Variable Pitch Accent and Prosodic Phrasing in Japanese Adjectival Complex DPs

Le Xuan Chan¹, Rina Furusawa², Seunghun J. Lee³, ¹

¹National University of Singapore, Singapore, ²International Christian University, Japan, ³IIT Guwahati, India

lxlinguistics@gmail.com, furusawaling@gmail.com, seunghunjlee@gmail.com

Abstract

This study investigated how complex DPs in Japanese are prosodically phrased, and whether an interaction occurs between pitch accent and prosodic phrasing. Using f0 cues, we found that complex DPs made up of two adjectival modifiers and a head noun can have two possible phrasings: a recursive structure (Adj1 (Adj2 N)) and non-recursive ternary phrasing (Adj1 Adj2 N). Out of four possible combinations of lexical pitch accent for each adjective, three combinations exhibited a non-recursive phrasing, whereas AU (accented Adj1 and unaccented Adj2) showed recursive phrasing. Our results show that for complex DPs in Japanese, a non-recursive flat phrasing seems to be the default prosody structure, but AU depicts a recursive pattern, which we attribute to phonetic restrictions on f0 realization.

Index Terms: prosodic phrasing, recursivity, pitch accent, complex DPs, Japanese

1. Introduction

This paper provides an investigation of prosodic phrasing in Japanese complex DPs and its interaction with lexical pitch accents. Unlike stress-accent languages such as English which mark prominence with postlexical pitch accents on the prominent “head” of a prosodic unit, pitch accents in Japanese are lexically specified, and prominence is marked through phrasing, which governs how these lexical pitch accents are realized [1].

Acoustically, F0 cues have been used to diagnose such prosodic phrasing in Japanese. The most well-documented example is downstream, whereby a sequence of accented words show downstream within a phonological phrase [2, 3]. [4] show that in coordinated clauses, the intonation phrase is marked by a final H tone, while in Fukuoka Japanese, an H tone plateaus throughout the intonation phrase in wh-questions [5].

Few studies examine prosodic phrasing involving complex DPs: a DP with more than one modifier. A DP with two modifiers can display an interplay between syntax and prosody because prosody is realized as prosodic recursion [6] with the closest adjective forming a phonological phrase φ(Adj1 φ(Adj2 N)φ), or the prosodic realization may be ternary φ(Adj1 Adj2 N)φ. Ternary phonological patterns are rare, but they have been metrical in reported in languages such as Estonian and Cayuvaha [7]. [8] points out that ternary prosodic structure still remains an empirical issue to be explored (but cf. [9]). The prosodic pattern of Japanese complex DPs examined in this paper suggests that the flat prosodic phrasing with a ternary phrase is preferred to the recursive prosodic phrasing, unless phonetic pressures to realize pitch accent exist.

The rest of the paper is structured as follows. Section 2 outlines the methods used for this study, while the analysis and results of prosodic phrasing and lexical pitch accents is provided in Section 3. Section 4 provides a discussion of the results, before a brief conclusion in Section 5.

2. Methods

Complex DPs made up of two adjectival modifiers (Adj1/Adj2) and a head noun (N) were elicited, as shown in (1) embedded in a carrier sentence.

(1) Gakkoo-de, omoi marui iruka dake hakkiri mieta

‘At school, I clearly saw heavy round dolphins only’

This structure provides speakers freedom to produce the DP as one entire phonological phrase, as in (Adj1 Adj2 N), or to place a left boundary between Adj1 and Adj2, as in (Adj1 (Adj2 N)), where Adj2 marks the beginning of a separate phonological phrase. Tokyo Japanese distinguishes between lexically accented (A) and unaccented (U) words, where accented words are specified for a H*L pitch accent while unaccented words do not have this specification. All possible accent combinations for Adj1-Adj2 are analyzed: accented-accented /AA/, unaccented-unaccented /UU/, unaccented-accented /UA/, and accented-unaccented /AU/.

Sixteen sentences were constructed for each accent combination: eight attached to an accented noun, and eight attached to an unaccented noun, yielding 64 items per speaker. A total of 768 tokens across two repetitions were elicited from 6 native speakers of Tokyo Japanese in their early twenties.

After elicitation, we observed that many unaccented adjectives in our data showed accentuation, i.e. realizations with a H*L pitch accent that is not lexically specified. Previous studies [10, 11] have observed these patterns for unaccented adjectives in Japanese, and noted that this is a variable phenomenon between speakers. To avoid skewing the analysis, we first manually coded for the surface pitch accent of each token, and categorized them into the following patterns: [AA], [UU], [UA], and [AU]. The subsequent analysis is based on these surface pitch accents. The results of the coding are in Table 1.
Table 1: No. of tokens by underlying and surface pitch accent

<table>
<thead>
<tr>
<th></th>
<th>underlying (intended)</th>
<th>surface realization</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>192</td>
<td>320</td>
</tr>
<tr>
<td>AU</td>
<td>192</td>
<td>198</td>
</tr>
<tr>
<td>UA</td>
<td>192</td>
<td>108</td>
</tr>
<tr>
<td>UU</td>
<td>192</td>
<td>142</td>
</tr>
</tbody>
</table>

After coding for surface pitch accent, each token was then manually coded for the presence (1 or 0) of a left boundary between Adj1 and Adj2. All items were coded by linguistically trained research assistants, and all discrepancies between raters were resolved. \( f_0 \) cues described in previous literature [12] were used to determine the presence of an Adj2 left boundary. The specific \( f_0 \) cues for each accent pattern are discussed in Section 3.

The research questions for the analysis are as follows: i) Do a noun and its closest modifier form a prosodic phrase, i.e. is there a left boundary between Adj1 and Adj2?, and ii) Do pitch accent combinations affect such DP internal prosodic grouping, i.e. the insertion of this boundary?

### 3. Analysis and Results

#### 3.1 Accented Adj2: AA and UA

Fig. 1 shows a sequence of accented words taken from [12]: Naoya-no ani-no wain-o ‘Naomi’s brother’s wine’, where each word is lexically-specified with a H*L pitch accent. When strung together, the pitch accent in each word is still realized, as indicated by the \( f_0 \) peaks, but the \( f_0 \) peak of each accented word is downstepped relative to its preceding word; In acoustic terms, max \( f_0 \) is lowered from Naoya to ani, and from ani to wain. This is referred to as the domain of downstep.

![Figure 1: Sequence of accented words showing downstep. Data from [12]; Recording is our own.](image)

Applying this to our data, when Adj2 is accented, we look at whether or not a pitch reset occurs at Adj2. A pitch reset at Adj2 cancels downstep, and in doing so resets the domain of downstep, signifying a left boundary before Adj2. This is illustrated in Fig. 2, which shows time-normalized GAM \( f_0 \) smooths. \( f_0 \) is normalized by speaker and expressed as cents. The black contour shows the realization of ternary grouping similar to the example provided in Fig. 1: Adj2 is downstepped relative to Adj1, and the Noun is downstepped relative to Adj1. The red contour shows the realization of a recursive grouping, when a left boundary is inserted between Adj2 and Adj1. Rather than being downstepped, a pitch reset occurs at Adj2, resulting in similar max \( f_0 \) values between Adj2 and Adj1.

![Figure 2: AA sequence. Black: ternary. Red: recursive. \( f_0 \) and time are normalized by speaker.](image)

For UA combinations shown in Fig. 3, downstep does not occur even in a ternary grouping. As shown in the black \( f_0 \) contour, an initial rise occurs at the unaccented Adj1, and the high tone is carried over to the accented Adj2, where we see the H*L accent realized. In contrast, the red contour indicates a pitch compression of Adj1 and a pitch reset in Adj2, indicating the insertion of a boundary between Adj1 and Adj2. Acoustically, this results in a substantial difference in max \( f_0 \) between Adj1 and Adj2, where a ternary non-recursive production (black) would show a minimal difference.

![Figure 3: UA sequence. Black: ternary. Red: recursive. \( f_0 \) and time are normalized by speaker.](image)

#### 3.2 Unaccented Adj2: UU and UA

What about when Adj2 is unaccented? Fig. 4 shows a sequence of two unaccented words – Naomi and ane attached to an accented head noun wain, ‘Naomi’s sister’s wine’. The focus here is on the \( f_0 \) movement in [Naomi-no ane]. Both words are unaccented and do not take on a H*L pitch accent. Naomi, however, takes on a %L boundary tone marking a left-boundary of the phrase, as it is in the phrase-initial position. This %L boundary tone is often referred to as “initial lowering”, and has the effect of a rising \( f_0 \) contour on Naomi. Crucially, ane, which is not adjacent to a phrase boundary, does not show this initial lowering, and is produced with a constant H tone.

![Figure 4: Sequence of unaccented words showing initial lowering. Data from [12]; Recording is our own.](image)
In our data, when Adj2 is unaccented, the presence of initial lowering was used to determine whether a left boundary is inserted between Adj1 and Adj2. Fig. 5 shows a UU pattern. The black f0 marks a ternary grouping without an Adj2 boundary, and is similar to the example provided in Fig. 4 – initial lowering occurs at Adj1 resulting in a rising f0 contour, and plateaus over to the end of Adj2. The red f0 contour, however, shows initial lowering occurring for both Adj1 and Adj2, and a similar f0 rise is seen in Adj2 is in Adj1. This indicates the insertion of a boundary between Adj1 and Adj2.

![Figure 5: UU sequence. Black: ternary, Red: recursive. f0 and time are normalized by speaker.](image)

For AU patterns in Fig. 6, the black contour shows that in a ternary grouping, the unaccented Adj2 undergoes pitch compression after the accented Adj1. As unaccented words are defined not by a certain pitch accent but by the absence of an accent specification, an unaccented Adj2 that is not adjacent to a boundary is free of both lexical and postlexical intonation (e.g. initial lowering) and is realized as an extension of the L tone following the H*L pitch accent from the preceding Adj1. In contrast, the red contour shows initial lowering and a rising f0 contour at Adj2 after the accented Adj1. After the H*L pitch accent in Adj1, f0 rises at Adj2, indicating the beginning of a separate grouping and an insertion of a left boundary.

![Figure 6: AU sequence. Black: ternary, Red: recursive. f0 and time are normalized by speaker.](image)

3.3 Coding results

After coding for phrasing, a mixed-effects logistic regression model in R was then fitted, with surface pitch accent and underlying pitch accent as fixed effects, while speaker, item, and repetition were included as random effects. The results of the coding is shown in Table 2. AU has more tokens with a left boundary of the Adj2, whereas the other three exhibit a ternary structure without the left boundary.

<table>
<thead>
<tr>
<th>Table 2: No. of tokens by pitch accent and phrasing</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(Adj1 Adj2 N)</td>
<td>(Adj1 (Adj2 N))</td>
</tr>
<tr>
<td>AA</td>
<td>126</td>
</tr>
<tr>
<td>AU</td>
<td>39</td>
</tr>
<tr>
<td>UA</td>
<td>115</td>
</tr>
<tr>
<td>UU</td>
<td>75</td>
</tr>
</tbody>
</table>

The bootstrap 95% CIs for each accent pattern (Fig. 7) and the logistic regression model confirm that with AA taken as the intercept, UA (z=1.5, p=.14) and UU (z=1.3, p=.18) do not show any significant difference, while an effect was found for AU (z=7.3, p<.001). AA, UA, and UU all are more likely to take on a non-recursive phrasing, while AU depicts a pattern of recursive phrasing, with (z=7.3, p<.001).

Numerically, however, AU tokens show a higher number of recursive phrasing compared to AA and UA.

![Figure 7: Bootstrap 95% CIs by accent pattern. Y-axis indicates the proportion of tokens realized recursively.](image)

Combining the results of phrasing with the f0 patterns described in Sections 3.1 and 3.2, we now have a better picture of how each accent combination is produced. Figure 8 illustrates the GAM f0 smooths of preferred realizations of each accent pattern.

![Figure 8: GAM f0 smooths for each accent pattern.](image)

In the AA pattern (top left), the H*L pitch accent in Adj2 is preserved but downstepped after Adj1; In the UA pattern (bottom left), f0 rises at Adj1 and plateaus over to Adj2 before culminating in the H*L pitch accent; In the UU pattern...
(bottom right), f0 rises at Adj1 and stays high throughout Adj2, although some effects of downtrend are observed. On the other hand, a left boundary is inserted between Adj1 and Adj2 for AU (top right), resulting in initial lowering and a rising f0 contour at Adj2. To sum up, AA, AU and UA exhibit ternary non-recursive phrasing, whereas AU shows a recursive phrasing.

4. Discussion

The main research question posed in Section 2 asked how complex DPs in Japanese are phrased. When a head noun is directly modified by two adjectival modifiers, does it form a prosodic phrase with its closest modifier (i.e. Adj2), or is the entire DP phrased in a ternary grouping? Evidence from AA, UA, and UU show a preference for a ternary grouping, without a left boundary between Adj1 and Adj2.

Recall that ternary grouping displays a flat prosodic structure that does not reflect the recursive syntactic structure. Our results suggest that a flat ternary prosodic grouping is preferred to syntactically matching prosodic structure unless there is a pressure to realize a less-flat prosodic structure.

The question arises as to why AU displays a recursive prosodic phrasing different from the other three accent combinations. One possible account is that some phonetic restriction plays a role: there is a minimum threshold preventing excessive f0 lowering in produced speech.

After the pitch falls on the penultimate mora in Adj1, speakers are presented with a choice: maintain the ternary grouping and extend the L tone from Adj1 to the unaccented Adj2, or insert a left boundary before Adj2 and realize Adj2 with a rising F0 contour. If speakers choose the former strategy, f0 will continue to downtrend towards the end of the phrase, which may cause difficulty in perceiving speech. This is in line with previous observations where f0 plays an important role in noise [13]. Thus, between the two realizations with or without the initial rising, the former is chosen in order to avoid excessive f0 lowering at an early point in the phrase.

For the other three accent patterns, the restriction of a minimum f0 threshold does not seem to constric the realization of a ternary grouping. For instance, the accented Adj2 of AA still contains a pitch rise even when downstepped, as f0 has to rise from the left edge in the realization of the H*L pitch accent. For UA and UU, the restriction does not apply, as the lack of a falling H*L accent in the unaccented Adj1 results in the H tone from Adj1 carrying over to Adj2. Hence, only the realization of the unaccented Adj2 in AU is constrained by this restriction, which inserts a left boundary at Adj2 to prevent excessive f0 lowering.

Physiological/phonetic restrictions have been reported to alter the realization of lexical accents such as speech rate in [14]. Our results with the AU phrasing extend these effects and inform us that the phonetic realization of lexical accents can also have an impact on how the elements are prosodically phrased, providing evidence that lexical pitch accents interact directly with prosodic phrasing.

This notion challenges the way we think of lexical and postlexical prosody. The realization of lexical prosody, such as stress in English or pitch accent in Japanese, is generally thought of as constrained by postlexical prosody, such as phrasing or discourse structure. An example of this is how phrasing induces downstep in accented Japanese words [2], or focus induces deaccentuation in English [1]. Under this understanding, lexical pitch accent should not have an effect on how complex DPs are phrased in our study, and any differences arising between accent combinations would only be in their realizations, not in phrasing. If the difference between AU and other accent combinations is due to phonetic restrictions, this would mean that prosodic phrasing is malleable and can be overridden by speakers’ intention to produce more perceivable speech.

Another question then arises – if rephrasing is involved, is the difference in AU a mere case of speakers putting focus on Adj2? In Japanese, focus has been conventionally thought of as inducing rephrasing [1], while more recent literature have argued that the effects of focus do not indicate rephrasing [12]. Yet, this still does not answer the question – why does AU pattern differently from the other accent combinations? Regardless of whether focus induces rephrasing, if we assume that the AU is backed by focus, there is still a need to account for the discrepancy between AU and the other accent combinations. This leads us back to the point of phonetic restrictions and speakers producing perceivable speech.

5. Conclusion

To sum up, this paper has shown that Japanese complex DPs generally show a preference towards ternary phrasings, contributing to existing research on how such structures are produced and phrased by speakers. On the other hand, the exception of AU challenges aspects of our understanding of prosody and intonation, and requires us to pay closer attention towards phonetic and physiological interactions with underlying prosodic structures.

6. Acknowledgements

This work was partially supported by JSPS KAKENHI Grant #20K00578 (POP), the ILCAA joint research project “Phonetic typology from cross-linguistic perspectives (PhonTyp)” (2021-2023). We thank Honoka Asai, Kotone Sato, Keitaro Mitsuhashi, and Rin Tsjita for the data collection and processing, and Michinori Suzuki and Toshio Matsuura for comments on the stimuli.

7. References


