Manifestation of downstep and intonation in Japanese: comparison of the Tokyo and Kochi dialects

Yasuko Nagano-Madsen
Department of Languages and Literatures, University of Gothenburg, Gothenburg

Abstract
This paper examines the manifestation of downstep and intonation in the Tokyo and Kochi dialects of Japanese by using three types of syntactically balanced material - adjective phrases, adverbial phrases, and sentence modifiers. The main conclusion is that Kochi speakers produce a smaller Major Phrase consisting of fewer lexical accents than in the Tokyo dialect, the Major Phrase being defined as the domain of downstep. In addition, the two dialects differ considerably as to the distribution of F0 peaks in downstepping the lexical accents, the distance being greater in the Tokyo dialect. These findings support a prosodic typology of Japanese dialects in which the weight of accent and intonation differ.

Introduction
When we hear the intonation of a given dialect, our perception may be that the pitch is very much ‘up and down’ or that it is very ‘flat’. Yamaguchi (1997) points out that one of the issues in Japanese dialect intonation research is to examine how words are distributed across larger units such as phrases and clauses, and how such units map to intonation units. Previous studies have pointed out dialectal differences in the manifestation of accent and intonation (Sugito 2001, Nagano-Madsen 2003, and Ishihara 2004 among others), but I am not aware of any studies that have compared two or more dialects with identical material directly.

This paper presents the preliminary results of downstep analysis in the Tokyo and Kochi dialects of Japanese which differ greatly as regards their auditory impression of intonation. The Kochi dialect belongs to the keihanshiki accent group and shares the basic features of the Osaka dialect.

Downstep
In Japanese, when two or more words make up a phrase or clause, the second and third accents are often realized in a lower pitch region than the first accent. This phenomenon has been referred to as jun-akusento ‘sub-accent’, catatheses, or downstep (henceforth downstep). Downstep in Japanese is triggered only by H*L lexical pitch accent and not by an unaccented word. Within the framework of intonational phonology, the domain of downstep in Japanese is said to be the Major Phrase (MaP) (Poser 1984, Pierrehumbert and Beckman 1988, Kubozono 1993).

![Figure 1. Schematic diagram of downstep. H*L=lexical pitch accent, MaP=Major Phrase.](image)

As for syntactic conditions, Kori (1997) suggests the following three categories as typical contexts where Japanese downstep occurs: (1) adjective phrases, (2) adverbial phrases, and (3) parallel expression. In addition, he also lists pragmatic focus as a typical downstep environment.

In the present paper, the syntactic conditions (1) and (2) above as well as a sentence modifier that contains either (a) or (2) are examined first in their mapping with intonation. Second the F0 peak value for each accent was measured and the difference in the F0 value between the accents was calculated.

Pitch range in semitone
In recent years, there has been a discussion concerning which acoustic unit most faithfully corresponds to the perception of pitch, the semitones or ERB (Equivalent Rectangular Bandwidth) (cf. Gussenhoven 2004). In the present study, semitone scale has been adopted to include a wide range of speakers differing in age and sex. In order to give an idea of the pitch range used in Japanese measured in
semitone, a comparison of the pitch ranges used in Chinese and Japanese speech is shown in Table 1. While the Chinese data is based on spontaneous speech, the Japanese data is based on the experiment on emotion produced by six voice actors. As such, the data presented in Table 1 has some limitations and the two languages cannot be directly compared. However, when measured in semitone, the similarity in pitch range between men and women is striking. For Japanese, it is around 14 semitones while in Chinese it is around 16 semitones.

Table 1. Pitch range in semitone in Chinese and Japanese. The data are adopted from Wang and Lin 2004 and Nagano-Madsen and Ayusawa 2011.

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>125.7 Hz</td>
<td>217.9 Hz</td>
</tr>
<tr>
<td></td>
<td>16.4 st</td>
<td>16.2 st</td>
</tr>
<tr>
<td>Japanese</td>
<td>13.9 st</td>
<td>14.3 st</td>
</tr>
</tbody>
</table>

**Experiment**

A list of 36 phrases/sentences containing adjective phrases, adverbial phrases, and/or sentence modifiers were constructed. Although the Tokyo and Kochi dialects differ greatly in their accent patterns, some basic words as well as loan words have the same accent type in the two dialects. Care was taken to choose as similar accent patterns as possible in order to make the experimental conditions essentially identical. Examples of phrases and sentence modifiers are as follows:

1. (1) adjective phrase
   
   *okaasan no prezento*
   
   ‘my mother’s gift’

2. (2) adverbial phrase
   
   *dondon arui-ta*
   
   ‘walked quickly’

3. (3) to have either (1) or (2) above as sentence modifier
   
   *okaasan-no prezento-ga todoi-ta asa-dat-ta*
   
   ‘(my) mother’s gift arrived’

The list was read at two speaking rates (slow and fast) by four Tokyo speakers and five Kochi speakers. Altering the speaking rate is a method that can be used in examining the invariant part of speech that remains constant regardless of the speaking rate. The subjects were men and women from their 20’s up to the age of 65. The peak F0 value of the lexical pitch accent was measured in both hertz and semitones but in the present paper the latter was adopted in order to facilitate a comparison between men and women.

**Results 1: mapping to syntactic units**

The F0 peak value of each H*L accent was measured. If the F0 value is higher than that of the preceding H*L accent, it is defined as pitch reset to start a new Major Phrase (MaP).

**Adjective and adverbial phrases**

Figure 2 compares the percentage of adjective phrases and adverbial phrases that were produced as single Major Phrases. As regards adjective phrases 98% (Tokyo) and 94% (Kochi) of the instances were produced as a single MaP while for adverbial phrases, 100% (Tokyo) and 90% (Kochi) of them were produced as a single MaP.

**Sentence modifiers**

Unlike in the case of adjective and adverbial phrases, the mapping between sentence modifiers and MaP in the two dialects differed considerably (cf. Figure 3). For the Tokyo dialect, over 70% of the sentence modifiers consisting of three to four accented words was
produced as single MaP in which downstep operates. In contrast, over 60% of the sentence modifiers were produced as two MaPs in the Kochi dialect.

This upstep rule (metrical boost) that raises the pitch value of the first accent when more and more accents follow has been pointed out by Kubozono (1993). By contrast, in the Kochi dialect, no such systematic F0 variation was observed. In other words, the Kochi dialect gives an auditory impression in which each H*L accent is evenly focused whereas in the Tokyo dialect, the contrast between the H*L accents is more clear-cut.

Figure 3. Number of intonation units (MaP) per single sentence modifier in Tokyo (light) and Kochi (dark) dialect.

Result 2: F0 characteristics

General observation

Figure 4 shows F0 contours of the adjective phrase ‘gift of mother’ in the Tokyo and Kochi dialects. In both dialects, the F0 value of the second H*L accent (‘present’) is lower than that of the preceding accent (mothers). However, the difference in the F0 value between the two accents in the Tokyo dialect is much greater than that in the Kochi dialect, i.e. the magnitude of downstep is greater.

Figure 5 shows F0 contours of the sentence ‘it was the morning when my mother’s gift arrived’ in which a sentence modifier is embedded. Even here, the difference between the F0 of lexical accents is greater in the Tokyo dialect.

In the Tokyo dialect, there is a strong tendency for the F0 value of the first H*L accent to rise as the number of subsequent accents increases. This upstep rule (metrical boost) that raises the pitch value of the first accent when more and more accents follow has been pointed out by Kubozono (1993). By contrast, in the Kochi dialect, no such systematic F0 variation was observed. In other words, the Kochi dialect gives an auditory impression in which each H*L accent is evenly focused whereas in the Tokyo dialect, the contrast between the H*L accents is more clear-cut.

Figure 4. Typical F0 contours for the adjective phrase okaasan-no purezento ‘gift from my mother’ in the Tokyo (light line) and the Kochi (dark line) dialects.

Figure 5. Typical F0 contours for the sentence okaasan-no purezento-ga todoita asa datta ‘It was the morning when the gift from my mother arrived’. The light line represents the Tokyo dialect and the dark line represents the Kochi dialect.

F0 values

The maximum F0 value of each H*L accent was measured in semitone and the difference between the successive accents was calculated within the domain of downstep (cf. Figure 6). The result shows that in both the Tokyo and Kochi dialects, the degree of downstep decreases in the order of adverbial phrase, adjective phrase, and sentence modifier. The average value for Tokyo and Kochi dialects within each category was 6.4 versus 4.2 semitones for adverbial phrases, 5.2 versus 3.5 semitones for adjective phrases, and 2.9 versus 2.5 semitones for sentence modifiers. For all three syntactic categories the values were lower in the Kochi dialect than in the Tokyo dialect.

In both dialects, a larger difference of F0 between successive accents was found in the adverbial phrase. This is because there is a tendency for speakers to emphasize the adverb even if they were asked to read it ‘neutral’ and not ‘focused’. The resulting intonation resembles focus intonation in the sense that the adverb has an expanded F0 while the F0 of the verb is compressed. However, even here, the difference in the F0 value between the successive accents is smaller in the Kochi dialect.
Figure 6. Difference in mean F0 values in semitone between two successive H*L accents. Tokyo (light) and Kochi (dark).

Conclusion
The correspondence between intonation units (MaP) and syntactic units such as adjective phrases, adverbial phrases and sentence modifiers was examined in the Tokyo and Kochi dialects of Japanese. Differences were found concerning the mapping between intonation units and syntactic units as well as in the F0 peak distribution for H*L accents within a single downstep unit. In the Kochi dialect, there is a tendency for a MaP involving downstep to have fewer H*L accents than in the Tokyo dialect. Tokyo speakers, on the other hand, have a more marked contrast between the H*Ls by showing a higher F0 value for the first H*L. Furthermore, there is a strong tendency for Tokyo speakers to raise the F0 value of the first H*L as more and more H*L follow. Particularly noteworthy is that the above-mentioned observations were not much affected by the speaking rate and therefore this tendency is best regarded as a social norm of the dialect.

These findings presented here support Sugito’s (2001) claims that intonation is more affected by syntactic structure in The Tokyo dialect while in the Osaka dialect, the distinctive feature of accent is more salient. It can be hypothesized that the manifestation of MaP in relation to syntax as well as to internal phonetic differences in F0 peak distribution represents a potentially rich dimension in the description of the prosodic typology of Japanese dialects.

References