Dimensions of Variation in Disfluency Production in Discourse

Scott H. Fraundorf
University of Illinois at Urbana-Champaign
603 E Daniel St.
Champaign, IL 61820 USA
sfraund2@uiuc.edu

Duane G. Watson
University of Illinois at Urbana-Champaign
603 E Daniel St.
Champaign, IL 61820 USA
dgwatson@uiuc.edu

Abstract

This study demonstrates that four common types of disfluency in discourse (fillers, silent pauses, repairs, and repeated words) differ from one another on two dimensions related to language production processes: their temporal relation to speech production problems and the level of production at which those problems occurred. Participants' speech in a storytelling paradigm was coded for the four disfluency types. Comparisons between types in their relation to story events, to clause boundaries, to utterance length, to utterance position, and to other disfluencies suggest the four types reflect different difficulties in language production. Temporally, fillers, silent pauses, and repeats represent difficulties in upcoming speech, while repairs represent past difficulties. Fillers were most associated with discourse-level problems, while silent pauses were more associated with grammatical and phonological difficulty.

1 Introduction

Human speech is fraught with interruptions, or *disfluencies.* Although several types of disfluency occur in speech, the ways in which these types differ from one another have not been well defined. In this paper, we propose that disfluency types systematically differ along at least two dimensions: (a) their temporal relationship to the underlying production difficulty and (b) the level of production at which the difficulty occurs.

Precise taxonomies of disfluencies vary, but most are derived from the four categories proposed by Maclay and Osgood (1959). *Fillers*, as in (1) below, are verbal interruptions that do not relate to the proposition of the main message—in English, most commonly *uh* and *um*. *Silent pauses*, as in (2), are periods of silence longer than the pauses that would be produced in an equivalent fluent utterance. *Repeats*, as in (3), are unmodified repetitions of a word or of a string of words. Finally, *repairs* are self-corrections or revisions of material already spoken. Repairs such as (4), called *error repairs* by Levelt (1983), simply correct errors of linguistic form. Repairs like (5), called *appropriateness repairs* and *message repairs* by Levelt, present a new or rephrased message.

(1) She grabs the fan and **uh** one pair of gloves.
(2) She sees **...** a small **...** box saying "EAT ME."
(3) Alice doesn't think **that cats that cats** grin.
(4) The cake **make Alices makes Alice** grow.
(5) And they sent Bill the lizard down the chimney **to find her er to see what was going on**.

Because these four types of disfluencies obey different distributional patterns and the frequency of use of each type correlates only weakly with that of other types, Maclay and Osgood argued that different types of disfluency represent different production problems or different strategies for correcting problems. But since this proposal, differences between disfluency types have received little examination. Experimental studies have often examined single types of disfluency without comparison across categories. Further exploration of the differences between disfluency types is necessary because many psycholinguistic studies have used disfluency to study language production (e.g. Levelt, 1983) or comprehension (e.g. Arnold et al., 2003; Ferreira and Bailey, 2004; Fox Tree, 1995). Without a generalized theory of the relationship between production and the various types of disfluency, it is not clear how well findings regarding a single type of disfluency generalize to others.

The present study investigates the differences between fillers, silent pauses, repairs, and repeats in an extended discourse. We argue that some of the differences in distribution between these disfluency types can be understood by considering them in the context of a model of the language production system. Most models of language production (for review, see Bock, 1995) posit at least three cascaded levels: a message level representing preverbal meaning, a grammatical level at which lexical items are selected and assembled into a morphosyntactic structure, and a phonological level at which an utterance's overall prosody and the phonological encodings of individual words are constructed. We provide evidence that disfluency types differ on at least two dimensions related to this system : (a) the temporal relation of the underlying problem to the current state of the production system, and (b) the level (or stage) of production at which the underlying problem occurred.

One dimension on which disfluency types may vary is when they occur relative to the underlying production difficulty that caused them. Some disfluencies may occur in response to a problem detected in already produced speech, while others may reflect problems in speech being planned. It is generally accepted that overt repairs are used when there is a problem in speech that was already produced (e.g., Levelt, 1983). Conversely, fillers have frequently been argued to reflect delays in planning or encoding of upcoming material (e.g. Arnold et al., 2003; Clark and Fox Tree, 2002), as have silent pauses (Butterworth, 1980). However, it has not been explicitly tested whether the distribution of these disfluencies differs from that of repairs.

The temporal properties of repeats are a matter of controversy. Clark and Wasow (1998) argue that repeats also reflect delays in planning and describe a *commit-and-repair* strategy: speakers commit to a partially planned utterance and, if planning delays prevent its initial fluent completion, they repeat the beginning so that the entire utterance can still be presented fluently. This theory predicts that the words repeated most often should be those likely to be produced during these early commitments, and Clark and Wasow find that function words, which tend to begin major constituents, are indeed repeated more often than content words. However, Levelt (1983) suggests that some repeats may actually result from "false alarms" of production monitoring systems. When the repair process is erroneously initiated in response to an acceptable utterance, the material in question ends up being reproduced without change, resulting in a repeat. This theory predicts that most repeats should share more properties with repairs. Since it is unknown whether either or both of these mechanisms underlie repeats, the present study examined the temporal properties of repeats as well.

In addition to their temporal relation to a production problem, disfluency types may also vary on a second dimension: the level of production at which the underlying problem occurs. If production involves a series of stages, as reviewed above, it is likely that errors and delays can occur at all the levels. Problems at different levels may give rise to disfluencies differing in form and time course.

Fillers may be particularly associated with delays in message-level planning because speakers may use them as deliberate linguistic signals of difficulty in their planning speech (Clark & Fox Tree, 2002). If fillers are indeed used deliberately, then they should require message-level planning. Such message-level revision should be easy when the difficulty arose on the message level, but may be more difficult when information about grammatical and phonological difficulties must first be sent back to the message level. Thus, fillers should be particularly apt to arise from delays in conceptualization rather than problems with grammatical or phonological planning. Evidence from the literature suggests that message-level planning can indeed play a role in filler production. Swerts (1998) observed that fillers occur more at stronger discourse boundaries than at weaker ones, because more planning is required to determine the next message. Similarly, speakers produce more fillers when answering questions about which they are less certain (Smith and Clark, 1993). These findings have been interpreted as indicating that planning demands associated with new or difficult topics result in a higher rate of fillers.

While fillers may be most apt to arise from message-level difficulties, silent pauses and repeats may arise from problems at all levels. Information about problems at the grammatical and phonological levels may not easily reach the message level to produce a revision or signal of difficulty for listeners. When the production system cannot easily produce any overt message-level signal of difficulty, then delays in production would instead be manifested as a silent pause or repeat. Again, evidence from the literature is inconclusive, but suggests that grammatical and phonological factors can play a role in silent pause production. Maclay and Osgood (1959) observed that, while fillers usually occur between phrase boundaries, silent pauses tend to occur *within* phrases. Because the unit of message level planning has been argued to be at least an entire phrase (e.g., Garrett, 1988), phrase-internal disfluencies may reflect mostly delays in grammatical and phonological planning processes such as lexical or phonological retrieval. However, Reynolds and Paivio (1959) argue, based on effects of noun concreteness on silent pause production, that silent pauses may operate on the message planning level as well. These hypotheses have not been directly compared.

2 Present Work

Two hypotheses about disfluencies have been proposed: (a) disfluency types differ in their temporal relation to the underlying production difficulty, and (b) disfluency types differ in the level of production at which the difficulty occurred. The present work examines whether these hypothesized dimensions reflect actual differences in disfluency form and distribution during language production, and where specific disfluency types fall on these axes.

On the temporal relation dimension, it was hypothesized that fillers and silent pauses reflect problems with upcoming speech. Repairs were expected to reflect prior problems, by definition. Two competing hypotheses regarding repeats were also compared: the commit-and-repair theory predicts repeats to be more associated with upcoming problems, while the false-alarm theory predicts repeats to be more associated with prior problems. On the level-of-production dimension, it was hypothesized that fillers and appropriateness repairs usually reflect problems at the message level, while silent pauses and repairs correcting errors of form usually reflect problems at the grammatical or phonological level.

These questions were investigated in the context of an extended monologue to provide both a naturalistic situation and a discourse context for examining potential message-level effects on disfluency. Language may be produced quite differently in an extended monologue, yet disfluencies have rarely been examined experimentally in these situations. Conversely, disfluency use has sometimes been examined via corpus studies, but these observational studies lack the controls available in experimental work. Investigation of disfluencies in the context of an extended monologue provided a balance between naturalism and experimental control.

A storytelling paradigm was used in which participants were presented with stories and retold them to audiotape. The stories were three passages from *Alice's Adventures in Wonderland* (Carroll, 1865). Each passage was centered around a set of fourteen key points. Each key point was either a single action or two related actions crucial to the plot of the passage, such as *Alice finds a cake marked "EAT ME."*

3 Method

Ten University of Illinois undergraduates participated for course credit. All were native speakers of English between the ages of 18 and 22.

Participants read three passages, each approximately 2000 words, excerpted from *Alice's Adventures in Wonderland* (Carroll, 1865). Each passage was chosen to represent a distinct section of the plot that involved a number of discrete actions and that had a specific beginning and end. The three passages involved Alice getting trapped in a cave, Alice visiting the White Rabbit's house, and Alice meeting the Duchess.

Each participant read all three stories, presented in randomized order. For each passage, the participant first read the printed copy of the passage. Participants were told to read at their preferred speed and not to memorize all the events of the story since they would be receiving a list of key points to include. After reading the full story, the printed story was taken away and participants received the list of fourteen key points, printed in bullet-point format on a separate sheet. When participants indicated they were ready, the experimenter turned on a digital recorder and recorded the participant telling the story. Participants could consult the list of key points while speaking but were required to retell the story in their own words. Each recording continued until the speaker indicated to the experimenter that he or she was finished. The participant then repeated the process with the next story.

3.1 Transcription

The first author transcribed each retelling from the recordings. The transcripts were then scored for the beginning and end of each of four types of disfluency. A *filler* was any use of *uh*, *um*, *ah*, or *er*; in the uncommon case of several fillers in a row, each was coded as a separate instance. *Silent pauses* were the perception of a disfluent gap in the fluent speech stream, based on the speaker's typical speech rate and the surrounding prosodic context. *Repeats* were one or more repetitions without modification of the same word, part of word, or string of words. *Repairs* were mid-utterance alterations of material already produced, including abandonment of the entire utterance (sometimes termed a *fresh start*; e.g., Bear et al., 1993). For reliability, a second observer also coded all the recordings for silent pauses. Only silent pauses coded by both observers were included in the final analysis.

Repairs were then subcategorized either as *error repairs*, which corrected identifiable lexical, phonetic, or syntactic errors, or as *appropriateness repairs*, which involved either a rewording of the same concept, the addition of a previously unstated fact, or the correction of a previous factual error. To assess reliability of these subcategorizations, a second observer, blind to the experimental hypotheses, scored all the repairs. Agreement between the two observers was good ($\kappa = .75$); where the observers disagreed, the first author's ratings were used.

Transcripts were also coded for the beginning and end of each key point. The beginning of a key point was coded at the first phrase introducing a fact from the printed bullet-point list. The key point continued through the last phrase regarding that bullet-point, at which point the end was coded. Typically, each key point was then followed by additional elaboration of the events or by explanations of how the key points related to each other in time within the story (e.g., "This went on for some time before..."). Such elaboration was not coded as part of the key point.

4 Initial Analyses

Participants were successful in retelling the stories, including a mean of 13.33 of the 14 key points ($SE = 0.84$) per passage.

Participants were also frequently disfluent. Collapsing across disfluency types, a mean ratio of 6.55 disfluencies per 100 words was observed. This ratio is close to past estimates of 6 disfluencies per 100 words (Fox Tree, 1995) suggesting that the present task yielded a typical sample of disfluencies. However, because the number of words spoken differed between participants, the raw frequency of disfluency is confounded with the total amount of speech. Consequently, we calculated the ratio of disfluency in proportion to the number of words.

5 Temporal Relation Analyses

5.1 Relation to Difficult Material

It was hypothesized that disfluency types would differ in their distribution relative to the demands of new key points. New key points were expected to be especially difficult at multiple levels of production: they introduce new plot elements to the story that may require additional discourse-level planning, and they often require access of new lexical, syntactic, and phonological forms. Thus, it was expected that fillers and silent pauses, hypothesized to reflect trouble with upcoming speech, should be more common before new key points than elsewhere. Since repairs reflect prior problems, however, they should be more common *after* new key points.

To test this hypothesis, transcripts were divided into four regions based on the key point codings (see section 3.1). The *Within* region comprised all words inside a key point. The *Before* region included the three words immediately before each new key point was introduced. The *After* region included the three words immediately after the end of each new key point. The *Between* region included all the words not inside or within three words of the introduction of a new key point. Mean rates of each type of disfluency in each region are presented in Table 1.

Type	Between	Within	Before	After
Filler	2.02	1.18	5.19	2.87
Silent Pause	2.47	1.81	5.18	4.15
Repair	1.24	1.16	1.26	2.48
Repeat	0.80	0.74	0.62	1.24

Table 1. Rate of disfluency per 100 words by location relative to key points.

A planned comparison indicated that, as predicted, fillers occurred at a higher rate in the Before region ($M = 5.19$ per 100 words) than in the three other regions ($M = 2.02$), $F_1(1,9) = 16.56$, $p < .001$, 97.5% CI of the difference.[1] = ± 1.85. Silent pauses were also more prevalent before new key points ($M = 5.18$) than elsewhere ($M = 2.81$), $F_1(1,9) = 11.22$, $p < .01$, 97.5% CI of the difference = ± 1.68.

[1] Because a second comparison (see section 6.1) was also conducted for the rate of each disfluency type, Bonferroni correction was applied to the 95% confidence intervals to avoid compounding the Type I error rate.

Repairs, as predicted, occurred at a significantly higher rate in the After region (M = 2.48 per 100 words) than in the three other regions (M = 1.22), $F_1(1,9) = 6.55$, $p < .05$, 97.5% CI of the difference = ± 1.18. Repairs were *not* more prevalent in the Before region (M = 1.24) than in the other regions (M = 1.63), $F_1(1,9) = 0.61$, $p = .44$, 97.5% CI of the difference = ± 1.11. These differences in distribution are consistent with the hypothesis that repairs reflect problems in prior speech but fillers and silent pauses reflect problems in speech being planned.

For repeats, no overall effect of region was observed, $F_1(3,9) = 0.80$, $p = .51$. Recall, however, that the commit-and-repair theory posits that repeats should be most common when a speaker has just begun production of problematic material. This theory predicts that repeats should be most common immediately after the start of key points, not immediately before. Consequently, an additional *Beginning* region was created, comprising the first three words after the beginning of each new key point. A planned comparison revealed that repeats occurred at a marginally significantly higher rate in the Beginning region (M = 1.29 per 100 words) than elsewhere (M = 0.77), $F_1(1,9) = 4.04$, $p = .10$, 97.5% CI of the difference = ± 0.62, supporting the commit-and-repair theory. These results suggest that while repeats, fillers, and silent pauses all relate to planning difficulties, they may differ in time course relative to this difficulty: fillers and silent pauses tend to reflect upcoming difficulty, while repairs may reflect more immediate difficulty.

5.2 Relation to Other Disfluencies

If fillers and silent pauses reflect upcoming problems, they should be more apt to occur *before* repairs, which reflect past problems, than *after* repairs. This pattern should hold even within a single utterance that does not cross key point boundaries, and provides an additional test of the temporal location dimension. To test this prediction, all utterances containing at least one repair were examined for other disfluencies within the same sentence. The rate of disfluency in the portion of the sentence before the *reparandum*—the problem being repaired—was compared to the rate of disfluency in the portion of the sentence after the conclusion of the repair. Rates of disfluency before and after repairs are presented in Table 2.

Within a sentence, the rate of fillers was significantly greater before repairs than afterwards,

Type	Before Repairs	After Repairs
Filler	2.44	1.81
Silent Pause	3.75	2.40
Repeat	0.79	0.62

Table 2. Rate of disfluency before the first repair and after the last repair within a sentence.

$t_1(9) = 3.15$, $p < .05$, as predicted. Silent pauses were also more common before repairs than after them, $t_1(9) = 3.29$, $p < .01$. Repeats were only marginally more prevalent before repairs, $t_1(9) = 2.00$, $p = .08$.

6 Level of Production Analyses

6.1 Clauses and Disfluency

Fillers and silent pauses were both observed to occur more frequently before new key points. Because new key points almost always begin new clauses, it is possible that this distribution simply reflects the grammatical and phonological planning demands of new clauses. Alternatively, the message-level and discursive demands associated with a new story event may create an additional burden on the production system beyond the effect of a new clause.

These hypotheses can be tested by comparing the Before region with another region that contains clause boundaries but does not precede new key points. The After region was expected to also contain clause boundaries because the end of a key point generally represented the boundary between the introduction of a point and its elaboration. A paired samples t-test confirmed that, by participants, the prevalence of clause boundaries did not significantly differ between the beginning and end of key points, $t_1(9) = 1.868$, $p = .10$. By participants, 97.30% of Before (SE = 2.40%) regions contained a clause boundary and 95.07% of After regions (SE = 3.08%) contained a clause boundary.

Because both the Before and After regions contain clause boundaries, they were used to compare the introduction of new key points to a clause boundary baseline. A planned comparison by key points revealed that fillers were significantly more common in the Before region (M = 5.19 per 100 words) than in the After region (M = 2.87), 97.5% CI of the difference = ± 2.26. However, no significant difference in the rate of silent pauses was found between the Before region ($M = 5.18$) and the After region ($M = 4.15$), 97.5% CI of the difference = ± 2.05.

These data suggest that planning new topics or points within a discourse may be especially difficult and may lead to more fillers than planning other clauses. However, since silent pauses occurred equally frequently at all clause boundaries, they may be more associated with the grammatical planning that should occur at all clause boundaries.

6.2 Distribution within Utterances

Because language production is incremental, planning at the grammatical and phonological levels continues until nearly the end of an utterance (Butterworth, 1980). Semantic and discourse-level planning, on the other hand, is thought to represent an earlier level of production that is completed sooner than grammatical and phonological planning (e.g. Bock, 1995; Butterworth, 1980). Thus, disfluencies reflecting message-level difficulties should be more prevalent early in an utterance, before message-level planning has been completed, whereas grammatical and phonological difficulties should be found throughout an utterance.

To test this hypothesis, a simple division of early and late locations within an utterance was constructed by dividing each utterance in half according to the number of words. Table 3 presents mean rates of disfluency for the first half of an utterance versus the second half. (Repairs and repeats that spanned the midpoint of an utterance were excluded from this analysis.)

Type	First Half	Second Half
Filler	3.19	1.24
Silent Pause	3.76	1.50
Appropriateness Repair	0.52	0.21
Error Repair	0.34	0.24
Repeat	0.82	0.55

Table 3. Rate of disfluency per 100 words by location with an utterance.

As predicted, fillers were significantly more common in the first half of an utterance than in the second half, by participants, $t_1(9) = 3.48$, $p < .01$, as were appropriateness repairs, $t_1(9) = 3.18$, $p < .05$. This suggests that both fillers and appropriateness repairs are associated with the message level. Also as predicted, the rate of error repairs did not significantly differ between the first half and second half, $t_1(9) = 1.54$, $p = .16$, consistent with the hypothesis that these repairs reflect grammatical and phonological processes that continue throughout utterance production. The frequency of repeats also did not significantly differ between the first half and second half, $t_1(9) = 1.68$, $p = .13$. Since silent pauses were hypothesized to be most associated with the grammatical and phonological levels, they were also expected to be equally prevalent in both halves of the utterance. Contrary to this expectation, however, silent pauses were significantly more common in the first half of utterances, $t_1(9) = 4.52$, $p < .01$.

6.3 Correlation between Types

A speaker having difficulty at a particular level is likely to produce many disfluencies associated with that level over the course of the task. Thus, it was expected that a speaker's overall rate of use of a particular disfluency type would correlate with the rate of use of other disfluency types that stem from problems at the same level. Pearson's correlation coefficients correlated filler and silent pauses with both types of repairs. Participants' overall rate of filler use was significantly positively correlated with the rate of appropriateness repairs, $r = .70$, $p < .05$, Bonferroni corrected, but not with that of error repairs, $r = .44$, $p = .42$, supporting the hypothesis that fillers are associated with the message level. As expected, silent pauses were uncorrelated with the frequency of appropriateness repairs, $r = .13$, $p > .99$, Bonferroni corrected, but contrary to expectations, silent pauses were also uncorrelated with the frequency of error repairs, $r = .37$, $p = .60$.

6.4 Utterance Length

It was expected that longer utterances would place greater demands on grammatical and phonological planning but not on message planning. Butterworth (1980) has argued that lexical retrieval takes time and can easily result in disfluency. The number of content words (nouns, verbs, adjectives, and adverbs) may thus index an utterance's grammatical and phonological planning demands. (Function words such as prepositions and determiners have frequently been argued to be placed by fast syntactic processes unlikely to result in disfluency; e.g., Butterworth, 1980). Thus, it was expected that silent pauses and error repairs, both hypothesized to reflect grammatical and phonological difficulties, would occur at a greater rate in utterances with more content words. However, utterance length does not necessarily index demands at the message level. The *Alice's Adventures in Wonderland* passages contain a number of semantically anomalous events such as a baby turning into a pig. Although conceptually difficult, these events can be expressed with lexically and syntactically simple utterances (e.g., "The baby turned into a pig.") that do not place great demands on grammatical and phonological planning. Since an utterance's length is likely to be less related to its difficulty at the message level, the rate of message-level disfluencies should be less related to utterance length.

A regression was conducted to determine whether greater grammatical and phonological planning demands, as reflected in utterance length, predicted rate of disfluency. Controlling for participants and number of function words, the number of content words was a significant predictor of the rate of silent pauses (standardized $\beta = .23$, $r = .11$, $t(576) = 2.75$, $p < .01$) and of error repairs ($\beta = .18$, $r = .09$, $t(576) = 2.09$, $p < .05$), but not of the rate of appropriateness repairs ($\beta = .10$, $r = .05$, $t(576) = 1.20$, $p = .23$), of fillers ($\beta = -.01$, $r < .01$, $t(576) = 0.14$, $p = .89$), or of repeats ($\beta = -.01$, $r < .01$, $t(576) = -0.10$, $p = .92$). These results support the hypothesis that silent pauses and error repairs tend to reflect difficulties at the grammatical and phonological level, whereas fillers and appropriateness repairs reflect difficulties at the message level and are less affected by increased grammatical and phonological planning demands.

7 Discussion

Observation of the distribution of disfluencies in a storytelling task supported both proposed dimensions. On the temporal relation dimension, fillers and silent pauses were found to be associated with problems in upcoming speech, repeats with more immediate upcoming difficulty, and repairs with past problems. On the level of production dimension, fillers and appropriateness repairs were found to be most associated with the message level, while error repairs were most associated with the grammatical and phonological level. Results generally suggest that silent pauses are associated with grammatical and phonological problems, though analysis of the position of silent pauses within an utterance also suggests possible message-level influences on silent pause production. Repeats were not found to be clearly associated with one level or another; it is possible that they can function at all levels. These results suggest that the distribution of disfluency types could be diagrammed in a model containing at least two dimensions, as in Figure 1.

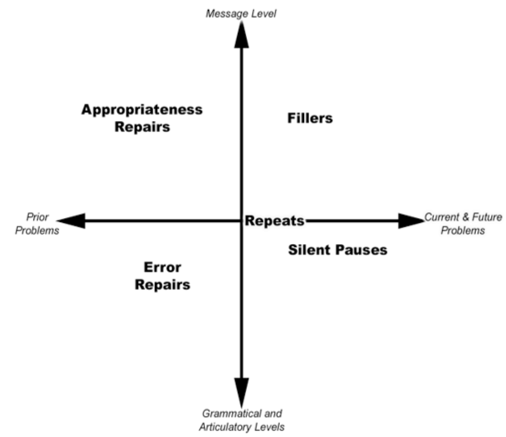

Figure 1. Schematic representation of the location of some disfluency types on dimensions of temporal location and level of production.

These findings suggest that current psycholinguistic work on language production provide a framework for understanding some aspects of disfluency. Of course, disfluency types may differ on other dimensions as well; in particular, some disfluencies may serve conversational purposes not captured by the present monologue task. For instance, fillers like *uh* and *um* have been argued to perform conversational functions like indicating a dispreferred response (e.g., Schegloff, 2006).

Nevertheless, the effects reported here of the message level of language production on disfluency are particularly important because message-level influences on disfluency are less frequently investigated. Prior findings that fillers occur more before objects with infrequent names (e.g. Schnadt and Corley, 2006) have often been interpreted as revealing an association between lexical access and filler production. However, uncommon objects differ from common ones in conceptual frequency as well as lexical frequency. Thus, prior studies confound the grammatical-level factor of lexical frequency with message-level semantic factors. The present study suggests that filler use may actually be more associated with message-level difficulty.

Why might fillers be most apt to arise from message-level difficulties? Recall that Clark and Fox Tree (2002) argue that speakers deliberately produce fillers to communicate the fact that they are having difficulty in production. Speakers may desire to signal their difficulty for a number of reasons, such as self-presentation (Smith & Clark, 1993) or “holding the conversational floor” and preventing an interlocutor from beginning a turn (Maclay & Osgood, 1959). All these purposes should require message-level planning.

The distributions of disfluency observed in the present study also bear on the debate about what mechanisms underlie repeated words. Recall that Levelt (1983) has argued that repeats arise when the repair system is mistakenly activated and ends up reproducing the original utterance unmodified. Alternately, Clark and Wasow (1998) argue that repeats are part of a commit-and-repair strategy used by speakers to allow fluent delivery of an utterance that encountered planning problems early in delivery. In the present study, repeats occurred most frequently at the beginning of a key point. The commit-and-repair strategy predicts this pattern but the false alarm hypothesis makes no *a priori* prediction that repeats should follow such a distribution. Thus, the present data suggest the commit-and-repair theory may best describe most repeats.

7.1 Conclusion

Disfluency is not a unitary phenomenon. Speech in a discourse is subject to several types of disruptions, which can represent different problems and different responses from the production system. The present work demonstrates that fillers, silent pauses, repairs, and repeats differ on two dimensions related to language production: their temporal relations to the problem that caused them and in the level of production with which they are associated. These dimensions provide a framework by which the differences between various kinds of disfluencies can be captured in future psycholinguistic work on disfluency.

Acknowledgement

Scott H. Fraundorf was supported by National Science Foundation Graduate Research Fellowship 2007053221.

References

Jennifer E. Arnold, Maria Fagnano, and Michael K. Tanenhaus. 2003. Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research*, 32(1): 25-36.

John Bear, John Dowding, Elizabeth Shriberg, and Patti Price. 1993. A system for labeling self-repairs in speech. SRI AI Center Tech Note #522.

Kathryn Bock. 1995. Sentence production: From mind to mouth. In Joanne L. Miller and Peter D. Eimas (Eds.), *Handbook of perception and cognition. Vol. 11: Speech, language, and communication* (pp. 181-216). Academic Press, Orlando, FL.

Brian Butterworth. 1980. Evidence from pauses in speech. In Brian Butterworth (Ed.), *Language Production, vol. 1: Speech and* talk (pp. 155-176). Academic Press: London, UK.

Lewis Carroll. 1865. *Alice's Adventures in Wonderland.* Retrieved September 15, 2006, from http://www.gutenberg.org/etext/11

Herbert H. Clark and Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition*, 84(1):73-111.

Herbert H. Clark and Thomas Wasow. 1998. Repeating words in spontaneous speech. *Cognitive Psychology*, 37(3):201-242.

Fernanda Ferreira and Karl G. D. Bailey. 2004. Disfluencies and human language comprehension. *TRENDS in Cognitive Science*, 8(5):231-237.

Jean E. Fox Tree. 1995. The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, 34(6): 709-738.

Merrill F. Garrett. 1988. Processes in language production. In Frederick J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey: Vol 3. Language: Psychological and biological aspects* (pp 69-96). Cambridge University Press, Cambridge, UK.

Willem J. M. Levelt. 1983. Monitoring and self-repair in speech. *Cognition,* 14(1):41-104.

Howard Maclay and Charles E. Osgood. 1959. Hesitation phenomena in spontaneous speech. *Word*, 14(1):19-44.

Allan Reynolds and Allan Paivio. 1968. Cognitive and emotional determinants of speech. *Canadian Journal of Psychology*, 22(3):164-175.

Emanuel A. Schegloff. 2006. On “uh” and “uhm” and some of the things they are used to do. Paper presented at BRANDIAL 2006, September 11-13, 2006, Potsdam, Germany.

Michael J. Schnadt and Martin Corley. 2006. The influence of lexical, conceptual and planning based factors on disfluency production. In *Proceedings of the twenty-eighth meeting of the Cognitive Science Society.*

Vicki L. Smith and Herbert H. Clark. 1993. On the course of answering questions. *Journal of Memory and Language*, 32(1):25-38.

Marc Swerts. 1998. Filled pauses as markers of discourse structure. *Journal of Pragmatics*, 30(4):485-496.