Proceedings of Disfluency in Spontaneous Speech, DiSS 2013

Acoustic and linguistics features related to speech planning appearing at weak clause boundaries in Japanese monologs

Hanae Koiso[1] & Yasuharu Den[2]

[1] National Institute for Japanese Language and Linguistics, Japan
[2] Faculty of Letters, Chiba University, Japan

Abstract

In this paper, we focus on weak clause boundaries in Japanese monologs in order to investigate the relationship of the length of constituents following weak boundaries to three acoustic and linguistic features: 1) occurrence rate of fillers, 2) occurrence rate of boundary pitch movements, and 3) degree of lengthening of clause-final morae. We found that all these features were significantly correlated with the length of following constituents. Most importantly, boundary pitch movements had an additional effect that can be distinct from the effect of clause-final lengthening. These results suggest that Japanese speakers have *earlier-occurring* items that help them deal with cognitive load in speech planning, in addition to fillers and other clause-initial disfluencies.

Index Terms: fillers, boundary pitch movements, clause-final lengthening, Japanese monologs

1. Introduction

In spoken Japanese, successive clauses are frequently linked up, resulting in a long stretch of them in a sentence or utterance [1]. Since Japanese is a predicate-final language, this clause chaining consists of more than one finite or non-finite predicate clause, each followed by a conjunctive particle, and terminates at a clause with a finite predicate. Thus, Japanese speakers can formulate utterances with an arbitrary and in principle unlimited number of clauses by manipulating the final elements of these clauses.

Clauses in Japanese can be classified into those with strong boundaries and those with weak boundaries, according to their degree of dependency on the subsequent clauses. Watanabe [2] has found that the rate of fillers at strong clause boundaries is higher than the rate at weak clause boundaries. In addition, Koiso [3] has found that the boundary pitch movements (BPMs) of accentual phrases occur more frequently at strong clause boundaries than at weak clause boundaries.[1] These facts suggest that, in the course of clause chaining, fillers and BPMs are used at similar locations, that is, where speech planning mainly takes place. However, this does not necessarily mean that BPMs have a relation to planning difficulty as fillers do, since BPMs generally involve the lengthening of segments, which is itself considered to be related to planning difficulty [4].

There is another fact that may indicate the possibility that clauses with weak boundaries are also locations for speech planning. Watanabe [2] has reported a tendency that the longer the following clause, the more frequently fillers occur at weak clause boundaries, but not at strong clause boundaries. Also, Watanabe and Den [5] have found that the duration of a filler *e* is positively correlated with the length of the following clause at weak clause boundaries only. These results suggest that weak clause boundaries may also serve as locations for speech planning when the following clause is long and complex.

In this paper, we focus on weak boundaries within utterances and investigate how the complexity of the following constituents is related to three acoustic and linguistic features: fillers, clause-final lengthening, and BPMs, any of which may serve to gain time for speech planning. For this purpose, we conduct a quantitative analysis of a large-scale corpus of spontaneous Japanese monologs, and show the relation between the length of the following constituents and these features.

2. Method

2.1. Data

We used the *Corpus of Spontaneous Japanese* [6] in the present analysis. We selected 177 monologs in its "Core" data set (CSJ-Core), which consists of "Academic Presentation Speech" (APS, 70 monologs) and "Simulated Public Speech" (SPS, 107 monologs). APS consists of live recording of academic presentations covering meetings of scholars in engineering and the humanities and social sciences, while SPS consists of public speeches of about 10–12 minutes given by laypeople on everyday topics like "my most delightful memory" and "the town I live in" in front of small, friendly audience. Including both APS and SPS, there were 78 female and 99 male speakers, ranging in age from their early 20s to their late 60s.

2.2. Annotation

CSJ-Core contains a variety of hand-corrected annotations, including clause units, *bunsetsu* phrases (see below), long- and short-unit words, phonetic segments, dependency structures, and prosodic information. Clause boundaries are automatically detected by the CBAP-csj program [7] and classified into one of the following three categories based on their degree of completeness as a syntactic and semantic unit and on that of dependency on the subsequent clause:

Absolute Boundary (AB) corresponds to the sentence boundary in the usual sense.

Strong Boundary (SB) is the boundary of a clause that is relatively independent of the subsequent clause.

Weak Boundary (WB) is the boundary of a clause that is relatively dependent on the subsequent clause.

[1] Note, however, that weak boundaries defined in Watanabe [2] and Koiso [3] are different from each other; weak boundaries in Koiso are all located within utterances, whereas those in Watanabe correspond to utterance-final position approved by some pragmatic criteria.

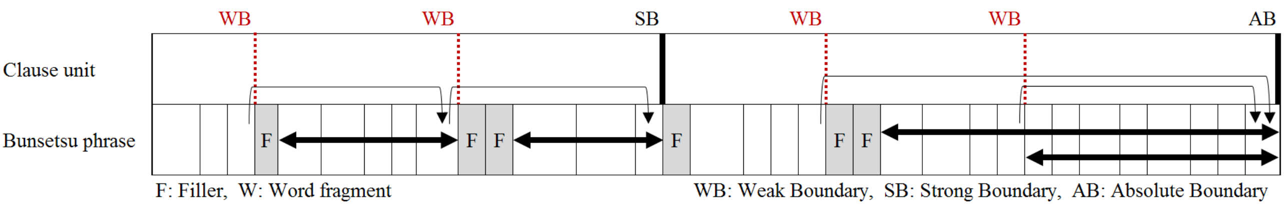

Figure 2: *The lengths of constituents following WBs. The arcs indicate dependency between the clauses ending with WBs and the clauses they modify. The length of the constituents was measured by the duration of the interval from the left edge of the phrase immediately following the WB in question to the right edge of the clause modified by the WB clause. Fillers (F) appearing at the beginning of this interval were excluded from this calculation.*

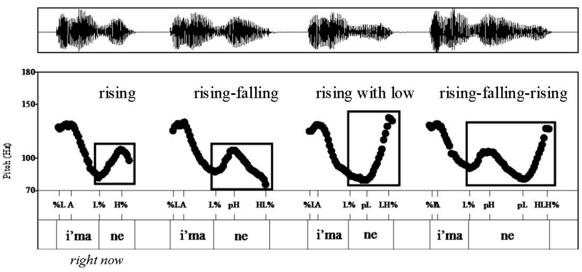

Figure 1: *Four main types of BPMs in Japanese: a simple rising tone (H%), a rising–falling tone (HL%), a rising tone with a sustained low (LH%), and a rising–falling–rising tone (HLH%). BPMs are indicated by squares. Note that BPM is always preceded by a low tone (L%) marking the right edge of an accentual phrase.*

On the basis of these categories, *clause units* are defined as clauses ending with either absolute or strong boundaries. The clause units identified were carefully checked by expert labelers. WBs located within these hand-corrected clause units are regarded as utterance-internal breaks – they are the target of the present study. Clause units are segmented in terms of *bunsetsu* phrases, which consist of one content word possibly followed by one or more function words. Dependency structures between *bunsetsu* phrases were also labeled, again by expert labelers.

In addition to these syntactic and morphological annotations, CSJ-Core was also annotated in term of prosody using the X-JToBI scheme [8]. Among the labels of X-JToBI, we focus on the final boundary tones of accentual phrases, which are a fundamental prosodic unit in Japanese. The right edge of an accentual phrase is always characterized by a low tone (L%), which can be followed by an additional tone such as a simple rising tone (H%), a rising–falling tone (HL%), a rising tone with a sustained low (LH%), or a rising–falling–rising tone (HLH%) (see Figure 1). These extra movements at the right edge of accentual phrases are called *boundary pitch movements* (BPMs).

2.3. Variables for stastistical analysis

The dependent variable of the analysis was the length of the constituent following a WB, which was taken to function as a rough estimate of the cognitive load imposed on speakers by speech planning at that boundary. The length of the constituent was measured by the duration between the start of the phrase immediately following the WB and the end of the clause modified by the WB clause (see Figure 2). Fillers appearing at the beginning of the interval were excluded from this calculation.

Three acoustic and linguistic features at WBs were considered as factors that may affect the dependent variable:

1. presence or absence of fillers that immediately follow the WB clause (Filler);
2. presence or absence of a BPM at the end of the WB clause (BPM); and
3. duration of the last mora of the WB clause (DurLastMora).

The duration variables, that is, the dependent variable and the third independent variable, were log-transformed and standardized before statistical analysis.

2.4. Target of the analysis

We excluded from the analysis clauses ending with WB that did not modify any other clause or that did not coincide with accentual phrase boundaries. Of 19,186 WBs in the data, 16,245 instances were retained for the analysis.

3. Results

The top row in Figure 3 shows the distribution of the duration of constituents following a WB depending on the presence or absence of following fillers and BPMs and the scatter plot between the duration of constituents and the duration of the last mora. The duration of constituents was longer in the presence both of following fillers and of BPMs, and also when the duration of the last mora of the WB clause lengthened.

Furthermore, when we restricted WB clauses to major subclasses, namely, causal clauses (1,390 instances), conditional clauses (1,613 instances), and *te*-marked clauses (5,520 instances), the same tendencies were replicated (see the second, third and fourth rows in Figure 3).

In order to statistically test the effects of the three variables, we applied linear mixed-effects models with random intercept for speakers. The (means of the) estimated parameters and the *p*-values were calculated using Markov Chain Monte Carlo (MCMC) sampling implemented in lme4 and languageR packages of the R language [9].

Table 1 shows the results. All three independent variables had significant effects on the duration of constituents following WBs. This result was consistent throughout the four data sets (that is, all WBs and causal, conditional, and *te*-marked subclasses.

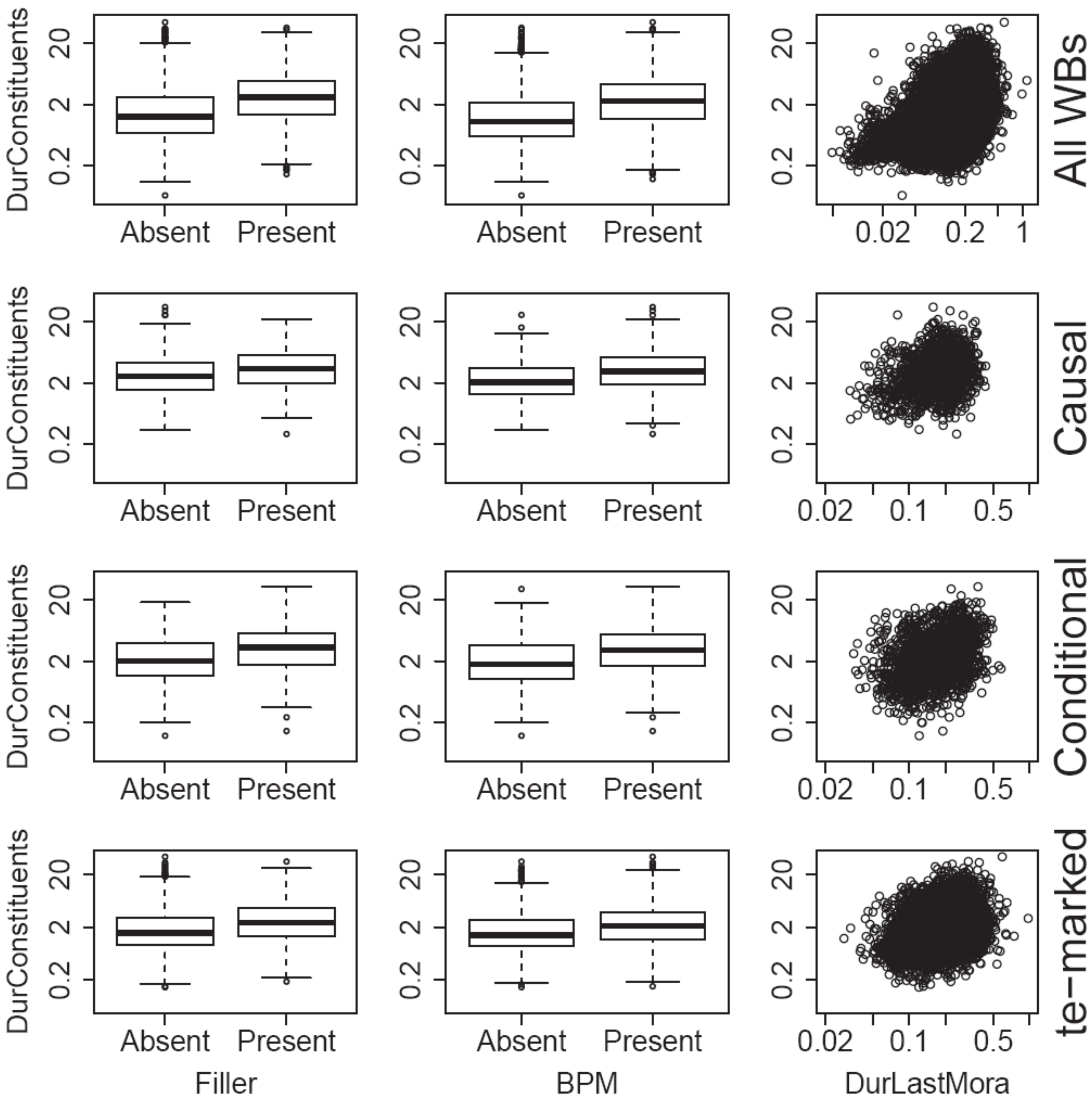

Figure 3: *The relations between the duration of constituents following WBs and the three independent variables (*`Filler`*, BPM, and DurLastMora). The top row shows the results for all WBs, and the other three rows show the results for causal, conditional, and te-marked subclasses, respectively.*

Table 1: *Estimated parameters of the linear mixed-effects models applied to WBs all together and to causal, conditional, and te-marked subclasses. σs indicates the standard deviation at the speaker level and σ indicates the residual standard deviation. The means and p values of the coefficients were obtained by using MCMC sampling.*

	Fixed-effect	MCMC Mean	MCMC p
All WBs	(Intercept)	−.200	.0001
	`Filler`	.333	.0001
	`BPM`	.293	.0001
	`DurLastMora`	.357	.0001
		σ_S = .112, σ = .861	

	Fixed-effect	MCMC Mean	MCMC p
Causal	(Intercept)	.370	.0001
	`Filler`	.199	.0001
	`BPM`	.192	.0002
	`DurLastMora`	.227	.0001
		σ_S = .119, σ = .725	

	Fixed-effect	MCMC Mean	MCMC p
Conditional	(Intercept)	.228	.0001
	`Filler`	.199	.0006
	`BPM`	.168	.0018
	`DurLastMora`	.308	.0001
		σ_S = .139, σ = .847	

	Fixed-effect	MCMC Mean	MCMC p
te-marked	(Intercept)	−.058	.0001
	`Filler`	.189	.0001
	`BPM`	.128	.0001
	`DurLastMora`	.336	.0001
		σ_S = .115, σ = .841	

4. Discussion

We found that all the three independent variables in this study had significant effects on the length of constituents following WBs. The duration of the constituents was longer when there were fillers immediately following than when there were not; duration also increased in the presence of BPMs at the boundaries; and finally, the duration of the constituents lengthened when the duration of the last mora of the WB clause did. These results suggest that fillers, BPMs, and clause-final lengthening are used when Japanese speakers are about to produce a long, complex clause and, hence, have an increased need to manage cognitive load in their speech planning.

The same results were replicated in all the analyses of the three major subclasses of the data, that is, in causal, conditional, and *te*-marked clauses. This means that these tendencies were widely and consistently observed across various types of WBs.

Fillers and word repetitions are reported to occur very frequently at utterance-initial positions, where they help speakers gain time for speech planning [10, 11]. In Japanese, these linguistic devices for speech planning have been found also to occur at clause-initial positions [12, 2]. Our result on fillers following WBs is consistent with these previous findings. However, the other two features found to be relevant to speech planning (that is, BPMs and clause-final lengthening) are not located at the beginning of a new clause but at the end of the preceding clause. In this sense, it can be stated that Japanese speakers have *earlier-occurring* items, as well as clause-initial ones, to deal with cognitive load in speech planning.

The result regarding BPM is the most surprising here. BPM generally involves the lengthening of the last mora of the clause, which bears the complex tones involved. Since this lengthening is considered in and of itself to be related to planning difficulty, it is tempting to infer that it may be the *duration*, and not the *presence,* of BPMs that is relevant to speech planning. However, this is not the case: *both* BPMs themselves and the lengthening of the morae bearing them have significant effects on the length of constituent after WBs. This suggests that the presence of BPMs *per se* has some relation to the complexity of the following constituents.

One possible interpretation of the correlation of BPMs to constituent complexity is that BPMs are in fact related to the continuity of utterances and that the continuity in turn is related to the complexity. As Koori [13] pointed out, rising tones in Japanese tend to occur at syntactically and semantically deep boundaries, and indicate continuing or on-going speech. When the constituent following the WB is long and complex, the boundary will be relatively deep. On these occasions, speakers may use BPMs to indicate the continuation of the utterance being construed. We will explore this possibility in future research.

5. Acknowledgements

This work is supported by Grant-in-Aid for Collaborative Research Project of NINJAL "Empirical study on the role of prosodic features in conversations" led by Hanae Koiso, and Grant-in-Aid for Scientific Research (B) "Constructing cognitive and communicative models for dialog based on utterance-unit annotations" led by Yasuharu Den.

6. References

[1] S. Iwasaki and T. Ono, "'Sentence' in spontaneous spoken Japanese discourse," in *Complex sentences in grammar and discourse*, J. Bybee and M. Noonan, (Eds), Amsterdam: Benjamins, 2001, pp. 175–202.

[2] M. Watanabe, *Features and roles of filled pauses in speech communication: A corpus-based study of spontaneous speech*. Tokyo: Hituzi Syobo, 2009.

[3] H. Koiso, "Nihongo-hanasi-kotoba-koopasu-o motii-ta hukugoo-kyookai-ontyoo-no hatugen-keizoku-hyoozi-kinoo-no kentoo (Continuation function of boundary pitch movements in the Corpus of Spontaneous Japanese)," in *Proceedings of the 2nd Workshop on Japanese corpus linguistics*, pp. 221–230, 2012.

[4] J. E. Fox Tree and H. H. Clark, "Pronouncing "the" as "thee" to signal problems in speaking," *Cognition*, vol. 62, pp. 151–167, 1997.

[5] M. Watanabe and Y. Den, "Utterance-initial elements in Japanese: A comparison among fillers, conjunctions, and topic phrases," in *Proceedings of the DiSS-LPSS Joint Workshop 2010*, Tokyo, 2010, pp. 31–34.

[6] K. Maekawa, "Corpus of Spontaneous Japanese: Its design and evaluation," in *Proceedings of the ISCA and IEEE Workshop on Spontaneous speech processing and recognition*, Tokyo, 2003, pp. 7–12.

[7] K. Takanashi, T. Maruyama, K. Uchimoto, and H. Isahara, "Identification of "sentences" in spontaneous Japanese: Detection and modification of clause boundaries," in *Proceedings of the ISCA and IEEE Workshop on Spontaneous speech processing and recognition*, Tokyo, 2003, pp. 183–186.

[8] K. Maekawa, H. Kikuchi, Y. Igarashi, and J. J. Venditti, "X-JToBI: An extended J_ToBI for spontaneous speech," in *Proceedings of the 7th International Conference on Spoken Language Processing (INTERSPEECH 2002)*, Denver, CO, pp. 1545–1548, 2002.

[9] R. H. Baayen, *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press, 2008.

[10] H. H. Clark, "Speaking in time," *Speech Communication*, vol. 36, pp. 5–13, 2002.

[11] H. H. Clark and T. Wasow, "Repeating words in spontaneous speech," *Cognitive Psychology*, vol. 37, pp. 201–242, 1998.

[12] Y. Den, "Are word repetitions really intended by the speaker?" in *Proceedings of the ISCA Tutorial and Research Workshop on Disfluency in Spontaneous Speech*, 2001, pp. 25–28.

[13] S. Koori, "Onsee-no tokutyoo-kara mi-ta bun (Sentence characteristics in terms of phonetic aspects)," *Nihongogaku*, vol. 15, no. 9, pp. 60–70, 1996.