aomori?’

b. [AU-Wh] ...*aómorí-de naomi-to náni-o...*

‘What did you see with Naomi in Aomori?’

---

<sup>3</sup> Selkirk & Tateishi’s (1991) top-down approach to intonational structure would predict downstep is blocked between *X-de* and *Y-to* as well as between *Y-to* and *náni-o* in (12) since these positions correspond to the left edge of XPs and, hence, introduce a Major Phrase boundary. This prediction cannot be borne out in our experiment, as we will see in section 4.2.

c. [UA-Wh] ...oomori-de náoko-to náni-o...

‘What did you see with Naoko in Oomori?’

d. [UU-Wh] ...oomori-de naomi-to náni-o...

‘What did you see with Naomi in Oomori?’

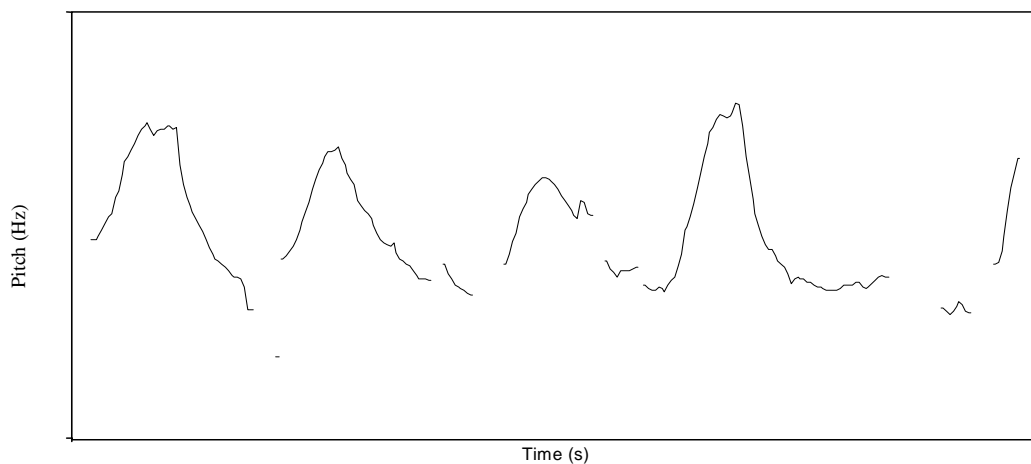
These four test sentences were mixed with dummy sentences in a random way. They were read by seven native speakers of Tokyo Japanese (five male and two female speakers), who pronounced the sentences eleven times at a normal tempo—one repetition per set for a total of eleven sets. This recorded eleven tokens of each test sentence, the first of which was discarded for analysis. A total of 280 tokens were subsequently analyzed (10 tokens x 4 sentences x 7 speakers).

In the analysis, we measured the pitch peak and valley of the three phrases comprising the sentences in (13), with particular focus on the peak of the Wh word. If the Wh element turns out to have the same peak value irrespective of the accentedness of its preceding phrases, then it will mean that downstep has been blocked by the Wh element with the pitch being reset in this position. This will be the result in accordance with the result reported by P&B. If, on the other hand, the pitch is not reset by the Wh element, this element will exhibit different pitch heights among the four test sentences, reflecting the differences in the phonological structure of its preceding materials. In particular, the Wh element will be realized at a considerably lower pitch level when following a sequence of accented phrases as in (13a) than when following a sequence of unaccented phrases as in (13d). If this were the case, it would be a clear case of downstep in the traditional (paradigmatic) sense of the term, indicating that the Wh element does not block this lowering process. This will be the result compatible with

Poser's (1984). As we will see in the next section, our experimental data support this second scenario.

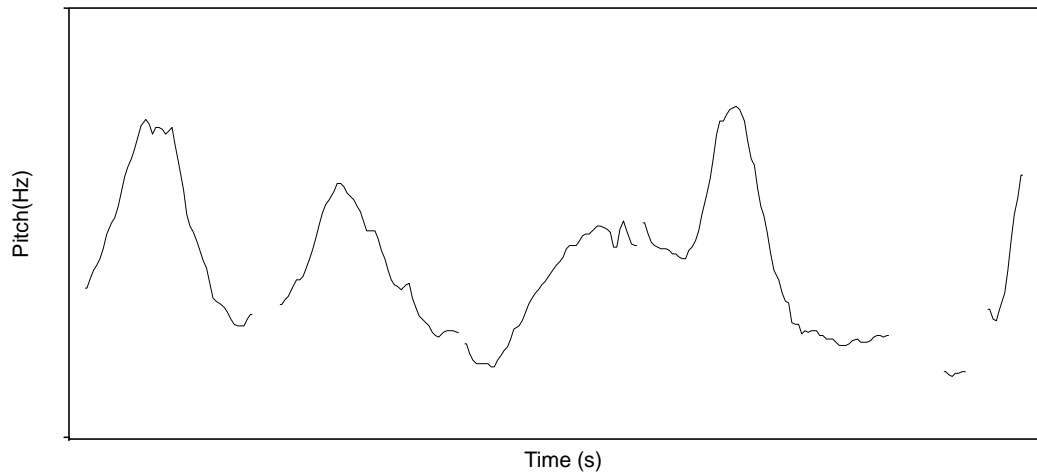
## 4.2 Results

First of all, the Wh element attains a very high pitch level in all the four sentences in (13). In fact, *náni-o* 'what' showed the highest pitch peak in every one of the 280 tokens that were analyzed, higher than the sentence-initial phrase and much higher than the phrase immediately preceding it. This is illustrated in Figures 2 and 3, which show the typical pitch contours of the sentences in (13a) and (13b), respectively. These contours indicate that the Wh phrase is the very focused element in the test sentences.



anáta-wa aómori-de náoko-to náni-o mimásita-ka?

Figure 2 A typical pitch contour of the sentence in (13a)



anáta-wa aómori-de naomi-to náni-o mimásita-ka?

Figure 3 A typical pitch contour of the sentence in (13b)

More significantly, the Wh element attained different peak levels among the four test sentences in (13). The biggest difference was found between (13a) and (13d), with the former showing a considerably lower peak than the latter. This tendency was shown by all seven speakers: for five of these speakers the difference was statistically significant (by a two-tailed t test), while the other two speakers exhibited a similar tendency although the difference did not reach a level of statistical significance (Table 1).

Table 1 Statistics of the peak F0 values (in Hz) in (13a) and (13d)

Speaker	Peak of Wh in (13a)	Peak of Wh in (13d)	T value	P value
TS (male)	146.2	153.7	2.630	p<0.05
TY (male)	139.7	156.3	2.297	p<0.05
JI (male)	205.9	222.6	3.393	p<0.01
AO (male)	184.3	196.2	4.005	p<0.01
AK (female)	277.8	289.4	2.202	p<0.05

NI (male)	179.4	182.5	0.415	p=0.683
MM (female)	271.9	281.9	1.445	p=0.165

Figure 4 schematizes the overall differences that Speaker AO showed between (13a) and (13d); the average peak and valley values of the three phrases comprising (13a) and (13d) are plotted, respectively.

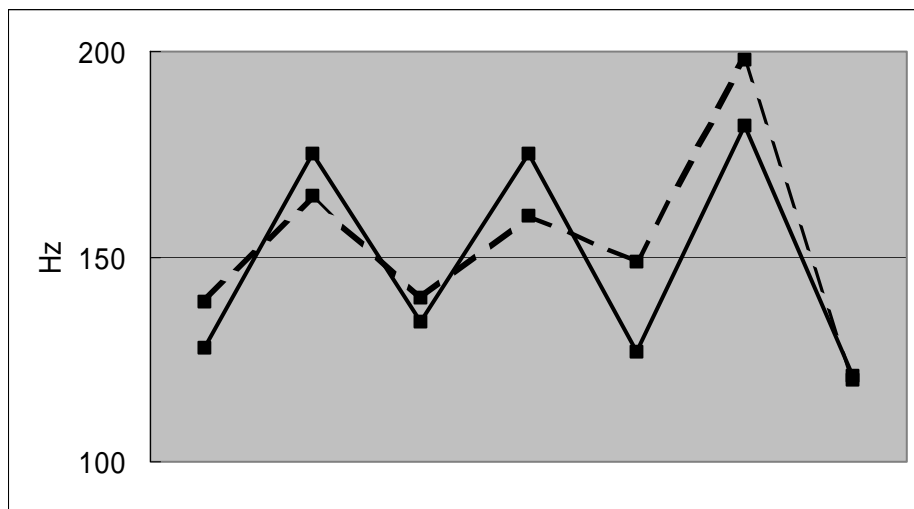


Figure 4 Schematic comparison of the pitch contours [AA-Wh] (solid line) vs. [UU-Wh] (dashed line) [Speaker AO]

The pitch contours in Figure 4 differ from each other in two crucial respects. First, the two phrases before the Wh phrase exhibit a higher peak in [AA-Wh] sequence than in [UU-Wh] sequence: *aómori-de* and *náoko-to* in [AA-Wh] have higher peaks than *oomori-de* and *naomi-to* in [UU-Wh], respectively. This is due largely to the presence or absence of lexical accent, or the effect of accentual boost described in section 2.2 above. A second and more important difference between the two contours in Figure 4 is that the Wh element, *náni-o*, has a considerably lower peak in [AA-Wh] than in [UU-Wh].

This difference was statistically significant for most speakers, as summarized in Table 1.

The overall difference between the two test sentences is clear. [AA-Wh] has a higher peak than [UU-Wh] in the pre-Wh context, but it has a lower pitch in the Wh element itself. This is the same situation that we saw in Figure 1 above, where the two definitions of downstep were explained. In other words, the Wh element in [AA-Wh] is lowered in pitch by the accent(s) of its preceding elements as compared to its counterpart in [UU-Wh]. This clearly shows that downstep has occurred between the Wh element and its preceding phrase in [AA-Wh] according to the original, paradigmatic definition of the term.

In fact, the only substantial difference between the solid line in Figure 1 and the [AA-Wh] contour in Figure 4 is that in the latter, the downstepped element (*náni-o*) is higher in pitch than its preceding phrase (*náoko-to*) despite that the former phrase has been lowered by the accent of the latter phrase in a paradigmatic sense. In this sense, the [AA-Wh] contour in Figure 4 presents a paradoxical case, a case where the downstepped element (Wh) is realized at a higher pitch level than the element whose accent has triggered the lowering process.

## 5 Discussion

### 5.1 Downstep or no downstep?

We are now faced with a puzzling case where the two definitions of downstep—syntagmatic and paradigmatic—do not agree with each other. In the syntagmatic dimension, it appears that the Wh phrase in the [AA-Wh] sentence in Figure 4 is not downstepped as relative to its preceding phrase since it is higher in pitch than the latter. On the other hand, the same Wh element is realized at a lower pitch region in the [AA-Wh] sentence than in its [UU-Wh] counterpart,

reflecting the difference in the accentedness of the pre-Wh material. This is a clear case of downstep in the paradigmatic sense of the term.

The question is how we can interpret this puzzling case. One thing that is worth serious attention here is that the paradoxical case in question is not an isolated phenomenon in Japanese prosody. There are at least two independent cases reported in the literature. One case has already been mentioned in section 3, where we discussed Poser's (1984) experimental data. A comparison of the pitch patterns exhibited by (2c) and (2d) showed that the focused adjective *AÓI* is significantly lower in pitch when preceded by a lexically accented word, *umái*, than when preceded by a lexically unaccented one, *amai*. Namely, the focused adjective is downstepped as relative to its preceding word in (2d). However, this downstepped adjective is realized at a higher pitch level than its preceding word as the values in (3d) clearly show.

A paradoxical case of the same nature has also been reported in my previous experiments (Kubozono 1988, 1989 and 1992). One case concerns the two sentences in (14).

(14) a. [AAAA]

[[[náoko-no] [áni-no]] [[aói] [erímaki]]]

‘(I saw) Naoko’s brother’s blue muffler’

b. [AUAA]

[[[náoko-no] [ane-no]] [[aói] [erímaki]]]

‘(I saw) Naoko’s sister’s blue muffler’

These two sentences have an identical syntactic construction, with four phrases constituting a binary branching structure. In phonological terms, they only differ in the accentedness of the second phrases, *áni-no* ‘brother’s’ vs. *ane-no* ‘sister’s’. These two sentences exhibit pitch contours as schematized in Figure 5.

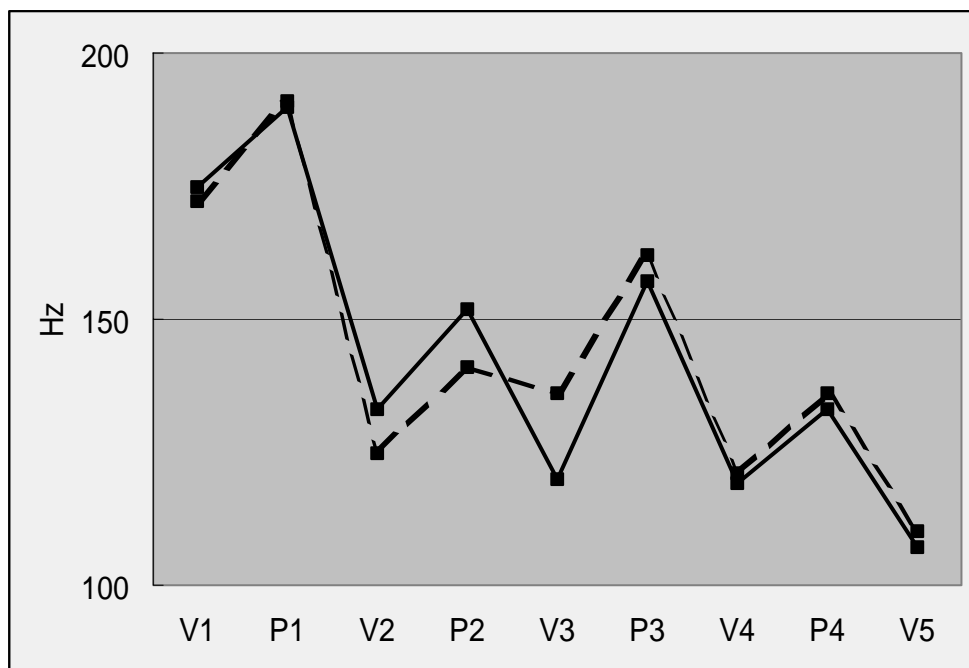


Figure 5 Comparison of the pitch contours shown by the two sentences in (14): (14a) (solid line) vs. (14b) (dashed line).

In every utterance of the two sentences in (14), the third phrase shows a higher peak than the second contour, suggesting that there is some prosodic boundary between the second and third phrases. On the other hand, the third phrase is realized at a lower pitch level when following the accented phrase, *áni-no* ‘brother’s’, than when following the unaccented one, *ane-no* ‘sister’s’. This indicates that the accent of the second phrase has exerted a lowering effect on the third phrase in (14a) although the lowered phrase is realized at a higher pitch

---

level than the phrase whose accent has triggered the lowering process. This is a situation that is identical to the one we saw in Figure 4 above.

What do the paradoxical cases in Figures 4 and 5 tell us? For one thing, they clearly show that the two definitions of downstep—syntagmatic and paradigmatic—do not always agree with each other. More specifically, they suggest that the lowering effect exerted by the accent of a phrase cannot be identified by comparing the pitch values of two consecutive phrases of a single contour. This speaks against the syntagmatic definition of downstep. In both cases illustrated in Figures 4 and 5, an accented phrase has exerted a lowering effect on the following material in a clear and objective way. This is a clear case of downstep according to the original and more objective definition of the term. There may be some prosodic boundary between the two relevant phrases, i.e., between the Wh phrase and its preceding phrase in Figure 4, and between the second and third phrases in Figure 5. However, this boundary cannot be a Major Phrase boundary. On the contrary, given that downstep is not blocked between the two relevant phrases, it follows that the two phrases do belong to one and the same Major Phrase as long as we adopt the general definition of this prosodic category described in section 2.3 above. Consequently, pitch reset has not taken place between the two phrases.

## 5.2 Implications

The data presented and analyzed in sections 4 and 5.1 have some important implications for the modeling of Japanese intonation while raising interesting questions for future research. Here we will focus on two issues, one concerning the interface between focus and intonation in general, and the other regarding the hierarchical structure of Japanese intonation.

### 5.2.1 Two types of focus?

We have seen in section 4 that Wh words do not block downstep and, hence, do not trigger pitch reset. How can we compromise this finding with the existing data about contrastive focus? As mentioned in section 3, Poser (1984) presented data suggesting that contrastive focus fails to block downstep and, hence, to trigger pitch reset. His analysis was supported by Shinya (1999), who also looked at the effect of contrastive focus on downstep. On the other hand, P&B (1988) claimed that contrastive focus does block downstep, introducing a Major Phrase boundary immediately before the focused element. However, P&B's analysis requires reconsideration, as pointed out in section 3. It seems to follow from these considerations that contrastive focus does not block downstep or trigger pitch reset. This analysis is compatible with our experimental data dealing with non-contrastive focus. This suggests that focus generally fails to block downstep in Japanese, whether it is contrastive or otherwise.

### 5.2.2 Intonation structure

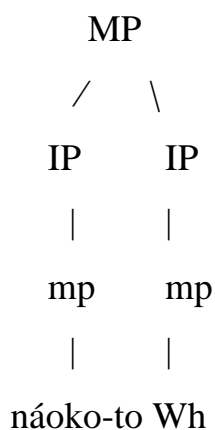
The data presented in the preceding sections have a significant implication for the prosodic organization of Japanese utterances, too. We have confirmed that Wh elements do not block downstep in the sense that their pitch height is influenced by the accentedness of the material immediately preceding them. On the other hand, Wh elements are realized in a higher pitch region than their preceding phrases. This latter fact suggests that there may be some prosodic boundary immediately before the Wh elements.

Assuming that this interpretation is correct, one may naturally ask what this prosodic boundary is. This cannot be a Major Phrase boundary since, as mentioned above, the Major Phrase is defined as the domain of downstep. As long as downstep is observed between the Wh element and its preceding phrase,

there cannot be a Major Phrase boundary between the two elements. The boundary in question cannot be a minor phrase boundary, either, since the Wh element and its preceding phrase clearly form two independent minor phrases, as can be seen from the pitch contours in Figures 2–4. This will raise a challenging question for the intonation model of Japanese since previous studies of Japanese intonation did not posit any intermediate phrase between the Major Phrase and the minor phrase.

One solution to this will be to posit an independent prosodic level/phrase between the two existing prosodic phrases. Let us tentatively call it an ‘intermediate phrase’ (IP), noting that it is different from what P&B called by the same name (by which they meant what we call the Major Phrase here). Restricting ourselves to the string of the Wh phrase and its preceding phrase in Figure 2, this analysis will assume a prosodic hierarchy as illustrated in (15), where the two phrases belong to one and the same Major Phrase, with each phrase constituting an intermediate phrase and a minor phrase on its own.

(15)



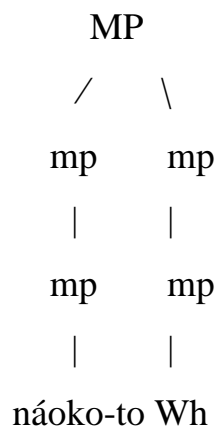
Under this analysis, the Wh phrase undergoes downstep because its preceding phrase is accented and belongs to the same Major Phrase. The Wh

phrase can have a higher pitch than its preceding phrase since it belongs to a different intermediate phrase from the latter.

While this appears to be an appealing solution, it falls into a problem regarding the definition of the new phrase. As mentioned in section 2, both the Major Phrase and the minor phrase have been defined in prosodic terms, as the domains of certain prosodic processes. However, the IP, which we have tentatively proposed in (15), does not have such an objective definition since it cannot be defined as the domain of any independent prosodic process. This seems to be a critical problem if one wants to propose a reasonably constrained model of intonation.

A solution to avoid using this unmotivated prosodic phrase is to appeal to the notion of ‘recursive’ category proposed by Ladd (1996) and supported by Kubozono (1988, 1989, 1992). This analysis allows a certain prosodic phrase/level to occur recursively in the prosodic hierarchy, as illustrated in (16).

(16)



Since this analysis introduces no new prosodic phrase/level, it is free from the kind of problem that the analysis in (15) poses.<sup>4</sup> Other merits as well as demerits of this analysis need to be explored in detail.

Another solution to the puzzling case in question might be to claim that focus intonation is independent of prosodic phrasing (cf. Ishihara 2005). This approach is certainly incompatible with the assumption generally adopted in the literature, i.e., that focus effects on intonation can be captured in terms of prosodic phrasing. However, this might allow us to solve our puzzle and, moreover, to understand the seemingly complicated interaction between focus and intonation in a reasonably principled way. We would like to leave this issue as a topic for future research.

## References

- Akinaga, Kazue. 1985. *Kyootuugo no akusento* (Accent of Standard Japanese). In the appendix to NHK (ed.), *NHK Pronunciation and Accent Dictionary*. Nihon Hoso Shuppan Kyokai, Tokyo.
- Ishihara, Shinichiro. 2003. *Intonation and Interface Conditions*. Ph.D. dissertation, MIT.
- Ishihara, Shinichiro. 2005. Prosody-scope match and mismatch in Tokyo Japanese *Wh*-Questions. *English Linguistics* 22–2: 347–379
- Kubozono, Haruo. 1988. *The Organization of Japanese Prosody*. Ph.D. dissertation, Edinburgh University. [Kurosio Publishers, 1993].
- Kubozono, Haruo. 1989. Syntactic and rhythmic effects on downstep in Japanese. *Phonology* 6–1: 39–67.

---

<sup>4</sup> A reviewer suggests another recursive model of intonation, where the Major Phrase rather than the minor phrase occurs recursively. In this analysis, *náoko-to* and the *Wh* element in (16) belong to different Major Phrases. This analysis would have to abandon the traditional definition of Major Phrase as the domain of downstep and, hence, require a new definition of this intonational phrase.

- Kubozono, Haruo. 1992. Modeling syntactic effects on downstep in Japanese. In G. J. Docherty and D. R. Ladd (eds.) *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. 368–387. Cambridge: Cambridge University Press.
- Kubozono, Haruo. 1996. Syllable and accent in Japanese: evidence from loanword accentuation. *The Bulletin* (Phonetic Society of Japan) 211, 71–82.
- Kubozono, Haruo and Misa Fukui. 1996. Phonological structure and unaccented nouns in Tokyo and Osaka Japanese. In Timothy J. Vance and Kimberly Jones (eds.) *Japanese/Korean Linguistics 14*, 39–50. Stanford: CSLI Publications & SLI.
- Ladd, D. Robert. 1986. Intonational phrasing: the case for recursive prosodic structure. *Phonology Yearbook* 3, 311–340.
- Nagahara, Hiroyuki. 1994. Phonological Phrasing in Japanese. Ph.D. dissertation, UCLA.
- Ogawa, Shinji. 2004. Sino-Japanese word accent and syllable structure. *Phonological Studies* 7, 41–48.
- Pierrehumbert, Janet and Mary Beckman. 1988. *Japanese Tone Structure*. Cambridge, Mass.: MIT Press.
- Poser, William. 1984. The Phonetics and Phonology of Tone and Intonation in Japanese. Ph.D. dissertation, MIT.
- Selkirk, Elizabeth and Koichi Tateishi. 1991. Syntax and downstep in Japanese. In Carol Georgopoulos and Roberta Ishihara (eds.) *Interdisciplinary Approaches to Language*, 519–543. Dordrecht: Kluwer Academic Publishers.
- Shinya, Takahito. 1999. Eigo to nihongo ni okeru fookasu ni yoru daunsuteppu no sosi to tyooon undoo no tyoogoo [The blocking of downstep by focus and articulatory overlap in English and Japanese]. In *Proceedings of Sophia Linguistics Society*, 14, 35–51.
- Shinya, Takahito. 2005. Lexical accent and the perception of intonational peaks in Japanese. Talk presented at the monthly meeting of the Phonological Association in Kansai (PAIK), June 25, 2005.

Truckenbrodt, Hubert. 1995. Phonological Phrases: Their Relation to Syntax, Focus and Prominence. Ph.D. dissertation, MIT.

Zubizarreta, Maria. 1998. Prosody, *Focus and Word Order*. Cambridge, Massachusetts, The MIT Press.

*Haruo Kubozono*  
*Department of Linguistics*  
*Faculty of Letters, Kobe University*  
*Rokkodai-cho, Nada-ku*  
*Kobe 657-8501*  
*Japan*  
*kubozno@lit.kobe-u.ac.jp*