Boundary-Driven Downstep in Japanese

Kei Furukawa and Yuki Hirose

The University of Tokyo
keifurukawa314@gmail.com, hirose@boz.c.u-tokyo.ac.jp

ABSTRACT

This study examines a structure-dependent F0 downtrend in coordinated unaccented words in Tokyo Japanese. Ten native speakers of Tokyo Japanese participated in the production experiments. Experiment 1 paradigmatically compared the F0 falls of single unaccented words such as *arakuremono* (villain) and coordinated unaccented words such as *momo-ya booru* (peaches and balls), assuming that a coordinate structure triggered the separation of domains. Experiment 2 syntagmatically compared the F0 falls in regions with a phonological boundary and in regions without a boundary. As expected, in both experiments, the conditions with boundaries showed larger F0 falls than the others. The results revealed another structure-dependent downtrend that is independent from downstep. We argue that this boundary-driven downstep is triggered by the boundary tone %L. It is further contended that downstep, initial lowering, and boundary-driven downstep can be unified taking into account that accent culminativity triggers the separation of domains.

Keywords: Japanese, Downstep, Unaccented, Boundary tone, Prosodic phrasing

1. INTRODUCTION

1.1. Background

Academic interest in F0 downtrend in Japanese has increased over the past decades [8]. There are two categories of F0 downtrend in Japanese: structure-independent and structure-dependent. The former is known as declination, which is the natural tendency for F0 to lower from the beginning to the end of an utterance. This is explained as a physiological phenomenon in which the gradual decrease of air pressure causes a gradual F0 decrease, yet it is also argued that declination is not just a by-product of physiological factors and has some aspects of being structure-sensitive [15]. Several researchers concur that this phenomenon is time-dependent [14].

In contrast, downstep is classified as a structure-dependent F0 downtrend. In Tokyo Japanese, words are classified into accented (A) and unaccented (U), and, conventionally, downstep is described as a pitch range compression triggered by the lexical pitch accent [10]. Two primary approaches help identify downstep in Japanese: syntagmatic and paradigmatic [5]. In the former, if the F0 peak following an accented word is clearly lower than the preceding F0 peak, it is identified as downstep [13]. Downstep differs from declination in that the F0 downfall is acute and the place of downfall is local to the phonological structure [14]. On the contrary, the latter approach claims that if the pitch height of X is significantly lower after an accented word than it is after an unaccented word, X is identified as downstepped [7], [6], [5]. Both approaches assume that downstep in Japanese is only triggered by accented words.

1.2. Research Objectives

This study primarily aims to demonstrate the existence of a structure-dependent F0 downtrend in coordinated unaccented words in Japanese, which is independent of downstep. Experiment 1 paradigmatically compares the F0 falls of single unaccented words and coordinated unaccented words provided that a coordinate structure triggered the separation of phonological domains. Experiment 2 adapts the syntagmatic design wherein the F0 falls in regions with and without a phonological boundary are compared within a sentence. The results demonstrate a structure-dependent downtrend independent of downstep. The study contends that the underlying factor of this boundary-driven downstep is the insertion of the boundary tone %L.

Further, the study aims to discuss the possibility that downstep, initial lowering, and boundary-driven downstep can be combined considering that lexical accent triggers the separation of domains because of accent culminativity and anti-lapse constraint [6].

2. EXPERIMENT 1

2.1 Experimental materials

Experiment 1 paradigmatically compared single unaccented words such as *arakuremono* (villain) and coordinated unaccented words such as *momo-ya booru* (peaches and balls) (Tables 1 and 2). The stimuli were constructed with the Domain Boundary factor comprising two levels: [+DB] and [-DB]. The hypothesis here is that a parallel coordinate structure, as in coordinated conditions, triggers the separation of domains, and phonological boundaries are inserted after the conjunction. This hypothesis can be supported by the fact that *Rendaku*, or sequential voicing, is blocked in compounds that have the semantic value of parallel coordination "A and B," as in *yama-kawa* (a mountain and a river) compared to *yama-gawa* (a river in a mountain) [9].

The tones that were associated with the target nouns in both conditions are assumed to be LHHHHH as the second noun of the coordinated condition begins with a heavy syllable, which normally blocks initial lowering [4].

Table 1: An example of [-DB] condition in Experiment 1.

(1a) single condition: [-DB]

item	arakuremono	ya	warumono-ga	daikiraida.
tone	LHHHHH	H		
gloss	villain	and	bad.guy-NOM	hateful
	'I hate villains and bad guys.'			

Table 2: An example of [+DB] condition in Experiment 1.

(1b) coordinated condition: [+DB]

item	momo-ya	booru	ya	nabebuta-ga	daikiraida.
tone	LHH	HHH	H		
gloss	peach-and	ball	and	pot.lid-NOM	hateful
	'I hate peaches, balls and pot lids.'				

2.2. Participants

Ten native speakers (7 females and 3 males, mean age 19.6 years, SD 1.20) of Tokyo Japanese from Kanto area (Tokyo, Kanagawa, Saitama, and Chiba) participated in the experiment as subjects. No participant had lived outside of Kanto area for more than two years. No participant reported a history of speech or hearing impairments.

2.3. Procedures and analysis

The recording was conducted in a sound-proof booth at the University of Tokyo using the Shure WH20XLR Dynamic Headset Microphone, which was connected to the Roland QUAD-CAPTURE audio interface; the audio was recorded to a computer. The sampling rate was 44.1 kHz.

Stimuli were displayed on a screen one after the other in a pseudo-random order. Participants were asked to read sentences aloud in the following order: twice at a normal speech rate, twice at a slow speech rate, and twice at a fast speech rate, in order to increase the variety of within-subjects duration. The order of reading at fast/slow speech rate was reversed in the second half of the stimuli. Although each speech rate was not fixed, participants were instructed to read at a rate that felt natural to them. Further, participants were instructed to read faster or slower compared to their normal speech rate. When subjects inserted an undesired pause or mispronunciation while reading a sentence, they were asked to repeat the production. A total of 6 items × 2 sentence types × 3 speech rates × 2 repetitions = 72 tokens were recorded. Sixty sentences (360 tokens) from other experiments, including Experiment 2, served as fillers.

Sound files were annotated using Praat [3]. Segmentation between the conjunction and following unaccented words was done on the basis of formants and waveforms. To avoid an undesired insertion of pause or emphasis, tokens at slow speech rate were removed from the analysis. One of two repetitions was analyzed at both normal and fast speech rate. Apparent errors by the algorithm in Praat, such as octave jumps, were manually checked and corrected.

The target words of both conditions contained seven moras, and the F0 peak differences between the first three moras (e.g., *araku*/*momo-ya*) and the last four moras (e.g., *remono-ya*/*booru-ya*) were statistically compared. The boundaries between these regions were annotated at the midpoint of the first consonant of the last four mora (e.g., the midpoint of the consonant [r] in *arakuremono*). The F0 differences were converted to semitones.

Two factors, Domain Boundary ([+DB]/[-DB]) and Duration, were considered as fixed effects. The factor Duration is the standardized duration of the target seven moras (e.g., *arakuremono-ya/momo-ya booru-ya*). This non-categorical factor was analyzed so as to observe the effect of declination, considering declination is affected by duration. The data were analyzed within the linear mixed-effects model (LME), using the lmer function within the lme4 package [2] in R [11], with subjects and items as random effects. The factor labels of Domain Boundary were centered to have a mean of 0 and a range of 1. The final models were obtained with backward selection [1].

2.4. Results

The results of Experiment 1 are presented in Table 3. A linear mixed-effects model revealed that the F0 fall is significantly greater in [+DB] condition than in [-DB] condition. Moreover, the main effect of duration was significant: the longer the duration, the larger the F0 fall. The interaction between the two factors was not significant.

Table 3: Results of mixed-effects models in Experiment 1.

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	0.33232	0.15077	13.07	2.204	<.05	*
Domain Boundary	0.5122	0.22268	11.79	2.3	<.05	*
Duration	0.22459	0.0753	206.46	2.983	<.005	**
Domain Boundary:Duration	0.01551	0.13614	214.5	0.114	n.s.	

2.5. Discussion

The significant main effect of Duration can be attributed to declination, assuming it is time-dependent.

However, the significant main effect of Domain Boundary suggests that a parallel coordinate structure triggers the separation of phonological domains, and the observed F0 downtrend must involve a mechanism that is dependent on phonological structures. If structure-independent downtrend is the only mechanism to cause the F0 downtrend for unaccented words, the F0 range should not substantially differ between [+DB] and [-DB] conditions. Additionally, these results cannot be explained by downstep, as the target words do not contain any accents.

Furthermore, these results cannot be attributed to initial lowering, as boundary tone %L is not associated with the first syllables when they are heavy. This point is further examined in Experiment 2.

3. EXPERIMENT 2

3.1 Experimental materials

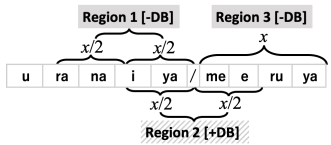
Table 4 shows an example of a total of six sentences used as stimuli. The targets were four-mora and three-mora unaccented words coordinated by the conjunction *ya*, as in *uranai-ya meeru* (fortune telling and e-mails). The initial syllables of the second unaccented words are heavy and consist of long vowels.

The hypothesis here is that a phonological boundary is yet to be inserted after the conjunction *ya*. The boundaries were annotated at the midpoint of the first consonant of the second noun (e.g., The midpoint of the consonant [m] in *meeru*). Each recorded utterance was divided into three regions so that the duration of each region is fixed to the duration of 4 moras (x) after the phonological boundary (e.g., *meeru-ya*). Region 1 is defined as x before the boundary ([-DB]), Region 2 is $x/2$ before the boundary plus $x/2$ after the boundary ([+DB]), and Region 3 is x after the boundary ([-DB]). The duration of 4 moras after the boundary is longer than the 4 moras before the boundary in all utterances. The duration of each region is fixed in order to eliminate the Duration factor from the statistic model. If we fixed the number of moras in each region instead, the duration of each region would covary.

Table 4: An example of the stimuli in Experiment 2.

item	uranai	ya	meeru	ya	marason-ga	zinsei-no-sasae-da
tone	LHHH	H	HHH	H		
gloss	Fortune.telling	and	e-mails	and	marathons-NOM	life-GEN-support-be
	'Fortune telling, e-mails and marathons are support for our lives.'					

Figure 1: Schematic of each region in Experiment 2.



3.2. Participants

The speakers in Experiment 2 were the same as in Experiment 1.

3.3. Procedures and analysis

The basic procedures were identical to those of Experiment 1. Given that the regions within a sentence are compared in a syntagmatic manner, sentence types are not relevant in Experiment 2. A total of 6 sentences × 3 speech rate × 2 repetitions = 36 tokens were recorded. Sixty-six sentences (396 tokens) from other experiments, including Experiment 1, served as fillers. To avoid an undesired insertion of pause or emphasis, tokens at a slow speech rate were removed from the analysis. One of two repetitions was analyzed at both normal and fast speech rates. The F0 average difference between the first and the latter halves of each region were measured. The F0 differences were converted to semitones. The statistical procedures were identical to those of Experiment 1, except that only the Domain Boundary factor was considered in Experiment 2: Domain Boundary ([+DB]/[-DB]).

3.4. Results

A linear mixed-effects model revealed that the F0 descent is significantly more extensive in Region 2 than in Regions 1 or 3 (Tables 5 and 6 and Figure 2).

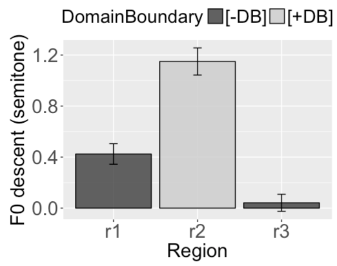
Table 5: Results of mixed-effects models between Region 1 and Region 2 in Experiment 2.

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	0.66024	0.06124	9.875	10.781	<.0001	***
Domain Boundary	0.72596	0.18455	10.036	3.934	<.005	**

Table 6: Results of mixed-effects models between Region 1 and Region 2 in Experiment 2.

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	0.59721	0.05966	9.817	10.01	<.0001	***
Domain Boundary	1.10151	0.18143	10.027	6.071	<.0005	***

Figure 2: Mean F0 descent. Error bars represent 95% confidence intervals.



3.5. Discussions

The significant effect of DB indicates that the observed F0 downtrend involves a structure-dependent mechanism that is triggered by a parallel coordinate structure. The fact that the F0 range differed substantially between Regions 1 and 2 or between Regions 2 and 3 means that the place of downfall is local to the phonological structure. In addition, downstep cannot account for the results of Experiment 2, since the target words are all unaccented.

Furthermore, these results cannot be due to initial lowering, since the association of boundary tone %L to the first syllables should be blocked when they are heavy. If the boundary tone %L is associated to the first mora and initial lowering occurs, then Region 3 should have shown the large negative value, which was not the case.

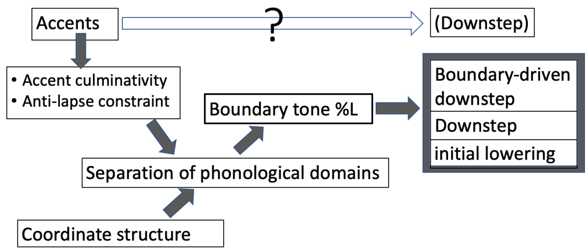
4. GENERAL DISCUSSIONS

Both the experiment results reveal a structure-dependent downtrend independent from downstep. Thus, the study contends that this "boundary-driven downstep" is triggered by a boundary tone %L, as boundary-driven downstep has been observed with the phonological boundaries. Boundary-driven downstep differs from declination in that (i) it is driven by a boundary tone and (ii) the place of downfall is local to the phonological structure.

This implies the possibility that downstep, which has been conventionally argued to be triggered by accents, can be theoretically unified with boundary-driven downstep. According to [6], there must be a phonological boundary after every accent owing to accent culminativity and the anti-lapse constraint. In other words, AA (accented + accented) sequence is separated because of accent culminativity, and AU (accented + unaccented) sequence is separated because of the anti-lapse constraint. Assuming that boundary-driven downstep is triggered by boundary tone %L, both AA and AU sequences meet the requirement of boundary-driven downstep, and there is a possibility that downstep is not directly triggered by accents but is rather triggered by the separation of domains (Figure 3); that is, downstep and boundary-driven downstep could essentially be the same phenomenon.

The study further argues that boundary-driven downstep and initial lowering can be unified, given that initial lowering is the association of boundary tone %L with the left edge of phonological phrases (Figure 3). Assuming that boundary-driven downstep is driven by the insertion of boundary tone %L, downstep, initial lowering, and boundary tone downstep can be seen as a unified phenomenon.

Figure 3: Unifying downstep, initial lowering, and boundary-driven downstep



Unifying downstep, initial lowering, and boundary-driven downstep has three advantages. First, the structure-dependent downtrend can be justified at the sequence of unaccented words as in Experiment 1, Experiment 2, and the structural downtrend in a sequence of unaccented words reported by [12]. Second, we can account for the structural downtrend without initial lowering. Initial lowering refers to the %L associated with the first syllable of the domain, and it can be blocked if the syllable is accented or heavy. By positing that every phonological phrase is the domain of the boundary tone %L, we can account for the downtrend without the associated %L, as shown in the two experiments. Third, we can simplify the theory by not assuming downstep or initial lowering as an independent mechanism.

There are several limitations to this study. It is possible that the coordination conditions used in the two experiments comprise an extremely special case, and as such, the results cannot be generalized with accounting for boundary tone insertion. However, the theory proposed by this study may provide useful information regarding the interaction between boundary tone and downstep. Further investigation concerning other factors of the separation of phonological domain is necessary.

5. REFERENCES

[1] Bates, D., Kliegl, R., Vasishth, S. and Baayen, H. 2015. Parsimonious mixed models. arXiv preprint arXiv:1506.04967.

[2] Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R.H.B., Singmann, H., Dai, B., Grothendieck, G., Green, P. 2015. Package "lme4" (Version 1.1-12).

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

[4] Gussenhoven, C. 2004. *The phonology of tone and intonation.* Cambridge University Press.

[5] Ishihara, S. 2016. Japanese downstep revisited. *Nat. Lang. Linguist. Theory* 34, 1389–1443.

[6] Ito, J., Mester, A. 2013. Prosodic subcategories in Japanese. *Lingua* 124, 20–40.

[7] Kubozono, H. 1993. *The organization of Japanese prosody.* Tokyo: Kurosio.

[8] McCawley, J.D. 1968. The phonological component of a grammar of Japanese.

[9] Ōtsu, Y. 1980. Some aspects of Rendaku in Japanese and related problems. *Theoretical issues in Japanese linguistics.* Dept. of Linguistics and Philosophy, Massachusetts Institute of Technology.

[10] Poser, W.J. 1984. The phonetics and phonology of tone and intonation in Japanese.

[11] R Core Team. 2016. R: *A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[12] Selkirk, E. Shinya, T., Sugahara, M., 2003. Degree of initial lowering in Japanese as a reflex of prosodic structure organization. *Proc. 15th ICPhS* Barcelona, Spain, 491–494.

[13] Selkirk, E., Tateishi, K. 1991. Syntax and downstep in Japanese. In: *Interdisciplinary Approaches to Language*. Dordrecht: Springer. 519–543.

[14] Shinya, T. 1999. Eigo to nihongo ni okeru fookasu ni yoru daunsteppu no sosi to tyooon-undoo no tyoogoo [The blocking of downstep by focus and articulatory overlap in English and Japanese]. *In Proc. of Sophia Linguistics Society,* 35–51.

[15] Yuan, J., Liberman, M. 2014. F0 declination in English and Mandarin broadcast news speech. *Speech Communication* 65, 67–74.