COMPARISON OF FACTORS RELATED TO CLAUSE-INITIAL FILLER PROBABILITIES IN ENGLISH AND JAPANESE

Michiko Watanabe* and Yusaku Korematsu**

*National Institute for Japanese Language and Linguistics
**The University of Tokyo
watanabem@ninjal.ac.jp, korematsu@gavo.t.u-tokyo.ac.jp

ABSTRACT

Factors related to probabilities of clause-initial fillers such as *um*s and *uh*s were analysed using parallel monologue corpora of English and Japanese. The factors considered in the analysis were boundary depth between the current and immediately preceding clauses, number of words in the clause, existence of clause-initial coordinate conjunctions and speakers' gender. Number of quotation clauses in a given clause was included as a predicting factor with English data. Results showed that boundary depth had significant positively-correlated effects on filler probabilities in both languages. However, the effect was stronger with English data. Number of words in the clause also had significant positively-correlated effects in both languages. Gender effects were found, but only in Japanese. Existence of conjunctions had no significant effect in either language. The results indicate that factors causing clause-initial fillers partly depend on the language.

Keywords: spontaneous speech, monologue, disfluency, filled pauses, contrastive study

1. INTRODUCTION

Fillers such as *um*s and *uh*s are prevalent in spontaneous speech. Fillers not only reveal speakers' troubles in timely speech delivery but help smooth communication by signalling the delay and eliciting interlocutors' cooperation ([1], [10], [12]).

In what kind of contexts are fillers frequent? Fillers have been found to be frequent at deep boundaries, such as discourse and sentence boundaries ([11], [14]). Speakers are likely to make decisions about what to talk and how to unfold the message at deep boundaries. These processes tend to take time and may result in silent and filled pauses. It was found that the deeper the boundary, the higher the filler frequency at discourse boundaries in Dutch monologues ([14]). We investigate the relationship between clause boundary depth and following clause initial filler probability in English and Japanese.

Some researchers observed that filler rate shortly before onset of finite clauses is not higher than that of basic clauses ([2], [3], [4]). Every finite clause contains one and only one tensed or finite verb, whereas every basic clause contains one and only one main verb, whether tensed or not [5]. From the findings, the authors argue that a basic clause is a major speech planning unit, not a finite clause.

Other researchers argue that clauses are major conceptualizing units, whereas phrases are major linguistic encoding units ([13]). Through their experiments, Smith and Wheeldon [13] found that speech onset latency is affected not only by complexity of the first noun phrase, but also by complexity of the remaining part of the sentence. Here, a sentence can be composed of one or more than one clauses. The authors argue that speakers spend time not only for lexical access of the first phrase but also for planning the following part, though not thoroughly, before starting their speech. The results can be interpreted that speakers are engaged in conceptual planning of the message to be conveyed in subsequent linguistic units. We surmise that speech onset latency is relevant to the occurrence of fillers because speakers are likely to use fillers to avoid unnaturally long silence. Thus, the question arises whether occurrence of fillers is also related to sentence complexity. The current study addresses this question. We regard sentence complexity as an index of message complexity, because sentence is a unit of a linguistically encoded message. We employed number of words in the sentence as an index of sentence or message complexity and examined correlations between sentence size in number of words and filler probability at sentence and clause initial positions.

We used parallel English and Japanese corpora as material to compare the results. The details of the corpora are given in the next section.

2. METHOD

2.1. Material

2.1.1. Japanese corpus

20 informal presentations excerpted from Simulated Public Speaking (SPS) in "The Corpus of Spontaneous Japanese (CSJ)" [8] were used for Japanese analysis. An excerpt is given in Example (1) below. The original transcriptions are in Japanese. Here, they are transcribed in Roman characters with

translation. The speakers are those of Tokyo Japanese. They are university graduates in their 20s or early 30s. Each speech lasts about 10 minutes. The topics were given to speakers beforehand. The corpus contains speech recordings and the transcriptions with rich linguistic annotations. Common gap-filling expressions such as *anoo* and *eeto* are identified as fillers. Fillers include 15 sound forms and their variants ([9] p.54). Round brackets with "F" inside are filler labels as can be found in line 1 and 3 in Example (1). We used these filler labels for analysis.

Sentence and clause boundary labels are contained in CSJ, as shown in Example (1) below. Subordinate clauses always precede main clauses, and connective particles are located subordinate clause-finally as "kedo (though)" and "node (because)" as in the example below.

Example (1)

1: (F eeto) (F n) (F maa) mazu ichiban
first the most
inshoobukakatta koto tte yuu n desu kedo
/concessive KEDO/
impressive event called is though
First, though (the given title is) "the most impressive event"
2: watashi wa doomo wasureppoi tachi mitaina nde
<reason NODE>
I-TOPIC probably forgetful character seem because
because I am probably a forgetful type
3: (F ano) saikin no koto shika omoiukabanakatta n desu ne [sentence-final]
recent event only came to mind the thing is
the thing is only a recent event came to mind, you know.

(From S05F1600, in CSJ)

Degree of syntactic dependency of subordinate clauses on the main clause is ranked based on connective type in Japanese syntax ([7], [15]). We used the connective type to estimate boundary depth between subordinate and main clauses. Boundaries marked with slashes (as in the first clause above) are strong clause boundaries, and those marked with ankle brackets (as in the second clause) are weak clause boundaries. We assumed more dependent subordinate clauses have a stronger connection to the main clause in the content as well as the form, and vice versa. Sentence boundaries, indicated by sentence-final morphemes, are marked with square brackets (as in the third clause). We assumed sentence boundaries to be the deepest, strong clause boundaries the second deepest and weak clause boundaries to be the shallowest among the three boundary types.

2.1.2. English corpus

We analysed speeches in a corpus called "The Corpus of Oral Presentations in English (COPE)" as a counterpart. The speeches were collected in a setting similar to that of the Japanese speeches to enable comparison between the two languages. 20 informal English presentations were recorded in Portland, Oregon and Los Angeles, California. The speakers were university students or university graduates in their 20s and early 30s. They were given a topic, "the most memorable event in my life", and instructed to give a talk for at least 10 minutes. Disfluencies are labeled, and clause boundaries are marked on the transcription. Only *um* and *uh* are identified as fillers in COPE. An excerpt from the transcription is given in Example (2) below.

Example (2)

1: <cb and when> and when we got to (l like) (r2 this little) this little food court area <ce and when>
2: /cb/ he stopped me /ce/
3: /cb and/ and he was like (qcb) so would you want (ncb to) to be my girlfriend (nce to) (qce) /ce and/
4: /cb and/ and I literally jumped on him /ce and/
5: /cb and/ and I have no idea (ncb why) why I did that (nce why) /ce1 and/

(From losF09 in COPE)

Transcriptions of COPE are partitioned at coordinate or adverbial finite clause boundaries as in Example (2). Ankle brackets, <cb> and <ce> indicate finite adverbial clause beginning and end, respectively. Connectives are included in the brackets. "/cb/" and "/ce/" in line 2 mean coordinate clause beginning and end, respectively. "(qcb)" and "(qce)" in the third line indicate quotation clause beginning and end, respectively. "(ncb to)" and "(nce to)" in the third line indicate nominal infinitive clause beginning and end, respectively. Similarly, "(ncb why)" and "(nce why)" in the fifth line indicate nominal clause beginning with "why" and its end, respectively.

In English, it is often difficult to tell whether a given clause boundary is also a sentence boundary or a boundary within a sentence, because no period or comma is used in speech. Unlike Japanese, there is no syntactic or morphological cues to tell boundary depth in English. In order to evaluate clause boundary depth, we had three labellers add sentence boundary labels to clause boundaries which they judged to be deep boundaries. The labellers were instructed to mark sentence boundaries based on the content and the prosody of speech. We estimated boundary depth based on the number of labellers who marked the boundary as a sentence boundary. We regarded boundaries marked by three labellers as sentence

boundaries to be the deepest, those marked by two labellers to be the second deepest, and so forth. Thus, clause boundaries were grouped into four types depending on perceived boundary depth.

An overview of the two corpora is shown in Table 1. Silent periods longer than 50ms were counted as silent pauses.

Table 1: Overview of two corpora.

Corpus	Number of presen -tations	Mean duration (sec) (SD)	Mean number of words (SD)	Mean number of fillers (SD)	Mean number of silent pauses (SD)
CSJ	20	701 (117)	2044 (501)	127 (73)	379 (74)
COPE	20	681 (66)	1991 (351)	79 (40)	368 (76)

2.2. Procedure

Besides the (1) boundary depth factor and (2) sentence complexity factor mentioned in the introduction section, the following factors were examined as those which are likely to have relevance to sentence or clause initial filler probability.

(3) Speaker's gender
Speaker's gender was included as a predictor, because previous studies have shown that male speakers use significantly more fillers than female speakers both in English and Japanese ([9], [11], [16]).

(4) Clause-initial coordinate conjunctions
Fillers sometimes appear adjacent to conjunctions as in "and um" in English and "de ee" in Japanese. We conjectured that use of conjunctions and fillers might have some relevance.

(5) Number of quotation clauses
Speech sometimes contains quoted remarks. The number of quotation clauses was found to have a negative effect on the containing clause initial filler probability in English ([16]). Because quotation clause labels are available only in English data, we included this variable only in English analysis.

We excluded clauses containing more than 30 words from analysis because samples were sparse for larger clauses. Table 2 and Table 3 show the number of boundary types in Japanese and in English, respectively.

We estimated clause-initial filler probability using a generalized linear mixed model, with maximum likelihood estimation of variance components. Because the response variable was binary, we conducted mixed-effects logistic regression. Fixed effects predictor variables were the factors (1) through (4) for Japanese and (1) through (5) for English. Speakers were treated as a random effects factor. Fillers and other disfluencies were not counted as words. We used 'glmer' function in 'lme4' package and 'MuMIn' and 'lmerTest' packages running under R version 3.5.1.

3. RESULTS AND DISCUSSION

Table 2 shows the number of boundary types in Japanese, and Table 3 in English.

Table 2: Number of boundary types in Japanese.

Weak clause boundaries	1663
Strong clause boundaries	932
Sentence boundaries	897
Total	3492

Table 3: Number of boundary types in English.

Type0	Boundaries marked by no labeler	719
Type1	Boundaries marked by one labeler	1521
Type2	Boundaries marked by two labelers	979
Type3	Boundaries marked by three labelers	806
	Total	4025

Table 4 shows the results of Japanese analysis, and Table 5 those of English. Estimates of categorical variables are given as the relative values to one of its levels whose estimate is zero in R. The reference variables are included in the table. Odds ratio indicates the degree of effect size of each factor. When the ratio is close to 1.0, the effect size is small. The more distant the ratio from 1.0, the larger the effect size.

First, clause boundary depth had significant effects in both languages. The deeper the boundary, the higher the clause-initial filler probability. However, boundary effect is larger in English than in Japanese. Particularly, the difference between strong clause boundaries and sentence boundaries is small in Japanese.

Second, the effect of number of words in clauses is significant in both languages. The more words clauses contain, the higher the clause-initial filler probability. The effect is larger in English than in Japanese.

Third, gender had a significant effect, but only in Japanese. Japanese male speakers use significantly more fillers than female counterparts. The results indicate the existence of a gender difference in language use at an unconscious level in Japanese. The results support previous studies on Japanese fillers, but not those on English ([6], [11]).

Existence of clause-initial coordinate conjunctions had no significant effect in either language. Fillers and coordinate conjunctions are likely to serve different purposes.

The number of quotation clauses had a significant negative effect on clause-initial filler probability in English. This result suggests that quoting others' speech hardly increases speakers' cognitive load even if the containing clause becomes longer because of quotation. Quotation is likely to be a cognitively easy task compared to message planning and encoding.

Table 4: Results of mixed-effects logistic regression for Japanese.

Variable	Estimate	Std. Error	z value	Pr(>|z|)		Odds ratio
(Intercept)	-1.844	0.135	-13.713	< 2e-16	***	0.158
Boundary type						
Weak clause boundary	0					
Strong clause boundary	0.573	0.098	5.872	4.30e-09	***	1.774
Sentence boundary	0.632	0.103	6.14	8.26e-10	***	1.882
Number of words	0.024	0.007	3.676	0.0002	***	1.024
Gender						
Female	0					
Male	0.767	0.149	5.155	2.54e-07	***	2.153
Initial coordinate conjunction						
Without	0					
With	-0.117	0.108	-1.079	0.281		0.890

Table 5: Results of mixed-effects logistic regression for English.

Variable	Estimate	Std. Error	z value	Pr(>|z|)		Odds ratio
(Intercept)	-3.676	0.324	-11.362	< 2e-16	***	0.025
Boundary depth	0.957	0.048	20.069	< 2e-16	***	2.605
Number of words	0.049	0.009	5.647	1.63e-08	***	1.050
Gender						
Female	0					
Male	0.380	0.416	0.913	0.361		1.463
Initial coordinate conjunction						
Without	0					
With	0.050	0.091	0.55	0.582		1.051
Number of quotation clauses	-0.581	0.141	-4.129	3.65e-05	***	0.559

We reanalysed 90% of the data only with significant variables, and estimated filler probability for the remaining data. Figure 1 (Japanese) and Figure 2 (English) illustrate estimated clause-initial filler probability for each boundary type as a function of the number of words in the clause. In both figures, filled circles indicate observed boundaries with fillers and pluses indicate observed boundaries without fillers. In Figure 2, Type0 and Type1 boundaries (see Table 3) are shown combined as shallow boundaries.

A comparison of the two figures demonstrates difference in effect size of boundary depth and clause size between the two languages. Both factors have larger effects in English.

As is seen from these figures, fillers appear quite frequently when the estimated probability is lower than 50%. Marginal R-squared values of these models are 0.072 for Japanese and 0.208 for English. The Japanese value indicates very low relevance of boundary depth and clause size to occurrence of clause-initial fillers. When we assume that the two factors are indices of cognitive load of message planning, clause-initial Japanese fillers have very little to do with clause level message planning compared with English. Difficulties that clause-initial fillers reflect are likely to depend on the language.

Our models do not explain a large amount of variance, but that is to be expected because there certainly are many things left out of our models that may also influence filler use. What is important to us is the relative differences in factor significance. This informs us of where the best bet is for future research. Then we can move on to more strictly controlled experimentation or more sophisticated models.

Figure 1: Estimated clause-initial filler probability for each boundary type as a function of number of words in the clause in Japanese.

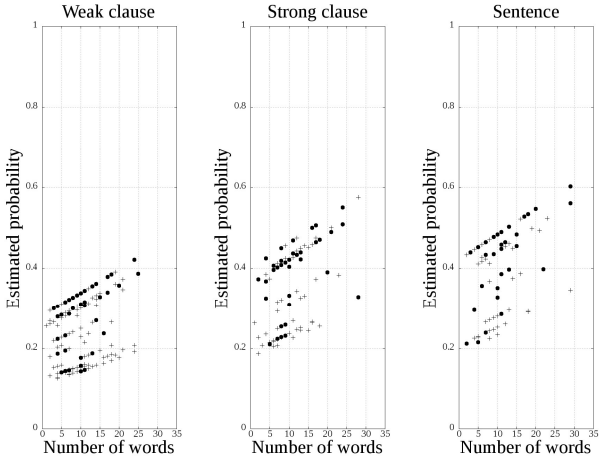
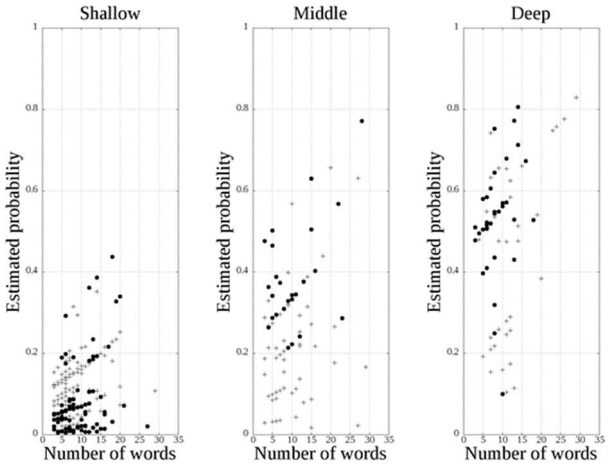


Figure 2: Estimated clause-initial filler probability for each boundary depth as a function of number of words in the clause in English.



4. ACKNOWLEDGMENTS

This research is supported by JSPS KAKENHI, Grant Numbers 15K02553 and 18K00559.

5. REFERENCES

[1] Clark, H. H. 2002. Speaking in time. *Speech Communication* 36, 5-13.

[2] Ford, M. 1982. Sentence planning units: Implications for the speaker's representation of meaningful relations underlying sentences. In J. Bresnan, ed. *The mental representation of grammatical relations*, 797-827. Cambridge, MA: MIT Press.

[3] Ford, M., Holmes, V. 1978. Planning units and syntax in sentence production. *Cognition* 6: 35-53.

[4] Holmes, V. 1988. Hesitations and sentence planning. *Language and Cognitive Processes* 3, 323-361.

[5] Levelt, W., J., M. 1989. *Speaking*. The MIT Press: Cambridge, Massachusetts.

[6] Maekawa, K. 2004. Design, compilation, and some preliminary analyses of the Corpus of Spontaneous Japanese. In K. Yoneyama and K. Maekawa eds.. *Spontaneous Speech: Data and Analysis* 87‐108. Tokyo: The National Institute for Japanese Language.

[7] Minami, F. 1974. *Gendai Nihongo no Kouzou* (Structure of Modern Japanese), in Japanese. Taishukan: Tokyo.

[8] National Institute for Japanese Language and Linguistics (NINJAL). Corpus of Spontaneous Japanese CSJ. https://pj.ninjal.ac.jp/corpus_center/csj/en/

[9] National Institute for Japanese Language and Linguistics (NINJAL). 2006. https://pj.ninjal.ac.jp/corpus_center/csj/k-report-f/CSJ_rep.pdf

[10] Sadanobu, T., Takubo, Y. 1995. The monitoring devices of mental operations in discourse – a case of eeto and anoo –. *Gengo Kenkyu* 108, 74-93, in Japanese.

[11] Shriberg, E. 1994. Preliminaries to a theory of speech disfluencie. Ph.D. thesis, University of California at Berkeley. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.443.7755&rep=rep1&type=pdf.

[12] Shriberg, E. 2005. Spontaneous speech: How people really talk and why engineers should care. *Proc. Interspeech 2005*, 1781-1784.

[13] Smith, M. C., Wheeldon L. 1999. High level processing scope in spoken sentence production. *Cognition* 73, 205‐246.

[14] Swerts, M. 1998. Filled pauses as markers of discourse structure. *Journal of Pragmatics* 30, 485-496.

[15] Takubo, Y. 1987. Tougo kouzou to bunmyaku jouhou (Syntactic structure and contextual information), *Nihongogaku* 65, 37-48, in Japanese.

[16] Watanabe, M., Korematsu, Y. 2017. Factors affecting clause-initial filler probability in an English monologue corpus. *Journal of the Phonetic Society of Japan*, 21-3, 24–32.