Proceedings of DiSS-LPSS Joint Workshop 2010,
25-26 September 2010, Tokyo, Japan, pp. 31-34.

Utterance-Initial Elements in Japanese: A Comparison among Fillers, Conjunctions, and Topic Phrases

Michiko Watanabe[1], *Yasuharu Den*[2]

[1]Graduate School of Frontier Sciences, The University of Tokyo, Japan
[2]Faculty of Letters, Chiba University, Japan
watanabe@k.u-tokyo.ac.jp, den@cogsci.l.chiba-u.ac.jp

Abstract

Speakers need to plan the following part of speech under the pressure of a temporal imperative at utterance-initial positions. Each language seems to have some devices to solve this problem, which we call *utterance-initial elements* (UIEs). We investigated effects of two factors, boundary strengths and complexity of the following constituents, on the durations of possible UIEs, such as fillers, conjunctions, and topic phrases. We found that the last mora of filler *e*, as well as *wa*-marked topic phrases, became longer as the complexity increased in certain conditions. Possible interpretations for the results are discussed.
Index Terms: utterance-initial elements, prolongation, boundary strengths, constituent complexity

1. Introduction

Utterance-initial position is an important place where speakers should resolve the *communication-cognition trade-off* problem. From a cognitive perspective, speakers may sometimes require much time in planning the content, and formulating the structure, of the constituent to be produced. This cognitive load may be particularly heavy at the utterance-initial position in which speakers should prepare for the whole utterance. From a communicative viewpoint, on the other hand, speakers are pressed by a temporal imperative; they may avoid pausing too long so that they prevent themselves from being heard as opting out or distracted or losing an opportunity to take, or keep, the turn. Thus, speakers at the utterance-initial position, should manage the trade-off between the time required by the cognitive process and the time suitable for the communicative process.

The grammar of a language seems to equip particular devices to solve this communication-cognition trade-off. For instance, fillers, such as *um* and *uh* in English, are considered to serve as such a device [1]. Word repetitions may also function as a strategy to secure time for speech planning without being suffered from communication problems [2].

Our previous studies have been seeking for such devices in Japanese [3, 4, 5]. Since the communication-cognition trade-off is most critical at the beginning of an utterance, these devices would be frequently used at utterance-initial positions. When used at utterance-initial positions, we call them *utterance-initial elements* (UIEs).

Two factors have been reported to be relevant to types, frequencies and acoustic features of UIEs. The one is the boundary depth in terms of discourse and syntactic structure. Swerts [6] found that the filler rate, particularly the *um* rate, is higher at deeper discourse segment boundaries than at shallower ones in Dutch monologs. Fillers at deeper discourse segment boundaries are acoustically more prominent than those at shallower boundaries as well. Watanabe et al. [5] classified clause boundaries into two types, strong and weak boundaries, according to the degree of dependency of the preceding clause on the following clause. The filler rate at strong boundaries was found to be higher than the rate at weak boundaries. However, the filler rate at sentence boundaries, which were assumed to be stronger than clause boundaries, was not higher than the rate at strong clause boundaries. At sentence boundaries conjunctions appear much more frequently than at strong clause boundaries: 41% at sentence boundaries and 10% at strong clause boundaries. Therefore, we infer that time for producing conjunctions may also be used for planning speech. In this paper, we hypothesize topic phrases to be another device to gain time for planning.

The other factor that is likely to affect UIEs is the complexity of the following constituents. Clark & Wasow [2] found that the more complex the following constituent is, the higher the repetition rates of articles and pronouns are in English. Watanabe et al. [5] reported that there was an interaction between the boundary strengths and the complexity of the following constituent. The filler rates at weak clause boundaries tend to be higher, the longer the clause is; the filler rates at sentence and strong clause boundaries did not differ depending on the length of the following clause. These findings indicate that the complexity of the following constituents is reflected in the features of UIEs only when speakers do not need to be heavily involved in planning the content of utterance.

In this paper, we investigate whether and how boundary strengths and complexity of the following constituents affect the durations of UIEs. We take three types of possible UIEs into consideration: fillers, conjunctions, and topic phrases. In Japanese, topic phrases most often appear at the beginning of sentences, or clauses, and are marked with particle *wa* at the end of phrases. We assume that topic phrases, or their boundaries, are locations where speech planning sometimes takes place.

2. Method

2.1. Data

A part of the *Corpus of Spontaneous Japanese* [7] was used for the current study. From among the entire data, 177 monologs in the *Core* data were selected, which come with hand-corrected annotation of clause units, 'bunsetsu' phrases, long- and short-unit words, and phonetic segments. The data were classified into two groups according to recording source: academic presentation speech (APS) and simulated public speech (SPS). APS is the live recording of academic presentations for several academic societies. SPS, on the other hand, is studio recorded speeches of paid layman speakers, of about 10-12 minutes, on everyday topics presented in front of a small audience. The

Table 1: Summary of the data

	APS	SPS	Total
No. of sessions	70	107	177
Duration	18.8 hrs	19.9 hrs	38.6 hrs
No. of clauses	8,516	9,675	18,191
No. of phrases	79,332	100,374	195,075
No. of long-unit words	176,848	197,751	374,599
No. of short-unit words	218,161	225,572	443,733
No. of fillers	15,567	14,545	30,112

speakers were 24 females and 46 males in APS and 54 females and 53 males in SPS, ranging in age from their early 20s to late 60s, with a median at the mid 30s. Table 1 shows the summary statistics of the data.

2.2. Annotation

For the selected 177 monologs, the boundaries of clauses, phrases, and words were provided in the corpus. Words were segmented in two different ways: *short-unit words* and *long-unit words*. Short-unit words correspond roughly to simple words, whereas long-unit words cover compound words. For each word, a part of speech and starting and ending times were described. Fillers, such as *ano* and *eeto*, were treated as genuine words. In particular, fillers *ee* and *maa* were regarded as prolonged forms of *e* and *ma*, respectively, since the shorter forms were also frequently used.

Phrases were segmented in terms of 'bunsetsu,' which is a widely used notion in Japanese linguistics. A 'bunsetsu' consists of one content, long-unit word, possibly followed by one or more function words including grammaticalized ones.

Clauses were identified as *clause units*. Clause units were segmented based mainly on syntactic criteria, but additional discourse factors were also taken into account in marking weaker syntactic boundaries as clause-unit boundaries [8]. For each clause unit, a boundary type according to the morpho-syntactic form of the final word in that clause was given. Three boundary types were distinguished:

Absolute Boundary (AB): The clause unit ends with a verb, adjective, or auxiliary verb in conclusive or imperative form, possibly followed by (a) final particle(s).

Strong Boundary (SB): The clause unit ends with a coordinate conjunctive particle such as *kedo* and *ga*.

Weak Boundary (WB): The clause unit, ending with a subordinate conjunctive particle such as *kara* and *node*, was perceived as exhibiting disjuncture due to some discourse factors, such as preface to a new topic, summary of the current topic, topic-shift, etc.

Note that weak boundaries are not necessarily regarded as weaker than absolute and strong boundaries. Although they are weaker than absolute and strong boundaries syntactically, the discourse factors that mark them as clause-unit boundaries may increase their boundary strength.

2.3. Target of the analysis

For the 18,191 clauses in the data, the initial phrases of the clauses were extracted, and the part of speech patterns of the long-unit words contained in these clause-initial phrases were collected. Conjunctions, fillers, and particles were subclassified according to their word forms. Table 2 shows the top 12 frequent part of speech patterns appearing in the clause-initial phrases. Although the places are slightly different between the data for the academic presentation speech and the simulated public speech, many patterns are common to both speech types. In particular, the following patterns are interesting from the viewpoint of the current study, since they indicate that specific lexical items are repeatedly used at utterance-initial positions: (a) fillers: *ano*, *e*, *eeto*, *ma*; (b) conjunctions: *de*; (c) topic phrases: "Noun + *wa*", "Pronoun + *wa*"; and (d) genitive phrases: "Noun + *no*".

Table 2: Top 12 frequent part of speech patterns appearing in the clause-initial phrases

	APS		SPS	
#	Part of Speech	%	Part of Speech	%
1	F_*e*	27.31	CNJ_*de*	20.01
2	CNJ_*de*	17.18	F_*e*	9.79
3	Adverb	4.37	F_*ma*	8.23
4	F_*eeto*	3.86	Adverb	6.58
5	F_*ma*	3.77	F_*ano*	5.24
6	Noun + CP_*no*	3.51	F_*eeto*	3.25
7	Adnoun	2.99	Noun	2.98
8	Pronoun + TP_*wa*	2.13	Adnoun	2.89
9	Noun	2.04	Pronoun + TP_*wa*	2.23
10	Noun + TP_*wa*	1.83	CNJ_*sorede*	2.13
11	F_*ano*	1.81	Noun + CP_*no*	1.86
12	CNJ_*sorekara*	1.77	Noun + TP_*wa*	1.80

F: filler, CNJ: conjunction, CP: case particle, TP: topic particle

In this paper, we focus on the 7 items from the first three patterns, i.e., *ano*, *e*, *eeto*, *ma*, *de*, "Noun + *wa*," and "Pronoun + *wa*," and investigate factors that determine the durations of the final morae in these items.

2.4. Measurements

The dependent variable of the study was the duration of the final mora of clause-initial phrases. For the selected 7 target items, the final mora consisted of *no*, *e*, *to*, *ma*, and *de* for the first 5 items, respectively, and *wa* for the remaining 2 items. All values were converted, after log-transformation, into z-scores on a per-speaker basis.

Two major independent variables were considered as possible factors that may have effect on the dependent variable: i) the strength of the clause boundary preceding the target phrase (`preCB`) and ii) the complexity of the constituents following the target phrase (`comp`). Their interaction (`preCB:comp`) was also taken into account. The boundary strength was a categorical variable with three levels, `AB`, `SB`, and `WB`, determined by the type of the preceding clause boundary. The complexity was measured by the (log-transformed) number of short-unit words contained in the clause after the target phrase. In calculating complexity, fillers were always excluded.

In addition to these two major factors, the following four variables were considered as possible covariates:

i. the presence of a pause immediately preceding the target phrase (`ifPrePau`)

ii. the (log-transformed) duration of the preceding pause (`durPrePau`)

iii. the presence of a pause immediately succeeding the target phrase (`ifSucPau`)

Table 3: Summary of the optimal models

	APS							SPS						
	ano	*e*	*eeto*	*ma*	*de*	N+*wa*	PRO+*wa*	*ano*	*e*	*eeto*	*ma*	*de*	N+*wa*	PRO+*wa*
preCB	ns	***	**	ns	+	ns	ns	ns	ns	ns	ns	ns	ns	+
comp	ns	ns	ns	ns	ns	**	ns	ns	ns	ns	ns	ns	ns	ns
preCB:comp	ns	***	ns	ns	ns	ns	ns	ns	*	ns	ns	ns	ns	ns
ifPrePau	ns	ns	ns	***	ns	***	ns	ns	ns	ns	***	ns	+	ns
durPrePau	ns	***	ns	+	***	+	ns	ns	**	ns	+	***	ns	*
ifSucPau	***	***	***	*	***	ns	***	***	***	***	***	***	***	***
ifSucFil	ns	***	ns	***	***	***	ns	***	***	*	+	***	ns	ns

***: $p < .001$, **: $p < .01$, *: $p < .05$, +: $p < .1$, ns: not significant or not survived in an optimal model

iv. the presence of a filler immediately succeeding the target phrase (ifSucFil)

2.5. Statistical analysis

In order to identify factors affecting the duration of the final mora of clause-initial phrases, regression models with different combinations of independent variables were fitted to the data, and the optimal model was chosen by using a model selection technique based on AIC. Since the data was clustered by speakers, linear mixed-effects regression, rather than ordinary linear regression, was applied using packages lme4 and languageR of the R software environment [9]. To remove the extreme outliers from the data, a full model with all independent variables was first applied, and the outliers with standardized residuals at a distance greater than 3 standard deviations from zero were removed. Then, model selection was conducted on the remaining data. All p-values were obtained using Markov Chain Monte Carlo (MCMC) sampling, and they are, thus, conservative.

3. Results

Table 3 summarizes the significance of the effect of each variable on each item in the academic presentation speech (APS) and the simulated public speech (SPS). When an independent variable was not survived in the optimal model for a particular item, symbol 'ns' is place on the corresponding cell. When a variable was survived in an optimal model but its MCMC p-value was not significant, it is indicated by 'ns' as well. Only variables remaining in an optimal model whose MCMC p-values were less than 10% are marked by '+', '*', '**', or '***' depending on their significance levels.

For all items, except for topic phrase "Noun + *wa*" in APS, the presence of an immediately succeeding pause had a significant effect on the duration of the final mora of a clause-initial phrase; the duration was longer when there was a succeeding pause. The presence of an immediately succeeding filler had a similar effect, although this effect was not observed uniformly. Over the items and speech types, there was a reliable effect of the duration of an immediately preceding pause for filler *e* and conjunction *de* in both APS and SPS and for noun phrase "Pronoun + *wa*" in SPS; the longer the preceding pause is, the longer the duration is.

The two major factors of the study, i.e., boundary strengths and complexity of the following constituents, were found to have significant effects only in a few cases. For filler *e*, a significant interaction between the boundary depth and the complexity were found in both APS and SPS, although the effect was less reliable in SPS (APS: $p < .001$; SPS: $p < .05$). Each graph in Figure 1 shows the scatter-plot between the complexity and the (log) duration, after removing the fixed effects of covariates and the random effect, for each clause boundary type. In APS, the regression line had a slightly negative slope for the clause-initial *e*s after SB type boundaries ($p < .05$), whereas the slope of the regression line for those after WB type boundaries was positive ($p < .001$); the durations for AB type boundaries showed no significant correlation with the complexity. In SPS, on the other hand, the data for both SB and WB type boundaries indicated positive correlations (SB: $p < .05$; WB: $p < .1$), but the data for AB type boundaries showed no correlation.

For filler *eeto*, a main effect of the boundary strengths was significant only in APS. Figure 2 shows the mean (log) duration, after adjusted for the fixed effects of covariates and the random effect, as function of the boundary depth. MCMC p-values revealed that the duration of *eeto* was longer after an SB type boundary than after an AB type boundary ($p < .05$) and that the duration was also longer after a WB type boundary than after an AB type boundary ($p < .05$); there was no significant difference between the SB and WB type boundaries.

For noun phrases "Noun + *wa*", a main effect of the complexity was significant only in APS. Figure 3 shows the scatter-plot between the complexity and the (log) duration, after adjusted for the fixed effects of covariates and the random effect, for the data pooled over three clause boundary types. The duration became longer as the complexity increased.

4. Discussion

The results show some similarities with the results of our previous study on the filler rates at clause boundaries [5]. The filler rates are positively correlated with the complexity of the following clause only at boundaries other than sentence and strong clause boundaries. A similar tendency was found in the duration of *e* at the beginning of clauses; positive correlation with the complexity of the following constituents was consistently observed, across speech types, at weak boundaries only.

It is unclear, at this stage, why such correspondence between the duration of clause-initial *e* and the complexity of the following constituents was observed only at weak boundaries. As noted in section 2.2, weak boundaries are syntactically weaker than absolute and strong boundaries. However, the strength of weak boundaries may be intensified by factors related to discourse structures involved in this boundary type, such as initiation, ending, and shift of topics. If syntactically-determined boundary strength is crucial, we would conjecture that prolongation of filler *e* is more relevant to the time required for linguistic encoding than for the planning of the content, the former being more likely to come to the fore at shallow bound-

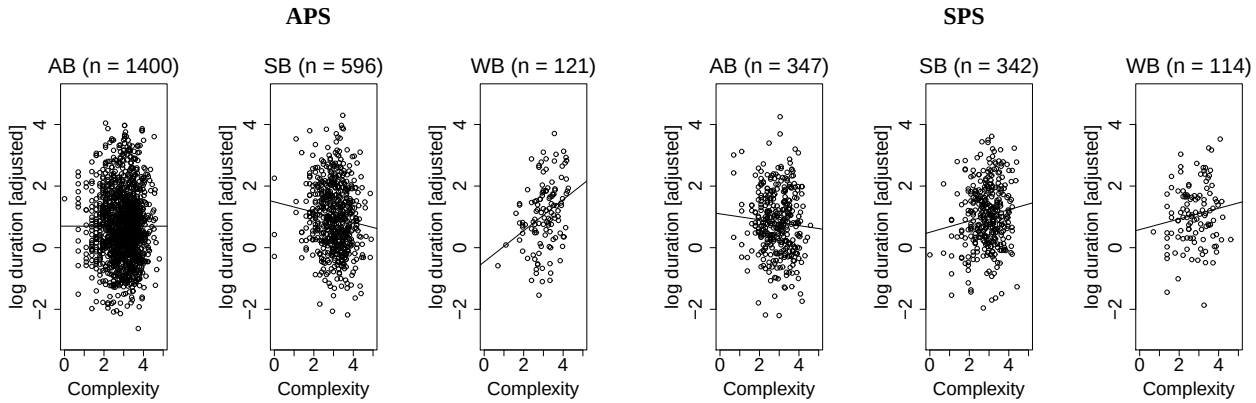

Figure 1: Scatter-plot between the complexity and the (log) duration, after adjusted for the fixed effects of covariates and the random effect, grouped by boundary strengths for filler *e* in APS and SPS

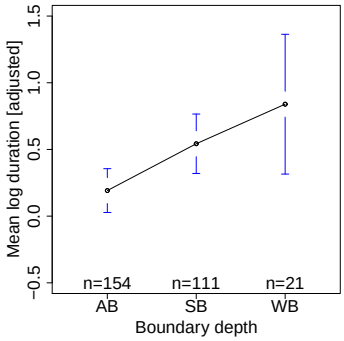

Figure 2: Mean (log) duration, after adjusted for the fixed effects of covariates and the random effect, as function of boundary strengths for filler *eeto* in APS. Error bars represent (anti-conservative) 95% confidence intervals obtained from the data pooled over speakers.

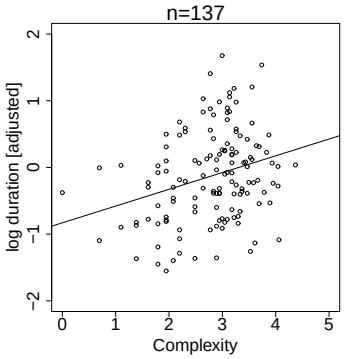

Figure 3: Scatter-plot between the complexity and the (log) duration, after adjusted for the fixed effects of covariates and the random effect, for topic phrase "N + *wa*" in APS

aries. If discourse structure, by contrast, is a decisive factor, the conclusion could go in the opposite direction; filler *e* would be concerned with deeper process, which is more likely to take place at or around discourse boundaries.

Quite interestingly, *wa*-marked topic phrases also showed correlation between the duration and the complexity, although the effect was constantly observed across clause boundary types. This suggests that topic phrases may sometimes be available for the speaker to gain time for deep cognitive processing; *wa* can be among the class of utterance-initial elements.

There are some discrepancies in the results from APS and SPS. The effect of boundary strengths on the prolongation of *eeto* was observed only in APS. Also, the effect of the complexity on "Noun + *wa*" was found only in APS. This may result from difference in difficulty producing APS and SPS. Speakers need to elaborate more in preparing the following utterance in APS than in SPS because they are required to be more logical and precise in APS than in SPS. Heavier cognitive load in APS may have affected the results for *eeto* and *wa*.

5. Acknowledgments

This work was partly supported by Grant-in-Aid for Scientific Research by JSPS (2009–2012, Grant No. 21520467).

6. References

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

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

[3] Y. Den, "Are word repetitions really intended by the speaker?," in *Proc. ISCA tutorial and research workshop on Disfluency in spontaneous speech*, pp. 25–28, 2001.

[4] Y. Den, "Prolongation of clause-initial mono-word phrases in Japanese," in *Linguistic patterns in spontaneous speech* (S.-C. Tseng, ed.), pp. 167–192, Taipei: Institute of Linguistics, Academia Sinica, 2009.

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

[6] M. Swerts, "Filled pauses as markers of discourse structure," *Journal of Pragmatics*, vol. 30, pp. 485–496, 1998.

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

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

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