BOUNDARY-DRIVEN ACCOUNT FOR DOWNSTEP IN JAPANESE

Kei Furukawa[1] & Satoshi Nakamura[1]

[1]Nara Institute of Science and Technology, Japan
furukawa.kei.fi4@is.naist.jp, s-nakamura@is.naist.jp

ABSTRACT

This study verifies the boundary-driven account for downstep in Japanese: downstep is triggered by phonological boundaries. The assumption widely adopted in the literature is that downstep is triggered only by accents. However, it remains unknown which account is more accurate. We propose that downstep is one example of the phonetic realization of boundary-driven downstep. We define boundary-driven downstep as a phonological mechanism that lowers the pitch of subsequent PPhrases when a PPhrase or a PClause directly dominates two or more PPhrases. Nine native speakers of Tokyo Japanese participated in a production experiment, which paradigmatically compared the F0 heights of a sequence of a final-accented word preceding an unaccented word to a sequence of unaccented words. The results showed that the condition with final accents without particles did not show a large F0 downtrend. Our finding indicates that accents do not directly trigger downstep.

Keywords: downstep, Japanese, prosodic phrasing, boundary-driven downstep, PPhrase

1. INTRODUCTION

1.1. Downstep in Japanese

Japanese is one of the most widely studied languages concerning intonation phonology, and downstep has been actively researched since the late 1960s [1–9]. In Tokyo Japanese, words are classified as either accented (A) or unaccented (U). Traditionally, downstep has been defined as a pitch range compression triggered by lexical pitch accents [2–5, 7–9]. Most theories of Japanese intonation assume two distinct prosodic categories, Minor Phrase (MiP) and Major Phrase (MaP), both of which can theoretically be unified into recursive phonological domains called PPhrases [10–12]. MiP is defined by accent culminativity and initial lowering, and MaP is defined as a domain of downstep [10]. In these two works [11, 12], the domains of initial lowering and downstep are considered recursive PPhrases.

Two primary approaches help identify downstep in Japanese: syntagmatic and paradigmatic [7]. In the former approach, if the F0 peak following an accented word is clearly lower than the preceding F0 peak, it is diagnosed as downstep [6]. On the contrary, the latter paradigmatic approach claims that if the pitch height of X is significantly lower after an accented word (A) than after an unaccented word (U), X is diagnosed as having been downstepped [5, 7, 12]. This approach has been adopted in many studies [2–5, 7–9, 12]. Both diagnoses assume that downstep in Japanese is triggered only by accents.

1.2. Accent-driven vs. boundary-driven accounts

In contrast to previous studies, perhaps downstep is triggered by phonological boundaries, not lexical pitch accents [13]. One work argued that a phonological boundary must be inserted after every accent owing to accent culminativity and anti-lapse constraints [12]. Another study [13] reported that parallel structures trigger the insertion of boundaries, which prompt a step-like F0 downtrend that resembles downstep, even without accents (gray arrows in Figure 1). Conversely, since accents lead to the insertion of phonological boundaries, downstep's direct cause may not be the accent but phonological boundaries [13].

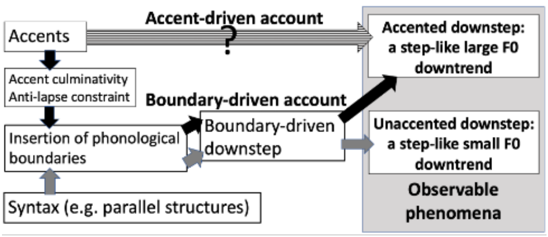

Figure 1: Accent-driven vs. boundary-driven accounts for downstep in Japanese.

In this paper, the accent-driven account argues that downstep is caused directly by accents, as shown in the horizontal striped arrow in Figure 1. In contrast, the boundary-driven account theorizes that downstep is caused by the insertion of phonological boundaries (black arrows in Figure 1).

1.3. Research objectives

This study primarily aims to verify the boundary-driven account for downstep in Japanese: downstep is caused by boundaries rather than directly by accents. To clarify the discussion, this paper redefines the terminology concerning downstep, which is a pitch range compression triggered only by lexical pitch accents, as diagnosed by the paradigmatic approach. Accented downstep is a step-like large F0 downtrend after an accented word. The term merely refers to observable phenomena without presupposing that accented downstep is triggered directly by accents. Unaccented downstep is a step-like small F0 downtrend after an unaccented word [13].

We propose that the large F0 step-like downtrend, which has traditionally been called downstep, is one example of the phonetic realization of boundary-driven downstep. We define boundary-driven downstep as a phonological mechanism that lowers the pitch of subsequent PPhrases when a PPhrase or a PClause directly dominates two or more PPhrases. Boundary-driven downstep phonetically shows a step-like F0 downtrend, unless neutralized by other pitch-rising phenomena. It can phonetically be realized as either accented or unaccented downstep.

2. EXPERIMENT

2.1. Experimental materials

Our experiment paradigmatically compared the sequence of a final-accented word preceding unaccented word to a sequence of unaccented words. Words with an accent on the final syllable are called final-accented or *odaka*-accented words. The difference between final-accented and unaccented words is clearer when a particle follows them. For instance, the final-accented word *hana* (LH*) "flower" and the unaccented word *hana* (LH-) "nose" in Tokyo Japanese do not show an acute pitch fall when pronounced in isolation. However, when followed by a nominative case marker -*ga*, the final-accented *hana-ga* shows the LH*L pattern. In contrast, an unaccented *hana-ga* shows no acute pitch fall, following LHH. Final-accented words in isolation and unaccented words cannot be distinguished in production [14–16] or in perception [17].

In the stimuli, three nouns (N1, N2, and N3) are given in a parallel structure with a conjunction *ya* or middle dots. Region 1 is defined as the area where N1 and the conjunction *ya* or middle dots are combined. The area between N2 and the following conjunction *ya* or the middle dots is defined as Region 2. N1 is final-accented or unaccented; N2 and N3 are unaccented words. One stimuli set is given in Table 1. 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, in contrast, N1 is final-accented. The second factor is the Particle factor, which has two levels: [-particle] and [+particle]. At the [-particle] level, N1, N2, and N3 form a parallel structure with middle dots. In Japanese, middle dots are not pronounced and are realized as silences or short pauses. In the [+particle] level, N1, N2, and N3 form a parallel structure with a conjunction *ya*. This study assumes that a parallel structure triggers the insertion of phonological boundaries after *ya* or the middle dots [13]. In the [+accent, +particle] condition, a final-accented N1 is followed by *ya*, so the last and the second to last moras are expected to show a H*L pattern. In the [+accent, -particle] condition, on the other hand, the final mora of Region 1 remains high, and no sharp pitch fall is expected to be observed.

Table 1: Sample stimuli used in experiment: Accented moras are underlined.

a. [-accent][-particle]						
item	hana,	mori,	ue	to-itta	kanji-ga	kaitearu
gloss	nose,	forest,	top	such as	chinese character-NOM	written
	'The Chinese characters such as a nose, forests, and top are written there.'					
b. [+accent][-particle]						
item	hana,	mori,	ue	to-itta	kanji-ga	kaitearu
gloss	flower,	forest,	top	such as	chinese character-NOM	written
	'The Chinese characters such as flowers, forests, and top are written there.'					
c. [-accent][+particle]						
item	hana-ya	mori-ya	ue	to-itta	kanji-ga	kaitearu
gloss	nose-and	forest-and	top	such as	chinese character-NOM	written
	'The Chinese characters such as a nose, forests, and top are written there.'					
d. [+accent][+particle]						
item	hana-ya	mori-ya	ue	to-itta	kanji-ga	kaitearu
gloss	flower-and	forest-and	top	such as	chinese character-NOM	written
	'The Chinese characters such as flowers, forests, and top are written there.'					

2.2. Participants

Nine native speakers (four females and five males, mean age 19.6 years, SD 0.96) of Tokyo Japanese from the Kanto area (Tokyo, Kanagawa, Saitama, and Chiba) participated in our experiment as subjects. No participant ever lived outside of Kanto area for more than two years. None reported a history of speech or hearing impairments.

2.3. Procedures and analysis

The recording was conducted in a soundproof booth at the University of Tokyo using a Shure

WH20XLR Dynamic Headset Microphone, which was connected to a Roland QUAD-CAPTURE audio interface; the audio was recorded to a computer at a sampling rate of 44.1 kHz. The stimuli were displayed on a screen one by one in a pseudo-random order. Participants read sentences aloud three times at a normal speech rate that felt natural to them. When they inserted an undesired pause or a mispronunciation while reading a sentence, they repeated it. We recorded a total of 6 items $\times$ 4 sentence types $\times$ 3 repetitions = 72 tokens. 120 sentences (360 tokens) from other experiments served as fillers. Sound files were annotated using Praat [18] and a script called ProsodyPro [19]. Segmentation between the conjunction and the following unaccented words was based on formants and waveforms. Two of the three repetitions were analyzed. Apparent errors by the algorithm in Praat, such as octave jumps, were manually checked and corrected.

For each utterance, we made the following three measurement variables. The first measurement is R1Fall: the maximum value of the pitch in N1 minus the minimum value of the pitch of the following conjunction *ya*. This measurement is only calculated for the [+particle] level. It is then converted to semitones. The second measurement is R1MaxF0: the normalized F0 maximum in Region 1. To factor out the pitch range differences among speakers, we converted the values of F0 maximum x to normalized values y with reference to two points, using the following formula (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 in Regions 1 and 2 across all the data points for the speaker. This normalization was previously used [7]:

$$(1) \quad y = \frac{x - R_2}{R_1 - R_2}$$

The third variable is R2MaxF0, which is the normalized F0 maximum in Region 2. This variable is also normalized using the formula (1). In the sequence of *hana-ya mori-ya* "flowers and forests" in the [+accent, +particle] condition, for instance, R1MaxF0 is calculated as the normalized maximum value of the pitch in the sequence *hana-ya* "flowers and", while R2MaxF0 is the maximum value of the pitch in the sequence of *mori-ya* "forests and." As for the [+accent, -particle] condition, in the sequence of *hana*, *mori* "flowers, forests," R1MaxF0 is the normalized maximum value of the pitch in the sequence *hana* "flowers", while R2MaxF0 refers to the maximum value of the pitch in the sequence *mori* "forests."

The data were analysed within the linear mixed-effects model (LME) using the lmer function within the lme4 package [20] in R [21], where the subjects and items are random effects. The factor labels of Accent and Particle were cantered to have a mean of 0 and a range of 1. The final models were obtained using backward selection [22].

3. RESULTS AND DISCUSSION

3.1. Results

Our results show two noteworthy findings. First, contrary to the prediction from the accent-driven account, the [+accent/-particle] condition, which has final accents without particles, did not show a large F0 downtrend from N1 to N2 (Figure 2). Second, the small step-like downtrend found in the [+accent/-particle] level is slightly larger than the downtrend in the [-accent/-particle] level.

The results for each variable are presented in Table 2 (for R1Fall), Table 3 (for R1MaxF0), and Table 4 (for R2MaxF0). Figure 2 displays the sample F0 contours of all conditions.

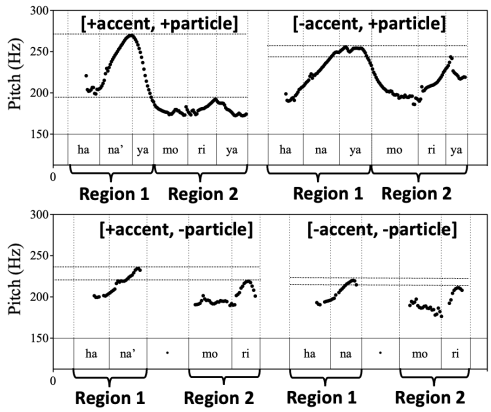

Figure 2: Sample F0 contours of all conditions.

For R1Fall, the main effect of the Accent factor was significant, where the [+accent] level showed a greater F0 fall than the [-accent] level. This suggests that the final accent is realized as a sharp F0 fall in the [+accent, +particle] condition.

Table 2: Results of mixed-effects models for R1Fall.

Predictor	β	t	p
(Intercept)	3.763	7.993	$<.001$
Accent	4.372	26.260	$<.001$

Regarding R1MaxF0, the difference between the [-accent, -particle] condition and the [+accent, -particle] condition was not statistically significant. Meanwhile, the [+accent, +particle] condition had a significantly higher value compared to the [-accent,

+particle] condition. These results suggest that pre-low raising causes accentual H* tones to have a higher pitch than phrasal H- tones [23], since a higher F0 peak at Region 1 was only observed when an accent was realized as a sharp F0 fall at the [+particle] level. Moreover, the Particle factor has significant effect on the R1MaxF0 in both [-accent] and [+accent] levels.

Table 3: Results of mixed-effects models for R1MaxF0.

Condition	Predictor	β	t	p
[-accent, -particle]	(Intercept)	0.858	37.17	<.001
vs. [+accent, -particle]	Accent	0.045	0.90	.372
[-accent, +particle]	(Intercept)	1.083	39.895	<.001
vs. [+accent, +particle]	Accent	0.274	8.547	<.001
[-accent, -particle]	(Intercept)	0.891	39.891	<.001
vs. [-accent, +particle]	Particle	0.110	3.734	<.001
[+accent, -particle]	(Intercept)	1.050	28.992	<.001
vs. [+accent, +particle]	Particle	0.339	6.884	<.001

The R2MaxF0 for the [+accent, -particle] condition was significantly smaller than that for the [-accent, -particle] condition. This does not satisfy the requirement for a paradigmatic diagnosis because the F0 difference between the two conditions is quite small, as illustrated in Figure 2. The R2MaxF0 for the [+accent, +particle] condition was also significantly smaller than that for the [-accent, +particle] condition. The [+accent, +particle] condition is diagnosed as under downstep because a large F0 difference satisfies the requirement for paradigmatic diagnosis. The R2MaxF0 for the [-accent, -particle] condition was significantly smaller than that for the [-accent, +particle] condition. Additionally, the R2MaxF0 for the [+accent, +particle] condition was significantly smaller than that for the [+accent, -particle] condition. The observation that the F0 peak of Region 2 for the [+accent, +particle] condition is considerably smaller than that of the [+accent, -particle] condition suggests that the F0 downtrend in the [+accent, -particle] condition should be differentiated from the paradigmatically diagnosed downstep found in the [+accent, +particle] condition.

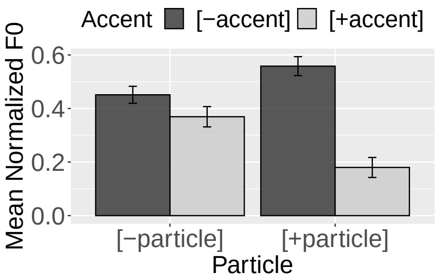

Figure 3: Normalized F0 peak of Region 2: Error bars represent 95% confidence intervals.

Table 4: Results of mixed-effects models for R2MaxF0.

Condition	Predictor	β	t	p
[-accent, -particle]	(Intercept)	0.410	11.497	<.001
vs. [+accent, -particle]	Accent	-0.081	-3.947	<.001
[-accent, +particle]	(Intercept)	0.369	12.919	<.001
vs. [+accent, +particle]	Accent	-0.378	-8.199	<.001
[-accent, -particle]	(Intercept)	0.504	18.758	<.001
vs. [-accent, +particle]	Particle	0.108	4.998	<.001
[+accent, -particle]	(Intercept)	0.275	7.570	<.001
vs. [+accent, +particle]	Particle	-0.189	-4.399	<.01

3.2. Discussion

Our results support the boundary-driven account. Accents do not directly trigger a large F0 fall, which is traditionally referred to as downstep. Two facts indicate that there is no downstep in the [+accent, -particle] condition: 1) the [+accent, -particle] condition did not exhibit a large F0 compression for meeting the requirements for a paradigmatic diagnosis; 2) the [+accent, -particle] condition had a sufficiently higher F0 peak at Region 2 compared to the [+accent, +particle] condition. One might argue that the phonological accents in the [+accent, -particle] condition are deleted rather than unrealized. However, the significant difference between the [-accent, -particle] and [+accent, -particle] conditions refutes the possibility because the Accent factor did create a slightly lower F0 peak at Region 2 in the [+accent, -particle] condition than in the [-accent, -particle] condition.

We believe that boundaries are inserted between N1 and N2 due to the Lapse-L constraint or the parallel structure. These boundaries trigger boundary-driven downstep, which is phonetically realized as accented downstep in the [+accent, +particle] condition and unaccented downstep at the [-accent] level. The slightly larger downtrend than the unaccented downstep at the [+accent, -particle] condition may be due to the unassociated tones of L from the final accent H*L.

A question remains concerning the differences of the F0 fall between AX and UX by a paradigmatic approach [2–5, 7–9, 12]. This study suggests that pre-low raising, phonological phrasing, and spillover effect of accents could be confounding factors, but additional research is necessary to determine their exact role.

4. CONCLUSION

Our experiment results reveal a step-like F0 downtrend that is smaller than downstep, despite the presence of final accents. Thus, our study contends that accented downstep in Japanese is caused by boundaries rather than directly by accents.

5. ACKNOWLEDGEMENTS

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

6. 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] ——, "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] S. Ishihara, "Japanese downstep revisited," *Nat. Lang. Linguist. Theory*, vol. 34, no. 4, pp. 1389–1443, Nov. 2016.

[8] M. Hirayama and H. K. Hwang, "Relative clause and downstep in japanese," *Proceedings of the Annual Meetings on Phonology*, vol. 6, no. 0, Jun. 2019.

[9] M. Hirayama, H. K. Hwang, and T. Kato, "Lexical category and downstep in japanese," *Languages*, vol. 7, no. 1, p. 25, Jan. 2022.

[10] S. Ishihara, "14 syntax-phonology interface," handbook of japanese phonetics and phonology," pp. 569–618, 2015.

[11] J. Ito and A. Mester, "Recursive prosodic phrasing in japanese," prosody matters: Essays in honor of elisabeth selkirk," pp. 280–303, 2012.

[12] ——, "Prosodic subcategories in japanese," *Lingua*, vol. 124, pp. 20–40, Jan. 2013.

[13] K. Furukawa and Y. Hirose, "Boundary-driven downstep in japanese," *Proceedings of 19th ICPhS, Canberra: ASSTA*, 2019.

[14] M. Sugito, "Comparison between "high-final" & "level" tone in tokyo accent [a study of final-accented and unaccented disyllabic words in tokyo using dynamic measurement," *Onseigakkai Kaihoo*, vol. 129, pp. 1–4, 1968.

[15] T. J. Vance, "Final accent vs. no accent: Utterance-final neutralization in tokyo japanese," *Journal of Phonetics*, vol. 23, no. 4, pp. 487–499, 1995.

[16] Y. Sugiyama, *Japanese pitch accent: Examination of final-accented and unaccented minimal pairs. Toronto Working Papers in Linguistics*, 2006.

[17] N. Warner, "Japanese final-accented and unaccented phrases," *J. Phon.*, vol. 25, no. 1, pp. 43–60, Jan. 1997.

[18] P. Boersma and D. Weenink, *Praat: A System for Doing Phonetics by Computer" Version 6.0. 39*, 2018.

[19] 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.

[20] D. Bates, M. Maechler, B. Bolker, S. Walker, R. H. B. Christensen, H. Singmann, B. Dai, G. Grothendieck, and P. Green, "Package "lme4," 2015.

[21] 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/

[22] D. Bates, R. Kliegl, S. Vasishth, and H. Baayen, "Parsimonious mixed models," Jun. 2015.

[23] A. Lee, S. Prom-On, and Y. Xu, "Pre-Low raising in japanese pitch accent," *Phonetica*, vol. 74, no. 4, pp. 231–246, Mar. 2017.