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

ITRW on Disfluency in
Spontaneous Speech (DiSS'01)
Edinburgh, Scotland, UK
August 29-31, 2001

Are Word Repetitions Really Intended by the Speaker?

Yasuharu Den

Faculty of Letters
Chiba University
den@L.chiba-u.ac.jp

Abstract

This paper compares, using our Japanese data, word repetitions with error repairs in terms of their temporal structures in order to examine whether or not the prolongation of first tokens in word repetitions, observed by Den and Clark (2000), is really an effect of the speaker's strategy. Analyses of 10 task-oriented Japanese dialogues reveal a difference between word repetitions and error repairs for the data involving cut-off in first tokens; in both types of disfluencies, the final phoneme of the first token is considerably prolonged, but the degree of the prolongation is much greater in word repetitions than in error repairs. These results support our view that prolonged first tokens in word repetitions are a product of a process under the speaker's control or intention.

1. Introduction

Spontaneous speech contains various disfluencies such as fillers, self-repairs, and repeated words. These disfluencies seem to reflect problems in speech production. When speakers cannot formulate an entire utterance at once, or when they change their minds about what to say, they may suspend their speech and produce fillers or replace words they have already produced.

There are two views of speech disfluencies. One is that they are merely accidents which are beyond speakers' control. For example, speakers may suspend a word for some unexpected reason and restart it from the beginning. The other view is that they are the results of certain strategies under speakers' control (Levelt, 1989; Clark & Wasow, 1998). Speakers may produce a filler to signal to their addressees that they are having trouble in speech production, or to inform their addressees of the kind of trouble they have. Clark and Wasow (1998), taking this 'strategic' view, proposed the *commit-and-restore* model of repeated words in English, insisting that speakers can make a *preliminary commitment* at the beginning of a major constituent, even when they are aware of its not having been well-formulated, and restart the constituent, after suspension, to restore continuity to its delivery.

Den and Clark (2000) showed that the theory of preliminary commitment also applies to Japanese, that has completely different syntax from English. They found that pauses immediately following repetitions are less frequent than those immediately preceding repetitions and those between repeated tokens. They also found that the first tokens of repetitions are prolonged whereas the second tokens are produced in the same speed as the fluent speech. These findings suggest that Japanese speakers, as well as English speakers, sometimes use word repetitions as a linguistic device to communicate to addressees their cognitive states and that they produce these tokens with the intention of making the addressees recognize as such.

These pieces of evidence, however, are still insufficient. It might be the case that the prolongation of first tokens is merely a general characteristic of the disrupted speech, having nothing to do with the speaker's strategy. In order for the prolongation to be evidence of preliminary commitment, it should be observed only when the speaker's strategy can be relevant. That is, the prolongation of tokens should not be observed when the disrupted speech cannot be viewed as a marker for intended communication.

In this study, we compare, using our Japanese data, word repetitions with error repairs in terms of their temporal structures. In error repairs, first tokens are erroneous items, which are produced by accident and cannot be viewed as a marker for intended communication. Thus, if the prolongation of first tokens is observed in word repetitions but not in error repairs, that would form strong evidence for the theory of preliminary commitment.

2. Word Repetitions in Japanese

Word repetitions in Japanese have different characteristics from word repetitions in English. First, most word repetitions in Japanese are repetitions of content words (Den, Ishizaki, & Haruki, 1997), as opposed to function words, which are frequently repeated in English. In Japanese, repetitions of function words alone are very rare. Second, as a consequence of the first characteristic, Japanese has no typical lexical items for repetitions, like *the* or *a* in English. Third, in Japanese, first tokens in repetitions are frequently cut off in the middle, which is less frequent in English. In our Japanese data, repetitions involving word cut-off amount to over 60% of the data. This is mainly due to long durations of content words, which are frequently repeated in Japanese.

A typical example of word repetitions in Japanese is the following one:

(1)	ano	Ya=	Yamaguti-to	Hirosima	ari-masu-ka
	uh	Ya=	Yamaguchi-and	Hiroshima	be-POLITE-Q
	uh, are there Ya= Yamaguchi and Hiroshima?				

In (1), the speaker suspended the word *Yamaguti* (a place name) after the initial mora *Ya*, and then restarted it from the beginning, resulting in a word repetition *Ya= Yamaguti*.[1] First tokens in repetitions are frequently cut off in the middle, as in (1), but in a few cases, speakers produce an entire word before repeating it, like *Yamaguti Yamaguti*. Both cases will be considered in this study.

This research has been partly supported by CREST of JST (Japan Science and Technology Corporation).

[1] The symbol '=' is used to indicate a word cut-off.

Den and Clark (2000) tested with Japanese data two of the three hypotheses of the commit-and-restore model, which had been proposed by Clark and Wasow (1998) to account for word repetitions in English. The two hypotheses were

The continuity hypothesis: All other things being equal, speakers prefer to produce constituents with a continuous delivery.

The commitment hypothesis: Some initial commitments to constituents are preliminary, with speakers already expecting, at some level of processing, to suspend speaking immediately afterward.

Den and Clark's findings that pauses are less frequent after restarts than before or during suspension and that the first tokens of repetitions are prolonged were consistent with the continuity hypothesis and the commitment hypothesis, respectively. The second piece of evidence, however, is not sufficient for supporting the commitment hypothesis, as mentioned in the previous section. Re-examination of the hypotheses is the topic of the present study.

3. Method

Data. The corpus used in the present study was the same as that used in Den and Clark's (2000) study.

The corpus consists of 30 task-oriented dialogues by 60 different native speakers of Japanese (total running time = apprx. 7.5 hours, number of words = 73,258). The task was collaborative route finding, in which two participants looking at slightly different railroad maps, each unseen by the other, were asked to find a connecting path from a given start to a given goal. The two maps were different in that some connections on one map were missing on the other and in that some station names were missing on either map. The participants had to find a path which was available on both maps. This setting induced the participants to exchange spontaneously and naturally utterances about connections and station names. All dialogues were digitized on a computer and transcribed with part-of-speech and disfluency annotations.

For the present study, the first 10 of the 30 dialogues were selected. These included 11 male and 9 female speakers.

Since the purpose of the analysis is comparison of word repetitions with error repairs, these two types of disfluencies were taken into account. The data for word repetitions were the same as those used in Den and Clark's (2000) study. The data for error repairs were carefully chosen from among those annotated as 'substitution repairs.' Substitution repairs are repairs in which the first tokens are replaced by the second tokens that are closely related, in syntax and/or semantics, to the first tokens (Shriberg, 1994). Substitution occurs on at least two occasions; (i) when the first token is wrong in the context and the second token corrects it, and (ii) when the second token is more appropriate in the context than the first token, which is not necessarily wrong. The first type is called *error repairs*, and the second type *appropriateness repairs* (Levelt, 1983).

For the present study, only error repairs were considered. Speakers may use a less appropriate, but not incorrect, word, as preliminary commitment, to signal their addressees that they have trouble in speech production, and replace it afterward with a more appropriate word, resulting in an appropriateness repair. Since the purpose of comparing word repetitions with other type of disfluency is to see the difference between cases where the speaker's strategy can be relevant and cases where the speaker's strategy cannot be relevant, appropriateness repairs, which might be similar to word repetitions in terms of the speaker's strategy, are not suitable for the target of the comparison. Error repairs, on the other hand, are suitable for the target, for first tokens in error repairs are erroneous items and seem to irrelevant to the speaker's strategy.

ano	⟨**P1**⟩	Ya=	⟨**P2**⟩	Yamaguti	⟨**P3**⟩	to	Hirosima
uh		Ya=		Yamaguchi		and	Hiroshima
		T1		**T2**			

- Presence of a pause at Pi, longer than 60 msec.
- Duration of Ti, normalized by the mean of the durations of its fluent counterparts by the same speaker.

Figure 1: Measurement used in the analysis.

The following is a typical example of error repairs in Japanese:

(2) a Na= Morioka-mo tunagatte-masu
uh Na= Morioka-as well is reachable-POLITE
uh, Na= Morioka is reachable as well.

In (2), the speaker suspended the word *Naha* (a place name) after the initial mora *Na*, and then replaced it with the word *Morioka* (a place name), resulting in an error repair *Na= Morioka*. Like word repetitions, error repairs also involve frequent cut-off in first tokens. The distinction between cases with word cut-off and cases without word cut-off will be taken into account in the analyses below.

Measurement. In the same way as Den and Clark's (2000) study, we counted the number of pauses around word repetitions and error repairs, and the durations of repeated/repaired tokens. The precise measurement is illustrated in Figure 1.

First, we counted pauses just before first tokens (P1), between first and second tokens (P2), and just after second tokens (P3). At P1 and P3, only silent pauses were considered, whereas at P2, filled pauses and editing expressions were also regarded as pauses with the corresponding durations. Silences, or fillers, of 60 msec or longer were counted as pauses.

Second, we measured the normalized durations of first tokens (T1s) and second tokens (T2s). The normalization was needed to compensate for the differences of speaking rate across speakers and of inherent phoneme length across items. For each speaker and for each item, we collected from the corpus the fluent versions of the same word by the same speaker, where a 'fluent' version means an occurrence of the word not involved in a disfluency of any type. Then, we calculated the normalized duration as the duration of a target token divided by the mean of the durations of its fluent counterparts. When a target token was cut off in the middle, we used the duration of the corresponding region in its fluent counterpart for normalization. Note that by this normalization, the value of 1.0 indicates that the token was produced at the same speaking rate as its fluent counterpart. We also calculated the normalized durations for the final phonemes of T1s and T2s by identifying these phonemes based on spectral cues.

Exceptions. The following cases were excluded from the analysis (cf. Den & Clark, 2000):

1. Repetitions/Repairs involving verbs and/or auxiliary verbs. Repetitions of (auxiliary) verbs were excluded for

they are likely to be for emphasis rather than disfluencies. Repairs of (auxiliary) verbs were excluded as well to limit the target tokens to noun phrases.

2. Repetitions/Repairs overlapping/overlapped with the other's speech. These were excluded for it was difficult to perform precise phoneme alignment for overlapping/overlapped speech (the speech signals of the two participants in a dialogue were not recorded on separate channels in our corpus).
3. Repetitions/Repairs having no fluent counterparts, or repairs in which the intended words for first tokens that are cut off in the middle are not identifiable. These were excluded simply because the normalization procedure could not be applied.

Predictions. The predictions about the difference between word repetitions and error repairs in terms of the continuity and the commitment hypotheses are as follows:

The continuity hypothesis: In both types of disfluencies, pauses at P3 are less frequent than those at P1 and P2.

The commitment hypothesis: In word repetitions, the normalized duration of T1 is longer than that of T2, whereas in error repairs, the normalized duration of T1 is as long as that of T2. Or, even if the normalized duration of T1 is longer than that of T2 in error repairs, the degree of the prolongation is smaller than that in word repetitions.

According to the continuity hypothesis, speakers prefer to produce constituents with a continuous delivery. They don't like to add a delay before every word when they have trouble in formulating an entire phrase or utterance. Rather, they are likely to suspend speaking at some point in a constituent and restore continuity to their delivery of the constituent after they have formulated it well enough. This does not depend on the source of the suspension. Whether the suspension is intended by the speaker or it is by accident, the speech after restart would be produced with a continuous delivery. Thus, we can expect, in both word repetitions and error repairs, pauses after restarts to be less frequent than pauses before or during suspensions.

According to the commitment hypotheses, on the other hand, speakers can make a preliminary commitment at the beginning of a major constituent, even when they are aware of its not having been well-formulated. They do so, being pressed by a temporal imperative; by initiating a constituent, even if prematurely, to inform their addressees that they are engaged in planning the constituent, they can escape from being heard, due to a long delay, as opting out, as confused, or as having nothing immediately to contribute. Preliminary commitments, being a signal to addressees, are marked as such by some linguistic means, e.g., prolongation (Shriberg, 1999). This is a 'strategic' use by the speaker of disfluency, and that is the case with (some) word repetitions.

The situation, however, is completely different in error repairs. In error repairs, first tokens are produced by accident, having nothing to do with the speaker's strategy. Thus, there is no reason to expect first tokens to be linguistically marked as a signal to addressees. The prediction on the normalized duration of first and second tokens has two possibilities. If the prolongation of first tokens in word repetitions, observed by Den and Clark (2000), is purely an effect of the speaker's strategy, it will not be observed in error repairs. Or, if the prolongation of first tokens observed in word repetitions is a combined effect of the speaker's strategy and of the phonological disturbance due to the disrupted speech, the prolongation will also be observed in error repairs, due to disruption, but the degree of the prolongation will be smaller than that in word repetitions.

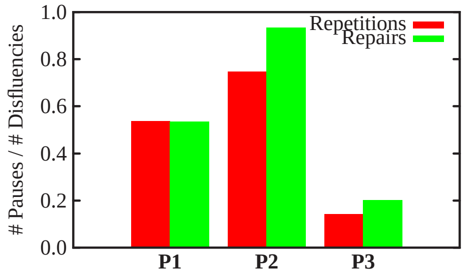

Figure 2: Frequencies of pauses at P1, P2, and P3. ($N = 43$ for repetitions and $N = 15$ for repairs)

4. Results

Pauses at P1, P2, and P3. The frequencies of pauses at P1, P2, and P3 are shown in Figure 2. For those cases where repetitions/repairs occur at the inside of an utterance, i.e., not at the beginning or the end of an utterance, a Cochran's Q test was applied separately for the repetition and the repair data.

For word repetitions, the numbers of pauses at the three locations were significantly different ($Q = 30.76$, $p < .001$). Multiple comparisons using the Ryan's procedure showed that pauses were significantly less frequent at P3 than both at P1 and at P2 ($ps < .001$). No significant difference was found between the numbers of pauses at P1 and at P2.

For error repairs, the numbers of pauses at the three locations were significantly different as well ($Q = 15.17$, $p < .001$). Multiple comparisons showed that pauses at P3 were significantly less frequent than pauses at P2 ($p < .01$), and that pauses at P1 were significantly less frequent than pauses at P2 ($p < .05$). No significant difference was found between the numbers of pauses at P1 and at P3.

Although the results for word repetitions and error repairs were slightly different, i.e., pauses were less frequent at P3 than both at P1 and at P2 in repetitions but than only at P2 in repairs, this would be due to the small data size for error repairs. The distributions of pauses for repetitions and repairs shown in Figure 2 are quite similar.

In any case, pauses after restarts are less frequent than those (before or) during suspensions, supporting the prediction from the continuity hypothesis.

Normalized Durations of T1 and T2. The mean normalized durations of T1s and T2s for word repetitions and error repairs are shown in Figure 3.

A two-factors ANOVA was applied to the data. The main effect of the token position (T1 vs. T2) was significant ($F(1, 75) = 19.36$, $p < .001$), but the main effect of the disfluency type (Repetitions vs. Repairs) nor the interaction between the two factors were not significant ($Fs < 1$). In both types, T2 had approximately the same duration as its fluent version (mean norm. dur. = 1.04 for repetitions and 1.02 for repairs), and T1 was considerable prolonged (mean norm. dur. = 1.60 for repetitions and 1.40 for repairs). The results for the final phonemes of T1s and T2s were similar.

These results seem to contradict with our prediction on the difference between word repetitions and error repairs in terms of the commitment hypothesis. The prolongation of first tokens

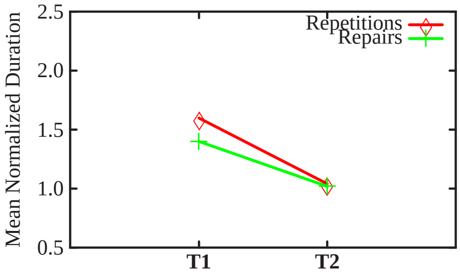

Figure 3: Mean normalized durations of T1s and T2s. ($N = 60$ for repetitions and $N = 17$ for repairs)

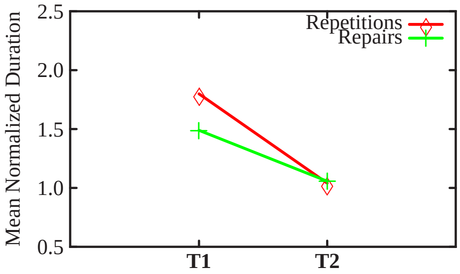

Figure 4: Mean normalized durations of T1s and T2s for the data involving cut-off. ($N = 38$ for repetitions and $N = 13$ for repairs)

was not peculiar to word repetitions, nor no additional prolongation was observed in word repetitions. However, when we focused only on the data involving cut-off in first tokens, the story changed dramatically.

Figure 4 shows the mean normalized durations of T1s and T2s for word repetitions and error repairs involving word cut-off, and Figure 5 shows the same data for the final phonemes of T1s and T2s. Although an ANOVA on the data shown in Figure 4 revealed only a significant main effect of the token position (token position: $F(1, 49) = 17.35$, $p < .001$; disfluency type: $F(1, 49) = 1.02$, $p = .32$; interaction: $F(1, 49) = 1.33$, $p = .25$), an analysis on the data shown in Figure 5 revealed a significant interaction between the token position and the disfluency type (token position: $F(1, 48) = 28.74$, $p < .001$; disfluency type: $F(1, 48) = 1.15$, $p = .29$; interaction: $F(1, 48) = 5.36$, $p < .05$). The final phoneme of T2 had approximately the same duration as its fluent version in both word repetitions (mean norm. dur. = 1.03) and error repairs (mean norm. dur. = 1.15), whereas the final phoneme of T1 was considerably prolonged and the degree of the prolongation was much greater in word repetitions (mean norm. dur. = 2.08) than in error repairs (mean norm. dur. = 1.56).

5. Discussion

In this paper, we have compared word repetitions with error repairs in terms of their temporal structures in order to examine whether or not the prolongation of first tokens in word repetitions, observed by Den and Clark (2000), is really an effect of the speaker's strategy. We have found that when we focus on the cases where first tokens are cut off in the middle, word repetitions and error repairs have different characteristics. Although in both types of disfluencies, the final phonemes of

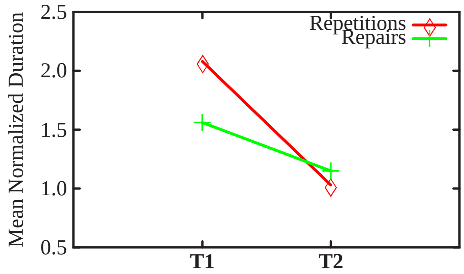

Figure 5: Mean normalized durations of the final phonemes of T1s and T2s for the data involving word cut-off. ($N = 38$ for repetitions and $N = 12$ for repairs)

the first tokens are prolonged, the degree of the prolongation is much greater in word repetitions than in error repairs. This means that, even if the prolongation can be partly attributed to the phonological disturbance due to the disrupted speech, something more is happening in word repetitions. Speakers sometimes use word repetitions as a linguistic device to signal to their addressees that they are having trouble in speech production, and produce these tokens with the intention of making the addressees recognize them as a signal.

The results of the present study seem to support our view that prolonged first tokens in word repetitions are a product of a process under the speaker's control or intention. Speakers intendedly use them as preliminary commitments in order to announce their engagement in the production of the following material, making an otherwise intolerable delay in the delivery of constituents permissible by addressees.

There are, however, still several points to be accounted for. The difference between word repetitions and error repairs was observed only in the cases involving word cut-off and only at the final phonemes of first tokens. Why the difference was not observed in other cases might be due partly to the small data size for error repairs used in the study. However, we need more detailed examination of the data and comparison with other types of disfluencies as well as disfluencies in other languages.

6. References

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, *37*, 201–242.

Den, Y., & Clark, H. H. (2000). Word repetitions in Japanese spontaneous speech. In *Proceedings of the 6th International Conference on Spoken Language Processing* (pp. 58–61). Beijing.

Den, Y., Ishizaki, M., & Haruki, Y. (1997). A corpus-based analysis of speech repairs in Japanese. Oral presentation at *Computational Psycholinguistics 1997*. Berkeley, CA.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Shriberg, E. E. (1994). *Preliminaries to a theory of speech disfluencies*. Unpublished doctoral dissertation, University of California, Berkeley.

Shriberg, E. E. (1999). Phonetic consequences of speech disfluency. In *Proceedings of the 14th International Congress on Phonetic Science* (pp. 619–622). San Francisco.