Proceedings of DiSS-LPSS Joint Workshop 2010,
25-26 September 2010, Tokyo, Japan, pp. 27-30.

On the characteristics of three types of Japanese fillers: *e*-, *ma*-, and demonstrative-type fillers

Takuya Kawada[1]

[1]Kyoto University, Japan
tkawada@gmail.com

Abstract

Japanese has various forms of fillers. However, the characteristics of each form have yet to be well understood. We use a large corpus of spontaneous Japanese speech and conversation and focus on three frequently observed types of fillers : *e*-, *ma*-, and demonstrative-type fillers. We show that it is possible to characterize Japanese fillers from the viewpoint of how a speaker concerns himself with the listener in the communicative setting. The type of discourse, way of speaking, and direction of gaze of the speaker influence the distribution of the types of filler.

Index Terms: Japanese, fillers, spoken settings, gaze.

1. Introduction

A wide variety of filler types is observed in spontaneous Japanese speech, such as *ee, aa, uu, ano, sono,* and *ma*. These fillers can be divided into categories. Some are simply prolonged vowels, such as *ee, aa,* and *uu*. *Ano* and *sono* are identical in form to demonstratives. The third type is originally an adverb that corresponds to *ma(a)*.

Early studies on Japanese fillers mainly took cognitive and linguistic approaches. In one example, Watanabe [1] showed that (1) Japanese fillers are more frequent at deeper syntactic and discourse boundaries, and (2) fillers appear at a higher rate at the beginning of complex constituents. In a linguistic approach, Sadanobu et al. [2] showed the difference between the fillers *ee(to)* and *anoo* based on artificial examples. They showed that both *anoo* and *ee* can be uttered during a speaker's considering something, while *anoo* is unnatural when a speaker does a numerical calculation. Few studies have been conducted to describe the various forms of Japanese filler through empirical analysis.

In this paper, we focus on *ee*, *ano*, *sono*, and *ma*, which are frequently observed in spontaneous Japanese speech, and describe the behavior and characteristics of each form of filler. We claim that each filler can be characterized from the viewpoint of the degree of the speaker's commitment to the listener in the communicative setting. Specifically, our conclusions will be as follows: (1) the type of discourse influences the distribution of the types of filler, (2) the way of speaking also affects the frequency of each type of filler, (3) the speaker's gaze direction synchronized with the uttering of a filler differs depending on the type of filler.

2. Method

2.1. Corpora

To analyze Japanese fillers we used two corpora. One is the "Corpus of Spontaneous Japanese" (CSJ), one of the largest recorded spontaneous Japanese speech corpora, constructed by the National Institute for Japanese Language in 2004. The second is the "Billboard Corpus" (BC), a video recorded corpus constructed by Kyoto University in 2006.

The CSJ consists mainly of monologues and dialogues. The monologues include 1,715 academic presentations and 987 simulated public speeches. The average recorded time is 10–25 minutes. The attributes of speakers of the academic presentations are biased toward males and graduate students. In the simulated public speeches, the participants talk about everyday topics, such as the happiest experiences of their lives. In the CSJ monologue data, interaction with the listener was seldom observed because of the characteristics of the speech style. The dialogues in the CSJ form a set of data in which 26 interviews, 16 free dialogues, and 16 task-oriented dialogues are recorded. All the data are two-participant dialogues. In the interviews, the interviewees were also the presenters of the academic presentations or simulated public speeches, and interviewed about their speech. In the task-oriented speech, two participants discuss and solve a given task: (1) Two participants are given a list of nine or ten Japanese entertainers. (2) They discuss how much it would cost to invite the entertainer to speak. (3) They rank the entertainers in terms of their speaking price. An overview of the CSJ is shown on table 1.

Table 1: Overview of the CSJ

Sound type	Speakers	Data	Time(h)
Academic presentation	819	987	274.4
Simulated public speech	594	1715	329.9
Interview	26	26	5.5
Free dialogue	16	16	3.1
Task-oriented dialogue	16	16	3.6

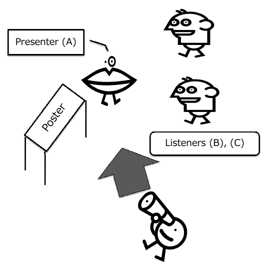


Figure 1: Billboard Corpus

The BC is a video corpus in which nine simulated poster sessions are recorded. Each session consists of three participants: one is the presenter and the others are listeners. All the presenters are graduate students. They talk about their research themes. Originally, the BC was constructed as basic engineering data for the development of an automated indexing system using audio and visual information. For that reason, the BC comprises not only transcription but is also rich in nonverbal information tags. For example, it contains annotations on the participants' gazes, nods, and pointing. The BC was recorded in a room adjusted for video recording. The scenery of the recording is as in figure 1. More than two cameras are in fact set.

Annotators can choose a camera angle to suit their purpose. Table 2 summarizes the kinds of data used.

Table 2: Summary of used data

	CSJ Monologue	Billboard	CSJ Dialogue
Data set	2,702	9	58
Time (s)	2,180,292	12,509	43,774

2.2. Spoken settings

There are various settings for spoken language according to how speakers concern themselves with their listeners. Clark [3] classifies spoken settings into the categories shown in Table 3. In this paper, we focus on "personal settings" and "nonpersonal settings". Personal settings are normal dialogues, such as face-to-face or phone conversations. They may have more than two participants and turn-taking is often observed. In nonpersonal settings, one of the participants is the primary speaker and is seldom interrupted by the listeners. For that reason, turn-taking is rarely observed. The CSJ monologues fit into nonpersonal spoken settings because interactions are seldom observed, while the CSJ dialogues are personal spoken settings because of their frequent interactions. Poster session conversations in BC are neither personal nor nonpersonal settings but intermediate between the two settings. In poster session conversation, listeners can freely ask questions as they wish and even take control of the conversation, though the speaker mainly provides information for listeners unilaterally like a lecture or speech [4].

Table 3: Settings of language use (Clark 1996: 8)

Spoken Settings	Example
Personal	**A converses face to face with B**
Nonpersonal	**Professor A lectures to students in class B**
Institutional	Lawyer A interrogates witness B in court
Prescriptive	Groom A makes ritual promise to bride B in front of witnesses
Fictional	A performs a play for audience B
Mediated	C simultaneously translates for B what A says to B
Private	A talks to self about plans

3. Results

3.1. Three types of Japanese fillers

In this section, we show the different distribution of Japanese fillers by their forms based on the type of discourse: monologue (speech), dialogue, and poster session conversation. Table 4 shows the top 10 frequencies of fillers by type of discourse. First, we introduce the frequency of fillers in the CSJ monologue data. The following forms are counted as fillers in CSJ [5]:

Table 4: Top 10 frequencies of fillers

Monologues		Billboard		Dialogs	
Filler	Freq.	Filler.	Freq.	Filler	Freq.
ee	117,160	*ma*	686	*anoo*	590
e	46,716	*maa*	684	*maa*	511
ma	44,768	*anoo*	275	*ee*	471
anoo	40,777	*ano*	259	*a*	447
ano	33,644	*etto*	199	*ma*	428
maa	31,430	*ettoo*	171	*aa*	381
sono	15,130	*ee*	141	*ano*	365
a	11,829	*n*	133	*n*	281
eeto	11,079	*e*	112	*sono*	206
n	9,897	*eeto*	88	*e*	173
others(87 forms)	66,449	others(34 forms)	467	others(55 forms)	964
Total	428,879	Total	3,215	Total	4,817

(1) Fillers in the CSJ
a(a), i(i), u(u), e(e), o(o), nn, to(o)**, ma(a)***, u(u)n, a(a)nno(o)***, so(o)nno(o)***, u(u)nnt(t)o(o)***, a(a)t(t)oo***, e(e)t(t)o(o)***, n(n)t(t)o(o)**

† The following particles can be connected with each filler : *-desune(e)*, *-ssune(e)*, (e.g. *ano desune*)

† Fillers marked with * above can connect with the following particles: *ne(e), sa(a)*. (e.g. *maa nee*)

† Parentheses are optional.

In the CSJ monologue data, fillers corresponding to those in (1) occurred about 430,000 times. Although in total 97 forms of fillers are observed, a limited number of filler types makes up the majority. Especially, *e*-type fillers like *e*, *ee*, or *eeto*, demonstrative-type fillers that are identical in form to demonstratives such as *ano(o)*, *sono(o)*, and *ma*-type fillers which have common form of adverb *maa* occupy more than 75% of all types of filler in the CSJ monologue data. The CSJ dialogue data shows the same distribution of fillers as the monologue data. *E*-type, demonstrative-type, and *ma*-type fillers constituted the majority (about 65% of all fillers). These three types were also the majority in BC. The rate of the three types of filler was about 87%.

3.2. Distribution of the three types of filler

In this section, we show the relationship between spoken settings and the types of filler. We first calculated the proportion of the frequencies of each filler type (*e*, demonstrative, and *ma*) to the frequencies of all kinds of filler per spoken setting: the CSJ monologues and dialogues, and poster session conversations in the BC. Next, we compared the mean of the frequency rate of filler type per spoken setting (Figure 2). We hypothesized that the frequency rate of fillers is the same regardless of the spoken setting, and tested the hypothesis. Because the variances of the rate of each filler frequency were not always homogenous, we conducted a nonparametric test (Kruskal-Wallis test). The results are as follows: *e*-type ($\chi^2 = 29.06, p < .05$) and *ma*-type ($\chi^2 = 15.27, p < .05$) fillers are significant in spoken settings. In addition, multiple comparison (Mann-Whitney test) was conducted in terms of the rate of *e*-type and *ma*-type fillers. *E*-type fillers showed the highest rate of occurrence in the monologue and the *ma*-type occurred most frequently in the BC. The results show that the frequency of some filler types is related to the spoken setting. Personal settings like dialogues are a relatively incompatible environment with *e*-type fillers, while the type of spoken setting does not affect the frequency of demonstrative-type fillers. *Ma*-type fillers are characteristically compatible with poster session conversation, which is a spoken setting intermediate between personal and nonpersonal.

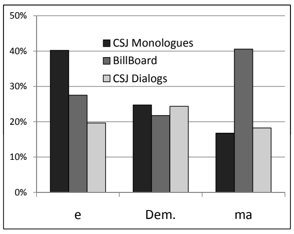


Figure 2: Filler rate

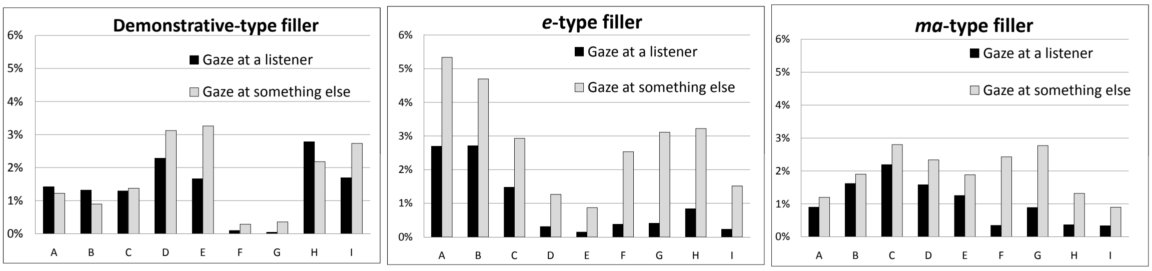
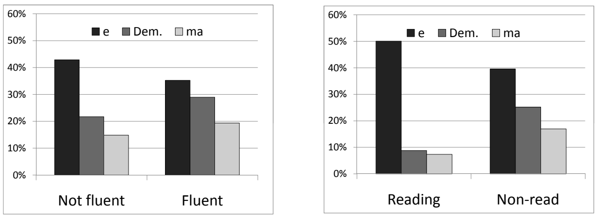


Figure 4: Rate of synchronization between speakers' gaze and filler

3.3. Listeners' impressions

In this section we analyze the relationship between the listeners' impression of a speech and the distribution of the forms of fillers. Information about the listeners' impression and observations on how the speakers made their speeches–such as the spontaneity, fluency, audibility, and whether or not the speaker read from a manuscript–is attached to all CSJ monologue data. If the frequency of the filler forms is affected by how speakers concern themselves with the listeners during their speech, some listeners' impressions of a speech may also relate to the distribution of a speaker's fillers. The data regarding the impression ratings stems from the judgments of 42 evaluators who were also recording staff when they listened to half of the speech. We focused on judgments of the speaker's fluency and whether the speaker read from a manuscript, and analyzed the distribution of the three forms of fillers per rating.



(a) Fluency

(b) Reading

Figure 3: Listeners' impressions and filler frequencies

The results are shown in Figure 3. The vertical axes in both figure 3(a) and 3(b) indicate the average ratio of each filler type frequency to all uttered fillers per speech. Figure 3(a) shows that *e*-type fillers are observed at significantly higher frequency in disfluent than in fluent speeches ($t = 6.76, df = 2482.3, p < .05$). Conversely, the frequencies of demonstrative-type ($t = 7.38, df = 2105.64, p < .05$) and *ma*-type fillers ($t = 6.35, df = 1988.30, p < .05$) are higher in fluent speeches than disfluent. Figure 3(b) compares speeches with a manuscript with those without. The same tendency is observed as with the fluency of speech. *E*-type fillers are more frequently observed in speeches with manuscript than in non-read ones ($t = 3.05, df = 101.64, p < .05$), while demonstrative-type ($t = 9.49, df = 110.73, p < .05$) and *ma*-type fillers ($t = 6.24, df = 105.43, p < .05$) are observed more often in non-read speeches than in read ones.

We thought that whether a speech is read from a manuscript is related to how much concern speakers have toward the listeners during their speeches. In read speech or disfluent speech it can be assumed that the speaker gives little attention to listeners during his/her speech but instead concentrates on reading the manuscript or on the speech going smoothly. Based on that assumption, *e*-type fillers have different characteristics compared to other fillers in the way speakers attend to listeners. Demonstrative and *ma*-type fillers are especially compatible with an environment wherein the speaker gives high attention to the listeners.

3.4. Gaze synchronization with filler

In this section, we introduce Kawada [6], which observed the synchronization of speakers' gazes with their fillers, and link this to the discussion above. Kawada [6] investigated the relationship between the direction of speakers' gazes and the form of filler using the BC. We especially focused on whether the speaker's gaze was on the listeners or on something other than the listeners, such as a poster, at the moment of uttering fillers. We calculated the duration of synchronizing a gaze at listeners and at something other than the listeners with the speaker's fillers. Because the duration of a gaze at a listener or another target differs for each data item, normalization of synchronizing duration is required. We calculated ratio between the duration of gazing at listeners or other targets synchronized with a filler and the total duration of gazing. The results are shown in Figure 4. The horizontal axis indicates each data item (speaker) and the vertical axis indicates the rate of synchronized time. As indicated by Figure 4, each form of filler has different characteristics in terms of the speaker's gaze. At the moment of uttering *e*-type fillers, speakers tend to look away from listeners, while for demonstrative-type fillers a stable tendency is not discerned. *Ma*-type fillers show almost the same tendency as *e*-types.

Kendon [7] points out that speakers uttering disfluencies tend to look away from the listener and that speakers can also concentrate on their own speech production during the time not looking at listeners. Although Kendon is not concerned with different types of disfluencies, particularly regarding differences in filler forms, we found that the hypothesis does not apply to all fillers, as is demonstrated by Japanese demonstrative-type fillers. The period of looking away from the listeners during speech implies the speaker's weak commitment to his listeners. Based on that assumption, demonstrative-type fillers differ from *e*-type and *ma*-type fillers in the degree of commitment to the listener.

4. Discussion

4.1. Speaker-/listener-oriented fillers

The above observations can be summarized as follows: *e*-type fillers are not compatible with active interaction and fluent speech. The speaker tends to look away from listeners while uttering *e*-type filler. This evidence leads to a possible generalization that *e*-type fillers are frequently observed in environments in which speakers do not have great concern toward listeners during their speech. In other words, *e*-type fillers are speaker-oriented. Demonstrative fillers, on the other hand, are still observed, even in dialogue. Although the gaze of the speaker also do not affect the frequency of demonstrative fillers, fluent and non-read speeches are more compatible with demonstrative fillers. This implies that demonstrative-type fillers have not only speaker-oriented but listener-oriented aspects. However, *ma*-type fillers are peculiar in that they are most frequently observed in the BC, which is an intermediate style between monologue and dialogue. On the other hand, fluent and non-read speeches are compatible with *ma*-type as well as with demonstrative-type fillers. In terms of gaze direction, the tendencies of *ma*-type fillers are similar to those of *e*-type fillers. So how do we characterize *ma*-type fillers? In the next section we analyze the *ma*-type filler through a linguistic (pragmatic) approach.

4.2. *Ma*-type filler

According to Kawada [8], the Japanese adverb *maa* has the discourse function of mitigating the utterance, and the following characteristics: its preferable environment is a situation where the speaker provides new information to the listeners. Conversely a situation where the source of the speaker's statement can be shared the listener is not compatible.

(2) a. (The speaker reads an alumni bulletin and notices that the family name of a friend has changed)
?? *Maa* kanozyo wa moo kekkon sita **rasii**.
'*Maa,* it seems that she's already married.'

b. (Hanako left the speaker three years ago)
Maa Hanako wa moo kekkon sita **daroo**.
'*Maa,* I guess she's already married'

(3) a. (Speaker is looking at listener's laptop on the desk)
?? *Maa* atarasii PC wo kaimasita **ne**.
'*Maa,* you bought a new PC.'

b. A: PC ga kowareta to kikimasita?
'I heard that your PC was broken.'
B: *Maa* Atarasii no kau koto ni simasita **yo**.
'*Maa*, I decided to buy a new PC.'

(2) compares the co-occurrence of *maa* and the auxiliary verbs, *rasii* and *daroo*. In this case, the co-occurrence of *maa* with *rasii* in (2a) is unnatural though *maa* can be interpreted as an interjection expressing the speaker's astonishment. (3) is the case of co-occurrence of *maa* with the sentence final particles *yo* and *ne*. This case indicates that *maa* is not compatible with *ne*. In fact, *rasii* and *ne* have a characteristics in common. Both require that information can be shared by listeners [9, 10]. In both (2a) and (3a), the speaker draws his conclusion based on visual evidence that is shared by the listeners. On the other hand, the speakers give new information to the listeners in (2b) and (3). Therefore, *maa* requires an environment where a speaker gives new information.

These characteristics of the adverbial *maa* are shared by *ma* type fillers. In terms of the CSJ dialogue data, *ma*-type fillers occur more frequently in interviews than in task-oriented dialogue (Table 5). Interviews are dialogues where the speaker tends to give new information unilaterally, while in the task oriented dialogues, the participants can easily share the information through the list of entertainers. Based on this fact, it is natural that the frequency of *ma* is high in the poster conversations because it is also the style in which the amount of information is asymmetrical between speakers and listeners. Unlike lectures or speeches, whose listeners are unspecified, poster session conversation and dialogue are typical joint actions where the participants build up common ground through interaction [3]. *Ma*-type filler is suitable for an environment with specific listeners where the speaker is required to concern listeners' current information during his speech. Therefore, *maa*-type fillers are also affected by the relationship between speakers and listeners.

Table 5: Filler frequencies in the CSJ dialogues

	All dialogues	Interviews	Task	Free
Data set	58	26	16	16
Time（s）	43,774	19,831	11,041	12,902
Demonstrative	1,450	911	71	468
e	922	535	180	207
ma	**944**	**626**	**48**	**270**
others	1,501	524	466	511
Total	4,817	2,596	765	1,456

5. Conclusion

In this paper, we focused on three types of fillers, *e*, demonstrative, and *ma*, which are frequently observed in Japanese discourse, and clarified the characteristics of each. The distribution of each type of filler is characterized from the viewpoint of how the speakers concern themselves with the listeners.

6. Acknowledgments

This work is partly supported by MEXT Grant-in-Aid for Scientific Research on Priority Areas "Cyber Infrastructure for the Information-explosion Era", Ministry of Education, Culture, Sports, Science and Technology, Japan.

7. References

[1] M. Watanabe, *Features and Roles of Filled Pauses in Speech Communication: A corpus-based study of spontaneous speech*. Tokyo: Hituzi Syobo, 2009.

[2] T. Sadanobu and Y. Takubo, "The monitoring devices of mental operations in discourse — a case of "eeto" and "ano(o)" —," *Gengokenkyuu*, vol. 108, pp. 74–93, 1995, (in Japanese).

[3] H. Clark, *Using language*. Cambridge: Cambridge University Press, 1996.

[4] H. Setoguchi, K. Takanashi, and K. Tatsuya, "Multi-modal recording of poster sessions and preliminary analysis," *IPSJ SIG Technical Report 2007-SLP-67*, pp. 31–36, 2007, (in Japanese).

[5] H. Koiso, Y. Mabuchi, K. Nishikawa, M. Saito, and K. Maekawa, *The specification of transcription*. The national institute for Japanese languages, 2004, (in Japanese).

[6] T. Kawada, "The synchronization between fillers and gaze in natural discourses – the case of japanese poster sessions," *Kyoto University linguistic research*, vol. 27, pp. 151–168, 2008, (in Japanese).

[7] A. Kendon, "Some functions of gaze-direction in social interaction," *Acta Psychologica*, vol. 26, pp. 22–63, 1967.

[8] T. Kawada, "The role and function of *maa* in japanese discourses," in *Linguistics and Japanese language education V*, 2007, pp. 175–192, (in Japanese).

[9] Y. Takubo, "Two types of modal auxiliaries in japanese: two directionalities in inference," in *Japanese/Korean Linguistics*, N. McGloin, Ed. Stanford: CSLI Publications, 2007, vol. 15, pp. 440–451.

[10] T. Masuoka, *The grammar of Japanese modality*. Tokyo: Kuroshio, 1991, (in Japanese).