

THE RELATIONSHIP BETWEEN PRECEDING CLAUSE TYPE, SUBSEQUENT CLAUSE LENGTH AND THE DURATION OF SILENT AND FILLED PAUSES AT CLAUSE BOUNDARIES IN JAPANESE MONOLOGUES

Michiko Watanabe^a, Yosuke Kashiwagi^b, Kikuo Maekawa^a

^aNational Institute for Japanese Language and Linguistics, ^bThe University of Tokyo
^awatanabem@ninjal.ac.jp, ^bkashiwagi@gavo.t.u-tokyo.ac.jp, ^akikuo@ninjal.ac.jp

ABSTRACT

Filled pauses (FPs) are claimed to occur when speakers have some difficulties and need extra time in speech production. This study investigated whether the following two factors affect silent pause (SP) and FP durations at clause boundaries, using a spontaneous speech corpus: 1) boundary strength and 2) subsequent clause length. First, whether SP and FP durations increase with syntactic boundary strength was examined. Second, whether subsequent clause length affects SP and FP durations at the boundaries was investigated. Results show SP duration increased with boundary strength and subsequent clause length, but FP duration did not, suggesting only SP duration is affected by the two factors.

Keywords: silent pause, filled pause, clause boundary, speech planning, disfluency.

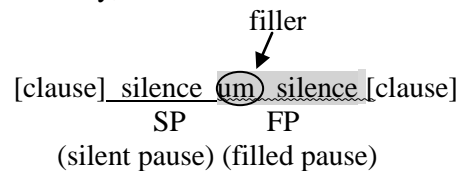
1. INTRODUCTION

Speech sounds such as “um” and “uh” in English and “eto” and “ano” in Japanese are rare in read speech, but common in spontaneous speech. These sounds are believed to be relevant to speech planning difficulties [1, 4]. We call such sounds “fillers”. The list of fillers in Japanese differs depending on researchers. In this study we employ the inventory of fillers in “The Corpus of Spontaneous Japanese (CSJ)” [6]. When the following sounds are used to fill in a gap, they are regarded as fillers:

- ano, sono (originating from demonstratives, meaning “that” and “the”, respectively)
- ato, eto, nto, to (some vowels or a nasal flowed by “to,” which is probably a quotation particle)
- ma (originating from an adverb)
- a, i, u, e, o, n (Japanese vowels and a nasal)

The items in the first three groups followed by “ne” (an interjectory particle) or “desune” (a copula + ne) are also included in the list.

Fillers are commonly observed at sentence and clause boundaries in spontaneous monologues in CSJ. They are often preceded and occasionally followed by silence at such locations. We call silences immediately after the end of clauses “silent pauses (SPs)”. If there is a filler between two consecutive clauses, we call the sound including the subsequent silence, if any, “filled pauses (FPs)” in this study, as shown below.



Speakers are likely to be engaged in conceptualizing the message to be conveyed and encoding some part of it into linguistic forms at deep syntactic boundaries. It is conceivable that speakers use FPs when they need some more time to resume speech after a certain period of silence and want to inform the listener of their current situation. In addition to FPs, there are several means to do so. Repeating words, prolonging a part of words, and making false starts are among others. FPs seem one of the most common devices in English and Japanese.

Previous research indicates that the probability of fillers appearing at clause boundaries is affected by the boundary strength and the subsequent clause length [7]. It tested two hypotheses, “the boundary hypothesis” and “the complexity hypothesis”. The boundary hypothesis is that speakers are more likely to use fillers at deeper syntactic boundaries because they need to plan larger units of information than at shallower boundaries. The complexity hypothesis is that speakers are more likely to use fillers when the subsequent clause is more complex. The number of words in a clause was used as its index of complexity, because the number of words in a linguistic unit is reported to be highly correlated with its complexity [2]. Fillers appeared more frequently at stronger boundaries than at weaker boundaries, supporting the boundary hypothesis. Longer clauses were more often preceded by fillers

than shorter clauses, also supporting the complexity hypothesis. As previous research investigated only the rate of fillers appearing at clause boundaries, in this study we considered the durations of SPs and FPs at clause boundaries. We adjusted the two hypotheses to SP durations and FP durations as follows:

- 1) The stronger the boundary, the longer SPs and FPs at the boundary (the boundary hypothesis).
- 2) SP and FP durations increase with the subsequent clause length (the complexity hypothesis).

We regarded clause length as the amount of information expressed by the clause. We tested the two hypotheses by measuring SP and FP durations as a function of clause boundary strength and subsequent clause length.

Adverbial clauses in Japanese

We focus on adverbial clause boundaries in this study. Adverbial clauses are grouped into three types according to connective particles or certain forms of verbs, adjectives or auxiliary verbs at the end of the clause [3, 5]. Japanese is a head-final language and adverbial clauses always precede the main clause as schematically shown below.

(adverbial clause <connective particle>), (main clause).

Table 1 shows the classification of adverbial clauses. It is known that type C clauses are syntactically more independent from the main clauses than the other types because type C clauses can have their own topics as well as their own subjects. Type A clauses are the most dependent on the main clauses because the topics and the subjects of the main clauses are automatically those of type A clauses. That is, they cannot exist without the subsequent main clause. The degree of dependency of type B clauses is in between those of type A and type C clauses. Type B clauses can have their own subjects, but not their own topics. It is also known that the three types have a hierarchical structure. Type C clauses can contain any of the three types. However, type B clauses can contain type A and type B, but not type C. Type A clauses can contain only type A clauses. Based on the degree of dependency and the structure, we regarded boundaries after type C clauses as the strongest, boundaries after type A clauses the weakest, and those after type B clauses to be in between the two types. We also included sentence boundaries in the analysis, but regarded them as stronger than any clause boundary type. The ends of sentences are marked with sentence-final particles or sentence-final forms of verbs, adjectives or auxiliary verbs. The boundary hypothesis predicts

the order of SP and FP durations at the boundaries as follows:

sentence boundaries > type C > type B > type A

Table 1: Classification of adverbial clauses

Boundary type	Connective	Meaning, usage
A	~tari, ~tari ~nagara, tsutsu	~ing and ~ing (listing actions) ~ing (expressing accompanying actions)
B	~to,ba,tara,nara	when ~, if ~, in case
	~te, de	~ and
	~te kara	since ~, after~
	~te mo	even if ~
	~yoo ni	so that ~
C	adverbial forms of verbs and adjectives	~ and
	~kara, node	as ~ (reason)
	~noni, ke(re)do	though ~
	~ga	although ~, ~but
	~si	~ and (listing actions or features)
sentence	~desite, masite	~ auxiliary verb for politeness + and
	~yo, ne, to	sentence-final particles
	~desu, masu, ta, n	sentence-final forms of auxiliary verbs

2. METHOD

2.1. Data

106 presentation speeches from the “Core” part of *The Corpus of Spontaneous Japanese* (CSJ) were used for analysis [6]. The presentations were given by paid volunteers to a small audience in an informal setting. In the corpus they are referred to as simulated public speaking (SPS). 53 of them were given by female speakers and 53 by male speakers. All the speakers spoke Tokyo dialect. They talked about general topics such as “the happiest memory in my life” or “my town” for about 10 minutes. The topics were given to the speakers beforehand. They were instructed to prepare notes for their speeches, but to not read from manuscripts. All the speeches were transcribed, and detailed linguistic information was given to the transcription.

2.2. Procedures

First, we identified clause boundaries and grouped them into three types according to the connectives. Sentence boundaries were also identified depending on sentence final forms or particles as shown in Table 1. We referred to a set of phrases between two consecutive boundaries a clause. Second, we counted the number of *Bunsetu*-phrases (*Bunsetu*, hereafter) in each clause. *Bunsetu* is composed of one content word with or without function words. We regarded the number of *Bunsetu* as an index of the amount of information in the clause. Third, we grouped boundaries according to the boundary type and the number of *Bunsetu* in the subsequent clause in each presentation. We measured SP and FP durations at each boundary as a function of the boundary type and the number of *Bunsetu* in the subsequent clause. An example of clause boundary grouping is shown below. The number of *Bunsetu* is given at the beginning of each clause. A single letter paired with a number at the end of each clause indicates a category of boundaries. For example, B2 means a type B boundary followed by a two-*Bunsetu* clause.

- 6: (F ano) ato kekkoo uchi-no ryooshin-wa
 um also quite our parents-TOPIC
 petto-ga shinu-to <conditional> B2
 pet-SUB die-CONDITIONAL
- 2: hekomu hito-na-node /reason/ C3
 depressed people-are-REASON
- 3: (F ano) zettai dame-to iwarete-i-te <te> B2
 um absolutely no-QUATATION had-been-
 told-AND
- 2: medaka-sura kae-masen-deshita [sentence
 final] D1
 killifish-even were-not-able-to-keep
- 1: sore-na-node /reason/ C2
 that-is-REASON
- (CSJ: S01F0006)

We took the mean value of SP durations, durations of SPs with subsequent FPs, where SPs were followed by FPs (SP + FP), and FP durations for each boundary group in each presentation, and averaged these durations over presentations.

3. RESULTS

The mean number of each boundary type per presentation was 5 for type A, 73 for type B, 39 for type C, and 42 for sentence boundaries. Type A boundaries were excluded from further analysis because of insufficient number of samples. Type A clauses were treated as parts of the larger clauses which contained them. We discuss SPs and FPs at type B, type C and sentence boundaries, hereafter.

Fig. 1 shows SP durations at the three boundary types as a function of subsequent clause length in the number of *Bunsetu*. First, the figure shows that SP durations are closely related to boundary strengths. SPs at sentence boundaries are longer than those at clause boundaries. Likewise SPs at type C boundaries are longer than those at type B boundaries. Fig. 1 also illustrates SP durations tend to increase with the subsequent clause length until the length reaches 8 *Bunsetu* at all boundary types.

Fig. 2 illustrates durations of SPs including following FPs (SP + FP) at the three boundary types as a function of subsequent clause length. Although SP + FP durations are naturally longer than SP only durations, the pattern of the duration change in Fig. 2 is very similar to that of Fig. 1. SP + FP durations are the longest at sentence boundaries, the shortest at type B clause boundaries, and in between at type C boundaries. SP + FP durations increase with the subsequent clause length until the length reaches 6 *Bunsetu* at all boundary types.

Fig. 3 shows FP durations at the three boundary types as a function of subsequent clause length. FPs at type C boundaries tend to be longer than those at type B and sentence boundaries. FP durations do not

Figure 1: SP durations at three boundary types as a function of subsequent clause length. Error bars indicate standard errors in all the figures.

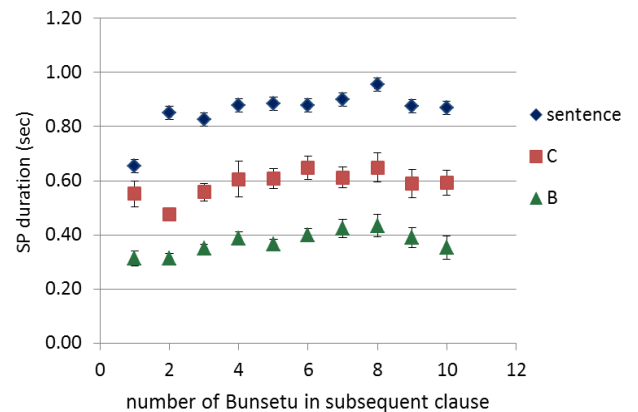
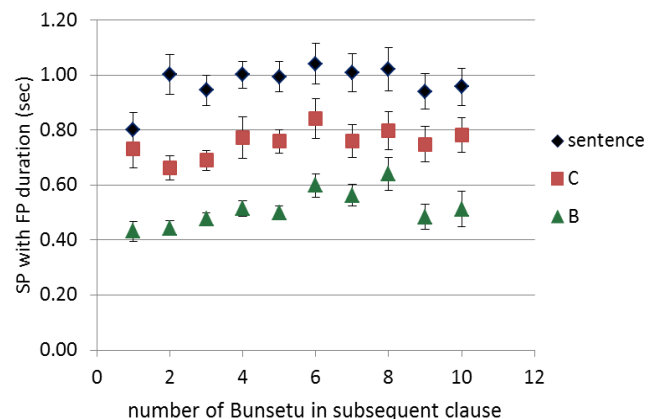
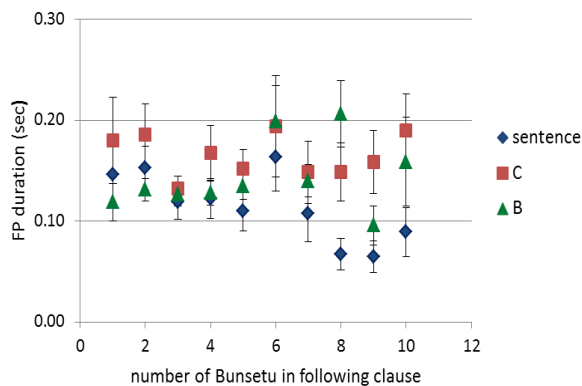


Figure 2: Durations SPs with following FPs at three boundary types as a function of subsequent clause length.



increase with the following clause length at any boundary type.

Figure 3: FP durations at three boundary types as a function of subsequent clause length.



4. DISCUSSION

Both SP durations and SP + FP durations were the longest at sentence boundaries, the second longest at type C boundaries, and the shortest at type B boundaries, supporting the boundary hypothesis. FPs alone did not support the boundary hypothesis, because FPs were the longest at type C boundaries, not at sentence boundaries. When FPs at type B and type C clause boundaries are compared, the hypothesis is supported because FPs at type C boundaries are longer than those at type B. SP durations are closely related to the boundary strength, whereas FP durations are related to the boundary strength only at clause boundaries.

As a general trend, both SP durations and SP + FP durations increased with the number of *Bunsetu* in the subsequent clause, supporting the complexity hypothesis. However, there were upper limits for both SP durations and SP + FP durations at all boundary types. SP durations were longest when the following clauses contain 8 *Bunsetu*, whereas SP + FP durations were longest when the subsequent clauses contained 6 *Bunsetu*. It is possible that speakers have the threshold value of appropriate pause duration for each boundary type in mind and try to avoid pausing longer than the value. Regarding SPs, the threshold value is likely to be about 900ms for sentence boundaries, about 600ms for type C, and about 400ms for type B clause boundaries. As for SP + FP durations, the value can reach up to about 1000ms for sentence boundaries, about 800ms for type C, and about 600ms for type B clause boundaries.

Another interpretation of the results is that speakers' maximum encoding span is about 8 *Bunsetu* at clause boundaries in casual presentations

like these. When the message to be conveyed is too rich to be expressed within 8 *Bunsetu*, or within the threshold pausing time, speakers are likely to encode a part of the message at the beginning of the clause and encode further later. It is interesting to note that SP + FP durations at sentence boundaries are almost constant, around one second, regardless of the following clause length, except when the subsequent clause is composed of one *Bunsetu* (Fig. 2). The complexity hypothesis is hardly supported at sentence boundaries. It is possible that pause durations at sentence boundaries reflect cognitive load of macro level message planning, which cannot be measured by the number of *Bunsetu* in the immediately following clause.

A conjunction “de” (an abbreviation for “sorede”, meaning “and”) is frequent at the beginning of sentences, often followed by SPs and/or FPs. It is possible that some conjunctions are used to gain time for planning at sentence and clause boundaries. As a next step, we plan to investigate the distribution of conjunctions and their relationship with SPs, FPs and the subsequent clause length at sentence and clause boundaries.

5. ACKNOWLEDGEMENTS

This research was conducted as a part of the project, “Creation of Corpus Japanese Linguistics” at National Institute for Japanese Language and Linguistics, and supported by JSPS KAKENHI Grant Numbers 15K02553 and 26284062.

6. REFERENCES

- [1] Clark, H. H. 2002. Speaking in time. *Speech Communication* 36, 5-13.
- [2] Clark, H. H., Wasow, T. 1998. Repeating words in spontaneous speech. *Cognitive Psychology* 37, 201-242.
- [3] Minami, F. 1974. *Gendai nihongo no kouzou* (The structure of modern Japanese) (in Japanese). Tokyo: Taisyukan syoten.
- [4] Sadanobu, T., Takubo, Y. 1995. The monitoring devices of mental operations in discourse – a case of ‘eeto’ and ‘ano (o)’ – (in Japanese), *Gengo Kenkyu* 108, 74-93.
- [5] Takubo, Y. 1987. Syntactic structure and contextual information (in Japanese), *Nihongogaku*, 6 (5), 37-48.
- [6] The National Institute for Japanese Language. http://pj.ninjal.ac.jp/corpus_center/csj/misc/preliminary/index_e.html
- [7] Watanabe, M. 2009. *Features and Roles of Filled Pauses in Speech Communication -A corpus-based study of spontaneous speech* (Hituzi Linguistics in English No.14). Tokyo: Hituzi Syobo Publishing.