Article

Lexical Category and Downstep in Japanese

Manami Hirayama 1,* , Hyun Kyung Hwang 2 and Takaomi Kato 3

1 Department of English, Seikei University, Tokyo 180-8633, Japan
2 Faculty of Humanities and Social Sciences, University of Tsukuba, Tsukuba 305-8571, Japan;
hwang.kyung.gu@u.tsukuba.ac.jp
3 Graduate School of Languages and Linguistics, Sophia University, Tokyo 102-8554, Japan; t-kato@sophia.ac.jp
* Correspondence: hirayama@fh.seikei.ac.jp

Abstract: In pursuing the mapping between syntax and phonology/prosody, little attention has been paid to the kinds of syntactic information that can affect prosody. In this paper, we explore Japanese downstep, a process in phrasal phonology. What syntactic information affects downstep and what does not? Specifically, do lexical categories affect downstep? We investigate the effects of nouns, adjectives, and verbs in different syntactic settings (e.g., [X1 [X2 N]], [[X1 X2] N], predicative X) through production experiments. We found that adjectives in [X1 [X2 N]] may block downstep, whereas adjectives in other structures as well as nouns and verbs generally do not block it. We analyze this phonological patterning as being derivative of an interaction between syntactic structures and lexical categories.

Keywords: downstep; Japanese; parts of speech; syntax-prosody mapping

1. Introduction

In the literature on the syntax–phonology interface, syntactic information is often considered visible in phonology (Nespor and Vogel 1986; Selkirk 1984; Truckenbrodt 1995, et seq.). This is particularly true with respect to sentence-level syntax, as reflected in prosody. For example, the contrast between a statement and question is often made using different intonation patterns. Since Japanese downstep is a process in phrasal phonology, it offers an excellent case for testing the hypothesis that syntactic information is mapped onto prosody. In Japanese downstep, an accented phrase triggers the phrase that follows to be rendered in a lower pitch register (e.g., Kubozono 1989; Pierrehumbert and Beckman 1988; Poser 1984), as in (1a). Figure 1a depicts the pitch contour. The accented word あく ‘blue’ triggers the word that follows it, ながい ‘long’, to be rendered in a lower pitch (the acute accent mark indicates a vowel in the accented syllable), and ながい further triggers the next word, ねぎ ‘leek’, to be rendered in an even lower-pitched register. In contrast, in (1b), as shown in Figure 1b, the word あく ‘sweet’ does not trigger downstep because it is unaccented; thus, the pitch peak in the word that follows it, ながい ‘long’, is not as low as the pitch peak of the word in the same position, ながい, in (1a).

The major phrase (MaP) is the domain of downstep (see Igarashi 2015; Ishihara 2015 for a review). For example, in (1a) and Figure 1, focusing on the noun phrase あくながいなげ ‘green and long leek’, the whole phrase constitutes a single MaP, with each word forming a minor phrase (MiP), a phrase that allows at most one accent: ( (あく)MaP (ながい)MiP (なげ)MaP)MaP. Moreover, as discussed extensively in Section 2, downstep is reportedly sensitive to certain syntactic information, including whether a given constituent is a maximal projection (i.e., an XP) (Selkirk and Tateishi 1991), branching structures (e.g., Kubozono 1989, 1992; Ito and Mester 2013), and the part of speech of a given word (Hirayama and Hwang 2016, 2019; Hwang and Hirayama 2021; Selkirk and Tateishi 1991). The left edges of relevant syntactic elements are presumably mapped onto the left edges of MaPs, which then block downstep.
The general question regarding the syntax-prosody interface we are concerned with in this paper is what kinds of syntactic information can affect phrasal phonology. An exploration of the literature on downstep shows that different syntactic structures and different kinds of syntactic boundaries have been discussed. What affects the process and what does not? In what follows, we address one specific aspect of this question on which the literature offers different views, namely whether and how parts of speech affect downstep. The main goal of this paper is to shed new light on this issue.

1. **Japanese Downstep and Syntax**

This section reviews the literature on the interaction between Japanese downstep and syntax. The review shows that different kinds of syntactic information may block downstep, while not all syntactic boundaries affect it. We discuss the effects of maximal projection boundaries (Selkirk and Tateishi 1991), the boundary between the subject noun phrase (NP) and predicate verb phrase (VP) (Hirayama and Hwang 2019; Ishihara 2016), relative clause boundaries (Hirayama and Hwang 2019), and parts of speech (Selkirk and Tateishi 1991; Hirayama and Hwang 2016, 2019; Hirayama et al. 2019; Hwang and Hirayama 2021) on downstep.

Selkirk and Tateishi (1991) argue that the left edges of maximal projections, or XPs, block downstep. In other words, the left edges of XPs are mapped onto the left edges of MaPs, affecting the realization of downstep by blocking it. Ishihara (2019) finds that the effect is larger when there are multiple left edges of XPs, although there is interspeaker variation. Ishihara (2016) reports that the left edge of a predicate VP variably blocks downstep.

Figure 1. Pitch curves of (1) [A female speaker]: (a) *ane-wa aóku nagái négi to itta*; (b) *ane-wa amaku nagái négi to itta*. (9).

This paper is organized as follows. Section 2 reviews the literature on downstep and factors that reportedly affect it. Section 3 presents the methodology of the production experiment that we conducted to test whether downstep realization is sensitive to parts of speech, or more specifically, whether adjectives, nouns, and verbs have different effects on the process. The results are reported in Section 4 and discussed in Section 5. Section 6 presents the study’s conclusions.
whereas Hirayama and Hwang (2019) (and this study) do not find that this particular type of boundary affects downstep in this way.

Clause boundaries have not been intensively studied with downstep, although they have been argued to support the presence of the Intonational Phrase in Japanese (e.g., Kawahara and Shinya 2008; Selkirk 2009; Ishihara 2019). Furthermore, Hirayama and Hwang (2019) tested whether relative clauses would affect downstep, finding that the left edge of the relative clause does not block downstep.

Lexical categories have been reported to affect downstep, but the literature does not agree on how they do this. Selkirk and Tateishi (1991) argue that downstep occurs in a phrase in which adjectives (A) are involved, whereas it is blocked when nouns (N) are involved: they find that downstep occurs between the adjectives in \([A_1 \ [A_2 N]]\) but not between \(N_1\) and \(N_2\) in \([N_1 \ [N_2 N_3]]\). They explain this result by postulating that downstep is blocked by XP boundaries and that (a) adjectives do not project XPs while nouns do, and (b) \([A_2 N]\) does not constitute a maximal projection. Downstep is not blocked in \([A_1 \ [A_2 N]]\) because there is no XP boundary to the left of \(A_2\). However, the process is blocked in \([N_1 \ [N_2 N_3]]\) because there is an XP boundary to the left of \(N_2\).

The empirical results of Hirayama and Hwang (2016), Hwang and Hirayama (2021), and Kubozono (1992) are opposite to those reported in Selkirk and Tateishi (1991). Kubozono (1992) reports that downstep occurs in \([N_1 \ [N_2 N_3]]\). Hirayama and Hwang (2016) and Hwang and Hirayama (2021), like Selkirk and Tateishi, compared \([A_1 \ [A_2 N]]\) and \([N_1 \ [N_2 N_3]]\), and investigated whether downstep occurred. However, they increased the number of speakers in their experiment and adopted the more traditional definition of downstep found in the Japanese literature (e.g., Kubozono 1988, 1989, 1993; Pierrehumbert and Beckman 1988; Poser 1984). The results suggest that downstep occurred in \([N_1 \ [N_2 N_3]]\), in particular at \(N_2\), the target of downstep, as in (2), but it was blocked at \(A_2\) in \([A_1 \ [A_2 N]]\). The target words/phrases of the process are underlined. Downstepped targets are indicated with ! in (2) to (5) below, although the head nouns are not marked, as that is not the focus of investigation here.

2. a. An example of \([N_1 \ [N_2 N_3]]\)

\begin{verbatim}
номо-но нара-но маем Nomo-GEN Nara-GEN bean

‘Nomo’s beans from Nara’
\end{verbatim}

b. An example of \([A_1 \ [A_2 N]]\)

\begin{verbatim}
shiroi nagai mamé

‘long white beans’
\end{verbatim}

Hirayama and Hwang (2019) tested adjectives and verbs in their past forms, but retained the right-branching structure, as in (3): \([VPAST]_{RC} [PASTE]_{RC} [N]_{NP}\) vs. \([VPAST]_{RC} [VPAST]_{RC} [N]_{NP}\). 2

3. a. An example of \([VPAST]_{RC} [PASTE]_{RC} [N]_{NP}\)

\begin{verbatim}
[ниринда]_{RC} [(0)дарукatta]_{RC} magol NP

stare.PAST tired.PAST grandchild

‘a grandchild who stared disfavourably and was tired’
\end{verbatim}

b. An example of \([VPAST]_{RC} [VPAST]_{RC} [N]_{NP}\)

\begin{verbatim}
[наижинда]_{RC} [ниринда]_{RC} magol NP

get adjusted.PAST stare.PAST grandchild

‘a grandchild who got adjusted and stared disfavourably’
\end{verbatim}

They also tested these parts of speech in the predicative position in non-relative clauses, as in (4): \(N-ga A, N-ga V\).
4. a. An example of N-ga *magó-ga* 渝 nemui
   grandchild-NOM sleepy
   "(Someone’s) grandchild is sleepy'.

   b. An example of N-ga *magó-ga* 渥 nirámu
   grandchild-NOM stare
   "(Someone’s) grandchild stares (at someone) disfavourably'.

Hirayama and Hwang (2019) report that downstep occurred in all conditions (i.e., (3) and (4)) but note that in (3), the pattern was much more robust in the verb condition (3b) than in the adjective condition (3a) in that all speakers showed downstep in the former while there was interspeaker variation in the latter.

Hirayama et al. (2019) extended the investigation of downstep to remaining possibilities researched in Hirayama and Hwang (2016, 2019) and Hwang and Hirayama (2021), and tested verbs in their nonpast forms in relative clauses, nouns accompanied by the past tense form of a copula in relative clauses, and nouns in the predicative position in non-relative clauses, as in (5): [[V_NONPAST RC [[V_NONPAST RC N]] NP, [[V_PAST RC [[N Copula_PAST RC N]] NP, N-ga *N*. They report having found downstep in all conditions.

5. a. An example of [[V_NONPAST RC [[V_NONPAST RC N]] NP
   get lost.NONPAST worry.NONPAST grandchild
   ‘a grandchild who gets lost and worries’

b. An example of [[V_PAST RC [[N Copula_PAST RC N]] NP
   worry.PAST no good Copula grandchild
   ‘a grandchild who worried and was no good’

c. An example of N-ga *N*
   magó-ga námi
   grandchild-NOM Nami
   ‘(My) grandchild is called Nami’.

   Table 1 recapitulates the results obtained in Hirayama and Hwang (2016, 2019), Hirayama et al. (2019), and Hwang and Hirayama (2021) with respect to the presence and absence of downstep and particular parts of speech (nouns, adjectives, and verbs). It can be observed that adjectives may block downstep when they are in attributive use, modifying the head noun, whereas adjectives in predicative use do not block downstep, and nouns and verbs do not block downstep regardless of the type of use (attributive or predicative).

Table 1. Presence (yes) and absence (no) of downstep in Hirayama and Hwang (2016, 2019), Hirayama et al. (2019), and Hwang and Hirayama (2021) (adapted from Hirayama et al. 2019, Table 3, with slight modification).

| Attributive ([X [X N]]) | Nonpast | Adjective | Verb |
|--------------------------|---------|-----------|------|
|                          | Yes     | No        | Yes  |
| Past                     | Yes     | Yes/No 1 | Yes  |

1 Interspeaker variation noted.

As a source of the different patterning between different lexical categories, or rather, the patterning of adjectives in attributive use thereby blocking downstep, Hirayama et al. (2019) consider the (un)naturalness of NPs containing adjectives as used in the experiments. For example, an NP that has two adjectives individually modifying the head noun ([A [A N]])
as in (2b) [shiroi [nagai mame]] ‘white, long beans’, does not sound quite natural in Japanese, although it is not ungrammatical. They propose that because of this unnaturalness, (some) speakers inserted a phonological phrase (i.e., MaP) boundary between the two adjectives, which resulted in the downstep being blocked. The same can apply to the other NPs that have an adjective in the past tense form, as in (3a). They point out that when two adjectives are used to modify a noun, the structure where the first one appears in the -te (or gerundive) form (i.e., [[A-te A] N]) sounds more natural (e.g., [shiroku-te nagai mame] ‘white and long beans’).\(^5\)

To summarize past findings on possible syntactic effects on downstep in Japanese, clause boundaries such as the relative clause boundary do not appear to affect it, whereas phrase-level information (the left edges of XPs) may affect it. Another line of investigation is concerned with parts of speech, although there is debate as to which categories block the process.

3. Materials and Methods

3.1. Speech Materials

Of the different types of syntactic information, we focus on the relation between lexical categories and downstep. Specifically, we test the effect of adjectives, nouns, and verbs on downstep by paying careful attention to the structures of the test sentences. Recall from Section 2 that Hirayama et al. (2019) point out that a particular structure (i.e., [X [X N]]) might have yielded unnaturalness, causing a MaP boundary to be inserted to block downstep, especially when adjectives are involved. In this experiment, we use structures in which the combination of two modifying items forms a constituent ([[X X] N]), rather than individually modifying the head noun, so that the sentences are more natural for the speaker. As shown in the following, the structure [[X X] N] involves a left-branching constituent [X X], but it is not recursive in that the first X does not modify the second X. We return to this structural aspect in Section 5.

We prepared two structures for nouns and verbs and one structure for adjectives. In all these structures, the target of downstep is X\(_2\) in [[X\(_1\) X\(_2\) N]]. First, for nouns, we prepared a structure in which X\(_1\) and X\(_2\) are nouns, and X\(_1\) is accompanied by the particle -to ‘and’, as in (6a), as well as a structure in which X\(_1\) and X\(_2\) are nouns, and X\(_1\) is accompanied by -de, the continuative form (or renyoukei) of the copula -da, as in (6b).

6. a. [[N\(_1\) N\(_2\) N\(_3\)]] where N\(_1\) is accompanied by -to
   [[nêgi-to rámù-no] nábe]
   ‘hot pot that has leek and lamb’
   
   b. [[N\(_1\) N\(_2\) N\(_3\)]] where N\(_3\) is accompanied by -de
      [[mûmî-de rámù-no] nábe]\(^6\)
      tasteless-DE lamb-GEN hot pot
      ‘hot pot that is tasteless and has lamb’

For verbs, we prepared a structure in which X\(_1\) and X\(_2\) are verbs, and X\(_1\) is in the continuative form, as in (7a), and a structure in which X\(_1\) and X\(_2\) are verbs, and X\(_1\) is in the -te form, as in (7b).\(^7\)

7. a. [[V\(_1\) V\(_2\) N]] where V\(_1\) is in the continuative form
   [[nómî nâyâmu] mâmî]
   drink worry Mami
   ‘Mami, who drinks and gets worried’

   b. [[V\(_1\) V\(_2\) N]] where V\(_1\) is in the -te form
      [[nôn-de nâyâmu] mâmî]
      drink-TE worry Mami
      ‘Mami, who drinks and then gets worried’

Finally, in the structure we prepared for adjectives, X\(_1\) and X\(_2\) are adjectives, and X\(_1\) is in the continuative form, as in (8).
8. \([A_1 A_2] \text{ N}\) where \(A_1\) is in the continuative form

\[
\begin{array}{c}
\text{blue} \\
\text{long} \\
\text{leek}
\end{array}
\]

‘green and long leek’

There are no adjective test phrases in which \(A_1\) in \([A_1 A_2] \text{ N}\) is in the -te form, because adjectives are always accented in this form; hence, we cannot test whether (paradigmatic) downstep has occurred without a comparable phrase with an unaccented item in \(X_1\).

To test whether downstep occurs, we follow the traditional (i.e., paradigmatic) understanding of Japanese downstep (e.g., Kubozono 1989; Pierrehumbert and Beckman 1988; Poser 1984; see Note 2) and compare these phrases, as in (6)–(8), with others containing an unaccented item in the position that precedes the target. Downstep is judged to be present if the pitch peak of the target position in a phrase with an accented trigger, as in (9a), is lower than the pitch peak of the target position in a phrase with an unaccented trigger, as in (9b).\(^8\)

Recall that downstep is triggered by an accented word. Therefore, a phrase with an accented trigger, as in (9a), constitutes a downstep environment, whereas downstep would not occur if the trigger position is occupied by an unaccented item, as in (9b).

9. a. Accented trigger (=8))

\[
\begin{array}{c}
\text{blue} \\
\text{long} \\
\text{leek}
\end{array}
\]

‘green and long leek’

b. Unaccented trigger

\[
\begin{array}{c}
\text{sweet} \\
\text{long} \\
\text{leek}
\end{array}
\]

‘sweet and long leek’

For each of the five structures seen above (two for nouns, two for verbs, and one for adjectives), there are two accented vs. unaccented pairs, totalling 20 (5 × 2 × 2) phrases (see Appendix A for a full list of these phrases). Eight filler phrases are not discussed in this paper because they are intended for other studies. All 28 phrases are inserted into the carrier phrase \(\text{ane-wa \_\_\_\_}\) ‘(my) sister said \_\_\_\_’. The sentences are pseudo-randomized five times, yielding five lists for participants to read aloud.

3.2. Recording and Speakers

Recording was done in a studio with sound-attenuated walls. We used a Marantz digital recorder (PMD661) with a 44.1 kHz sampling rate and 24-bit quantization. The microphone was a unidirectional dynamic headset microphone (SHURE SM10A, frequency response: 50–15,000 Hz). There was a practice session at the beginning, in which sentences in the structures discussed in Section 3.1 were used, although they included lexical items different from the test items.

Twelve Japanese speakers from Tokyo and nearby areas participated in the study. They were all university students (1 male and 11 female speakers, mean: 19.2 years old, range: 18;3–21;0). Their dialect (i.e., Tokyo dialect) is comparable to the one discussed in previous studies on Japanese downstep. One speaker was removed from the analysis because her utterance was generally creaky, particularly during the test phrases. Tokens read with unexpected accentuation that went unnoticed during the recording were also removed from the analysis \((n = 14)\). One speaker’s recording was clipped for one list, such that there are four repetitions rather than five for that speaker. This left us with 1066 sentences for analysis.

3.3. Acoustic Analysis and Statistics

The peak f0 in each phrase including the carrier’s subject/topic phrase \(\text{ane-wa \_\_\_\_}\) ‘(my) sister-TOP’ was measured, as shown in (10), where the phrase boundaries are marked with a vertical bar (\(|\)). Figure 1 also illustrates the boundaries. The f0 measurements were performed using Praat (Boersma and Weenink 2020) and running a script called ProsodyPro (Xu 2013) on the phrase intervals made manually.
To examine whether the target phrases were downstepped, linear mixed-effects analyses were conducted on the relationship between the peak f0s (Hz) of the target and trigger accentuation (accented vs. unaccented) using R (ver. 3.6.2, R Core Team 2019) and the lmerTest package. Speakers were included as random effects (random intercepts) in the model. Items were not included as random effects because when models with and without items as random effects were compared using anova, the results did not show any significant differences ($p > 0.05$) in any of the conditions reported below.

### 4. Results

Figure 2 shows the representative f0 contours of $[[V_1 V_2] N]$ (7a) where $V_1$ is in the continuative form. The peak f0 of the target in the accented condition (solid line) is quite lower than that in the unaccented condition (dotted line) (see more contours in Figure 1 for (9)).

![Figure 2](image.png)

**Figure 2.** Representative f0 contours (Hz) for the continuative form of verbs with an accented (solid line) and unaccented trigger (dotted line).

Table 2 provides the results. As seen, accentedness has a significant effect on the peak f0 values in all conditions examined, with the f0 peaks in the target position being higher when the trigger is unaccented than when it is accented. These results can be interpreted as downstep being present in all the conditions examined in the study.

| Expression                  | β (Hz) | t     | p     |
|-----------------------------|--------|-------|-------|
| $[[N_1\text{-}te N_2] N]$ (6a) |        |       |       |
| (intercept)                 | 215.590| 16.48 | <0.001|
| Trigger unaccented          | 29.569 | 25.87 | <0.001|
| $[[N_1\text{-}de N_2] N]$ (6b) |        |       |       |
| (intercept)                 | 219.688| 16.88 | <0.001|
| Trigger unaccented          | 22.547 | 17.45 | <0.001|
| $[[V_1 V_2] N]$ (7a)        |        |       |       |
| (intercept)                 | 215.88 | 17.66 | <0.001|
| Trigger unaccented          | 23.25  | 19.70 | <0.001|
| $[[V_1\text{-}de V_2] N]$ (7b) |        |       |       |
| (intercept)                 | 211.403| 16.79 | <0.001|
| Trigger unaccented          | 26.102 | 23.20 | <0.001|
| $[[A_1 A_2] N]$ (8)         |        |       |       |
| (intercept)                 | 219.684| 17.36 | <0.001|
| Trigger unaccented          | 23.430 | 16.01 | <0.001|

Table 2. Results of the linear mixed-effects analyses.
Figures 3–5 show the mean f0 peak values for the adjective, verb, and noun tokens, respectively. The topic phrase is the subject *ane-wa* ‘(my) sister-TOP’ in the carrier phrase. The figures indicate that in mean terms, the peak pitch of the target position is always higher when the preceding word (i.e., the one in the trigger position) is unaccented (dashed lines), compared to when it is accented (solid lines), indicating that there is downstep in all cases. When we looked at individual graphs for each speaker (not given here), unlike in Hirayama and Hwang (2019), there does not seem to be interspeaker variation in the adjective patterns (or in the patterns of the other categories), and downstep is robustly found.

![Figure 3](image3.png)

**Figure 3.** Mean f0 peaks (Hz) for adjectives (solid line: accented trigger, dashed line: unaccented trigger).

![Figure 4](image4.png)

**Figure 4.** Mean f0 peaks (Hz) for verbs (solid line: accented trigger, dashed line: unaccented trigger).
Before testing the syntax-prosody mapping proposals against the patterns in Table 3, we present the prosodic phrasing suggested by the data in Table 3 based on the assumption that the downstep domain is MaP (Igarashi 2015; Ishihara 2015 and references therein). First, the left edge of the intermediate constituent in a right-branching structure (i.e., the left edge of [X X] in [X [X X]]) is variably mapped onto the left edge of a MaP, as shown in (11). When X₂ is an adjective, a MaP boundary can be inserted before it, or the whole phrase makes a single MaP (11a).¹⁰ When X₂ is a noun or verb, the whole phrase makes a MaP (11b).

Table 3. Presence (yes) and absence (no) of downstep in Hirayama and Hwang (2016, 2019), Hirayama et al. (2019), Hwang and Hirayama (2021), and this study.

| Attributive      | Noun | Adjective | Verb |
|------------------|------|-----------|------|
| [X [X N]]        | Nonpast | Yes | No | Yes |
|                  | Past | Yes | Yes/No ¹ | Yes |
| [[X X] N]        | Yes | Yes | Yes | Yes |

¹ Interspeaker variation noted.
11. MaP phrasing for \([X_1 [X_2 N]]\)
   a. \(X_2 = A\)
   Variation between \(((X_1)_{\text{MiP}})_{\text{MaP}} ((X_2)_{\text{MiP}} (N)_{\text{MiP}})_{\text{MaP}}\) and \(((X_1)_{\text{MiP}} (X_2)_{\text{MiP}} (N)_{\text{MiP}})_{\text{MaP}}\)

   b. \(X_2 = N, V\)
   \(((X_1)_{\text{MiP}} (X_2)_{\text{MiP}} (N)_{\text{MiP}})_{\text{MaP}}\)

On the other hand, when no right-branching structure is involved, i.e., in the left-branching structure \([[X_1 X_2] N]\) (12a) and \([N_{=\text{ga}} X]\) (12b), the whole phrase makes a single MaP.

12. MaP phrasing for \([[X_1 X_2] N]\) and \([N_{=\text{ga}} X]\)

   a. \([[X_1 X_2] N]\)
   \(((X_1)_{\text{MiP}} (X_2)_{\text{MiP}} (N)_{\text{MiP}})_{\text{MaP}}\)

   b. \([N_{=\text{ga}} X]\)
   \(((N_{=\text{ga}})_{\text{MiP}} (X)_{\text{MiP}})_{\text{MaP}}\)

As reviewed in Section 2, Selkirk (2009) and Selkirk and Tateishi (1991) argue in line with Align Theory that the left edge of maximal projections (XPs) is aligned with the left edge of MaP, blocking downstep. This is not readily tenable given the presence of downstep in our data, for example, in the N2 in \([N_1-no [N_2-no N_3]]\) (see Table 3). A standard syntax would project a maximal projection NP for the N2. Thus, the left edge of this XP would be mapped onto the left edge of a MaP and downstep would be blocked. However, downstep was robustly found there.

Kubozono (e.g., 1989, p. 59; 1992, p. 385) proposes a recursive MiP (not MaP) in his prosodic representation, reflecting the syntactic difference between phrases with the left-branching \([[X X] N]\) and right-branching \([X X N]\) structures. In his data, downstep occurs in both phrases, and the prosodic representations proposed are \(((X)_{\text{MiP}} (X)_{\text{MiP}})_{\text{MiP}} (N)_{\text{MiP}})_{\text{MaP}} for \([[X X] N]\) and \(((X)_{\text{MiP}} (N)_{\text{MiP}})_{\text{MiP}})_{\text{MaP}} for [X X N]. Note the recursive MiPs in these representations. However, a recursive MiP account does not explain our data as the \([X N]\) in the right-branching \([X X N]\) needs to have a MaP boundary to the left of it, as in (11a), to explain the downstep blockage when the second X is A. (Kubozono uses the branching structure in the explanation of the phenomenon called the metrical boost, which he analyzes as occurring in addition to downstep at the beginning of an intermediate constituent in the right-branching structure (Section 2)).

The idea of recognizing syntactic recursivity as reflected in recursive prosody in Japanese (and other languages) is discussed elsewhere as well (e.g., Ito and Mester 2012, 2013). In particular, Ito and Mester (2013, p. 34ff.) analyze the prosodic phrasing of left-branching NPs \([[N_{-\text{no}} N_{-\text{no}}} N]\) (containing recursive NPs) and right-branching NPs \([N_{-\text{no}} [A\text{-i} N]]\) as an interaction of the syntax-prosody mapping that uses Match Theory (e.g., Selkirk 2009), where syntactic XPs are mapped onto recursive Phonological Phrases (\(\phi\)), with phonological constraints (such as requirement for binarity for the Phonological Phrase and prohibition of recursivity). In their model, there is a possibility of prosodic phrasing that can assume additional bracketing for the right-branching structure. In fact, in their syntactic bracketing (p. 34ff.), the NP with the right-branching syntax \([[X_1] [X_2] N]\) is more complex than the NP with the left-branching one \([[X_1] X_2 N]\) in that it has more structure (note the additional bracket pair for the former). With this structural difference in mind, their syntax-prosody mapping would yield additional prosodic bracketing for the constituent \([[X_2] N]\) in the right-branching NP: \(((X_1)_{\text{\phi_{min}}} (X_2)_{\text{\phi_{min}}} (N)_{\phi_{min}})_{\phi_{max}}\). Recursive Phonological Phrases like this may explain the contrast between the left-branching and right-branching difference with downstep as seen in Table 3. Here, the intermediate Phonological Phrase \(((X_2)_{\phi_{max}} (N)_{\phi_{max}})\), in particular, the left boundary of this prosodic constituent, may block downstep. The variable nature of downstep blocking in Table 3 can be accounted for if the level of this constituent (intermediate \(\phi\)) is acknowledged and the claim that prosodic effects become cumulatively strong as more boundaries coincide at higher levels (Fougeron 2001, et seq.;
Ishihara 2019) is adopted. The boundary here is not as strong as the boundary at the maximal PhPhrase (ϕ_{max}), resulting in variable blocking of downstep.

Note, however, that in Table 3 downstep is not always (variably) blocked at the beginning of an intermediate constituent in a right-branching structure. It is blocked only when A is involved. This can be explained if we explore syntax-prosody mapping and adopt the architecture of prosodic hierarchy where a lower-level constituent is exhaustively contained in an immediately higher-level constituent as in the Strict Layer Hypothesis (e.g., Selkirk 1984; Nespor and Vogel 1986). In Table 3, the right-branching structure with A may actually involve an embedded clause, here a relative clause: [X]_{RC} [[A]_{RC} N]. A in the past tense form, as in (3a), projects a relative clause. A in the nonpast tense form (with the suffix -i), as in (2b), may also do so (e.g., Kuno 1973; Yamakido 2005; see Yamakido 2000 for other references). If embedded clauses are mapped onto Intonational Phrases or PClauses (Ishihara 2019 and references therein), which is a level higher than the MaP/Phonological Phrase, the left edge of the embedded clause that houses the A is mapped onto the left edge of an Intonational Phrase/PClause. Assuming that the edges of a higher-level prosodic category coincide with those in an immediately lower level prosodic category, the left edge of the Intonational Phrase/PClause is at the left edge of a MaP, which would block downstep. This is also compatible with the fact that the A in the predicate position [N-

The above accounts are all based on the syntax-prosody mapping hypotheses. Different syntactic branching structures are mapped onto different recursive PhPhrase structures. A syntactic clause is mapped onto an Intonational Phrase/PClause, and the syntactic difference between A and N in terms of their behaviour in clause projection provides different PClause phrasing. However, the pattern with V in Table 3 still cannot be explained, since V, having inflection regarding tense, would project a clause just like A if we assume a standard syntax. Then, the abovementioned accounts would predict that downstep would be blocked when V is involved in the right-branching structure [X [V N]], which is not the case in the data and downstep is robust there.

How are the differences in parts of speech, in particular the anomaly of A in the downstep data in Table 3, explained? The above discussion reveals that syntax may not be enough. That itself is not surprising because there are analyses in the literature on prosody that do not rely on syntax. One example is the information structure: focus is often said to affect prosodic phrasing. Other pragmatic cues have also been argued to be reflected in prosody such as illocutionary force (e.g., Selkirk 2009). Below, we explore accounts in terms of semantics, another field that deals with meaning, pragmatics, and an interaction with the baseline condition (i.e., phrases with an unaccented trigger).

N, A, and V differ in terms of their denotation. Placed in the right-branching structure [X₁ [X₂ N]], in which the two Xs individually modify the head noun and thus do not form a constituent, the semantic properties of the categories in X₂ may cause a conflict with X₁ in parsing from X₁ to X₂, in which case a MaP is created beginning with X₂, blocking downstep. If the Xs are verbs, X₁ and X₂ are interpreted to have a certain semantic relation. For example, in [mayōu [nayám̩u magó]] (5a), although mayōu ‘get lost’ and nayám̩u ‘worry’ individually modify magó ‘(one’s) grandchild’, one can easily identify a causal relation between the actions the verbs denote, i.e., the grandchild gets lost and as a result, gets worried. We can generalize the semantic relation in question as being temporal as the verbs typically denote actions/events: the actions/events conveyed by V₁ and V₂ are interpreted in such a way that one of them temporally precedes the other. This temporally ordered relation ensures a natural flow between the two verbs in [V₁ [V₂ N]], which results in a single phonological phrase (MaP). In contrast, when X₂ in [X₁ [X₂ N]] is an adjective
(i.e., \([A_1 \ [A_2 \ N]], \ [V.\text{past} \ [A.\text{past} \ N]])\), since adjectives denote a state (or a property) and not an event, \(X_1\) and \(X_2\) cannot forge a temporal relation of the kind observed with verbs. For example, \([\text{shirōt} \ [\text{naagāi \ mamē}]]\) ‘white, long beans’ (2a), an example of \([A_1 \ [A_2 \ N]],\) is difficult (if not impossible) to interpret in such a way that the two states described by the adjectives are temporally ordered. Rather, since adjectives convey a state, when the two Xs are adjectives (i.e., \([A_1 \ [A_2 \ N]],\) the states of \(A_1\) and \(A_2\) temporally coexist without being related in terms of any temporal precedence. In order to refer to such a situation (e.g., beans that are white and long), native speakers would probably disfavor the right-branching structure \([A_1 \ [A_2 \ N]]\) (2a). They would rather prefer to use another construction, for example, the one in which the first adjective is in the continuous form, \([[\text{shirō-ku \ naagāi] mamē}]]\), as in (8). The reason for preferring a particular structure over the other could be to avoid a sequence of forms with the same -i ending (presumably a type of Obligatory Contour Principle (OCP) or identity avoidance). In fact, in a study, wherein we asked 61 Japanese participants their preference among four NPs with two adjective endings \([A-i \ [A-i \ N]], \ [A-na \ [A-na \ N]], \ [A-i \ [A-na \ N]],\) and \([A-na \ [A-i \ N]]\) in terms of the naturalness scale from 1 (not sounding Japanese) to 6 (natural as Japanese), \([A-i \ [A-i \ N]]\) sounded least natural to them to a statistically significant degree compared to others. In \([V.\text{past} \ [A.\text{past} \ N]]\) (e.g., \([\text{niránda} \ [\text{darákatta \ magō}]\) ‘a grandchild who stared disfavourably and was tired’ (3a)), since verbs are involved, a temporal relation is expected between the verb action/event and adjective state. However, \([V.\text{past} \ [A.\text{past} \ N]]\) fails to be interpreted in such a way that the state of \(A\) begins after the event of \(V\) begins and ends or that the event of \(V\) begins after the state of \(A\) begins and ends. Crucially, however, it can be interpreted to mean that the event of \(V\) begins after the state of \(A\) has begun. For example, in (3a), the event of the grandchild staring disfavourably began after the start of their state of being tired. In this interpretation, there is no temporal precedence relation between the event of \(V\) and state of \(A\) in a strict sense, but it is still possible to claim that there is a partial temporal precedence relation between the two. The existence of this relation can be the source of the interspeaker variation in the presence of downstep in \([V.\text{past} \ [A.\text{past} \ N]]\) (see Table 3): Some speakers needed a strict temporal precedence relation between the two Xs in \([X_1 \ [X_2 \ N]]\) in order to make a single MaP over them. As such, they inserted a MaP boundary at the left edge of \(A\) in \([V \ [A \ N]]\). If a noun appears in \(X_2\) in \([X_1 \ [X_2 \ N]]\) (i.e., \([N_1-\text{no} \ [N_2-\text{no} \ N]], \ [V.\text{past} \ [N \ \text{datta} \ N]]\) when a verb is involved, i.e., in \([V.\text{past} \ [N \ \text{datta} \ N]]\), a semantic relation similar to the one found in verbs exists between the two Xs. When accompanied by datta, the past tense form of a copular, a noun can denote an event (e.g., dame-datta can mean ‘failed’), and thus it is verb-like, which creates a certain temporal precedence relation between the 

\(V\) and \(N\). In \([N_1-\text{no} \ [N_2-\text{no} \ N]]\), the semantically underspecified nature of the particle -no and frequency may play a role. It has been argued (e.g., in den Dikken and Singhapreecha 2004) that Japanese -no, like, for example, English of and Spanish de, is a linker that encodes a variety of semantic relations for the nouns it links. Because of this, and presumably with the help that the construction \(N-\text{no} \ N\) is very frequent in Japanese, speakers can smoothly parse from \(N_1\) to \(N_2\) in \([N_1-\text{no} \ [N_2-\text{no} \ N]]\), creating a single downstep domain.

Another possible factor for the MaP boundary to the left of \(A\) in \([X \ [A \ N]]\) is focus. One possible interpretation of the blocking of downstep is that the speakers somehow emphasized the \(A\), which resulted in placing the focus on that element. (The source of the focus placement could be the unnaturalness created by the structure and category \(A\) as discussed above.) Since focus can be realized even when the item is downstepped (e.g., Ishihara 2016), a study is necessary to test this account in which we carefully control for focus in examining downstep in \([X \ [A \ N]]\).

Another possible line of explanation for the absence of downstep in \([X \ [A \ N]]\) is based on variable phrasings in the baseline condition, i.e., phrases with an unaccented trigger. In Figure 3, the f0 peaks from the Trigger to Target show a downtrend in the baseline condition ([AA] [N] U), which is not observed, at least not to the same degree, in Figure 4 (verb) and Figure 5 (noun). This can suggest that there is a MiP boundary between \(X_1\) and \(X_2\) in [[X_1
When $X_1$ is unaccented and adjectives are involved, while this is usually not the case and the two Xs would make a single MiP when $X_1$ is unaccented. If there is indeed a MiP boundary with adjectives but not with verbs and nouns, this may partly explain the different patterns that adjectives demonstrate compared to nouns and verbs. This should also be further investigated.

6. Conclusions

We investigated questions with respect to the syntax–phonology interface with particular focus on downstep in Japanese. What kinds of syntactic information can affect phrasal phonology such as downstep? Do particular parts of speech affect downstep? If so, what does that mean linguistically? The results of the production experiment in this study, together with past research results, suggest that adjectives may block downstep if they are in a particular syntactic position of right-branching $[X_1 \ [X_2 \ N]]$, that is, with the two Xs individually modifying the head noun. However, adjectives did not affect downstep in other structures such as $[[X_1 \ X_2] \ N]$, where the two Xs form a left-branching constituent to the exclusion of N, or in the predicate position of non-relative clauses (N- ga $X$). Nouns and verbs did not affect downstep when they were located in any of these positions.

We explored accounts in terms of syntax-prosody mapping hypotheses discussed in the literature, finding that the potential site for downstep blocking, i.e., the beginning of an intermediate constituent in the right-branching structure, can be accounted for with recursive PhPhrases: an intermediate level PhPhrase may block downstep. The contrastive pattern between the N and A can be explained as the clause projection that A makes being mapped onto the PClause as opposed to no such projection with the N. However, syntax-prosody mapping cannot explain the difference between A and V. We then explored several possible factors in terms of semantics and pragmatics. Different parts of speech have different kinds of denotations, and an unsuccessful semantic relation held between $X_1$ and $X_2$ results in a prosodic phrasing such that a MaP boundary is inserted between the two Xs. Thus, the interaction between the syntactic structure and parts of speech explains the phonological patterning in the phrasal phonology.

Author Contributions: Conceptualization, M.H., H.K.H. and T.K.; methodology, M.H., H.K.H. and T.K.; formal analysis, M.H., H.K.H. and T.K.; data curation, M.H.; writing, M.H., H.K.H. and T.K.; funding acquisition, M.H., H.K.H. and T.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by JSPS (Japan Society for the Promotion of Science) KAKENHI grant JP19K00613.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by Seikei University Research Ethics Committee (protocol code SREC 17-17, approval date 6 October 2017).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author with the permission of the participant. The data are not publicly available due to the permission status of the data.

Acknowledgments: We very much thank two anonymous reviewers and the Academic Editors for their valuable comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.
## Appendix A. List of Test Phrases

### 1. Adjectives

| Trigger accent | Trigger | Target | Head N |
|----------------|---------|--------|--------|
| Accented       | aoku    | nagái  | négi   |
|                | green   | long   | leek   |
| Unaccented     | umaku   | nagái  | négi   |
|                | good    | long   | leek   |
| Unaccented     | maruku  | nagái  | négi   |
|                | round   | long   | leek   |
| Unaccented     | amaku   | nagái  | négi   |
|                | sweet   | long   | leek   |

### 2. Verbs

| Trigger accent | Trigger | Target | Head N |
|----------------|---------|--------|--------|
| Accented       | nómi    | nayámú|mámi   |
|                | drink   | get worried | Mami |
| Accented       | yómi    | nayámú|mámi   |
|                | read    | get worried | Mami |
| Unaccented     | umi     | nayámú|mámi   |
|                | give birth to | get worried | Mami |
| Unaccented     | yobi    | nayámú|mámi   |
|                | call    | get worried | Mami |
| -te            | nónde   | nayámú|mámi   |
| Accented       | yónde   | nayámú|mámi   |
|                | read    | get worried | Mami |
| Unaccented     | unde    | nayámú|mámi   |
|                | give birth to | get worried | Mami |
| Unaccented     | yonde   | nayámú|mámi   |
|                | call    | get worried | Mami |

### 3. Nouns

| Trigger accent | Trigger | Target | Head N |
|----------------|---------|--------|--------|
| Accented       | négi-to | rámú-no| nábe   |
|                | leek-and | lamb-NO | hot pot |
| Accented       | násu-to | rámú-no| nábe   |
|                | eggplant-and | lamb-NO | hot pot |
| Unaccented     | nira-to | rámú-no| nábe   |
|                | Chinese chive-and | lamb-NO | hot pot |
| Unaccented     | momo-to | rámú-no| nábe   |
|                | peach-and | lamb-NO | hot pot |
| Accented       | múmi-de | rámú-no| nábe   |
|                | tasteless-and | lamb-NO | hot pot |
| Accented       | múmi-de | négi-no| nábe   |
|                | tasteless-and | leek-NO | hot pot |
| Unaccented     | mee-de   | máma-no|mámi   |
|                | (my) niece-and | (a) mother-NO | Mami |
| Unaccented     | mee-de   | áná-no | mámi   |
|                | (my) niece-and | news caster-NO | Mami |
Notes
1 The categories MaP and MiP are referred to as an intonation phrase and accentual phrase, respectively, in the ToBI frameworks (see Ishihara 2015, p. 570 for the variation in the terminology in the literature).
2 In Selkirk and Tateishi (1991), downstep is defined differently than it is in most studies on Japanese downstep. They use the so-called syntagmatic diagnostic, in which the presence/absence of downstep is determined within a sentence (e.g., Kubozono 2007; Ito and Mester 2013). Other scholars have also taken the syntagmatic approach (e.g., Hirotani 2005; Nagahara 1994). Other researchers use paradigmatic—as opposed to syntagmatic—diagnostic, in which the presence/absence of downstep is judged by comparing sentences with an accented phrase before the target phrase and those with an unaccented phrase in the same position, as in this study (see Section 3). See Ishihara (2015, pp. 585–86) for other methodological issues with Selkirk and Tateishi (1991).
3 The different results between Selkirk and Tateishi (1991) and other works discussed here (Hirayama and Hwang 2016; Hwang and Hirayama 2021; Kubozono 1992) are not due to dialectal differences. In all the works, the participants are speakers of Tokyo Japanese, i.e., the dialect that has the Tokyo accentuation system.
4 Attributive adjectives in the nonpast tense forms as in (2b) may also project relative clauses (see, e.g., Kuno 1973). We return to this in Section 5.
5 It would be important to confirm this native speaker intuition with an experiment of naturalness ratings of native speakers. We leave the investigation in future work.
6 The word *mimi ‘tasteless’ may be ambiguous between a noun and an adjectival noun (aka nominal adjective). Some dictionaries (e.g., Yamada et al. 2012) treat it as a noun and others as both categories. Here, we treat adjectival nouns as a subclass of nouns because they pattern with nouns rather than with adjectives in downstep (Hirayama and Hwang 2019). It is important to further investigate the patterns that adjectival nouns demonstrate with respect to downstep in future works.
7 We follow Takano (2004) in translating the phrase in (7b) using ‘and then’.
8 Although unaccented items do not trigger downstep, we call them triggers in this paper if they are in the position before the targets.
9 The result showing the statistically significant effect of accentedness remains the same if semitone is used instead of Hz as the unit of pitch measurement.
10 The prosodic phrasing in (11) and (12) is possible when words are all accented. If they are unaccented, the phrasing should be made differently. See, for example, Ito and Mester (2013) for a review of prosodic phrasing in Japanese. Also in the phrasing in (11) and (12a), we assume that the head nouns are downstepped.
11 We thank an anonymous reviewer for pointing this out.
12 An analysis involving clausal boundaries in German can be found in Féry (2015). We thank the editors for drawing our attention to this work.
13 Selkirk (2009) uses a theory in which the illocutionary force of the sentence is represented in a functional head Force₀, projecting a Force Phrase. To this extent, the account refers to the syntax, not directly to the pragmatics.
14 Given this analysis, several predictions can be made, as pointed out by an anonymous reviewer. For example, if the two verbs are in the -teiru forms, the verbs can denote states rather than events, in which case the temporal precedence relation would not be expected to hold between the two verbs (see the discussion on adjectives below). If so, the flow from V₁ to V₂ may be disturbed and downstep may be blocked as a result. Furthermore, if the two verbs in the non-teiru forms are reversed and if a temporal precedence relation is not held with that order, downstep may also be blocked as a consequence. We leave these predictions to test in future research.
15 This predicts that if the two As in [A₁ [A₂ N]] are in different tense forms, downstep is not blocked within the whole NP. The temporal precedence relation is held between A₁ and A₂ because one is in the past and the other not in the past. Since there is not a conflict in terms of the semantic relation, there will not be a MaP boundary at the beginning of the A₂ and prosodic phrasing would be ((A₁)MaP (A₂)MaP (N)MiP)MaP.
16 The mean scores were 3.68 for [A-i [A-i N]], 4.31 for [A-na [A-na N]], 4.88 for [A-i [A-na N]] and 5.50 for [A-na [A-i N]]. A mixed-effects analysis (with speaker as random effects) revealed that the scores for [A-na [A-na N]] (β = 0.6230, p < 0.01), [A-i [A-na N]] (β = 1.1967, p < 0.001) and [A-na [A-i N]] (β = 1.8197, p < 0.001) are individually higher than the score for [A-i [A-i N]].
17 We thank an anonymous reviewer for drawing our attention to the focus and MiP boundary in an unaccented trigger discussed in the next paragraph.

References
Boersma, Paul, and David Weenink. 2020. Praat: Doing Phonetics by Computer [Computer Program]. Available online: http://www.praat.org/ (accessed on 20 January 2022).
den Dikken, Marcel, and Pornsiri Singhapreecha. 2004. Complex noun phrases and linkers. Syntax 7: 1–54. [CrossRef]
Féry, Caroline. 2015. Extrapolation and prosodic monsters in German. In Explicit and Implicit Prosody in Sentence Processing. Studies in Honor of Janet Dean Fodor. Edited by Lyn Frazier and Edward Gibson. Amsterdam: Springer, pp. 11–37.
