Downstep and prosodic structure in Japanese

Midori Hayashi
University of Toronto

This paper examines the relationship between the syntax and the prosodic structure of Japanese. According to Kubozono (1989), the syntactic branching direction determines the pitch pattern in downstep; in the right-branching condition, the left edge of a syntactic constituent results in an extra pitch boost (Metrical Boost, MB), while in the left-branching condition, the utterance is restructured into two prosodic constituents whose left edge is boosted (Rhythmical Boost, RB), regardless of the syntactic configuration. In this study, an experiment is conducted with two Tokyo Japanese speakers using sentences in both branching directions which contain three to six minor phrases (mps). The result suggests that Kubozono’s proposal needs to be modified; the prosodic structure has an access to the syntax when there are a smaller number of mps. When there are more, the individual weight balancing strategies become more important, which results in a varied grouping configuration inter- and intra-individually.

0 Introduction

Japanese downstep is a phonologically conditioned pitch-lowering process, which is triggered by a lexical accent (Poser, 1984; Kubozono, 1989). It applies to a sequence of H tones; the second H tone is lowered with respect to the first one, and the third one is lowered with respect to the second one. It has generally been assumed that the domain of downstep is the Major Phrase (MP), a phrase characterized by certain intonational properties (Pierrehumbert and Beckman, 1988; Kubozono, 1989).

Within this domain, however, tone is not lowered consistently, but is ‘boosted’ at certain points. It has been agreed that the boosts reflect the prosodic organization in such a way that the left edge of a prosodic constituent receives an extra boost in pitch height (Selkirk and Tateishi, 1991; Venditti, 1994; Pierrehumbert and Beckman, 1988; Kubozono, 1989, 1992). However, the question of how phonological words are grouped into constituents is still a matter of debate. I will argue that previous accounts are not

---

1 Declination (or downdrift) also refers to a pitch lowering process; however, it should be distinguished from downstep in that declination is a phonetic process whereby pitch is gradually lowered towards the end of the utterance with no phonological trigger. The magnitude of a declination effect is much smaller than that of downstep (Pierrehumbert and Beckman, 1988).
empirically adequate, and that downstep domains are organized into smaller units by means of a combination of syntactic constituency and weight-balancing strategies which vary across individuals.

A number of studies have examined the relationship between phonological configurations of downstep domains and syntactic constituents (Poser, 1984; Pierrehumbert and Beckman, 1988; Kubozono, 1989, 1992; Venditti, 1994; Selkirk and Tateishi, 1991). It has been agreed that there is a correlation between them (Selkirk, 1986; Selkirk and Tateishi, 1991; Kubozono, 1989; Venditti, 1994). However, it has also been argued that prosodic structure can deviate from syntax in ways that are characteristic of prosodic representations (Dresher, 1994; Kubozono, 1989, 1992; Ghini, 1993; Gee and Grosjean, 1983). In his study on Italian prosody, Ghini (1993) claimed that two phonological words are grouped into a unit under certain conditions, regardless of syntactic constituency. Gee and Grosjean (1983) demonstrated that there is a prosodic demand on speakers which divides the whole utterance by a pause into two equally-weighted chunks.

The questions are, what are the correlation and the deviation between downstep domains and syntax, and what do they tell us about the structure of downstep domains?

An empirical study was conducted in order to investigate these issues. The subjects were two female native speakers of Tokyo Japanese in their twenties. Their productions of sentences with multiple adjectives both in the right- and the left-branching conditions were examined. The rationale is Kubozono’s (1989) proposal of ‘metrical’ and ‘rhythmic’ boost (MB and RB, respectively) in downstep. It states that, in the right-branching structure, the left edge of syntactic constituents receives an extra pitch boost (MB). In the left-branching condition, one has to take into account the grouping of minor phrases (mp), which are phonological domains with a single accentual core. They are restructured into two groups in the left-branching condition regardless of the syntactic configuration. Each of these groups forms a prosodic constituent whose left edge is boosted in pitch height (RB).

The results supported a weak version of the syntactic edge-based theory; prosodic constituents were delimited by the syntactic left edge of XP when the sentences contained threemps in the right-branching condition. However, when there were four or more phonological words, individual weight distribution became more important. This results in no correlation between prosody and syntax in both right- and left-branching structures.

The organization of this paper is as follows. In the first section, Japanese tonology is briefly sketched, and the lexical tone assignment rules are illustrated. In the second section, two previous studies on Japanese downstep, Kubozono (1989) and Selkirk & Tateishi (1992), are reviewed. In the third section, the method and the results of the experiment are presented. Their theoretical implications are discussed in the same section. In the last section, the study is summarized and perspectives for future research are discussed.
1 Japanese tonology

1.1 Lexical tone assignment in Tokyo Japanese

Japanese is a pitch-accent language (Tsujimura, 1996; Roca and Johnson, 1999; Shibatani, 1990; McCawley, 1968). It is characterized in such a way that the accent is lexically specified, and that where the accent falls determines the whole tonal pattern of the word. The tone-bearing unit is the mora in Tokyo Japanese\(^2\), because CVV (and CVC) sequences are assigned two distinct tones as follows:

\[
\begin{align*}
\text{(1) } & \quad \begin{array}{ll}
\text{a. } & \text{ka o (‘face’) } \quad \text{b. } & \text{ho n (‘book’) } \\
\text{L } & \quad \text{H } \\
\end{array} \\
(1a) & \quad \text{and (1b) are CVV and CVC and are assigned LH and HL contour tones, respectively. This demonstrates that the tone bearing unit is not the syllable, but the mora in this dialect.}
\end{align*}
\]

The tone assignment algorithm in Tokyo Japanese is still controversial. The autosegmental approach to tone assignment used to be common (McCawley, 1968; Haraguchi, 1976), characterized by full specification of tones. However, I do not take this approach since it does not predict the gradual tone decline in unaccented phrases. Instead, I take Pierrehumbert and Beckman’s (1988) approach, which has been employed in a number of studies (Venditti, 1994, 1997; Sugahara, 2003; Takahashi, Selkirk and Kawahara, 2003). The specified tones are shown in 2:

\[
\begin{align*}
\text{(2) } & \quad \begin{array}{ll}
\text{a. } & \text{H*L sequence } \\
\text{b. } & \text{Boundary tone L\% } \\
\text{c. } & \text{Phrasal tone H } \\
\end{array} \\
(2a) & \quad \text{H*L contour tone is assigned to the accented phrase, with H* associated to the accented mora (*) indicates an accented tone). The boundary tone L\% is associated with the first mora of the following phonological phrase (H\% is assigned at the last mora when the sentence is a question. For finer distinctions about the kinds of boundary tones, please consult Venditti (1997)). The phrasal tone is assigned to the second mora of the phrase. According to these partial tone specifications, the lexical tone assignment to an mp in Tokyo Japanese is illustrated as follows:}
\end{align*}
\]

\[
\begin{align*}
\text{(3) } & \quad \begin{array}{ll}
\text{a. } & \text{ya ma za* ku ra (‘mountain cherry blossom’) } \\
\text{L\% H(phrasal) H* L } \\
\end{array} \\
\end{align*}
\]

---
\(^2\) The tone bearing unit can be the syllable in some other dialects. For more details, see Kubozono and Honma (2002).
b. u e* ru mo no (‘the ones that are starved’) (Venditti, 1994)
   \[\text{L}\% \ H^* \ L\]

c. (no lexical accent is assigned)
   u e ru mo no (‘something to plant’) (Venditti, 1994)
   \[\text{L}\% \ H(\text{phrasal})\]

In (3a), the lexical accent is on the third mora, to which the H*L sequence is associated. The boundary L% is associated to the first mora, and the phrasal H tone to the second mora. (3b) and (3c) are examples where the morphological forms are the same but are different in meaning and in intonation. In (3b), the lexical accent is on the second mora to which the H*L is associated. The second mora is occupied by the H*, therefore no phrasal H is assigned. On the other hand, (3c) has no lexical accent, hence no H*L. Appendix 1 (Kubozono, 1989, pp194, reproduced by informant A in my experiment) displays the pitch range for (3b) (figure 1.a) and (3c) (figure 1.b). These show that H* is higher in pitch than the phrasal H, and that the L tone in H*L lowers the pitch range dramatically as compared to (3c).

1.2 Japanese tonology beyond the lexical level

It has been argued that the prosodic structure beyond the lexical level has four levels of representation (Selkirk, 1986; Nespor and Vogel, 1983; Pierrehumbert and Beckman, 1988). The first level (the bottom level) is for phonological words. A phonological word in Japanese typically consists of one lexical word plus a functional word, such as case markings. It is delimited by L% and contains one H peak (either phrasal H or accented H*).

The next higher level is for mps. One or more phonological words are grouped into an mp. When there is more than one phonological word, H peak(s) in the subsequent phonological word(s) is (are) suppressed. Therefore, the characterization of mp ends up being the same as that of phonological words; it is delimited by L% and contains one H peak as follows:

\[
\text{(4) } \text{u*mi-de (sea-LOC) + o yo*gi (swim) } \rightarrow \text{ u*mi-de o yo gi (swim in the sea)}
\]

\[\text{H}* \ L \ \text{L}\% \ \text{L}\% \ H^* \ L \ \text{H}\% \ H^* \ L \ \text{L}\%\]

The next higher level is for Major Phrases (MP). This is defined as the domain for downstep. It consists of a grouped set of mps. Within an MP, H* in each mp keeps going lower with respect to the preceding peak until the end of the MP, which is delimited by the beginning of another MP.
The highest level is for the utterance. Below is a model of prosodic hierarchy in the Japanese proposed by Pierrehumbert and Beckman\(^3\) (1988):

\[\text{(5)}\]

\[\text{U} \quad \text{MP} \quad \text{MP} \]

\[\text{wp} \quad \text{wp} \quad \text{wp}\]

\[L\% \quad H \quad L\%H \quad H*L \quad LH \quad L \quad H\%\]

\[| \quad | \quad | \quad | \quad | \quad | \quad | \quad |

\[\text{A ne-no} \quad \text{a ka i se*eta a} \quad \text{do*ko} \quad \text{de su-ka}\]

\[\text{Sister-GEN}^5 \quad \text{red} \quad \text{sweater-TOP} \quad \text{where} \quad \text{AUX- Q} \]

‘Where is my sister’s red sweater?’

The two phonological words (shown as \(w\)), \text{akai} (‘red’) and \text{se*eta a} (‘sweater’), constitute one mp. This is evidenced by the fact that the H peak in the second phonological word \text{seeta a} is suppressed (Please refer to Appendix 2 (figure 2) for the pitch display, which is produced by informant A in my experiment). The first MP consists of two mps, and the H peaks are expected to be subject to the downstep process. However, this is not the case; this can be due to the Metrical Boost (MB) effect (Kubozono, 1989), which was briefly explained earlier: when there is a left edge of XP, the pitch peak in the left mp in the XP is boosted\(^6\).

The two influential previous studies, Kubozono (1989) and Selkirk and Tateishi (1991), are discussed in the next section.

---

\(^3\) The model proposed by Pierrehumbert and Beckman (1988) uses Intonational Phrase (IP) and Accentual Phrase (AP) which are equivalent to MP and mp, respectively. In addition, their model has a level for syllables level between phonological words and morae, but I omit it for the sake of simplicity.

\(^4\) As I previously mentioned, H\% is assigned on the sentence final mora for an interrogative sentence. H\% results in a dramatic raise in pitch.

\(^5\) GEN is an abbreviation for Genitive.

\(^6\) The model by Pierrehumbert and Beckman (1988) in (5) does not actually conform to the pitch display (Appendix 2). Since the last mp constitutes an MP on its own, there must be an extra pitch boost at the third pitch peak, which was not the case. Therefore, the correct prosodic representation for this sentence should have all have the three mps organized into one MP. The other thing is that the AUX should be organized with the preceding noun into one phonological word, since AUX is a functional morpheme. Otherwise the noun and AUX would have to be treated as a syntactic constituent, which is subject to MB.
2 Syntax-phonology interface: previous studies

In this section, I first discuss Kubozono’s (1989) study, which proposes a strong correlation between syntax and prosody on one hand (MB) and a prosody-specific organization independent of syntactic structure on the other (RB).

2.1 Kubozono (1989)

Kubozono (1989) examined the productions of noun phrases with three mps in the left-branching and the right-branching condition.

(6) a. Right-branching phrases with 3 mps

\[
\text{[ao*i [o*okina me*ron]]}
\]

\[
\text{blue big melon ('blue big melon')}
\]

b. Left-branching phrases with 3 mps

\[
\text{[[ao*i re*mon-no] nio*i]}
\]

\[
\text{blue lemon-GEN smell ('smell of blue lemon')}
\]

Kubozono (1989) noticed that, in the right-branching condition, the pitch peak of the second word was boosted in pitch (7a), which did not occur in the left-branching condition (7b). He attributed this phenomenon to Metrical Boost (MB). The pitch gets boosted at the left edge of syntactic constituents as follows:

(7) a. U

MP

mp

[mpp mp]

ao*i o*okina me*ron

↑

MB

b. U

MP

mp

[mpp mp ] mp ]

ao*i re*mon-no nio*i
Therefore, Kubozono (1989) claims that the prosodic representation has to make reference to the syntactic structure; hence he suggests that the mps are multilayered and are mapped onto the syntactic constituents. For this reason, he explicitly rejected the Strict Layer Hypothesis, which states that prosodic units of type Xn-1 are exhaustively grouped into units at the next higher level Xn (Selkirk, 1984) as (7c).

Kubozono (1989) also examined noun phrases consisting of four mps in the right-branching (8a) and the left-branching condition (8b), and a noun phrase branching both leftward and rightward (8c):

(8) a. Right-branching phrase with 4 mps
[ma*riko-no [o*okina [ao*i eri*maki]]]
Mariko-GEN big blue scarf ('Mariko’s big blue scarf')

b. Left-branching phrase with 4 mps
[[[ma*riko-ga no*n-da]              wa*in-no]  nio*i]
Mariko-NOM drank-past-REL wine-GEN smell
('smell of wine that Mariko drank')

c. Phrase branching both leftwards and rightwards
[[na*oko-no a*ni-no] [ao*i eri*maki]]
Naoko-GEN brother-GEN blue scarf ('blue scarf of Naoko’s brother')

The principle of MB predicts that, in (8a), MB applies twice, at the second peak and at the third peak, since those are in the leftmost mps in the constituents. In (8c), MB applies at the third peak. These predictions were confirmed in his experiment. On the other hand, in the left-branching condition, the prediction is that no boost should occur if MB is applicable because no left edge of XP is available. However, this was not the case; the pitch pattern for (8b) was the same pattern as that for (8c), where the third peak had an extra boost. Kubozono (1989) attributed this phenomenon to Rhythmic Boost (RB), which states that where there are more than three mps, restructuring occurs in such a way that the mps are grouped into two. Then, the pitch peak of the leftmost mp of the constituent gets boosted, as follows:

---

7 This is an abbreviation for relative clause.
In short, Kubozono’s (1989) observation indicates that the prosodic structure directly reflects syntactic constituencies in the right-branching condition, but it deviates from the syntactic structure in the left-branching condition in such a way that mps are grouped into two no matter what the syntactic configurations are.

Now I briefly discuss Selkirk and Tateishi’s (1991) study.

2.2 Selkirk and Tateishi (1991)

Selkirk and Tateishi (1991) also investigate the correlation between the syntax structure and downstep patterns by examining the productions of sentences with various syntactic configurations. The sentences used in their experiment consisted of three mps, either in the left-branching condition or in the right-branching condition, and with three mps branching in both directions. The example sentences follow (each mp is numbered for easy reference):

(10) a. Left-branching phrase with 3 mps

\[
\begin{array}{ccc}
\text{N1} & \text{N2} & \text{N3} \\
\text{[[[aoyama-no] yama*guchi-no] ani*yome-ga] inai]}
\end{array}
\]
aoyama-GEN Mr.Yamaguchi-GEN sister-in-law-GEN absent
‘The sister-in-law of Mr. Yamaguchi from Aoyama is absent.’
b. Phrase branching in both directions

\[
\begin{array}{|c|c|c|}
\hline
N1 & N2 & N3 \\
\hline
[[[ao\*yama-no] yama\*guchi-ga]] [[][ani\*yome-o] yonda]] \\
\hline
\end{array}
\]

aoyama-GEN Mr.Yamaguchi-NOM sister-in-law-ACC called
‘Mr.Yamaguchi from Aoyama called his sister-in-law.’

Their experiment was also conducted with four native Tokyo Japanese speakers. Their results supported Kubozono’s proposals. In (10a), the H peak in N3 was boosted, which conforms to the RB principle. In (10b), the H peak in N2 was boosted, which follows the MB principle. In (10c), the H peak in N3 is boosted, which again is predictable according to the principle of MB.

Hence, their results reproduced Kubozono’s (1989). The major difference between Selkirk and Tateishi (1991) and Kubozono (1989) is that the former assume the Strict Layer Hypothesis, while the latter proposes a binary branching prosodic hierarchy. Consequently, for Selkirk and Tateishi (1991), each extra boost is the indication for the onset of a new downstep domain, MP. Therefore, for (10c), Selkirk and Tateishi assume two MPs, whereas Kubozono would assume only one MP.

The questions are, firstly, how do Selkirk and Tateishi (1991) accommodate structures such as (10b), where the right-branching phrase is in the larger left-branching phrase? On the other hand, Kubozono’s (1989) binary branching hierarchy does not have a problem representing the structure (10b). However, the challenge for him would be to determine whether an extra boost should be interpreted as the onset of MP or the onset of a higher mp; how much higher should the H peak be to be considered as an onset of MP?
2.3 Research questions

Considering all these issues, I pose three questions:

i. Can MB and RB be demonstrated to exist? In other words, do the phrases in the right- and in the left-branching conditions differ in the patterns of downstep in the way the MB and RB principles predict?

ii. In the restructuring process, is there any principled manner in which prosodic words are grouped together?

iii. Which representation is more likely, the binary branching hierarchy in Kubozono (1989), shown in (iv), or The Strict Layer Hypothesis (v) in Selkirk and Tateishi (1991)? Their models are represented as follows respectively (the representations are the ones that apply to the left branching phrases with three mps):

![Diagram](iv). ![Diagram](v).

These are the questions that I pursue in my experiment. The methods, the results, and the discussion are presented in the following section.

3 Experiment

3.1 Methods

The informants were two native Tokyo Japanese speakers: one in her late twenties (informant A), and the other in her early twenties (informant B). There were 8 target sentences: 4 were left-branched (with 3 phonological words (ws), 4ws, 5ws, and 6 ws) and another 4 are right-branched (with 3ws, 4ws, 5ws, and 6 ws). Almost all the lexical words used in the experiment consisted of 3 morae with the lexical accent on the first mora. There were 12 dummy sentences.

Informants were first instructed to read a sentence silently in order to understand its meaning. Then they were asked to read the sentence aloud into the microphone plugged into a computer. Their speech was analyzed with Speech Analyzer (sixteen bit, 22,050 Hz sampling frequency), a program available on the internet from SIL International.
3.2 Results: Left-branching condition (Appendix 3)

Productions of left-branching sentences are first analyzed in this section. RB is expected to occur according to Kubozono (1989).

First of all, I briefly mention the guideline which I followed in reading and analyzing the pitch displays. For the criterion to determine where the MP onset is, I follow Venditti’s (1997) J_ToBI labelling guideline. J_ToBI is a tool which provides a standard for prosodic labeling of speech data in Tokyo Japanese. For the most part their labeling system follows Beckman and Pierrehumbert (1988). For the entire J_ToBI labeling system, please consult Venditti (1995, 1997).

I needed the guidelines to distinguish between the MP onset and the higher mp onset. In J_ToBI, the labeling to mark the MP boundary is “finality”. According to her, “finality” is characterized as “a strong sense of disjuncture, stronger than that of a non-final intonation phrase boundary” (Venditti 1997:17). It is subjective by nature for there is no single decisive phonetic cue on which the labeler can depend; the possible cues are not limited to final F0 lowering, segmental lengthening, creaky voice, amplitude lowering and so on. At the same time, there is no specific phonetic value (e.g. pitch height, amplitude, pause length) set up for those cues to be considered as an indication for an MP onset. What counts is the labeler’s intuition about whether the value gives him or her “a strong sense of disjuncture”.

In this study, I limit the range of phonetic cues to determine the MP onset to pitch height. If the pitch of H is much higher or higher enough to make me feel a strong disjunction than the preceding H tones, it is considered an MP onset. At the same time, if the pitch height does not exceed the preceding one, or if it does exceed the preceding one but only slightly, then it is considered as being in the same MP. How high a pitch should be raised in order to be considered as only slightly exceeded is again determined in an impressionistic way. When a problem arises concerning the way of determining the MP onset, it is discussed along the way. If there are any other phonetic cues in the data which give me a sense of strong disjunction, they are also discussed in the relevant section.

3.2.1 3ws (+ 1w as a verb)

Each phonological word is numbered for easy reference. Each pitch peak (P) is numbered accordingly (e.g., P1 means the pitch peak of w1). The sentence used is as follows:

(11) w1                  w2                   w3                (w4)
    To*ruko-no Mi*kan-no u*masa-ni odoro*ita.
    [[[Turkey-GEN orange-GEN] taste-DAT] surprised]
    ‘I was surprised at the taste of Turkish orange’

Informant A (Figure 3a)

P1 (320Hz), P2 (215Hz), P3 (185 Hz), and P4 (195Hz). P4 is boosted.

8 ‘Intonational Phrase’ is basically the same concept as MP.
Informant B (Figure 3b)

P1 (320Hz), P2 (290Hz), P3 (235Hz), and P4 (230Hz). P4 is boosted. The final vowel of w1 was lengthened.

For both informants, the pitch is boosted at P4. It can be assumed that phonological words are organized into two groups, [w1 w2 w3], and another, [w4].

The question to ask now is whether the prosodic structure should be represented hierarchically or not. If the Strict Layer Hypothesis is correct, MP [mp1 mp2 mp3] + MP [mp4] is assumed. If it is organized hierarchically, MP [[mp[mp[mp1 mp2]] + mp3] + mp4]. It amounts to the question of whether there is a strong sense of finality between mp3 and mp4.

It is more plausible to analyze that the H peak in mp4 is under an influence of a downstep effect from the larger constituent since the difference in pitch height between P3 and P4 is not as prominent as in other cases which were determined to mark an MP boundary. Hence, I propose that the prosodic structure should be hierarchically organized and assume the prosodic representation that both informants had in their mind at the moment of utterance is as follows (▲ indicates a boost):

(12)

```
  /\  
 /   
MP   
  /\  
 /   
mp   mp
  /\  
 /   
mp   
  /\  
 /   
mp1 mp2 mp3 mp4
  /\  
 /   
(w1) (w2) (w3) (w4)
```

From now on I take the position that the prosodic structure should be hierarchically organized, hence, I hierarchically represent the possible prosodic structures following Kubozono (1989).

Another question is whether RB occurred. The principle of RB cannot explain the boost; mp3 was the one which was supposed to be boosted, not mp4. It can be hypothesized at this point that RB occurs only in a constituent which belongs to a single syntactic category. This explains why there was no boost in mp3, since the number of mps which constitutes NP is three, mp1, mp2, and mp3 (RB applies only when there are 4 mps or more in the utterance). However, the presence of boost in mp4 still needs to be explained.

---

9 I look into organization of phonological words, instead of mps, since I consider that the grouping of phonological words into one mp also reflects prosodic organization in speakers’ minds.
3.2.2 4ws (+ 1w as a verb)

The sentence used was the following:

(13) \[\text{ni*gatsu-no ka*nada-no a*sa-no sa*musa-o taikenshita.}\]
    
    ‘I experienced the cold temperature in the morning of February in Canada’

**Informant A (Figure 4a)**

P1 (290Hz), P2 (220Hz), P3 (210Hz), P4 (185Hz), and P5 (170Hz). P3 is boosted.

**Informant B (Figure 4b)**

P1 (290Hz), P2 (315 Hz), P3 (285Hz), P4 (270Hz), and P5 (not applicable). P2 is boosted. P1 is considerably low with respect to P2. This can be due to the fact that this informant pronounced no lexical accent in w1. P5 was not observed, which implies that w4 and w5 are possibly united into one mp. It should also be noted that she lengthened the last vowel$^{10}$ of w2. This vowel lengthening occurred for this informant in the previous sentence as well. The function of vowel lengthening will be briefly discussed later.

The pitch pattern for informant A follows the RB principle, which divides mps into two groups. This division does not map onto the syntactic structure and also the hypothesis that RB applies only within the same syntactic category does not hold here. In the utterance of informant B, there is no clear sign of RB. The prosodic representations for the informant A and B can differ as follow:

(14) A:  

\[
\begin{array}{c}
\text{MP} \\
\text{mp1} \text{ mp2} \text{ mp3} \text{ mp4} \text{ mp5} \\
\downarrow \text{w1} \text{ w2} \text{ w3} \text{ w4} \text{ w5}
\end{array}
\]

B:  

\[
\begin{array}{c}
\text{MP} \\
\text{mp1} \\
\downarrow \text{w1} \\
\text{mp2} \text{ mp3} \text{ mp4} \\
\downarrow \text{w2} \text{ w3} \text{ w4} \text{ w5}
\end{array}
\]

There is no parallel between the prosodic structure and the syntactic structure. At the same time, (14) shows that the prosodic representation can differ from individual to individual. In addition, as I mentioned earlier, informant B has a tendency to divide the sentence in a certain way by lengthening a final vowel of a certain constituent. This division by the lengthened vowel again does not exactly map onto either the syntactic constituency or the prosodic one determined based on the pitch patterns. In the previous example, (12) the mps are grouped based on pitch height as [[mp1 mp2] mp3] and [mp4], whereas the lengthened vowel divides the utterance as [mp1] and [mp1 mp2 mp3]. In this

---

$^{10}$ The lengthening of the last vowel of mp (in other words, the last vowel of case markers) is common among the young generation.
example, (14b), the pitch heights predicts the grouping [mp1] and [[mp2 mp3] mp4], but the lengthened vowel divides it differently as [mp1 mp2] and [mp3 mp4].

Based on these facts, it can be assumed that, firstly, different individuals can resort to different phonetic cues to organize a prosodic structure; secondly, different phonetic cues can demonstrate different prosodic organizations.

3.2.3 5ws (+1w as a topic)

<table>
<thead>
<tr>
<th>(16)</th>
<th>w1</th>
<th>w2</th>
<th>w3</th>
<th>w4</th>
<th>w5</th>
<th>w6</th>
</tr>
</thead>
<tbody>
<tr>
<td>sore-wa, kyo<em>nen-no shi</em>gatsu-no i<em>ppi-no sho</em>ugo-no ji*ken-desu.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[It-TOP, [[[last-year-GEN April-GEN] first-day-GEN] noon-GEN] incident-AUX]]

‘It is an incident that happened at noon on the first day of last year’s April.’

**Informant A (Figure 5a)**

P1 (215Hz), P2 (270Hz), P3 (255Hz), P4 (170Hz), P5 (220Hz), and P6 (150Hz). P2 and P5 are boosted. It may be more reasonable to say that P4 and P6 are suppressed and are united into one mp with P3 and P5 respectively, provided that the differences between P3 and P4, and between P5 and P6, are much greater than the pitch-height differences between the other downstepping peaks such as between P2 and P3. At the same time, there is also a huge gap in pitch height and a long pause between P1 and P2. Hence I take this to be a case of finality following Veneditti (1997), which implies that P2 marks an MP onset.

**Informant B (Figure 5b)**

P1 (280Hz), P2 (320 Hz), P3 (250 Hz), P4 (255 Hz), P5 (265 Hz), and P6 (215 Hz). The pitch is boosted at P2, P4 and P5. The accent\(^\text{11}\) of w3, *shigatsu* (‘april’), is not pronounced. There is a pause between w1 and w2 as well as a huge gap in pitch-height between P1 and P2. Thus I take w1 and w2 as belonging to the separate MPs. P4 and P5 are boosted. The plausible prosodic representations for A and B are shown in (16):

(16) A:       U                        B:       U
     MP                        MP
      |                        |          |
  1mp1 mp2 mp3 mp4          mp1 mp2 mp3 mp4 mp5 mp6
    |                      |       |
   w1 w2 w3 w4 w5 w6       w1 w2 w3 w4 w5 w6

---

\(^{11}\) This informant has a tendency to pronounce the name of the months without a lexical accent.
The syntax and the prosodic structure match in that there is a MP boundary between w1 and w2 for both informants; w1 is syntactically a topic of the sentence. Other than that there is no correlation between the syntax and the possible prosodic structure.

It should be noted that informant B again lengthened the final vowels of w2 and w4. This divides the utterance in equally weighed three groups as [w1 w2] [w3 w4] and [w5 w6].

3.2.4 6 ws (2ws as a predicate)

(17) w1   w2   w3   w4   w5   w6
ku*gatsu-no  Ha*waii-no  bi*ichi-no  ku*uki-no  o*ndo-no  su*gosa-wa,
(w7)
[[[taerare*nai-mono-ga] arimasu]].
unbearable-something-NOM exist.

‘The height of the temperature of the air in September’s Hawaiian beach is unbearable.’

Informant A (Figure 6a)

P1 (310Hz), P2 (250 Hz), P3 (220 Hz), P4 (180 Hz), P5 (180Hz), P6 (155Hz), P7 (215 Hz), and P8 (not applicable). P5 and P7 are boosted. There is no sign of P8, which indicates that w7 and w8 are grouped into one mp. The difference in pitch height between P6 and P7 is much greater than the other downstepping peaks. Hence I assume that w7 and w8 constitute an MP. The final vowel lengthening of w6 supports this assumption.

Informant B (Figure 6b)

P1 (320 Hz), P2 (285Hz), P3 (250Hz), P4 (270 Hz), P5 (245Hz), P6 (240 Hz), P7 (250 Hz) and P8 (not applicable). P4, P6 and P7 are boosted. P8 is not observed, which indicates that w7 and w8 are grouped into one mp.

The plausible prosodic representations for A and B follow:

(18) A:                  B:                  U
MP             MP
mp mp mp mp mp mp
|     |     |      |     |     |                                |     |      |     |     |     |
| w1 w2 w3  w4 w5 w6 | w1 w2  w3 w4 w5 w6 | w7 w8 |

Again, in the utterance of informant B, vowel-lengthening was observed in w4, which divides the utterance into two constituents, both of which have an equal number of mps, as [mp1, mp2, mp3, mp4] and [mp5, mp6, mp7]. There is no correlation between the
way the lengthened vowel divides the utterance and the way the syntax and the pitch height do.

3.3 Results: Right-branching condition (Appendix 4)

In the right-branching condition, MB is expected at the left edge of each constituent. The prosodic constituents are supposed to match the syntactic structure. Hence no restructuring occurs according to the MB principle.

3.3.1 3ws

(19) \( w_1 \quad w_2 \quad w_3 \)

\[ [A^*tsukute [o^*okina te^*nohira-desu]]. \]

thick big palm-AUX

‘It is a thick, big palm’

Informant A (Figure 7a)

P1 (240Hz), P2 (250Hz), and P3 (200Hz). P2 is boosted.

Informant B (Figure 7b)

P1 (270Hz), P2 (255Hz), and P3 (245Hz). No peak is boosted.

The MB principle predicts a pitch boost at P2. However, this is the case only with informant A. The plausible prosodic structures for A and B follow:

(20) A: MP

MP

mp

mp

| w1 w2 w3 |

B: MP

MP

mp

mp

mp

| w1 w2 w3 |

3.3.2 4ws

(21) \( w_1 \quad w_2 \quad w_3 \quad w_4 \)

\[ [Hi^*rokute [hu^*kakute [ki^*reina pu^*uru-desu]]]. \]

large deep clean pool-AUX

‘It is a large, deep, clean pool’
Informant A (Figure 8a)
P1 (285Hz), P2 (220Hz), P3 (225Hz), and P4 (175Hz). P3 is boosted. P2 is expected to be boosted if MB applies; however, it is not the case.

Informant B (Figure 8b)
P1 (325Hz), P2 (265Hz), P3 (280Hz), and P4 (225Hz). P3 is boosted but P2 is not, which is the same as informant A.

The MB principle does not explain the pitch pattern observed here; rather, it conforms to the RB principle, which states that four or more mps are grouped into two.

The plausible prosodic representation for both A and B is as follows:

(22) \[
\begin{array}{c}
\text{MP} \\
\text{mp} & \text{mp} & \text{mp} & \text{mp} \\
| & | & | & |
\end{array}
\]

w1 w2 w3 w4

3.3.3 5ws

(23) [shi*zukade [hi*rokute [ha*yakute [o*okina je*tto-desu]]]]
quiet spacious fast big jet-AUX
'It is a quiet, spacious, fast and big jet'

Informant A (Figure 9a)
P1 (310Hz), P2 (265Hz), P3 (180Hz), P4 (190Hz), and P5 (not applicable). The pitch is boosted at P4. P5 is not observed, which indicates that w4 and w5 are grouped into one mp. The pitch boost was expected at all the left edges, but this was not the case.

Informant B (Figure 9b)
P1 (330Hz), P2 (265Hz), P3 (275Hz), P4 (255Hz), and P5 (245Hz). The pitch is boosted at P3.

Neither of the structures conforms to the MB principle. The plausible prosodic structures for A and B follow:

(24) A: \[
\begin{array}{c}
\text{MP} \\
\text{mp} & \text{mp} & \text{mp} & \text{mp} & \text{mp} \\
| & | & | & | & | \\
\text{w1} & \text{w2} & \text{w3} & \text{w4} & \text{w5} \\
\end{array}
\]

B: \[
\begin{array}{c}
\text{MP} \\
\text{mp} & \text{mp} & \text{mp} & \text{mp} & \text{mp} & \text{mp} \\
| & | & | & | & | & | \\
\text{w1} & \text{w2} & \text{w3} & \text{w4} & \text{w5} \\
\end{array}
\]
3.3.4 6ws

(25)  w1  w2  w3  w4  w5  w6
[hu*rukute [se*makute [ku*rakute [ku*sakute [chi*isana ma*nshon-desu]]]]]
old  narrow  dark  stinky  small  apartment-AUX
‘It is an old, narrow, dark, stinky, small apartment’

Informant A (Figure 10a)

P1 (295Hz), P2 (225Hz), P3 (195Hz), P4 (165Hz), P5 (175Hz) and P6 (145Hz).
Only P5 is boosted.

Informant B (Figure 10b)

P1 (310Hz), P2 (300Hz), P3 (260Hz), P4 (265 Hz), P5 (285Hz) and P6 (225Hz).
P2, P4 and P5 are boosted. The possible prosodic structures for A and B follow:

(26) A:       B:
               U               U
               |               |
               MP             MP
               |
               |
               mp mp   mp   mp  mp  mp  mp
w1 w2  w3  w4   w5  w6
               mp

It should be noted that informant B again lengthened the last vowel of w2 and w4. This gives an impression that the utterance is divided into 3 groups each of which consists of 2 mps as [mp1 mp2], [mp3 mp4], and [mp5 mp6].

3.3.5 Discussion

Before going into the discussion, I repeat my research questions:

i. Can MB and RB be demonstrated to exist?
ii. In the restructuring process, is there any principled manner in which prosodic words are grouped together?
iii. Which representation is more likely, the binary branching hierarchy in Kubozono (1989) or the Strict Layer Hypothesis in Selkirk and Tateishi (1991)?

As for question (a), MB is observed once with the informant A in the sentence with 3 right-branching mps (Figure 7a). On the other hand, RB occurred in all the cases where there are 4 or more mps, in a sense that the plausible prosodic structure does not
match the syntax. This is problematic for Kubozono’s (1989) proposal, since the prosodic structure must coincide with syntax in the right-branching condition.

In addition, I claim that the distinction between MB and RB is unnecessary. The concepts of RB and MB are based on the assumption that the prosodic representation and syntax have the same structure, and restructuring occurs from there. However, I do not see a compelling reason why it should be the case given the fact that the prosodic structure matches syntax only when there are 3 mps in the utterance and, furthermore, this does not always have to be the case as shown in Figure 7b.

Therefore, it is more reasonable to assume that the prosody has its own structure which is independent of syntax, and speakers can make reference to syntax when there is a smaller number of mps in their utterance, presumably three. However, speakers resort more to their own weight distribution strategy when there is a larger number of mps.

As for the question (b), I propose that, besides syntax, there are different phonetic means that speakers depend on to organize their utterance prosodically, and downstepping in pitch height is one of them. Lengthening the last vowel of a phonological word is one of them; informant B often lengthened the final vowel of certain phonological words, which broke the utterance into groups, which were often distributed into an equal number of mps (Figure (4b), (5b), (6b), (10b)). Informant A, on the other hand, did not lengthened a vowel; however, she had a tendency to put the last two phonological words into one constituent or the last two mps into one prosodic constituent, as in Figure (4a), (6a) (P8 was not observed), (7a), (8a), (9a), and (10a).

Lastly, to answer question (c), I follow Kubozono (1989) and claim that the prosodic representation should be hierarchically organized because of the reasons I mentioned earlier; there is a downstep domain whose left edge is boosted but still under the influence of the higher downstep domain. This cannot be accounted for by the Strict Layer Hypothesis.

4 Conclusion

The Japanese downstep was examined and the relationship between syntax and the prosodic structure was investigated. The experiment was conducted with two native Tokyo Japanese speakers, and their productions of sentences with multiple pronominal modifiers in the left-branching and the right-branching condition were analyzed. The rationale behind this experiment is Kubozono’s (1989) proposals on MB and RB.

Contrary to Kubozono (1989), this study demonstrated that the correlation between syntax and the prosodic structure is weak in that the prosodic structure may make reference to syntax only when there are three mps in the utterance, regardless of syntactic branching direction. Hence, I proposed that prosody has its own structure independent of syntax, and that syntax is more accessible when there is a smaller number of mps in the utterance; with more mps, individual weight balance strategies become more important. In addition, I suggested that there is no need to distinguish between MB and RB, and that the prosodic structure should be represented hierarchically following Kubozono (1989).

The following questions remain unanswered. First, can this weak correlation between syntax and prosody be confirmed with syntactic configurations other than a sentence consisting of NP with multiple modifiers? Second, why do speakers depend on
two (or more) different phonetic means to organize their utterance prosodically? Third, why and how do the representations obtained by those different phonetic means demonstrate different organizations? These issues are open for future research.

References

Appendix 1

Figure 1a.

Figure 1b.

Appendix 2

Figure 2
Appendix 3  The left-branching condition

Figure 3a. Informant A
Figure 3b. Informant B

Figure 4a. Informant A
Figure 4b. Informant B

Figure 5a. Informant A
Figure 5b. Informant B

Figure 6a. Informant A
Figure 6b. Informant B
Appendix 4 The right-branching condition

Figure 7a. Informant A

Figure 7b. Informant B

Figure 8a. Informant A

Figure 8b. Informant B

Figure 9a. Informant A

Figure 9b. Informant B
Figure 10.a. Informant A

Figure 10.b. Informant B