Fillers as Indicators of Discourse Segment Boundaries in Japanese Monologues

Michiko Watanabe

College of Arts and Sciences, The University of Tokyo, JST/CREST
watanabe@gavo.t.u-tokyo.ac.jp

Abstract

We investigated distribution of fillers (filled pauses) in the vicinity of boundaries of different strengths in Japanese monologues, to understand whether fillers may convey information about the location and the strength of boundaries. Consistent with the results of studies on Dutch monologues, fillers tend to increase as the boundary strength grows. It has also been revealed that fillers tend to occur phrase-initially, more strongly at deeper boundaries than at shallower ones. Regarding filler types, the frequency of *eto* grows most sharply as boundary strength increases, as does *e* to a lesser degree. These findings indicate that occurrence of fillers, particularly phrase-initial *eto* and *e*, provide contributory evidence to discourse boundaries.

1. Introduction

1.1. Research on fillers

Fillers are such utterances as *uh* and *um* in English and *ano* and *eto* in Japanese. They are often categorised as pauses. However, one feature distinguishing them from silent pauses is that they are observed only in spontaneous speech, not in previously prepared, read speeches. Therefore, it is assumed that fillers have some relevance to speech generation processes, and that they are uttered when speaker has some trouble in on-line speech production processing.

Research on fillers may be grouped into four types of approach; the first group regards fillers as defects in communication and tries to find out ways to decrease them (e.g. [1]); the second group assumes correspondence between occurrence of fillers and the speaker's mental processes or emotional states, and aims to find factors that increase or decrease fillers (see a review in [2]); the third group claims that fillers convey information about discourse structure and/or speaker's attitudes toward interlocutors, and that they contribute to smooth communication (e.g. [3]); the last group tries to find out acoustic characteristics of fillers to allow automatic detection for effective speech recognition (e.g. [4]). The present research is mainly relevant to the second and the third groups.

Research in the second group claims that presumed difficulty of the tasks affects the occurrence of fillers; they tend to increase as the task difficulty escalates and the speaker's cognitive load becomes heavier [2]. Christenfeld [5], for example, has found in his experimental study using mazes that fillers increase when a speaker is confronted with more complex mazes. He claims that fillers escalate when the speaker has more choices in contents and words to be uttered.

Lounsbury [6] has linked the speaker's cognitive processes with the occurrence of pauses (including fillers) and the discourse structure. He has argued that pauses should serve as clues to the strength of associations between two linguistic events, and that longer pauses are supposed to occur where transition probability is low, reflecting the presence of weak associations between linguistic events, and thereby marking the beginning (or end) of speaker units.

Concerning the relationship between occurrence of fillers and discourse segment boundary strength, Swerts [7] found correspondences between them in Dutch monologues; phrases immediately after major boundaries contained more fillers than those after minor ones. Another conclusion in his research was that fillers after stronger breaks tended to occur phrase-initially, while those after weaker ones phrase-internally. It was also pointed out that *um* tended to occur at phrase-initial positions, whereas *uh* at phrase-internal positions. The author concluded that fillers seemed to carry information about discourse segment boundaries, and that difference in types might reflect different planning processes.

1.2. Purpose of the study

Based on the above arguments, we tested the following hypotheses in our previous research, where an excerpt from university lectures and two speeches on an academic conference were used as material [8, 9].

1) Fillers appear more frequently at deeper discourse segment boundaries than at shallower ones.
2) Fillers occur phrase-initially more often after stronger discourse segment boundaries than after weaker ones.
3) *Eto* and *e* tend to appear more frequently at discourse segment boundaries than *ano* and *sono*.

Eto and *e* have no semantic meanings, whereas *ano* and *sono* have functions as demonstrative adjectives, similar to *that* and *the* in English respectively, as well as those as fillers. Hypotheses 1) and 2) were based on the assumptions that the likelihood of the occurrence of fillers varies with the speaker's cognitive load. It was hypothesised that speaker's cognitive load is heavier at deeper boundaries, because of the additional planning needed at discourse as well as sentence and phrase levels. The third hypothesis was based on our earlier finding that *eto* and *e* tended to occur more frequently at stronger syntactic boundaries such as sentence- and clause- boundaries than *ano* and *sono* [10]. As a discourse segment is usually larger than a sentence or a phrase, we inferred that the former group of fillers tended to occur more frequently at discourse segment boundaries than the latter.

The results of our previous studies were as follows; hypothesis 1) was supported only by one speech sample out of three. Hypothesis 2) was not supported. Contrary to Dutch cases, fillers tended to appear phrase-internally more often

after stronger boundaries than weaker ones. Hypothesis 3) was amended to *eto* because this tends to appear more frequently at discourse segment boundaries than *ano* and *sono*. We argued that the discrepancies in the results of studies on Dutch and Japanese fillers might have derived from the difference in the degree of spontaneity of speech material. Swerts used description of paintings, while we used academic monologues. Lectures and speeches on academic conferences are usually well prepared and rehearsed. Therefore, they may be much less spontaneous than Swerts' material. In the present study we tested the same hypotheses, using more spontaneous speech.

2. Method

2.1. Speech samples

Speech samples were taken from the *Corpus of Spontaneous Japanese* (CSJ, Monitor Version 2001, [11]). They were presentations on given topics by voluntary subjects. Six speech samples produced by three subjects were analysed. The speaker attributes, topics and duration of speech samples are given in Table 1. The subjects were informed of the topics and of the approximate duration of the speech beforehand. Therefore, concerning the contents and the structure, the talks are not completely spontaneous. However, [11] points out that they have more features in common with speeches in everyday conversation than talks on academic conferences do. Therefore, we assumed that these samples were more spontaneous than our former material.

The transcription of the samples was divided into intonational phrases (IP) in terms of J-ToBI labelling [12] by the author. IP was employed as a unit of analysis, because empirically it seemed the smallest unit that composes a discourse segment. A discourse segment (DS) was defined as a unit based on the speaker's purpose of utterances.

Table 1: Description of speech samples

Speech ID	Speaker ID	Topic ID	Sex	Age	Duration (sec)
A1	A	1	f	30s	711
B1	B	1	m	30s	811
C1	C	1	m	20s	921
B2	B	2	m	30s	805
A3	A	3	f	30s	815
C3	C	3	m	20s	627

Topic ID: Topic 1: The happiest memory of my life.
Topic 2: The worst memory of my life.
Topic 3: My city.

2.2. Definition of boundaries

We agreed with [13]'s assumption that there will be less disagreement among labellers about stronger breaks since they present clearer transitions in the flow of information, and employed his method of defining discourse segment boundary strengths. That is, places where more labellers marked boundaries are stronger boundaries. We also used the kappa coefficient to access inter-labeller reliability [14].

Three labellers segmented the text. They were instructed to mark boundaries at the beginning of IPs where the speaker fulfilled one purpose of discourse and moved on to talk about another. They were told to choose a smaller unit if they wavered in their judgment. An example of a segmented text (part of another speech) with discourse purposes labelled was given before they started segmentation. The task was individually carried out at a workstation.

Inter-labeller reliability between three labellers for each speech sample was measured using the kappa coefficient. The scores are shown in Table 2. Although there is no consensus for valid inter-labeller reliability at the moment, [14] suggested the following tentative categorization of kappa values for labelling by more than two labellers:

.40 < kappa < .60 : fair
.60 < kappa < .80 : good
.80 < kappa : excellent

Referring to this criterion, all the values obtained here, except for that of C1, can be regarded as good. Therefore, segmented texts, except for C1, were taken for further analysis.

Table 2: Kappa values, which indicate inter-labeller reliability between three labellers for each speech sample.

Speech	A1	A3	B1	B2	C1	C3
kappa	0.7	0.73	0.66	0.66	0.52	0.61

IP-boundaries immediately before the IPs where three labellers marked boundaries are called Bn3. Likewise, those before the IPs where two labellers marked boundaries are called Bn2, those before the IPs where one labeller marked boundaries, Bn1, and those before the IPs where no labeller marked a boundary, Bn0. Based on [13], Bn3 are regarded as the strongest boundaries, and Bn0 as the weakest. Bn3 are sometimes called discourse segment boundaries (DSB) in the following part.

2.3. Method of analysis

First, we calculated the numbers and ratios of IPs with fillers immediately after each type of boundary (Bn0 ~ Bn3). Second, we examined locations of fillers in IPs after four types of boundaries; whether they appeared phrase-initially or not. Third, we analysed the distribution of each type of filler in relation to the boundary strength.

3. Results

3.1. The ratio of IPs with fillers

Table 3 illustrates the total number of IPs immediately after each type of boundary, the number of IPs with fillers after each type of boundary, and the ratio of the latter to the former. Figure 1 shows the ratio of the IPs with fillers to the total number of IPs, immediately after each type of boundary, for each speech sample.

Figure 1 demonstrates that the ratio of IPs with fillers tends to grow, as the boundary strength increases. If we compare the ratios of IPs with fillers at Bn0 with those at Bn3, there is statistical difference between them in all the samples except for A1. In four out of five speeches, IPs after Bn3 contain fillers more often than those after Bn0. This result indicates that the frequency of fillers corresponds to the boundary strength. The deeper a boundary is, the higher the possibility of immediately following phrase's containing filler

is. Therefore, in four out of five speeches, hypothesis 1) was supported by the result.

Table 3: Total numbers of IPs immediately after each type of boundary (Bn0 ~Bn3), numbers of IPs with fillers immediately after each type of boundary, and the ratios of the latter to the former (in %).

ID		Bn0	Bn1	Bn2	Bn3	Total
A1	Number of IPs	354	33	11	19	417
	IPs with fillers	118 (33)	10 (30)	6 (55)	7 (37)	141
B1	Number of IPs	457	24	17	24	522
	IPs with fillers	144 (32)	12 (50)	10 (59)	18 (75)	184
B2	Number of IPs	473	24	10	27	534
	IPs with fillers	139 (29)	10 (42)	5 (50)	23 (85)	177
A3	Number of IPs	422	17	9	28	476
	IPs with fillers	130 (31)	4 (24)	7 (78)	17 (61)	158
C3	Number of IPs	370	23	8	22	423
	IPs with fillers	148 (40)	12 (52)	5 (63)	15 (68)	180

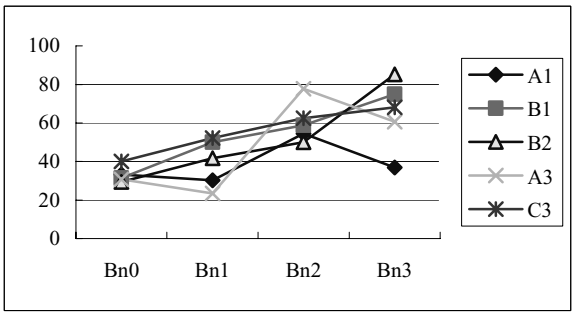

Figure 1: Ratios of IPs with fillers to the total number of IPs, immediately after each type of boundary (Bn0 ~ Bn3), for each speech sample.

3.2. The location of fillers in IPs

Table 4 gives the number of IPs with IP-initial fillers and the ratios of those in (%) to the total number of IPs immediately after each type of boundary, for each speech sample. Figure 2 demonstrates the ratios as a function of boundary strength. In Figure 2, slopes and shapes of lines are very similar to those of Figure 1, which shows the ratios of IPs with fillers at any location in them. This fact implies that it is mainly the IP-initial fillers that increase as the boundary strength grows.

Table 5 gives numbers of IPs with IP-non-initial fillers and the ratios of those in (%) to the total number of IPs, immediately following each type of boundary, for each speech sample. Figure 6 displays the ratios as a function of boundary strength. Compared with the cases of IPs with IP-initial fillers shown in Figure 2, the difference in ratios of IPs with IP-non-initial fillers at Bn0 and Bn3 is minimal. This fact supports our inference that it is mainly IP-initial fillers that escalate as the boundary strength grows. Both at Bn0 and Bn3 fillers occur more often IP-initially. This is inconsistent with the case of Dutch speech, where fillers appeared more often phrase-initially only at stronger boundaries. However, the tendency of fillers appearing at IP-initial positions is stronger at Bn3, because frequency of fillers increases mainly IP-initially at Bn3. This tendency is true of Dutch cases. If we amend hypothesis 2) to *the tendency of fillers appearing at phrase initial positions is stronger at deeper boundaries*, it is supported both by Dutch and Japanese cases.

Table 4: Numbers of IPs with IP-initial fillers and the ratios of those in (%) to the total number of IPs, immediately after each type of boundary (Bn0 ~ Bn3), for each speech sample.

IP-initial	Bn0	Bn1	Bn2	Bn3
A1	107 (30)	7 (21)	4 (36)	6 (32)
B1	113 (25)	11 (46)	7 (41)	18 (75)
B2	124 (26)	8 (33)	5 (50)	23 (85)
A3	120 (28)	2 (12)	5 (56)	14 (50)
C3	138 (37)	10 (43)	5 (63)	13 (59)

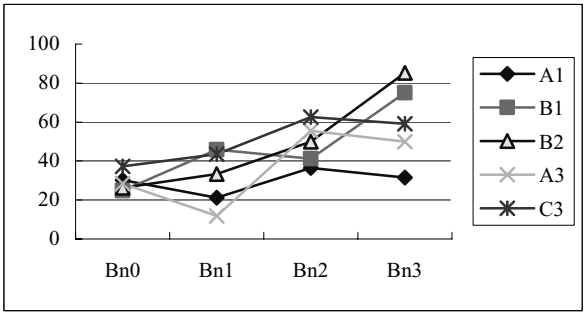

Figure 2: Ratios of IPs with fillers at IP-initial positions to the total number of IPs, immediately after each type of boundary (Bn0 ~ Bn3), for each speech sample.

Table 5: Numbers of IPs with IP-non-initial fillers and the ratios of those in (%) to the total numbers of IPs, immediately after each type of boundary (Bn0 ~ Bn3), for each speech sample.

IP-non-initial	Bn0	Bn1	Bn2	Bn3
A1	15 (4)	3 (9)	2 (18)	3 (16)
B1	33 (7)	2 (8)	5 (29)	1 (4)
B2	20 (4)	2 (8)	1 (10)	1 (4)
A3	12 (3)	2 (12)	2 (22)	4 (14)
C3	10 (3)	3 (13)	0 (0)	3 (14)

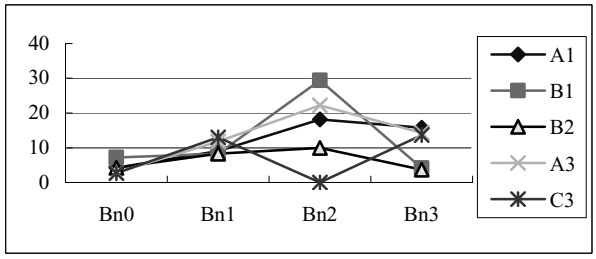

Figure 3: Ratios of IPs with IP-non-initial fillers to the total number of IPs, immediately after each type of boundary (Bn0 ~ Bn3), for each speech sample.

3.3. The distribution of each type of filler

Table 6 describes the average ratios among five speech samples of the IPs with each type of filler to the total number of IPs, immediately after each type of boundary. The ratios are also given in Figure 4. Figure 4 clearly shows that *eto* and *e* increase as boundary strength grows, *eto* sharply, and *e* more

gently. In academic presentations, only *eto* occurred more frequently than *ano* and *sono* at DSB. Here, to a lesser degree, *e* was also seen to increase at DSB, which supports hypothesis 3). However, as for *sono*, the sample size is too small to draw any conclusion. The frequency of each type of filler in five speeches, and the ratio of each to the total number of fillers, are given in Table 7.

Table 6: Average ratios among five speech samples of IPs with each type of filler to the total number of IPs, immediately after each type of boundary (in %).

	Bn0	Bn1	Bn2	Bn3
eto	7.8	13.7	22.3	39.1
e	7.8	8.2	14.1	18.0
ano	7.4	9.4	17.2	4.4
sono	7.0	0.0	0.0	1.8
ma	5.6	5.8	7.3	3.2

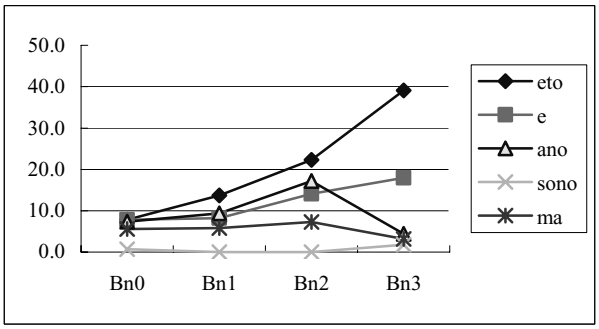

Figure 4: Average ratios among five speech samples of IPs with each type of filler to the total number of IPs, immediately after each type of boundary (in %).

Table 7: Frequencies of fillers in five speeches and ratios of each type of filler, to the total number of fillers, given in % in the bottom line.

	eto	*e*	*ano*	*sono*	*ma*	others	total
A1	43	18	47	7	12	21	148
B1	62	48	22	4	30	27	193
B2	51	72	14	0	30	17	184
A3	57	6	68	3	5	22	161
C3	28	59	26	1	54	14	182
sum	241	203	177	15	131	101	868
%	28	23	20	2	15	12	100

4. Discussion

The results described above are more consistent with those from studies on Dutch monologues than with those from our previous research on academic speeches in Japanese. Correspondences between boundary strength and the occurrence of fillers seem to decrease, as speech becomes more prepared, and less spontaneous. Fillers may signal DSB only in speeches with spontaneity above a certain degree. In less spontaneous speech such as academic presentations, the speaker usually knows well how to advance his talk. Therefore, it may be because he does not need extra time for speech planning or because he uses appropriate connectives rather than fillers at deeper boundaries, that fillers do not occur more frequently there than at shallower ones.

5. Conclusions

The present research has uncovered that phrase-initial *eto* and *e* tend to increase as boundary strength grows in relatively spontaneous speeches. This finding indicates that these fillers provide contributory evidence to the location and the strength of boundaries at the discourse level. At the next stage, perceptual experiments will be required to ascertain whether listeners actually make use of fillers to detect boundaries.

6. References

[1] Christenfeld, N., 1996. Effects of a metronome on the filled pauses of fluent speakers. *Journal of Speech and Hearing Research* 39 (6), 1232-1238.

[2] Rochester, S. R., 1973. The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research* 2, 51-81.

[3] Schiffrin, D., 1987. *Discourse markers*. Cambridge: Cambridge University Press.

[4] Goto, M.; Itou K.; and Hayamizu, S., 1999. A real-time system detecting filled pauses in spontaneous speech. *Information Processing Society of Japan SIG Notes* 99 (64), 9-16.

[5] Christenfeld, N., 1994. Options and ums, *Journal of Language and Social Psychology,* 13 (2), 192-199.

[6] Lounsbury, F. G., 1954. Transitional probability, linguistic structure, and systems of habit-family hierarchies. In *Psycholinguistics: A Survey of Theory and Research Problems,* Osgood, C. E., and Sebeok,.T.A. (eds.). Baltimore: Waverly Press, 93-101.

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

[8] Watanabe, M., 2001. The usage of fillers at discourse segment boundaries in Japanese lecture-style monologues, *Disfluency in spontaneous speech, ISCA Tutorial and research workshop*, Edinburgh, Scotland, 89-92.

[9] Watanabe, M., 2001. The distribution of fillers at discourse segment boundaries in academic monologues in Japanese. *Proceedings of the 15th General Meeting of Phonetic Society of Japan,* Kobe, 85-90.

[10] Watanabe, M.; Ishi, C. T., 2000. The Distribution of Fillers in Lectures in the Japanese Language. *Proceedings of the 6th ICSLP.* Beijing: Vol. 3, 167-170.

[11] Maekawa, K. 2001. Compiling *The Corpus of Spontaneous Japanese*. *Proc. the Spontaneous Speech Science and Technology Workshop*. Tokyo, 7-12.

[12] Venditti, J. 1995. Japanese ToBI labeling guideline. http://ling.ohio-state.edu/phonetics/J_ToBI/

[13] Swerts, M. 1997. Prosodic features at discourse boundaries of different strength. *Journal of Acoustical Society of America* 101 (1), 514-521.

[14] Araki, M., Itoh, T., Kumagai, T., and Ishizaki, M. 1999. Proposal of a standard utterance-unit tagging scheme. *Journal of Japanese Society for artificial intelligence*, Vol.14 No.2, 251-260.