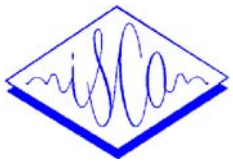
ISCA Archive
http://www.isca-speech.org/archive



INTERSPEECH 2004 - ICSLP
8th International Conference on Spoken Language Processing
ICC Jeju, Jeju Island, Korea
October 4-8, 2004

# Clause Types and Filled Pauses in Japanese Spontaneous Monologues

*Michiko Watanabe[1], Yasuharu Den[2], Keikichi Hirose[3] and Nobuaki Minematsu[4]*

[1]Macquarie Centre for Cognitive Science, Macquarie University, Australia, JST/CREST
[2]Faculty of Letters, Chiba University, Japan
[1, 3] Graduate School of Frontier Sciences, University of Tokyo, Japan
[4]Graduate School of Information Science and Technology, University of Tokyo, Japan
watanabe@gavo.t.u-tokyo.ac.jp

## Abstract

Hesitations are prevalent in spontaneous speech and believed to be relevant to on-line speech planning. We tested the complexity hypothesis that speakers are more likely to need to suspend speaking, the more complex the constituent, by examining ratios of filled pauses (fillers) at clause and case boundaries with following constituents of different degrees of complexity, using ì the Corpus of spontaneous Japaneseî. The filler ratios were constantly higher as the following constituents were more complex, supporting the hypothesis. The result indicates that fillers can be clues about complexity of the upcoming constituents.

10.21437/Interspeech.2004-748

## 1. Introduction

### 1.1. Hesitations and speech planning

Hesitations are ubiquitous in everyday speech. About 6 per 100 words are disfluent in Switchboard corpus, a corpus of telephone conversations in American English [1]. In Japanese about 50 % of sentences include redundant words (fillers) and 10 % of sentences contain repairs in simulated telephone conversations [2]. In spite of their frequency, surprisingly little is known about their features and roles in communication. The present paper aims at finding correspondence between hesitations and speech planning processes.

As hesitations appear only in spontaneous speech, not in read speech, they are considered to be relevant to on-line speech planning. Namely, when speakers have some difficulties in speech planning or speech execution, they may hesitate. Three main stages are assumed in speech planning: conceptualizing a message, formulating the appropriate linguistic expressions, and articulating them [3]. Hesitations are claimed to occur at any of these stages [4]. Major constituents such as sentences, clauses, and phrases are regarded as principal units of planning [3]. Concerning relationship between planning units and hesitations, it is claimed that the more information speakers want to convey, the more complex, the constituents become. Consequently the more time speakers need to formulate them and the more likely they produce hesitations. This view is called ì the complexity hypothesisî [5]. In brief, ì speakers are more likely to need to suspend speaking, the more complex the constituent.î Grammatical weight is employed as an index of complexity. It can be measured by the number of words, syntactic nodes, or phrasal nodes in the constituent and these numbers are reported to correlate each other at .94 and beyond [6].

Shribergís corpus based study on dialogues in English demonstrates that hesitation ratio can be expressed as a function of the sentence length in number of words excluding hesitations: the longer sentences, the higher the ratio of the sentences containing disfluencies, such as fillers, repeated words and speech segments [1]. Her study also shows that hesitations are significantly more probable at sentence beginning than elsewhere: about 20% at the beginning of sentences whereas less than 5% at other positions. These findings support the complexity hypothesis. Clark and Wasow tested the complexity hypothesis by comparing repetition rates of definite and indefinite articles in simple and complex noun phrases in four locations in sentences; in the positions of topics, subjects, objects and prepositional objects. The results revealed that the repetition rates were constantly higher in complex noun phrases than in simple noun phrases. The location of noun phrases also had an effect on the repetition rates. The closer the phrases were located to the left edge of sentences, the more frequently the articles in them were repeated; the repetition rate was highest when noun phrases were in topic position, next highest in subject position, then in object position, and lowest in objects of prepositions. The results support the complexity hypothesis, because the repetition rates of articles correspond to the complexity (measured by numbers of words) of constituents following them [5].

Speech planning processes must have language universal and language specific aspects. Although the complexity hypothesis may be applied language universally, ways of speakers suspending speech must differ depending on the language. Therefore, what to be measured should also differ depending on the language. It is, for example, impossible to measure repetition rates of articles in Japanese, because Japanese language does not have articles.

In the present study we test the complexity hypothesis by examining the ratios of filled pauses (or fillers, in other words) at several types of syntactic boundaries with following constituents of different degrees of complexity. As fillers are one of the most common types of hesitations [1], [2], frequencies of fillers should correspond to complexities of upcoming constituents if the complexity hypothesis is applied language universally. We formulated the complexity hypothesis of filler version as follows: the more complex the following constituent, the more likely speakers are to utter fillers. Hereafter we will call this the complexity hypothesis.

It is pointed out that ì eî, ì etoî, ì maî, ì anoî and ì sonoî are the most frequent fillers in Japanese monologues. ì Eî means ì yesî when it is pronounced with HL accent pattern. ì Eî is pronounced with low flat tone, when it is uttered as filler. ì Etoî is solely used as filler. ì Maî has functions as interjection expressing surprise and as modal adverb, as well as filler. ì Anoî and ì sonoî have usages as demonstrative adjectives,

similar to English ìthatî and ìtheî respectively, as well as fillers. These five types of fillers consist about 90 % of all fillers in Japanese academic monologues [7].

### 1.2. Clause types and the complexity

It is known that Japanese subordinate clauses have a hierarchical structure. Minami [8] pointed out that Japanese clauses can be classified into three groups based on the degree of independency from main clauses. Clauses categorized into *type A* are the most dependent on the main clauses. They can have neither their own topics nor subjects. Topics and subjects of the main clauses automatically become their topics and subjects. Clauses classified into *type B* can contain their own subjects, but not their own topics. Clauses belonging to *type C* can have both their own topics and subjects. Therefore, *type C* are the most independent of the main clauses. It is claimed that *type C* clauses should be regarded as independent sentential units in spontaneous speech [9].

The three types of clauses compose a hierarchical structure. *Type C* can contain all the three types of clauses, whereas *type B* can include *type A* and *B* but not *type C*. Clauses in *type A* can contain only *type A* clauses. Subordinate clauses are marked by connective particles or certain conjugations of verbs, adjectives and copula markers at the end of the clauses, and they always precede main clauses. Clause order of Japanese complex sentences is shown below.

[(subordinate clause <connective particle>), (main clause)].

Regarding complexity of the three types of clauses, *type C* clauses are assumed to be the most complex because they can contain any types of subordinate clauses and their own topics and subjects. *Type B* are supposed to be the second most complex because they can comprise *type A* and *B* clauses and their own subjects, but not their own topics. *Type A* are assumed to be the least complex because they can contain neither their own topics nor subjects. Therefore, it seems reasonable to assume that filler ratios are the largest at the beginning of *type C* clauses, smallest at the beginning of *type A* clauses and in between at the beginning of *type B* clauses.

However, there are some problems in defining the following constituents in complex sentences. As is obvious from the clause order described above, the beginning of subordinate clauses is also the beginning of sentences. Particularly when subordinate clauses are strongly dependent on main clauses, it is more likely that speakers plan whole sentences rather than only subordinate clauses when they start utterances. Then, it will be more reasonable to regard whole sentences, not subordinate clauses alone, as the following constituents. However, as we do not know for sure how far speakers plan sentences at the beginning of utterances, we have no clear criterion to decide up to which constituent we should include in ìthe following constituentsî in complex sentences. And it is impossible to measure complexity of ìthe following constituentsî without defining what ìthe following constituentsî are. In other words, we do not know which constituents filler ratios at the beginning of sentences are related to.

On the other hand, if we consider complexity of clauses immediately after subordinate clauses, following assumption is possible: clauses after *type C* tend to be more complex than clauses after *type A* and *type B*, because new topics and subjects can be introduced there as *type C* can have their own topics and subjects. Clauses after *type A* tend to be less complex than clauses after *type B*, because new subjects can appear in clauses after *type B* as *type B* can have their own subjects, whereas it is not possible after *type A*. Therefore, it is hypothesised that filler ratios are largest after *type C*, next largest after *type B* and smallest after *type A* clauses. This hypothesis is first tested.

### 1.3. Location of cases

The Japanese is a SOV language and cases are marked by case particles after nouns or noun phrases. Although the case order is not fixed, the most typical order is as follows:

Topic Nominative Dative Accusative

Topics tend to be the closest to the left edge of clauses and sentences, nominatives next closest, datives the third closest and accusatives the farthest from the left edge. As a consequence it is assumed that the following constituents tend to be most complex after topics, second most complex after nominatives, third most complex after datives and least complex after accusatives. Therefore, it is hypothesised that filler ratios are highest after topics, second highest after nominatives, third highest after datives and lowest after accusatives. This is the second hypothesis to be tested in this study.

## 2. Method

### 2.1. Corpus

*The Corpus of Spontaneous Japanese* (CSJ, Monitor Version 2002) was used to test the hypothesis. The corpus comprises speech, transcripts and morphological analyses of 134 academic presentations and 189 casual presentations. The speech data amounts to 86.5 hours, and the morphological data to 800 thousands words. It is reported that casual presentations are more like everyday conversations than academic presentations are in terms of length of sentences, speech rates and ratios of parts of speech [10]. Fillers are treated as words in the morphological analysis of the corpus. In the present research fillers are not regarded as words. To enable comparisons with our previous studies [11], 81 presentations (34 academic and 47 casual) comprising all the five most common types of fillers were excerpted from the corpus.

In the transcripts of CSJ, speech is divided into sentential units according to the following criteria: a stretch of speech either (a) delimited by silent pauses longer than 200ms or (b) ending with sentence final elements such as verbs in finite form and final particles [12]. As the segmentation primarily depends on pauses, the units do not precisely correspond to syntactic boundaries. Therefore, sentence and clause boundaries were first identified, and then subordinate clauses were classified into three groups depending on types of connective particles, based on classifications of [8], [13] and [9]. The classification of subordinate clauses employed in the present study is shown in table 1.

### 2.2. Measurement of filler ratios at clause boundaries

As the number of *type A* clauses was small (4 per speech on average), *type A* were excluded from the analysis of clause boundaries and treated as adverbial phrases.

As both preceding and following clauses may have effects on filler ratios at clause boundaries, clause boundaries were divided into nine cases depending on preceding and following clause types. The classification of boundary types is shown in Table 2. We call main clauses and simple sentences *type D* clauses for convenience. The nine cases will be referred to with combinations of preceding and following clause types hereafter. ìCDî, for example, are boundaries preceded by *type* C clauses and followed by *type* D (main) clauses. When preceding clause type is *D*, the boundaries are sentence boundaries. Ratio of boundaries containing fillers was calculated in each type of boundary in each speech and averaged over 81 speeches. Fillers following conjunctions immediately after boundaries were also counted as fillers at the boundaries.

### 2.3. Measurement of filler ratios after case markers

Ratios of four types of case particles (topic, nominative, dative, accusative) immediately followed by fillers were calculated in each speech and averaged over 81 speeches.

*Table 1*: Classification of subordinate clauses

| Types | connectives | meanings, usage |
|---|---|---|
| A | ~nagara | expresses accompanying actions |
| | ~tsutsu | expresses accompanying actions |
| | ~tari, dari | used in listing actions |
| | ~naku | without being ~ |
| B | ~to | when ~ |
| | ~ba | if ~ |
| | ~tara | if ~ |
| | ~nara | in case ~ |
| | ~te | ~ and |
| | ~te kara | since ~ |
| | ~te mo | even if ~ |
| | ~mae ni | before ~ |
| | ~ato de | after ~ |
| | ~toki ni | when ~ |
| | ~zu, ~zuni | without ~ |
| C | ~kara | as ~ (reason) |
| | ~node | as ~ (reason) |
| | ~noni | though ~ |
| | ~kedo, ~keredo | though ~ |
| | ~ga | although ~ |
| | ~si | ~ and (list similar actions or features) |
| | ~de | ~ and |
| | ~masite, ~desite | ~ polite auxiliary verbs + and |

*Table 2*: Classification of clause boundary types

| Preceding clause types | Following clause types | Boundary types |
|---|---|---|
| main clause (D) | simple sentence (D) | DD |
| | *type C* | DC |
| | *type B* | DB |
| *type C* | main clause (D) | CD |
| | *type C* | CC |
| | *type B* | CB |
| *type B* | main clause (D) | BD |
| | *type C* | BC |
| | *type B* | BB |

## 3. Results

### 3.1. Ratios of boundaries containing fillers

Ratios of nine types of boundaries containing fillers are shown in Figure 1. ì After_Bî, ì after_Cî and ì after_Dî along x axis indicate preceding clause types and simple ì Bî, ì Cî and ì Dî show following clause types. There is no significant difference between the ratios of three kinds of boundaries, BB, BC and BD, which are all preceded by *type B* clauses ($F = .162$, $p = .851$, ns.). There is no significant difference between the ratios of three kinds of boundaries, CB, CC and CD, which are preceded by *type C* clauses, either ($F = .298$, $p = .743$, ns.). Between ì after_Bî and ì after_Cî groups there are significant differences in all the three pairs having the same types of following clauses (BB vs. CB: $F = 30.409$, $p < .001$, BC vs. CC: $F = 28.346$, $p < .001$, BD vs. CD: $F = 30.680$, $p < .001$). These results support the complexity hypothesis, because ratios of clause boundaries containing fillers are constantly higher after *type C* clauses, where both topics and subjects may appear, than after *type B* clauses, where only subjects are allowed. The result indicates that filler ratios at clause boundaries can be predicted by the preceding clause types, not the following clause types, as we expected.

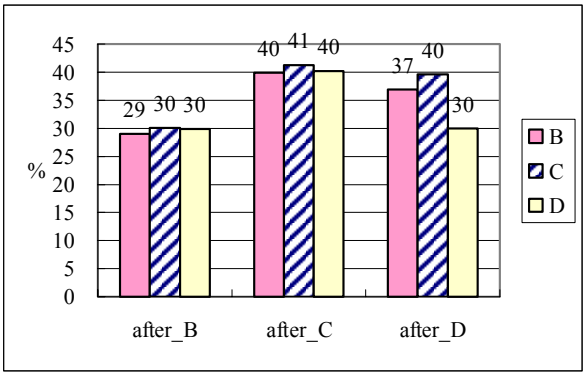



*Figure 1*: Ratios of 9 types of boundaries containing fillers. ì after_Bî, ì after_Cî and ì after_Dî along x axis indicate preceding clause types and simple ì Bî, ì Cî and ì Dî mean following clause types.

The complexity hypothesis holds true at sentence boundaries. Simple sentences are generally less complex than complex sentences. Filler ratios at sentence boundaries are significantly lower when the boundaries are followed by simple sentences, namely in case of DD, than when the boundaries are followed by complex sentences, in cases of DB and DC (DD vs. DB: $F = 6.740$, $p < .05$). There is no significant difference between the ratios of DB and DC, both of which are followed by complex sentences (DB vs. DC: $F = 1.368$, $p = .246$, ns.).

### 3.2. Ratios of case particles followed by fillers

Figure 2 illustrates ratios of particles immediately followed by fillers in four types of case particles. The closer the particles are located to the left edge of clauses or sentences, the more frequently the particles are followed by fillers; the ratio of fillers is highest after topic particles, next highest after nominative particles, then after dative particles and lowest after accusative particles. This result supports the complexity hypothesis, because the ratios of particles followed by fillers correspond to the complexity of constituents following them. This result shows the same tendency as the repetition rates of English articles in [5] referred to in 1.1. Different types of hesitations seem to be serving in similar ways for the same purpose of suspending speech.

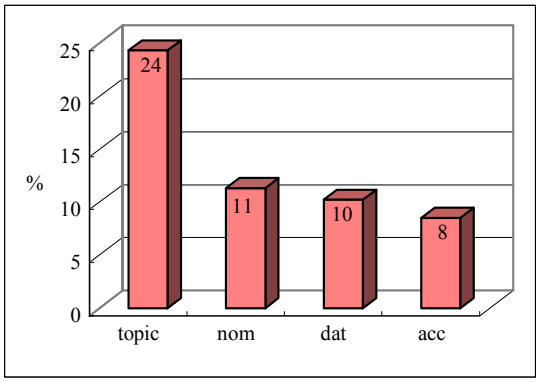



*Figure2*: Ratios of four types of case particles followed by fillers.

## 4. Discussion

As was mentioned in 3.1, filler ratios at clause boundaries did not show correspondence to the following clause types which differed in complexity. This finding indicates that speakers plan constituents larger than the immediate subordinate clauses when they start complex sentences. However, whether speakers do so with any type of complex sentences and how far and how exactly they plan the following constituents in advance remains to be investigated.

## 5. Conclusion

The present research tested the complexity hypothesis that the more complex the following constituent, the more likely speakers are to utter fillers, by examining ratios of filled pauses at clause and case boundaries with following constituents of different degrees of complexity. The filler ratios were constantly higher as the following constituents were more complex both at clause and case boundaries, supporting the hypothesis. The results indicate that the complexity hypothesis can be applied language universally and regardless of speech types, whether speech is dialogue or monologue, as long as speech is spontaneous. Fillers may be providing clues about complexity of the following utterances to listeners. We plan to investigate the effects of fillers on listenersí expectation and comprehension of upcoming phrases to better understand mechanisms of successful human speech communication.

## 6. Acknowledgements

This research is partly supported by the Japan Science and Technology Corporationís ìCore Research for Evolutional Science and Technology Programî, ìExpressive Speech Processing Projectî.

## 7. References

[1] Shriberg, E., E., *Preliminaries to a Theory of Speech Disfluencies*, PhD thesis submitted to the University of California at Berkeley, 1994.

[2] Murakami, J. and Sagayama, S., ìA discussion of acoustic and linguistic problems in spontaneous speech recognitionî, *Technical report of IEICE*, SP91-100, NLC91-57: 71-78, 1991.

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

[4] Clark, H. H. and Fox Tree, J. E., ìUsing *uh* and *um* in spontaneous speakingî, *Cognition* 84: 73-111, 2002.

[5] Clark, H. H., and Wasow, T., ìRepeating words in spontaneous speechî, *Cognitive Psychology*, 37: 201-242, 1998.

[6] Wasow, T., ìRemarks on grammatical weightî, *Language Variation and Change*, 9: 81-105, 1997.

[7] Watanabe, M., ìAn analysis of usage of fillers in Japanese Lecture-style speechî, *Proceedings of the Spontaneous Speech Science and Technology Workshop*: 69-76, Tokyo, 2001.

[8] Minami, F., *Gendai Nihongo no Kouzou* (The structure of modern Japanese) Taisyukan syoten, Tokyo, 1974.

[9] Takanashi, K., Maruyama, T., Uchimoto, K., and Isahara, H., ìIdentification of ìsentencesî in spontaneous Japanese ñDetection and modification of clause boundariesñì, *Proc. of ISCA & IEEE workshop on spontaneous speech processing and recognition*, Tokyo, Japan: 183-186, 2003.

[10] Maekawa, K., ìCompilation of the Corpus of Spontaneous Japanese: A status Reportî, *Proceedings of the Second Spontaneous Speech Science and Technology Workshop*, Tokyo: 7-10, 2002.

[11] Watanabe, M., ìThe constituent complexity and types of fillers in Japaneseî, *the Proceedings of the 15th International Congress of Phonetic Sciences*: 2473-2476, Barcelona, Spain, 2003.

[12] Koiso, H., ìTranscription criteria for the Corpus of Spontaneous Japaneseî, *Proceedings of the Spontaneous Speech Science and Technology Workshop,* Tokyo: 13-20, 2001.

[13] Noda, H., ìFukubun ni okeru *wa* to *ga* no kakarikata (Scopes of *wa* and *ga* in complex sentences)î, *Nihongo-gaku* vol.5/2: 31-43, 1986.