Investigation of Validity of Paradigmatic Diagnosis for Downstep in Japanese*

Kei Furukawa
Division of Information Science
Nara Institute of Science and Technology
Ikoma, Japan
furukawa.kei.fi4@is.naist.jp

Satoshi Nakamura
Division of Information Science
Nara Institute of Science and Technology
Ikoma, Japan
s-nakamura@is.naist.jp

***Abstract*—This study investigates the validity of paradigmatic diagnosis for downstep in Japanese and seeks to ascertain whether downstep is directly triggered by accents or not. If the F0 peak of X is lower after an accented word (A) than after an unaccented word (U), X is paradigmatically diagnosed as downstepped. However, the F0 differences between the AX and UX sequences before X may affect the F0 difference at X. Furthermore, if a phonological boundary results in another downtrend, it is crucial to control the phonological phrasing in order to accurately assess the isolated effect of downstep. We conducted a production experiment, which compared (A)(X) with both single domain (UX) and separated domains (U)(X). Our results yield two notable findings. First, we observed a more pronounced F0 rise in the downstep conditions than in the non-downstep conditions. Second, the paradigmatic F0 difference between AX and UX between AX and UX decreases when X is boosted. These findings suggest that downstep is triggered by phonological boundaries, not by accents.**

***Index Terms*—downstep, paradigmatic diagnosis, boundary-driven downstep, phonological phrase**

I. Introduction

A. Downstep in Japanese

Since the late 1960s, Japanese has been widely studied for its intonation phonology and has been the subject of active research on the phenomenon of downstep [1]–[12]. Words in Tokyo Japanese are categorized as either accented (A) or unaccented (U). Traditionally, downstep in Japanese has been defined as a pitch range compression triggered by lexical pitch accents [2]. Most Japanese intonation theories distinguish between two separate phonological categories: Minor Phrase (MiP) and Major Phrase (MaP), both of which can be merged into a recursive phonological domain called PPhrases [8], [9], [13]. Accent culminativity and initial lowering define MiP; the downstep domain defines MaP [8]. In these two works [8], [9], recursive PPhrases serve as the domains of initial lowering and downstep.

There are two main methods for identifying downstep in Japanese: syntagmatic and paradigmatic [13]. In the syntagmatic diagnosis, if the pitch peak of the subsequent word is lower than the preceding accented word, it is considered downstep [6]. In contrast, the paradigmatic approach maintains that if the F0 peak of word X is considerably lower after A compared to after U, then X is diagnosed as being under the effect of downstep [9]. The latter approach has been widely used in various studies [2]–[5], [7]–[10], [12]. Both methods recognize that downstep in Japanese is solely triggered by accents.

B. Two contrasting assumptions regarding the target of downstep

Previous studies have presented two contrasting assumptions regarding the target of downstep in Japanese. Assumption 1 states that downstep is a register shift that occurs at the boundary between the accentual H and L tone, and the accentual L tone is the first tone which is under the effect of downstep [2], [3]. In other words, in AX sequences, not only X but also the L tone within the A is the target of downstep [2], [3]. These studies consider only a single version of the L tone present in the Japanese tone inventory: post-accentual low tones are pronounced at a lower pitch compared to the phrase-initial L tones due to the effect of downstep on accentual low tones [2], [3].

On the other hand, Assumption 2 proposes that the target of downstep is only X in the AX sequence [9], [10], [12], [14]. Given the validity of this assumption, downstep is visible only in the region shaded in gray in Fig. 1. In fact, numerous studies have examined downstep in Japanese and consistently reported a lower pitch height of X in AX sequences than in UX sequences, as measured by the F0 peak of X [2]–[5], [7], [10], [12]. However, there are two possible explanations for this difference: (1) the lower peak of X in the AX sequence is solely due to downstep, or (2) the F0 in the AX sequence is lowered not only by downstep but also by the realization of accents before reaching X, and the spill-over effect of these accents lowers the peak of X. In other words, when there is a large pitch difference between the AX and UX conditions before reaching the X point, measuring the pitch at X alone may not accurately reflect the downstep's strength. To accurately measure the strength of the phenomenon that occurs within X, the proper measurement must be the differences between the AX and UX conditions that only develop after entering the X point, as shown in the arrows in Fig. 1.

C. Accent-driven vs. boundary-driven accounts

Existing research on downstep in Japanese consistently identifies the H*L accent as the trigger for this phenomenon

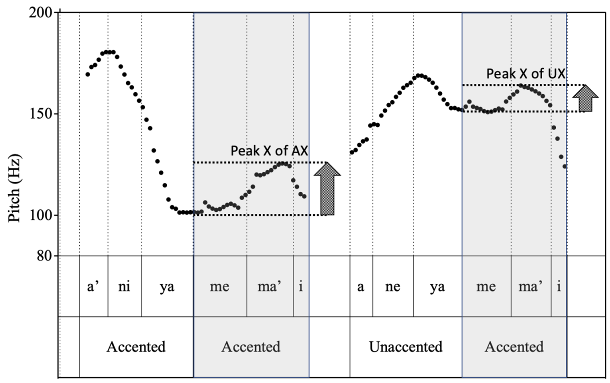

Fig. 1. Measurement points in paradigmatic diagnosis.

[2]–[10], [12]. However, previous findings indicating a step-like downward trend in sequences of unaccented words raise doubts about the hypothesis that accent is the trigger for downstep. One study reports that the F0 peak for each MiP decreases in a stair-like manner without accented words [15]. In addition, another study [11] found that parallel structures can trigger boundary insertions, resulting in a step-like F0 downtrend resembling downstep, even in the absence of accents. The results of these two studies suggest that the insertion of phonological boundaries can result in a stair-step pitch descent.

On the other hand, [9] posited that a phonological boundary must follow every accent because of accent culminativity and the anti-lapse constraint. Specifically, the AA sequence is separated as (A)(A) due to accent culminativity, and the AU sequence is divided into (A)(U) due to the anti-lapse constraint [9]. Since accents lead to phonological boundary insertions, it is plausible that the direct trigger of downstep is not the accent itself but the phonological boundary insertion. In this paper, the accent-driven account posits that downstep is directly caused by accents, as shown by the horizontal striped arrow in Fig. 2. In contrast, the boundary-driven account suggests that downstep is triggered by the insertion of phonological boundaries, as shown by the black arrows.

Interpretations of the paradigmatic diagnosis vary depending on which account is considered accurate. Contrary to the AX sequence, which must be separated, as in (A)(X), the UX sequence can form a single domain (UX) or a separate domain (U)(X). According to the accent-driven account, downstep and boundary-driven downstep are separate phenomena. We can evaluate the downstep effect by comparing (A)(X) and (U)(X), since both have a phonological boundary and thus exhibit boundary-driven downstep. Additionally, (A)(X) is under the effect of downstep. In contrast, when comparing (A)(X) and (UX), both downstep and boundary-driven downstep occur in the (A)(X) condition, complicating differentiation between the two factors.

D. Research objectives

Since traditional paradigmatic diagnosis is plagued by numerous confounding factors and fails to evaluate downstep

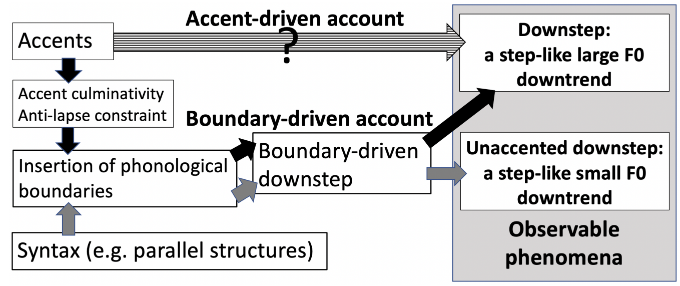

Fig. 2. Accent-driven vs. boundary-driven accounts for downstep in Japanese.

TABLE I
SAMPLE STIMULI USED IN THE EXPERIMENT

a. [-accent][genitive]

item	ane-no	mema’i-to	kaka’rete-ita
gloss	sister-GEN	dizziness-that	written-PAST
	’It was written as ”My sister’s dizziness.”’		

b. [-accent][coordinated]

item	ane-ya	mema’i-to	kaka’rete-ita
gloss	sister-and	dizziness-that	written-PAST
	’It was written as ”sister” and ”dizziness.”’		

c. [+accent][genitive]

item	a’ni-no	mema’i-to	kaka’rete-ita
gloss	brother-GEN	dizziness-that	written-PAST
	’It was written as ”my brother’s dizziness.”’		

d. [+accent][coordinated]

item	a’ni-ya	mema’i-to	kaka’rete-ita
gloss	brother-and	dizziness-that	written-PAST
	’It was written as ”brother” and ”dizziness.”’		

accurately, this study addresses this issue by modifying the measurement position and controlling phonological phrasing, enabling a more accurate assessment of downstep.

II. Experiment

A. Experimental materials

Our experiment compared sequences of two accented words and unaccented words preceding an accented word. The stimuli consisted of two nouns (N1 and N2) combined with a particle. We defined Region 1 as the area where N1 and the following particle are combined, and Region 2 is the area where N2 occurs. Table I displays a set of stimuli, with the accented vowel preceding the apostrophe.

The stimuli were constructed with two factors: Accent and Particle. The Accent factor is comprised of two levels: [-accent] and [+accent]. At the [-accent] level, N1 is unaccented. At the [+accent] level, it is accented. In the [+accent] level, a phonological boundary has to be inserted, as in (A)(A), owing to accent culminativity and anti-lapse constraints [9]. The second factor, the Particle factor, has two levels: [genitive] and [coordinated]. In the former, N1 is followed by a genitive case marker *-no*. When N1 is unaccented, N1 and N2 can form a single PPhrase, as in (U-*no* A). In the [coordinated] level, N1 and N2 form a parallel structure with a conjunction, *ya*. We assumed that a parallel structure triggered the insertion of phonological boundaries after *ya* [11]. Therefore, the [-accent, coordinated] condition formed a separate phonological structure, as in (U-*ya*)(A).

B. Participants

Eight native Tokyo Japanese speakers (four females and four males, mean age 19.75 years, SD 0.83) from the Kanto region (Tokyo, Saitama, Chiba, and Kanagawa) took part in our experiment. None ever lived outside of the Kanto area for more than two years. None reported a history of speech or hearing impairments.

C. Procedures and analysis

The audio recording took place in a soundproof booth at the University of Tokyo. It was captured using a Shure WH20XLR Dynamic Headset Microphone linked to a Roland QUAD-CAPTURE audio interface. The recording was then saved on a computer at a sampling rate of 44.1 kHz. The stimuli appeared individually on a screen in a pseudo-random sequence. Participants orally recited each sentence three times, maintaining a speech rate that was comfortable and natural for them. When they inserted an undesired pause or a mispronunciation while reading a sentence, they repeated it. We recorded a total of 72 tokens: 6 items $\times$ 4 sentence types $\times$ 3 repetitions. 120 sentences (360 tokens) served as fillers. We annotated the sound files using Praat [14] and a script called ProsodyPro [17]. Segmenting the conjunction from the subsequent unstressed words relied on analyzing formants and waveforms. Two of the three repetitions were analyzed. We manually checked and corrected apparent errors by the algorithm in Praat, such as octave jumps.

For each utterance, we made the following two measurement variables. The first measurement is R2MaxF0: the normalized F0 maximum in Region 2. To factor out the F0 range variations among speakers, we converted the values of F0 maximum x to normalized values y with reference to two points, using the following (1). R_1 was the mean value of the F0 maximum in Regions 1 and 2 across all the data points for the speaker, and R_2 is the mean value of the F0 minimum across all the data points for the speaker. This normalization was previously used [10]:

$$y = \frac{x - R_2}{R_1 - R_2} \tag{1}$$

The second measurement variable, R2RiseSize, captures the F0 difference between the maximum F0 of Region 2 minus the final F0 of Region 1, as indicated by arrows in Fig. 1. It is then converted to semitones. This measurement aligns with Assumption 1 and most likely captures the direct impact of downstep by removing the influence of Region 1. The accent-driven account predicts that the [+accent] level will show a smaller F0 rise than the [-accent] level since only the [+accent] level triggers downstep, which is effective solely in Region 2.

We analyzed the data through a Linear Mixed Effects (LME) model using the lmer function within the lme4 package [18] in R [19], where the subjects and items are random effects. We centered the factor labels of Accent and Particle to have a mean of 0 and a range of 1. The backward selection [20] was used to obtain the final models.

TABLE II
RESULTS OF MIXED-EFFECTS MODELS FOR R2MAXF0

Condition	Predictor	β	t	p
[-accent, genitive]	(Intercept)	0.653	16.21	<.001
vs. [+accent, genitive]	Accent	-0.678	-18.58	<.001
[-accent, coordinated]	(Intercept)	0.728	19.2	<.001
vs. [+accent, coordinated]	Accent	-0.438	-14.73	<.001
[-accent, genitive]	(Intercept)	0.970	21.98	<.001
vs. [-accent, coordinated]	Accent	-0.044	-1.327	n.s.
[+accent, genitive]	(Intercept)	0.411	9.711	<.001
vs. [+accent, coordinated]	Accent	0.196	6.168	<.001

TABLE III
RESULTS OF MIXED-EFFECTS MODELS FOR R2RISESIZE

Condition	Predictor	β	t	p
[-accent, genitive]	(Intercept)	1.024	6.092	<.001
vs. [+accent, genitive]	Accent	0.377	2.836	<.01
[-accent, coordinated]	(Intercept)	2.017	8.212	<.001
vs. [+accent, coordinated]	Accent	1.697	11.877	<.001
[-accent, genitive]	(Intercept)	1.002	7.685	<.001
vs. [-accent, coordinated]	Accent	0.333	3.198	<.01
[+accent, genitive]	(Intercept)	2.039	7.409	<.001
vs. [+accent, coordinated]	Accent	1.653	10.174	<.001

III. RESULTS

The results for each variable are presented in Table II (for R2MaxF0) and Table III (for R2RiseSize). The β refers to the fixed effects estimates. Fig. 3 displays the mean R2RiseSize for each condition. Fig. 4 displays the normalized mean F0 for each mora, based on (1). Fig. 1 in Section I displays the sample F0 contours for the [-accent, coordinated] and [+accent, coordinated] conditions.

For R2MaxF0, the [+accent, genitive] condition had a significantly lower value compared to the [-accent, genitive] condition, and the [+accent, coordinated] condition had a significantly lower value than the [-accent, coordinated] condition. Although the mean value of R2MaxF0 for the [-accent, coordinated] condition was lower than that for the [-accent, genitive] condition, the difference was not statistically significant. Additionally, the R2MaxF0 value for the [+accent, genitive] condition was significantly lower than that for the [+accent, coordinated] condition.

Regarding R2RiseSize, as shown in Fig. 3, the [+accent, genitive] condition had a significantly larger value than the [-accent, genitive] condition. The estimated difference between the [+accent, genitive] and [-accent, genitive] conditions was only 0.377 semitones, which indicates a small although statistically significant difference. The R2RiseSize for the [+accent, coordinated] condition was also significantly larger than that for the [-accent, coordinated] condition, and the R2RiseSize for the [-accent, coordinated] condition was significantly larger than that for the [-accent, genitive] condition. Additionally, the R2RiseSize for the [+accent, genitive] condition was significantly smaller than that for the [+accent, coordinated] condition.

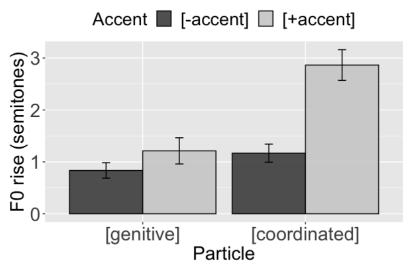

Fig. 3. F0 difference between maximum F0 of Region 2 minus final F0 of Region 1 (R2RiseSize): Error bars represent 95% confidence intervals.

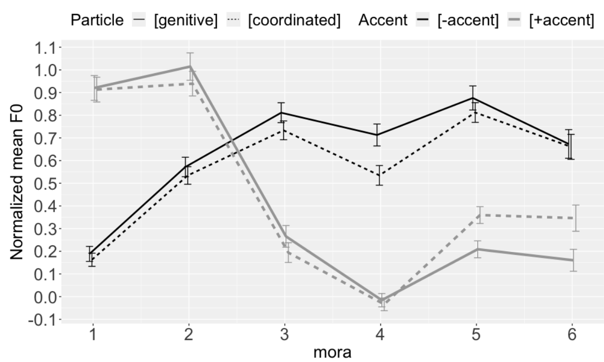

Fig. 4. Mean of normalized F0 means for [-accent, genitive], [+accent, genitive], [-accent, coordinated], and [+accent, coordinated] conditions represented by solid black, solid gray, dashed black, and dashed gray lines, respectively: Error bars indicate 95% confidence intervals.

IV. Discussion

Our results show two noteworthy findings. First, concerning R2RiseSize, we observed a greater F0 rise for the [+accent] conditions than for the [-accent] conditions, contradicting the prediction from the accent-driven account of downstep. Second, regarding R2MaxF0, the paradigmatic difference in F0 height between AX and UX decreases when X is boosted. These findings support the boundary-driven account, which posits that accents do not directly trigger downstep.

A. *Invalidity of traditional paradigmatic diagnosis*

Our study identified the inadequacy of using the F0 height of N2 as an indicator for a paradigmatic diagnosis of downstep in Japanese due to the presence of two confounding factors. The results showed that the R2MaxF0 of the [+accent] level was significantly lower than that of the [-accent] level in both the [genitive] and [coordinated] levels. According to the traditional paradigmatic diagnosis, these results suggest that downstep occurs in Region 2 in [+accent] conditions. However, two confounding factors obfuscate whether downstep exists independently of boundary-driven downstep. The first is phonological phrasing, because the [+accent, genitive] and [-accent, genitive] conditions are assumed to form (A)(A) and (UA). Since there is a phonological boundary in (A)(A), the lower F0 peak in Region 2 could be due to boundary-driven downstep, rather than downstep. In contrast, the [+accent, coordinated] and [-accent, coordinated] conditions are assumed to form (A)(A) and (U)(A). Unlike at the [genitive] level, phonological boundaries and boundary-driven downstep are not confounding factors at the [coordinated] level. Note that the [-accent, genitive] level may form (U)(A), depending on the speech rate or other factors.

The second confounding factor is the spill-over effect of accents in Region 1. As shown in Fig. 1, the pitch has already decreased considerably before entering Region 2 in the [+accent] conditions. Even without assuming downstep, the lowered F0 peak in Region 2 in the [+accent, genitive] and [+accent, coordinated] conditions can be attributed to the spill-over effect of the accent in Region 1. If downstep's scope encompasses the accent's L tone itself, the spill-over effect of the accent is not a confounding factor as the lowered F0 in Region 1 is attributed to downstep. However, if we posit that downstep excludes the accentual L, the conventional paradigmatic diagnosis comparing F0 peaks in AX or UX cannot faithfully capture downstep's genuine effect due to these two potential confounding factors.

A comparison of the [-accent, genitive] and [-accent, coordinated] conditions corresponds to a comparison of (UA) vs. (U)(A). According to the boundary-driven account, the insertion of a phonological boundary in the [-accent, coordinated] condition is predicted to cause boundary-driven downstep, lowering the pitch in Region 2 than in the [-accent, genitive] condition. However, our results do not strongly support this prediction, possibly because the notation of quotation marks led to an F0 boost by emphasizing N2, which neutralized the boundary-driven downstep. Consistent with this explanation, an F0 boosting effect was also observed in the [-accent, coordinated] condition. In contrast, since the comparison of the [+accent, genitive] and [+accent, coordinated] conditions corresponds to a comparison of (A)(A) vs. (A)(A), the boundary-driven account does not predict any differences between them. The difference between these two conditions is likely due to the notation of quotation marks, which emphasized N2 and caused a pitch boost.

B. *Paradigmatic difference in F0 height between AX and UX decreases when X is boosted*

The F0 boosting effect of quotation marks leads to a noteworthy generalization: the paradigmatic difference in F0 height between AX and UX decreases when X is boosted. In other words, the paradigmatic F0 difference between the X in AX and X in UX becomes smaller when X is boosted by focus or additional insertion of a boundary, compared to when X is not boosted. This observation aligns with prior research [10], [12], although this specific generalization had not been previously established. As for our results, the estimated paradigmatic F0 difference is 21.789 Hz in the coordinated level with the boosting effect by quotation marks, whereas it increases to 35.136 Hz in the genitive level. However, the accent-driven explanation falls short in elucidating this generalization, as it anticipates that the sole distinction between AX and UX lies in downstep, which would nullify the boosting effect through paradigmatic diagnosis.

C. Evidence that the accentual L tone is not under the effect of downstep

The experimental findings of R2RiseSize contradicts the accent-driven account under Assumption 1, which proposes that downstep takes place at an H*L transition and that the initial target of downstep is the accentual L in the downstep trigger. According to the model proposed by [3], the value of R2RiseSize in the AX sequence, R2RiseSize(AX), is expressed by (2). In contrast, the value of R2RiseSize in the UX sequence is represented by (3). In these equations, h represents the high-tone line, r represents the reference line, T(H) and T(L) represent relative tonal prominences between 0 and 1, and c is a downstep constant. According to the model, (3) should be larger than (2) since c is less than 1. However, our experimental results contradict this prediction, showing that R2RiseSize is larger in the downstep condition.

$$R2RiseSize\,(AX) = H_{new} - L_{new} = c \cdot (h_{old} - r) \cdot \{T\,(H_a) - [1 - T\,(L)]\} \tag{2}$$

$$R2RiseSize\,(UX) = H_{old} - L_{old} = (h_{old} - r) \cdot \{T\,(H_a) - [1 - T\,(L)]\} \tag{3}$$

D. Evidence that downstep does not target X in the string AX

The finding that R2RiseSize is larger in the [+accent, coordinated] condition than in the [-accent, coordinated] condition contradicts the accent-driven account even under Assumption 2: the target of downstep is X in the AX sequence. If it is assumed that downstep is a mechanism that only targets the X in AX and lowers its pitch more in the AX sequence than in the UX sequence, the results of R2RiseSize suggest that downstep does not exist.

The results showed that the R2RiseSize of [+accent] level was significantly larger than that of the [-accent] level in both the [genitive] and [coordinated] levels. The accent-driven account predicts that the [+accent] conditions should have a much smaller pitch rise than the [-accent] conditions, although this is contradicted by the results. The fact that it did not meet the requirements of paradigmatic diagnosis, where the downstep should have compressed the F0 peak of Region 2 at the [+accent] level, suggests a flaw in the traditional paradigmatic diagnosis or Assumption 2.

E. Boundary-driven account: trade-off between boundary-driven downstep and edge boost

Thus far, it has been shown that the findings of this study do not align with the accent-driven account, irrespective of whether Assumption 1 or Assumption 2 is employed. In this subsection, we will explore an alternative explanation for the results of this study using the boundary-driven account.

The boundary-driven account encompasses two key mechanisms: boundary-driven downstep and edge boost. We define *boundary-driven downstep* as a phonological mechanism that causes a decrease in pitch for succeeding PPhrases when there are two or more PPhrases directly governed by a PPhrase or PClause. To illustrate, consider the expression $({}_{\text{PPhrase1}}({}_{\text{PPhrase2}}X)({}_{\text{PPhrase3}}Y)({}_{\text{PPhrase4}}Z))$. X, Y, and Z can be either accented or unaccented. Since PPhrase1 directly governs PPhrase2, PPhrase3, and PPhrase4, it satisfies the conditions for boundary-driven downstep to occur. As a result, boundary-driven downstep in PPhrase1 causes the pitch of PPhrase2, PPhrase3, and PPhrase4 to lower in a stair-like manner.

In addition, we define *edge boost* as a phonological mechanism in which the number of left edges of PPhrases is proportional to the degree of the realization of and H tones and the boundary tone %L. In simpler terms, when there are multiple phonological left edges in certain situations, the F0 of the H tone increases, and the initial lowering has a greater F0 range. Edge boost is a theoretical concept that builds upon earlier ideas, including the metrical boost [4], [5] and the cumulative rise in recursive φ [8], as well as the notion that prosodic effects intensify cumulatively as multiple boundaries coincide at higher levels [21]. Edge boost differs from metrical boost in that it can occur not only at right-branching structures but also, unlike metrical boost, it affects not only F0 height but also the extent of initial lowering. Studies have reported that the effect of F0 boost increases as the number of phonological left boundaries increases both for accented sequences [5], [22] and for unaccented sequences [15].

The considerably lower values of R2MaxF0 in the [+accent, coordinated] condition, compared to the [-accent, coordinated] condition, can be attributed to two factors: the spill-over effect of accent and the trade-off between edge boost and boundary-driven downstep. Additionally, the larger R2RiseSize in the [+accent, coordinated] condition, compared to the [-accent, coordinated] condition, can also be attributed to these same factors.

First, the spill-over effect refers to a situation where the pitch of the onset of a subsequent PPhrase is phonetically lower due to the accent being realized as a sharp decrease in F0 in the preceding accented PPhrase. In simpler terms, the initial pitch in Region 2 should be similar to the final pitch in Region 1. The term "spill-over" suggests that the spill-over effect of accents is independent of the phonological lowering effect on the subsequent PPhrase.

The second factor involves a trade-off between boundary-driven downstep and edge boost. Put simply, when comparing AX and UX sequences, the same level of edge boost in X is insufficient to balance the effect of boundary-driven downstep between the AX and UX sequences. On the other hand, if the degree of pitch lowering caused by boundary-driven downstep in X is identical for both AX and UX sequences, the F0 rise at the beginning of X in AX becomes excessive in relation to the level of edge boost, assuming an equal number of boundaries between AX and UX. Consider, hypothetically, the case where the degree of edge boost is equal between the AX sequence and the UX sequence, as well as the case where the degree of boundary-driven downstep is equal between the AX sequence and the UX sequence. Assuming that the X in both AX and

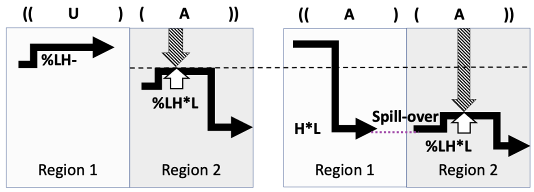

Fig. 5. Schematic illustration of excessive boundary-driven downstep. The length of the striped arrows represents the strength of the boundary-driven downstep, while the length of the white arrows represents the strength of the edge boost.

UX sequences have the same number of left boundaries, such as (A)(X) and (U)(X), the edge boost hypothesis predicts the same degree of F0 rise at X. In the AX condition, the initial F0 at X is already lowered due to the spill-over effect of accents. If the degree of edge boost realization is the same between (X) in (A)(X) and (X) in (U)(X), then the F0 peak at (X) in (A)(X) is considerably lower than the observed F0 contours in our results. In other words, when the same degree of edge boost is applied to X in both AX and UX sequences, it results in an excessive lowering of the F0 peak in the AX sequence, as shown in Fig. 5. Conversely, assuming the case where the degree of lowering F0 peak at X by boundary-driven downstep is equal between the AX sequence and the UX sequence. In this case, the deree of F0 rise by edge boost in AX sequece is excessively large compared to the number of edges, because the initial F0 at X is already lowered due to the spill-over effect of accents in the AX condition.

To strike a balance between the requirements of edge boost and boundary-driven downstep, considering the pitch already lowered due to the spill-over effect of accents, the F0 peak at X in AX should be lower than the F0 peak at X in UX. Simultaneously, the F0 rise at the beginning of X in AX should be greater than the F0 rise at the beginning of X in UX. If spill-over is blocked by another factor, such as the insertion of a pause, the pitch at Region 2 should recover to a similar extent. Alternatively, if additional boundaries are inserted to justify the large amount of edge boost or if there are other factors contributing to F0 boost, the F0 peak will be similar in both conditions. To put it differently, the boundary-driven account predicts a decrease in the paradigmatic F0 difference between AX and UX when X is boosted. This aligns with the results reported in [10], [12] as well as the finding of this study.

V. Conclusion

Our experimental design employed a paradigmatic comparison between (A)(X) and both single-domain (UX) and separated-domain (U)(X) conditions. Our results yield two notable findings. First, we observed a more pronounced F0 rise in the [+accent] conditions than in the [-accent] conditions, challenging the accent-driven explanation of downstep; the F0 rise experiences greater compression under downstep conditions compared to non-downstep conditions. Second, the paradigmatic F0 difference between AX and UX between AX and UX decreases when X is boosted. These findings lend support to the boundary-driven account, suggesting that accents do not directly trigger downstep. We propose that the lower F0 peak in the AX sequence results from the two factors: the spill-over effect of accent and the trade-off between edge boost and boundary-driven downstep.

Acknowledgment

Part of this work was supported by JST SPRING Grant Number JPMJSP2140 and JSPS KAKENHI Grant Number JP21H05054.

References

[1] J. D. Mccawley, “The phonological component of a grammar of Japanese.” Mouton, 1968.
[2] W. J. Poser, “The phonetics and phonology of tone and intonation in Japanese” (Doctoral dissertation), 1984.
[3] J. B. Pierrehumbert and M. E. Beckman, “Japanese Tone Structure.” Number 15 in Linguistic Inquiry Monographs, 1988.
[4] H. Kubozono, “The Organization of Japanese Prosody,” 1988.
[5] H. Kubozono, “Syntactic and rhythmic effects on downstep in Japanese,” Phonology, vol. 6, no. 1, pp. 39–67, May 1989.
[6] E. Selkirk and K. Tateishi, “Syntax and downstep in Japanese,” in Interdisciplinary Approaches to Language. Dordrecht: Springer Netherlands, 1991, pp. 519–543.
[7] H. Kubozono, “Focus and intonation in Japanese,” Interdisciplinary Studies on Information Structure: ISIS; Working Papers of the SFB, vol. 632, pp. 1–27, 2007.
[8] J. Ito and A. Mester, “Recursive prosodic phrasing in Japanese,” in Prosody Matters: Essays in Honor of Elisabeth Selkirk, pp. 280–303, 2012.
[9] J. Ito and A. Mester, “Prosodic subcategories in Japanese,” Lingua, vol. 124, pp. 20–40, Jan. 2013.
[10] S. Ishihara, “Japanese downstep revisited,” Nat. Lang. Linguist. Theory, vol. 34, no. 4, pp. 1389–1443, Nov. 2016.
[11] K. Furukawa and Y. Hirose, “Boundary-driven downstep in Japanese,” Proceedings of 19th ICPhS, Canberra: ASSTA, 2019.
[12] M. Hirayama, H.K. Hwang, and T. Kato, ”Lexical Category and Downstep in Japanese,” Languages, vol. 7, no. 1, p. 25, 2022.
[13] S. Ishihara, “14 syntax-phonology interface,” in Handbook of Japanese Phonetics and Phonology, 2015, pp. 569–618.
[14] Y. Igarashi et al., ”Phonological theory informs the analysis of intonational exaggeration in Japanese infant-directed speech,” J. Acoust. Soc. Am., vol. 134, no. 2, pp. 1283–1294, 2013.
[15] E. Selkirk, T. Shinya, and M. Sugahara, “Degree of initial lowering in Japanese as a reflex of prosodic structure organization,” in Proc. 15th ICPhS, Barcelona, Spain, 2003, pp. 491–494.
[16] P. Boersma and D. Weenink, “Praat: A System for Doing Phonetics by Computer,” Version 6.0.39, 2018.
[17] Y. Xu, “ProsodyPro — a tool for large-scale systematic prosody analysis,” in Tools and Resources for the Analysis of Speech Prosody. Aix-en-Provence, France: Laboratoire Parole et Langage, France, 2013, pp. 7–10.
[18] D. Bates, M. Maechler, B. Bolker, S. Walker, R. H. B. Christensen, H. Singmann, B. Dai, G. Grothendieck, and P. Green, “Package “lme4,” 2015.
[19] R Core Team, “R: A Language and Environment for Statistical Computing,” R Foundation for Statistical Computing, Vienna, Austria, 2022. [Online]. Available: https://www.R-project.org/
[20] D. Bates, R. Kliegl, S. Vasishth, and H. Baayen, “Parsimonious mixed models,” Jun. 2015.
[21] C. Fougeron, ”Articulatory properties of initial segments in several prosodic constituents in French,” J. Phon., vol. 29, no. 2, pp. 109–135, 2001.
[22] T. Shinya, E. Selkirk, and S. Kawahara, ”Rhythmic boost and recursive minor phrase in Japanese,” in Speech Prosody 2004, International Conference.