

# Disfluency Patterns in Dialogue Processing

Etsuko Yoshida<sup>1</sup>, Robin Lickley<sup>2</sup>

<sup>1</sup> Department of Humanities, Mie University, Japan

<sup>2</sup> Department of Speech and Hearing Sciences, Queen Margaret University, UK

tantan@human.mie-u.ac.jp, RLickley@qmu.ac.uk

## Abstract

Spontaneous speech abounds with disfluencies such as filled pauses, repairs, repetitions, false start and prolongations, all of which are significant but easily overlooked features of speech communication. Based on the comparable corpora of English and Japanese dialogues, we argue that disfluency features can have a positive effect on turn-taking issues and the establishment of common referring expressions in dialogue processing. We examined the occurrence of ten types of filled pauses in Japanese and investigated how they interact with discourse entities and the sharing of common ground. The results indicate that two patterns of disfluency features contribute to on-line speech planning of the participants and their four functions serve to construct the collaborative process of speech communication.

**Index Terms:** dialogue, disfluency, referring expressions, corpus, common ground

## 1. Introduction

This paper focuses on disfluency in spoken language, and it presents a pilot study, investigating how and why disfluency occurs, especially in dialogue. Spontaneous speech is abundant with disfluencies such as filled pauses, repairs, repetitions, false start and prolongations, all of which are significant but easily overlooked features of speech communication. Why are these features of disfluency important and how do they contribute to dialogue processing? In this paper, considering the theoretical and methodological problems that face analyses of dialogue, we present a general hypothesis: dialogue participants collaborate to create their common ground for introducing a new discourse entity and disfluency can have a positive effect on turn-taking issues and the establishment of common referring expressions. Then we intend to clarify what interpersonal interaction can affect the speech fluency and why disfluency has a positive effect on dialogue processing.

First, in section 2, we set out the background and the general issues in the study of spoken discourse and section 3 gives a detail of dialogue data we obtained and the method of analysis. In section 4, we show the results of analysis of Japanese dialogues into the special focus on the patterns of disfluency. Furthermore, in section 5, based on the comparison between two languages, we will discuss disfluency function of interpersonal interaction that can affect the sharing of common ground in dialogue. Section 6 is the conclusion.

## 2. Dialogue and Disfluency : A General Pattern

Speaking is the most basic form of language use and requires on-line understanding and exchanging information between the participants in a more dynamic way than writing does. Speaking has two manners of talking about things: narratives and dialogues. The approaches to the problem of social meaning from the discourse analyst's point of view give rise to the investigation of the discourse structure of spoken narratives

(e.g. [1]) and conversation (e.g.[2]). The sequence of the patterns and structures in spoken discourse has been clarified. However, the discourse of dialogues is more structured than narrative discourse, because dialogue is highly interactive, and dialogue processing is more dynamic and complex in terms of the conversational strategy of assuming 'common ground' in 'a collaborative process in discourse ([3] and [4])'. The structural complexity of discourse processing of dialogues can directly affect the patterns of use of the disfluency and anaphoric device of discourse entities. To clarify this view, we conduct a quantitative analysis based on the spontaneously occurring dialogues in two different languages, and the result will be provided with a qualitative interpretation supported by linguistic evidence.

Disfluency is generally represented as filled pauses or fillers, and 'repaired' utterances. Filled pauses are typically characterised as non-propositional fragment such as *ano* or *etto* in Japanese, and *um* and *uh* in English [5]. It is only recently that fillers in Japanese spontaneous monologues have been studied extensively from the discourse-grammatical perspectives [6], and there has been no systematic research on disfluencies in Japanese dialogues so far. Linguistically, according to the survey of the National Institute for Japanese Language, disfluency rate in Japanese is reported to be higher (around 6 % of the total utterances) than other languages, e.g., English and Dutch (roughly 1 %) and in *The Corpus of Spontaneous Japanese* (CSJ) the total rate of fillers and word fragments (results of repetitions, deletions, or substitutions) is 7.2 % [7]. Among the features of disfluency, filled pauses (fillers) have been reported to be the most frequent, and Watanabe [8] found that filled pauses occupied 8% of the time of the university lectures examined.

In Japanese simulated dialogue, on the other hand, fillers are noted to occur in about 40% of the sentences and that *e*, *etto*, *ano*, *a*, and *ma* are the most frequent type [9]. In naturally occurring dialogues, it is expected that disfluency rate is higher and more interactional than that in presentational speech. According to Yamane [10], types of filled pauses may be categorised into two functions: speaker-oriented fillers such as *ano* and *eto*, and addressee-oriented fillers such as *iya*, *nanka*, *ne*, and *ma*. Sadanobu and Takubo [11] argue that Japanese fillers can have a function of monitoring the speaker's mental operations in discourse and provide a clue to predicting a subsequent utterance for the side of addressee. Furthermore, it is found that 60 % of *eto* occurred at the beginning of sentences, and *ano* and *ma* were the most frequent at clause boundaries (40 % each) in simulated Japanese dialogues [12]. Based on the naturally occurring dialogue, a cross-linguistic study on English and Japanese repairs is provided in Fox et al. [13] Moreover, using a task-oriented dialogue corpus, Branigan et al. [14] showed a relationship between the rate of disfluency and non-linguistic factors, e.g. gender and conversational roles and familiarity of speakers, eye-contact. Lickley [15] showed how disfluency rates vary with the demands of the dialogue task that the speaker is performing.

Paying special attention to the fact that disfluencies are more frequent at discourse boundaries (i.e. between the chunks of

utterances), and the hypothesis that the deeper the boundary, the higher the disfluency rate (called the boundary hypothesis) proposed by Watanabe, we will focus on the correlation between the disfluency rate and two stages of discourse boundaries, that is, the stage of introducing discourse entities at the initial utterance of the discourse and the stage of repairing discourse entities at the middle of discourse.

What kind of information do disfluencies convey in dialogue? Disfluency provides useful information to comprehenders because it is more likely to occur in some situations than others, due to the fact that disfluency occurs when speakers are having some kind of production difficulty [16]. Presumably, disfluency can play a special role of interaction between the participants, providing that a given discourse entity involves new and unfamiliar information. Arnold [17] claims that ‘the informativeness of disfluency can be explained in terms of its distribution; i.e., disfluency occurs more frequently in the context of reference to new or unfamiliar objects’ (509). Let us consider example 1:

- (1)  
 A: That tree has, *uh, uh*, ...  
 B: Tentworms.  
 A: Yeah.  
 B: Yeah.  
 (Clark and Wilkes-Gibbs 1986)

This suggests that the filled pauses of speaker A, *uh, uh*, is not only a speaker’s prolongation for keeping his or her floor, but functions as a clue to indicate that the subsequent discourse entity involves new and unfamiliar information. Therefore, a further request for a response to elicit from the addressee is necessary by the processing effort of collaboration between the participants to establish the common ground.

This fact reflects a disfluency feature at the location of middle stage of dialogue, occurring as repairs.

- (2)  
 come down until you’re *two in- ...uh two to three inches above that*  
 (Lickley and Bard 1998)

- F: yama ga *e* yama ja nai  
 mountain SUBJ um mountain COP NEG  
 ‘a mountain, um, it’s not a mountain’  
 (dc.p.18)

- (3)  
 F: de dokode tomareba ii < 290 > \*dosha no migigawa  
 then where stop-if ok clay of right hand side  
 ‘Then where should I stop right hand side of the sand’

- G: *\*etto*  
 um  
 G: dosha no < 270 > hidari-kawa de  
 sand of left hand side PP  
 ‘at the left hand side of the sand’  
 (dc.p.15)

In example (2), we provided two self-repairs from English and Japanese data. These filled pauses, *uh* in English and *e* in Japanese function as a process of insertion immediately followed by the same speaker’s repairs, or practically, substitutions: *two to three* instead of *two* in English, and *yamajana* (‘not a mountain’) instead of *yama* (‘a mountain’). Example (3) illustrates a case of interactional phenomena in the sense that the repair is processed by the collaborative exchange between the participants. Here, the interlocutor G is taking his

or her turn by interrupting with a filled pause *etto* followed by the repair *dosha no hidari-kawa* ‘the left hand side of the sand’. These instances suggest that the modified information added by the repair is a discourse-new element.

In section 3, we provide the data description and the method of analysing disfluencies.

### 3. Data and Analysis

The data we use for the present analysis is dialogue corpora called English and Japanese Labelless Map Task Corpus (MTC), which is an experimentally collected small comparable corpus of English and Japanese. This is based on the original MTC that was compiled at Human Communication Research Centre (HCRC), University of Edinburgh, UK (English data) and at Chiba University, Japan (Japanese data).

#### 3.1. English and Japanese Labelless Map Task Corpus

The labelless MTC is also based on the same maps with the same landmarks as the original ones. The corpus consists of recordings and transcriptions of eight dialogues. English and Japanese MTC use the same labelless maps and the same experimental design: The task involves two participants each of whom has a map to work with; familiar and unfamiliar pair of speakers; each participant plays a role as giver twice on the same map with different followers, and then as follower twice with different givers and different maps. However, the condition concerning familiar / unfamiliar pair of speakers are not considered in this paper. The maps of the instruction giver and the instruction follower for both English and Japanese data are shown in Figure.1

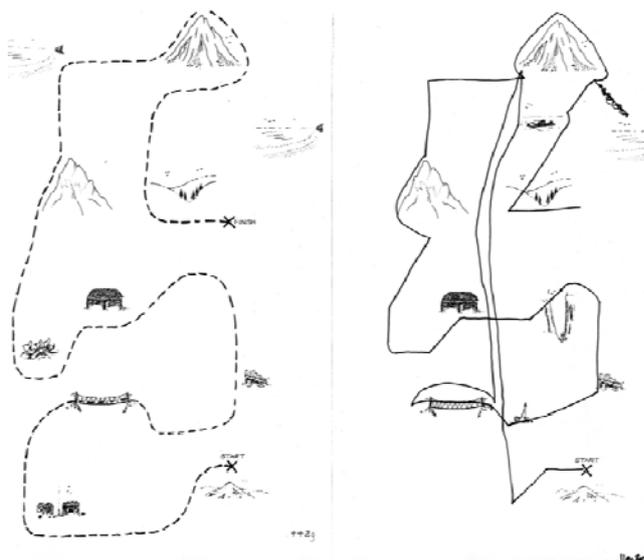


Figure. 1 English and Japanese Map Task Dialogue Corpus (Labelless): Instruction Giver’s Map(left) and Instruction Follower’s Map (right)

Based on this specific data, our current hypothesis is that the participants in the labelless MTC, especially the giver, tend to use a larger variety of referring expressions and disfluency more frequently to describe the landmarks and to explain the routes between landmarks than the participants in the labelled MTC. Therefore, the labelless MTC can provide us with relevant data in the sense that the lack of ‘ready-made’ written labels on the maps encourages the participants to construct

their own descriptions to identify entities of landmarks and to employ a number of referring and disfluency [18].

Furthermore, the labelless map task dialogue is more complicated than the original MTC due to the additional task design: naming the landmark. This task can require more effort into the participants' cooperation, especially at the initial stage of the dialogue. This makes the dialogue more complex and generates a different process of collaboration [19]. This innovation also makes grounding a more difficult task and should create a greater amount of disfluency, at least in the form of hesitation phenomena, like filled pauses and prolongations.

### 3.2. Method of Analysis

Based on the the set of eight dialogues of Japanese labelless MTC, we investigated the occurrence of filled pauses and non-lexical prolongations. Eight dialogues of the Japanese data contain 334 occurrences of disfluency in total. The analysis conducted here is based on the list of disfluency types and their distribution in the dialogue situation.

First, we have considered the general tendency of filler distribution in Japanese with respect to discourse situation in which filled pauses tend to occur mainly at two stages: discourse-initial and discourse-middle (Analysis 1).

Then, we analysed the types of fillers occurring at these two stages. The fillers we considered are divided into ten types: *ano*, *etto(-)* (so-called speaker-oriented fillers [10]), *e*, *a(t)* (based on vowel sounds), *(u)n* (a reply-type filler), *(n)to/(a)to* (based on conjunctive particle *to* 'and'), *so(no)*, *kou* (based on medial and proximal demonstrative), *m(a)*, and *nanka* 'something' (so-called addressee-oriented filler [10]) (Analysis 2). We do not count the fillers occurring in the utterances that are not directly related with the dialogue processing in producing, establishing, and maintaining discourse entities.

Finally, according to the participants' orientation to the speaker-role or the hearer-role, we divided all the fillers into four types: Type a. the speaker-role orientation at discourse-initial; Type b. the speaker-role orientation at taking turns; Type c. the hearer-role orientation as reply; Type d. the hearer-role orientation as clarification (Analysis 3). The two dialogues for each language are chosen out of 16 English and Japanese MTC dialogues.

## 4. Results

### 4.1. General tendency (Analysis 1)

In both English and Japanese data, it is confirmed that the filled pauses typically occur at the beginning of utterances at the initial stage of discourse and very rare at the end of the discourse.

The result of the analysis 1 is shown in Figure.2:

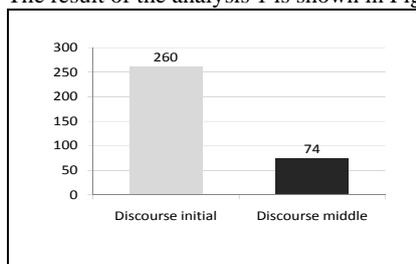


Figure. 2 Distribution of Fillers: occurring at two discourse stages

Among the 334 occurrences, 260 (78 %) of the filled pauses occur at the beginning of discourse (Discourse-initial) and 74 (22%) occur at the middle stage of discourse (Discourse-middle).

### 4.2. Features of fillers (Analysis 2)

The distribution of ten types of fillers is illustrated in Figure.3:

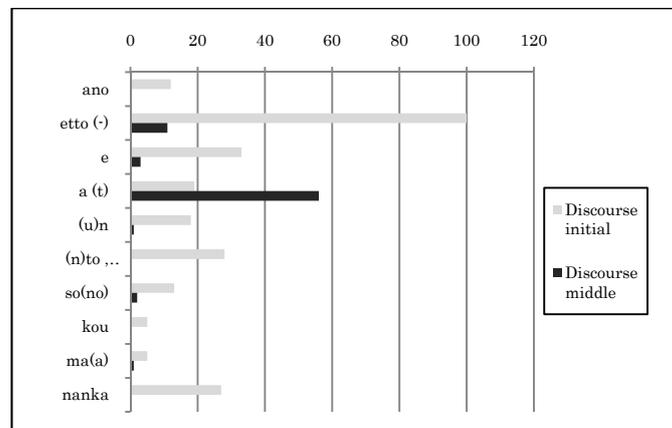


Figure. 3 Frequencies of filler types: Occurring at two discourse stages

As is shown here, at the initial stage of discourse, *etto* is the most frequent type of fillers (38%), and is subsequently followed by *e* (33.13%) and *(n)to* and *(a)to* (28.11%). As seen in section 2, *etto* is a speaker-oriented type of fillers and tend to occur with the introduction of the discourse entities. *Nanka* is frequently used to delay the production of a noun phrase that is introduced into a given discourse.

At the discourse middle, on the other hand, vowel prolongation *a(t)* is the most frequent (76%) and seem to be an overwhelming majority of the tokens. These fillers occur immediately before the repair as a clue to modification, substitution, and repetition.

In both stages, interestingly, *ano* is infrequent: 12.5% in discourse initial and almost none in discourse middle.

### 4.3. The function of fillers (Analysis 3)

As seen in Figure 4, as the speaker-role orientation, Type a is the most frequent, and Type b suggests a possible turn-taking, which occurs only a few in both sets. In the hearer-role orientation, compared with Japanese counterpart, English data indicates more frequent use of Type c and Type d.

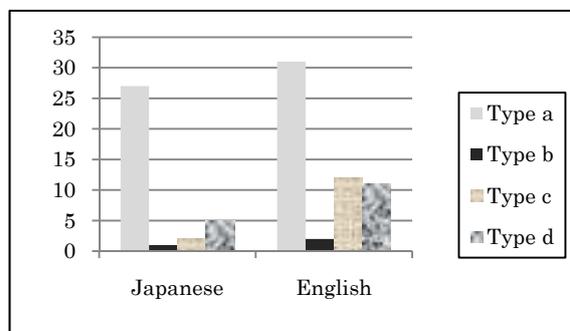


Figure. 4 Frequencies of filler functions: four types

## 5. Discussion

The findings in section 4 indicate two patterns of disfluency features and four patterns of disfluency functions. First, disfluency tends to occur at the discourse initial and it signals that an immediately subsequent utterance can include unfamiliar and new information. In dialogue, participants collaborate to introduce this information, i.e. usually a discourse-new entity, to establish common referring expressions for them. For example, *etto* is used not only as a prolongation of the speaker to keep his or her floor but also a clue to provide a helpful information for the interlocutor to keep the dialogue go forward as in example (4):

- (4)  
G: \*kuruma no hidari  
car of left 'at the left side of the car'  
F: hai  
Yes 'Yes'  
G: gawani nanika arimasu ka  
side at something is POL Q 'Is there something?'  
F: e  
F: to 'um'  
G: koyano younamono wa arimasu ka  
hut of like something TOP is POL Q  
'Is there something like a hut?'  
(bd, p.9)

The follower's *eto* does have a positive effect on dialogue processing. Here the follower's filled pause can signal to the giver whether the dialogue needs to proceed by taking his or her turns for further utterances.

Next, in the middle of discourse, many repairs are a speaker's modification, i.e. self-repair, but there are cases that the participants collaborate to share the discourse entity or the route description to establish the common ground for their better achievement of the task.

Furthermore, some speech context e.g., delaying NPs and generating possible NPs, can affect the speech fluency at the initial stage of discourse. In Analysis 3, the follower's utterance tends to contain fillers of the hearer-role orientation as clarification and as replies in English data. These types of fillers can modify the giver's planning of dialogue processing, i.e. confirming and up-dating the shared information.

Lastly, disfluency is prevalent in the whole process of discourse, and participants keep using a variety of fillers to send a pragmatic cue to each other. In dialogue, the participants collaborate to take turns and try a repaired utterance to avoid the suspension of speaking or the interruption of the dialogue flow.

## 6. Conclusions

We have provided some evidence that disfluency signals an immediately subsequent utterance containing unfamiliar information at the discourse initial, and a speaker's modification for a better achievement of the task in the discourse middle. The results also indicate that collaborative interaction (i.e. the participants' orientation to the speaker-role or the hearer-role) can affect the speech fluency and the sharing of common ground in dialogue, which can lead to better turn-taking. However, it still remains unknown how to clarify what collaborative interaction can affect the speech fluency. At the same time, more data analysis on disfluency in English and other languages should be conducted in further study to explicate why disfluency features are useful to understand the mechanism of successful communication.

## 7. Acknowledgements

This research is partly supported by the Daiwa Anglo-Japanese Foundation (Ref: 7102/7524) and the Great Britain-Sasakawa Foundation (Ref: J590).

## 8. References

- [1] Labov, William (1972) *Language in the Inner City: Studies in the Black English Vernacular*, Oxford: Blackwell.
- [2] Sack, H., Schegloff E. A. and Jefferson, G. (1974) 'A Simplest Systematics for the Organisation of Turn-taking for Conversation,' *Language* 50, 696-735.
- [3] Clark, H.H. and Wilkes-Gibbs, D. (1986) 'Referring as a Collaborative Process,' *Cognition*, 22, 1-30.
- [4] Brennan, S.E. and Clark, H.H. (1996) 'Conceptual Pacts and Lexical Choice in Conversation,' *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22 (6), 1482-1493.
- [5] Lickley R. J. and E.G. Bard (1998) 'When can Listeners Detect Disfluency in Spontaneous Speech?' *Language and Speech*, 41(2), 203-226.
- [6] Watanabe, M. (2009) *Features and Roles of Filled Pauses in Speech Communication: A Corpus-based Study of Spontaneous Speech*, Tokyo: Hituzi Syobo Publishing.
- [7] Maekawa, K. (2004) 'Design, Compilation, and Some Preliminary Analyses of the Corpus of Spontaneous Japanese,' In Yoneyama, K. and Maekawa, K. (Eds.). *Spontaneous Speech: Data and Analysis*, Tokyo: The National Institute for Japanese Language, 87-108.
- [8] Watanabe, M. (1999) 'The Distribution of Filled Pauses in the University Lectures,' *Proceedings of the 13th General Meeting of the Phonetic Society of Japan*, 143-148. (in Japanese)
- [9] Murakami, J. and S. Sagayama (1991) 'A Discussion of Acoustic and Linguistic Problems in Spontaneous Speech Recognition,' *Technical Report of IEICE, SP91-100, NLC91-57*, 71-78. (in Japanese)
- [10] Yamane, C. (2002) *Nihongo no Danwa ni Okeru Fira* (Fillers in Japanese Discourse), Tokyo: Kuroshio Syuppan. (in Japanese)
- [11] Sadanobu, T and Y. Takubo (1995) 'The Monitoring, Devices of Mental Operations in Discourse – A Case of 'etto' and 'ano(o)' –', *Gengo Kenkyu*, 108, 74-93. (in Japanese)
- [12] Nakagawa, S. and S. Kobayashi (1995) 'Phenomena and Acoustic Variation on Interjections, Pauses, and Repairs in Spontaneous Speech,' *Journal of the Acoustical Society of Japan*, 51(3), 202-212. (in Japanese)
- [13] Fox, B.A., M. Hayashi, and R. Jasperson (1996) 'Resources and repair: A cross-linguistic study of syntax and repair,' In Ochs, E., E.A. Schegloff, and S. A. Thompson (Eds.) *Interaction and Grammar*, 185-237.
- [14] Branigan, H., R. Lickley, and D. McKelvie (1999) 'Non-linguistic influences on rates of disfluency in spontaneous speech,' *Proceedings of the 14th International Conference of Phonetic Sciences*.
- [15] Lickley, R. J. (2001) 'Dialogue moves and disfluency rates,' In *DISS'01*, 93-96.
- [16] Clark, H. H. and J. E. Fox Tree (2002) 'Using *uh* and *um* in Spontaneous Speaking,' *Cognition*, 84, 73-111.
- [17] Arnold, Jennifer E. (2008) 'Reference production: Production-internal and addressee-oriented processes,' *Language and Cognitive Processes*, 23 (4), 495-527.
- [18] Yoshida, E. (2002) 'Nihongo Meishonashi Chizukadai Taiwa Corpus no Gaiyou to Tenki Text no Sakusei: Houkoku,' (An Outline and Transcription Text of Japanese Unlabelled Map Task Dialogue Corpus: A Report) *Jinbun Ronso (Bulletin of the Faculty of Humanities and Social Sciences, Mie University)* 19, 241-249. (in Japanese)
- [19] Yoshida, E (2008) *Patterns of Use of Referring Expressions in English and Japanese Dialogues*, PhD Thesis, University of Edinburgh. (Ch.8)