OCP and downstep in Japanese
Manami Hirayama1, Hyun Kyung Hwang2, Takaomi Kato3
1Seikei University, Japan
2University of Tsukuba, Japan
3Sophia University, Japan
hirayama@fh.seikei.ac.jp, hwang.kyung.gu@u.tsukuba.ac.jp, t-kato@sophia.ac.jp

Abstract
This paper investigates the effect of the Obligatory Contour Principle (OCP) on downstep in Japanese. The hypothesis was that if the OCP is at work, a sequence of adjectives with the same endings would inhibit downstep compared to that of adjectives with different endings. A production experiment was conducted using ten speakers. The results indicate the presence of downstep in all four phrase conditions. The effect of accentuatedness differed among phrase types when item was removed from the model as a random factor, while the interaction was not significant when item was included in the model. Although it is not conclusive because of the small number of items per condition, the OCP might partially explain the smallest degree of pitch lowering in [Adjective-i [Adjective-i Noun]].

Index Terms: downstep, Japanese, OCP, adjective

1. Introduction
The Obligatory Contour Principle (OCP), originally proposed in tonal phenomena (e.g., [1, 2], et seq.), has been argued to account for other phenomena in phonology and also in the syntax-phonology interface (e.g., [3]), where identical elements are avoided to occur adjacently within a domain. In Optimality Theory, an output-driven theory, this is expressed as Identity Avoidance constraints.

The OCP has been said to be at work in Japanese segmental phonology as well, explaining cases such as the avoidance of multiple voiced obstructions in Renadaku ‘sequential voicing’ (e.g., [4]) and optional obstruent devoicing in loanword phonology ([5]).

In Japanese prosody, OCP was proposed to be a potential blocking factor in downstep (e.g., [6]) to account for the fact that the process was blocked in a sequence of two adjective (A) phrases with the same ending -i independently modifying a head noun ([A-i [A-i N]])) in one study [6] while a combination of two adjacent adjectives with different endings in the same structure ([A-i [A-na N]]) did not block downstep in another study [7].

Downstep in Japanese is a process in phrasal phonology in which an accented word triggers the pitch register of the following phrase to be lowered within the domain of a Major Phrase (e.g., [8, 9, 10]), as shown in the schematic pitch curves in Figure 1. In (a), the first accented phrase shirōi (the acute accent mark indicates lexical accent) triggers the next phrase nagāi to be downstepped, whereas in (b), the second phrase nagāi is not downstepped because the first phrase amai is unaccented. Consequently, the pitch peak in nagāi in (a) is higher than that of the downstepped nagāi in (b).

Figure 1: Schematic pitch curves in downstep (adopted from [6, 11]).

[6] proposed that part of the account of this downstep blocking in [A-i [A-i N]], but not in [A-i [A-na N]], may be related to the phrase having or not having two adjectives with the same endings adjacently. It may be the case that speakers found [A-i [A-i N]] somewhat unnatural so the downstep discontinued at the second adjective. If so, the blocking of downstep is accounted for as an OCP effect concerning forms with the same endings, preventing two successive adjectives with the same endings from being uttered over a single stretch of pitch lowering, i.e., downstep.

To test whether [A-i [A-i N]] was found to be unnatural by Japanese speakers, [11] conducted an experiment. They examined the acceptance of four phrases with all combinations of two adjectives, with endings -i and -na, modifying the head noun (i.e., [A-i [A-i N]], [A-i [A-na N]], [A-na [A-i N]], and [A-na [A-na N]]). [11] asked native Japanese speakers how natural the test phrases were to them on a scale of 1 to 6. The words included ooki ‘big’ and kashikoi ‘wise’ for A-i and ookina ‘big’ and rikōna ‘wise’ for A-na. Kashikoi and rikōna differ in terms of lexical strata: kashikoi is a native word and the root of rikōna is Sino-Japanese. The head noun was inu ‘dog’ (e.g., ooki [kashikoi inu]) or [A-i [A-i N]]). The stimuli were presented in Japanese orthography. The results of 61 participants supported the hypothesis that a phrase in which adjectives with the same endings are used adjacent would be less favoured than a phrase in which adjectives with different endings are used. In the mean score terms, [A-i [A-i N]] (3.68) was the least preferred, and [A-na [A-na N]] (4.31) was second least preferred, followed by [A-i [A-na N]] (4.88) and [A-na [A-i N]] (5.50). Furthermore, in addition to the report in [11] on the results of a linear mixed-effects analysis, our post-hoc pairwise comparisons with Tukey’s adjustment indicated that the mean scores were significantly different for all six pairwise comparisons. This suggests that noun phrases with the same adjective endings in a row are in fact found less natural by native speakers of Japanese than those with different adjective endings. If the unnaturalness causes speakers to stop downstepping in [A-i [A-i N]] in [6], OCP with respect to forms with the same endings can account for the blocking of downstep in this phrase.
This study investigates whether the OCP is in effect in downstep in Japanese in a production experiment, in particular using the four phrase types in [11] (i.e., [A-i [A-i N]], [A-i [A-na N]], [A-na [A-i N]], and [A-na [A-na N]]). If the OCP is a factor that can inhibit downstep, a phrase with the same ending in a sequence ([A-i [A-i N]] and [A-na [A-na N]]) might disfavor downstep, whereas downstep would not be blocked in other phrases ([A-na [A-i N]] and [A-i [A-na N]]). We will show below that our production experiment may indicate an OCP effect in [A-i [A-i N]] but that this is also not conclusive at this point.

2. Methodology

We conducted a production experiment to test whether a phrase in which two adjectives with the same endings modify the head noun shows an inhibitory effect on downstep.

2.1. Test phrases

The test phrases used all combinations of -i and -na adjectives in the [A [A N]] structure, as in (1)-(4). If downstep occurs, the target phrase in the accented trigger phrase (e.g., hirói in (1a)) is lower in terms of the pitch (f0) than the target phrase in the unaccented trigger phrase (e.g., hirói in (1b)).

(1) ii: [A-i [A-i N]]
   a. Accented trigger
      Trigger: aoi Target: hirói Head N: ié ‘blue large house’
   b. Unaccented trigger
      Trigger: kurai Target: hirói Head N: ié ‘dark large house’

(2) nana: [A-na [A-na N]]
   a. Accented trigger
      iyana Target: jimína Head N: ié ‘bad modest house’
   b. Unaccented trigger
      jamína Target: jimína Head N: ié ‘disturbing modest house’

(3) ina: [A-i [A-na N]]
   a. Accented trigger
      kurói Target: jimína Head N: ié ‘black modest house’
   b. Unaccented trigger
      kurai Target: jimína Head N: ié ‘dark modest house’

(4) nai: [A-na [A-i N]]
   a. Accented trigger
      hadéna Target: hirói Head N: ié ‘showy large house’
   b. Unaccented triggers
      jamína Target: shirói Head N: ié ‘disturbing white house’

There are two items each, totalling 16 phrases, put into a carrier phrase ‘ane-wa ___ to ita ‘(my) sister said __.’ Together with 16 distractor sentences that used sentence structures different from that of the test sentences, the list was near-randomized such that no two test sentences appeared one after another. Eight versions were prepared; thus, there were eight repetitions of the same sentence per speaker.

2.2. Recording, participants and measurements

Recordings were performed in a room with sound-attenuated walls. A Marantz recorder (PMD661), with a sampling rate of 44.1k Hz and 24-bit quantization, and a unidirectional dynamic headset microphone (SHURE SM10A, frequency response: 50-15,000 Hz) were used.

Ten speakers (two males and eight females) with the Tokyo accent system participated in the experiment. Their ages ranged from 20 to 22 years, and all received small monetary compensation for their participation. There was a small practice session before the main session.

After excluding 17 sentences that were mispronounced (in terms of, for example, accentuation) and went unnoticed during the recording sessions, 1,263 sentences remained for analysis.

Trigger phrases, target phrases, head noun phrases, and the subject phrases in the carrier (i.e., ane-wa) were manually annotated for their boundaries using Praat [12]. The pitch peak of each phrase was automatically measured using ProsodyPro [13], after which any f0 perturbations were manually taken care of.

3. Results

Figure 2 shows boxplots of the mean peak f0s in the target phrase for the four phrase types with different combinations of the adjectives in (1)-(4). The f0 values were normalized per speaker by using the scale function in R [14]. In each phrase type, the boxplot to the left (red) shows tokens for the accented trigger (e.g., (1a) for ii) and the one to the right (blue) unaccented trigger (e.g., (1b) for ii). On average, the peak f0 of the target phrase with the accented trigger (red) was lower than that with the unaccented trigger (blue), suggesting that downstep occurred in all four types.

![Figure 2: Boxplots of peak f0s (y-axis) in the four phrase types (x-axis).](image)

However, the difference in the target peak f0s between accented and unaccented triggers was not quite the same across the four types. The difference was slightly smaller for [A-i [A-
To test whether the accentedness affects the four phrase types in the same way, we conducted a statistical analysis using a linear mixed-effects model with R [14] and the lmerTest package [15]. The model is shown in (5), with the dependent variable being the peak f0 value (not z-score) and the predictors being accent (two levels: accented vs. unaccented), phrase type (four levels: ii vs. ina vs. nai vs. nana), and their interactions. Speaker and item were included as random effects in the model (random intercept and random slope for speaker and random intercept for item). The factors accent and type were sum-coded.

(5) peak f0 ~ accent*type + (1+accent|speaker) + (1|item)

A Wald chi-square test, carried out by using the Anova function of the car package [16], determined that the factors accent and type were significant, although the significance was marginal with type, while the interaction was not significant (Table 1).

<table>
<thead>
<tr>
<th>predictor</th>
<th>$\chi^2$</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accent</td>
<td>96.110</td>
<td>1</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Type</td>
<td>7.7724</td>
<td>3</td>
<td>0.050</td>
</tr>
<tr>
<td>Accent*type</td>
<td>2.4325</td>
<td>3</td>
<td>0.487</td>
</tr>
</tbody>
</table>

Table 2 presents the coefficient estimates and other relevant results. The analysis was carried out using the lmerTest package in R [15]. Together with the results in Table 1, overall, accent had a significant effect: the peak f0 for the unaccented trigger was significantly higher than that for the accented trigger (p < 0.001). In the accented trigger sentences, the peak f0 was significantly lower in the nana sentence than in the ii sentence (p < 0.05), but the interaction terms were not significant.

<table>
<thead>
<tr>
<th>predictor</th>
<th>Estimate</th>
<th>S.E.</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(intercept)</td>
<td>220.283</td>
<td>15.94</td>
<td>13.819</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>accentUnacc.</td>
<td>18.8894</td>
<td>2.73</td>
<td>6.910</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>typeina</td>
<td>-1.5560</td>
<td>2.00</td>
<td>-0.777</td>
<td>0.460</td>
</tr>
<tr>
<td>typenamei</td>
<td>-3.0839</td>
<td>2.00</td>
<td>-1.540</td>
<td>0.163</td>
</tr>
<tr>
<td>typenana</td>
<td>-5.5786</td>
<td>2.00</td>
<td>-2.786</td>
<td>0.024</td>
</tr>
<tr>
<td>acctU:typenain</td>
<td>0.2834</td>
<td>2.84</td>
<td>0.100</td>
<td>0.923</td>
</tr>
<tr>
<td>acctU:typennana</td>
<td>3.1438</td>
<td>2.84</td>
<td>1.106</td>
<td>0.301</td>
</tr>
<tr>
<td>acctU:typennnana</td>
<td>3.3817</td>
<td>2.84</td>
<td>1.190</td>
<td>0.268</td>
</tr>
</tbody>
</table>

When we tested how the accentedness affects in different phrase types by conducting a post hoc test between phrase types (e.g., ii vs. ina), no pairs were significantly different. This test was done by using the testInteractions function of the phia package (pairwise = accent). Overall, the above statistical examinations indicate that downstep occurred in the same manner for all phrase types. This suggests that the OCP was not in effect.

However, an important aspect of the statistical analysis should be reported. If we remove item from the model (6) and run with speaker only as the random effects, the interaction of accent and type is significant in a Wald chi-square test (so are accent and type as main effects), as shown in Table 3.

(6) peak f0 ~ accent*type + (1+accent|speaker)

Table 3: Wald chi-square test of predictors in the linear mixed-effects model (6).

<table>
<thead>
<tr>
<th>predictor</th>
<th>$\chi^2$</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accent</td>
<td>113.2320</td>
<td>1</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Type</td>
<td>22.2844</td>
<td>3</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Accent*type</td>
<td>7.0271</td>
<td>3</td>
<td>0.071</td>
</tr>
</tbody>
</table>

This indicates that the effect of accent differences is different for different phrase types. As mentioned previously, the f0 peak difference between the accented and unaccented triggers was smaller in [A-i [A-na N]] and [A-i [A-na N]] (both 20 Hz) than in [A-na [A-i N]] (22 Hz) and [A-na [A-na N]] (23 Hz). Therefore, if this result with model (6) is more representative of a larger population, the OCP may partially explain the smallest degree of pitch lowering in [A-i [A-na N]]. However, given the different statistical results depending on the treatment of item, it is inconclusive at this point whether this truly represents the behaviour of the phrase types in general.

4. Discussion

This study examined whether there was an OCP effect on downstep in Japanese, particularly OCP in terms of a sequence of adjectives with the same ending, -i or -na. The results indicate that downstep occurred in all four phrase types. The degree of downstep, measured as the effect of accent, significantly differed when item was removed from the model. However, no such effect was observed when item was treated as a random factor. Although an experiment using a larger dataset is necessary, we discuss two issues regarding our results.

The first is the presence of downstep in [A-na [A-na N]]. (Recall that this phrase was the second least preferred (after [A-i [A-na N]]) by Japanese speakers in [11].) Although the OCP could explain the results in [A-i [A-na N]] in this study, it could not explain [A-na [A-na N]] because the degree of downstep was found to be the largest in this phrase. If OCP were the explanation, [A-i [A-na N]] and [A-na [A-na N]] would pattern similarly. Here, we propose that the different patterns between A-i and A-na may be due to the different morphological behaviours between these two types of adjectives. In Japanese morphology, unlike A-i adjectives, A-na adjectives show a noun-like pattern and are thus called the nominal adjectives or adjectival nouns. Because downstep occurs robustly in the noun phrase [N-no [N-no N]] [6, 7], it is not surprising to find [A-na [A-na N]] patterning in a manner similar to [N-no [N-no N]] if A-na is atn (adjectival) noun.

The second discussion involves [A-i [A-i N]], which showed blocking of downstep in [6] but did exhibit downstep in this study, albeit to the smallest degree among the phrase types investigated (the other three types were not examined in [6]). Thus, we observed interspeaker variation. A similar variation was found in the downstep patterns when the test phrases included modifications with adjectives in the past tense forms (see [11] for references). In general, there is robust evidence of interspeaker variation in prosody. For example, sentences with high and low attachments are prosodically different in their production (and the listeners can perceive the
cues) when the referential contexts are identified by the speaker; however, without the awareness of such contexts, speakers may not produce the syntactic differences in their prosody (e.g., [17]). It can be the case that participants in [6] were more aware of the denotation of the phrases than those in this study.

5. Conclusion

This study investigated whether there is an OCP effect concerning adjective forms in Japanese downstep. The hypothesis was that if OCP in terms of adjective endings was at work, downstep would be inhibited in a phrase where there are two adjectives in a sequence with the same endings (e.g., [A-i [A-i N]]) whereas downstep would not show such an inhibition in a phrase where adjectives with different endings are adjacent to each other (e.g., [A-na [A-i N]]).

The results showed that overall, downstep occurred for all four phrase types examined in this study. However, the significance of the degree of downstep (the degree of peak f0 differences between phrases with accented and unaccented triggers) was not the same depending on whether item was included in the model as a random factor.

Given the small number of items used in this study (two in each condition), it is safer to conclude at this point that it is not conclusive which analysis is true for the population. In the future, it will be important to test this hypothesis by including additional items included in the experiment.

6. Acknowledgements

We thank the audience and the anonymous reviewers of TAI 2023 for their comments. We also thank Yoonjung Kang for her suggestions regarding the statistical analyses. All errors are our own.

This research is funded by JSPS KAKENHI grants 19K00613 and 21K00498.

7. References