

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**  
Progress Report No. 28 (2007)  
*Indiana University*

**Frequency of Use Leads to Automaticity of Production:  
Evidence from Repair in Conversation<sup>1</sup>**

**Vsevolod Kapatsinski**

*Speech Research Laboratory  
Department of Psychological and Brain Sciences  
Indiana University  
Bloomington, Indiana 47405*

---

<sup>1</sup> I would like to thank the NIH for financial support through Training Grant DC-00012 and Research Grant DC-00111 and to Adam Albright, Tessa Bent, Joan Bybee, Dan Jurafsky and the audiences at HDLS VII and the Workshop on Gradience and Frequency Effects in Phonology for helpful comments.

## **Frequency of Use Leads to Automaticity of Production: Evidence from Repair in Conversation**

**Abstract:** Investigation of spontaneous replacement repairs found in the Switchboard Corpus (Godfrey et al., 1992) shows that low-frequency repaired words are more likely to be interrupted prior to replacement than high-frequency words are. These results provide novel empirical support to the hypothesis that the production of high-frequency words is more automatic than the production of low-frequency words (Bybee, 2002; Logan, 1982). The relationship between the effects of frequency on interruptibility is argued to be partially mediated by the effect of frequency on duration. In addition to testing the link between frequency and automaticity, the present paper reports that replaced words tend to be more frequent than the words that replace them, providing support for the hypothesis that high-frequency words are easier to access in word production, which has been criticized on the basis of not observing this frequency asymmetry in semantic substitution errors (Garrett, 2001). Finally, whether a word is interrupted is found to depend strongly on the length of the word, with long to-be-replaced words being more likely to be interrupted than produced completely. Thus, while speakers prefer to produce constituents with a continuous delivery (Clark & Wasow, 1998), the drive to produce a continuous constituent competes with the drive to interrupt as soon as possible (Main Interruption Rule, Levelt, 1983, 1989).

### **Introduction**

#### **Theoretical Background**

Bybee (2002) suggests that the production of high-frequency words and phrases is more automated than the production of low-frequency words and phrases. Under this hypothesis, high-frequency words are more cohesive than low-frequency words: the parts forming a high-frequency word are more tightly linked together than the parts forming a low-frequency word.

Previous evidence for a link between cohesion and frequency has come from studies showing that high-frequency words are more likely to undergo reductive sound change (Bybee, 2002; Hooper, 1976). Mowrey and Pagliuca (1995; Pagliuca & Mowrey, 1987) go as far as claiming that all internally-motivated regular sound changes in progress that have been attested can be explained by an increase in gestural compression. Bybee (2001: 79-83) and Phillips (2001) suggest that there are other sources of sound change but that Mowrey and Pagliuca's claim holds for sound changes that involve lexical diffusion from high-frequency to low-frequency words.

An increase in the temporal overlap between successive gestures and temporal compression of the sequence of articulatory goals corresponding to a word is expected to result from automatization of word production (Bybee, 2002). Assuming that in a sequence of articulatory goals, a goal gains control of articulation when it is activated sufficiently, and that activation spreads from earlier goals to later ones, a goal will receive control of articulation earlier when it is strongly connected to the preceding goal. Thus, the preceding goal is less likely to be completely reached when the following goal is highly predictable in the context. In addition, when the gestures called for by successive goals do not interfere with each other, which could cause undershoot, articulatory overlap between gestures implementing successive goals is more likely in a high-frequency sequence. Under this account, a high-frequency word is a more cohesive unit than a low-frequency word.

However, the finding that reductive sound changes start in high-frequency words has also been interpreted as indicating that speakers do not expend as much articulatory effort in such words because of their high contextual predictability for the listener (e.g., Bybee 2002: 269; Gregory et al., 2000; Lindblom, 1990). Fowler (1988) shows that words that have already been mentioned in the course of the conversation are shorter than words that are mentioned for the first time (see also Fowler and Housum, 1987) but only if the two tokens are co-referential. Words are not shortened if a homonym has recently been pronounced but are shortened if preceded by a synonym. Fowler (1988: 317) writes that “production of a homophone of a target... is not sufficient to yield shortening... even though the word’s articulatory routine has recently been used. Apparently the shortening reflects the talker’s estimate that a listener has other information available to help identify the word”. Gregory et al. (2000) support this interpretation by showing that semantic relatedness to the discourse topic influences word duration even when repetition is controlled: words related to the discourse topic are shorter than unrelated words.

Under this alternative interpretation, word frequency does not directly influence gestural compression, automaticity of production, or word cohesion. Rather, frequency is simply one of the factors that influences contextual predictability, which serves as a constraint on how much reduction the speaker thinks s/he can get away with.

In the present paper, we investigate a hitherto untested prediction of the hypothesis that the production of high-frequency words is more automatic than the production of low-frequency words. As Anderson (2000: 99) puts it, “automaticity occurs when practice eliminates most of the need for central cognition”, which leads to the behavior becoming relatively impervious to cognitive influences. In particular, the more automatic a behavior, the harder it should be to interrupt. Thus, if the production of a high-frequency word is more automatic than the production of a low-frequency word, the production of a high-frequency word should be harder to interrupt than the production of a low-frequency word.

To address this issue, we will analyze a corpus of conversations among native English speakers (Switchboard, Godfrey et al., 1992), which has been tagged for disfluencies. The working hypothesis is that when the speaker interrupts his/her production to replace the word s/he has just produced or started producing, the interruption is more likely to be delayed until the end of the to-be-repeated or to-be-replaced word if the word is frequent than if it is rare.

## The Phenomenon

In a replacement repair, the speaker replaces the word s/he has just produced or started producing by a different word. Examples of replacement repairs from the Switchboard Corpus are shown in (1)-(4). The **replaced word** is shown in bold while the **replacement** is italicized. We will call the observed part of the replaced word, e.g., *wa* in (3), the **remainder**, reserving the term **replaced word** for the inferred complete lexical item, e.g., *watch* in (3). Examples in (1)-(4) show that the speaker has a choice of producing the replaced word completely or interrupting its production. The present paper is restricted to cases of replacement repair in which the replaced word and the replacement word are semantically related because it is nearly impossible to guess the identity of an interrupted replaced word if it is not semantically related to the replacement.

- (1) It was **pathe-**, I mean, it was *horrible*.
- (2) That’s why we were surprised to see ‘Toyota’ **written**, I mean, *imprinted* on the engine
- (3) I will intentionally buy newspaper to **wa-**, to *look at* the news.
- (4) They don’t want to become a state for fear of losing **Spanish**, uh, *Hispanic* heritage.

## Cohesion as an Influence on Disfluency Location

While there have been no studies of frequency effects in replacement repair, previous work on repetition repair and other disfluencies has shown that the location of interruption and how much material is repeated are influenced by constituency. Boomer (1965), Clark and Wasow (1998), Levelt (1983), and Maclay and Osgood (1959) found that interruption of speech production is more likely to occur at word boundaries than within words and between major syntactic constituents, such as subject and object, rather than within them. Beattie and Butterworth (1979), Goldman-Eisler (1958, 1968), Tannenbaum et al. (1965) and Cook (1969) demonstrated that hesitations tend to occur in between-word transitions of maximum uncertainty, as indicated by low transitional or Cloze probability. These results suggest that interruption is sensitive to cohesion: speech production is more likely to be interrupted at the boundary between cohesive units than within a cohesive unit. Thus, if high-frequency words are more cohesive than low-frequency words, speakers should be less likely to interrupt speech production in the middle of a high-frequency word than in the middle of a low-frequency word.

Several studies found that speakers tend to start repetition from the nearest major constituent boundary (Maclay & Osgood, 1959; DuBois, 1974; Nootboom, 1980; Levelt, 1983; Fox & Jasperson, 1995; Clark & Wasow, 1998; Kapatsinski, 2005). Definitions of major constituent boundaries differ somewhat across studies, with most researchers taking such boundaries to include clause, object, and oblique boundaries (Clark and Wasow, 1998; Fox and Jasperson, 1995; Kapatsinski, 2005; Maclay and Osgood, 1959).<sup>2</sup> Based on this work, Clark and Wasow (1998: 206) proposed the **Continuity Hypothesis**, which states that speakers prefer to produce syntactic constituents with a continuous delivery. For instance, if speech production is interrupted somewhere in a prepositional phrase, speakers tend to repeat everything they have produced after starting the phrase as in (5) below.

(5) I was really familiar **with a lot, with a lot of, of** the AOR type music

In (5), the speaker repeats three words s/he has already produced despite an overall preference to repeat as little as possible (in the sample of Kapatsinski 2005, 79% of repetitions are one-word repetitions, 18% are two-word repetitions, and only 3% are three-word repetitions). The likely reason, according to the Continuity Hypothesis, is that the speaker wants to produce the entire prepositional phrase without interruption.<sup>3</sup> Importantly, while English speakers often repeat prepositions, Japanese speakers do not repeat postpositions, which would involve restarting speech from the middle of a postpositional phrase (Fox et al., 1996). Finally, the Continuity Hypothesis is supported by the fact that if word production is interrupted within a word, the speaker almost always restarts the word, rather than continuing from the point of interruption.<sup>4</sup>

Kapatsinski (2005) found that how much is repeated in a repair is influenced by between-word transitional probability. Speakers do not start repeating from the nearest constituent boundary if that constituent boundary is a high-probability transition. Kapatsinski tried to predict how many words will be involved in each repetition found in the Switchboard corpus depending on the location of the nearest constituent boundary and on which of the three nearest between-word transitions has the lowest transitional probability. The location of the nearest constituent boundary correctly predicted 44% of the three-word repetitions in the Switchboard corpus. Then transitional probability was added as a predictor. The two predictors were combined so that if transitional probability at some nearby word boundary is much lower than at the nearest constituent boundary, subjects were predicted to start repeating from the

<sup>2</sup> The status of the subject-verb boundary is questionable (Fox and Jasperson, 1995). In addition, Levelt (1983) argues for an alternative criterion for where disfluencies should occur, according to which one should be able to continue the constituent interrupted by the disfluency in such a way that it would be conjoinable with the constituent following the disfluency.

<sup>3</sup> Alternatively, speakers may have difficulty initiating production from the middle of a cohesive unit.

<sup>4</sup> I have been able to find only one example of the latter on Switchboard.

transition with the lowest transition probability. Otherwise, they were predicted to start from the nearest constituent boundary. This modification of the Continuity hypothesis improved the predictability of three-word repetitions to 57%, while maintaining 70% accuracy on one-word and two-word repetitions, where chance performance is 33% (Kapatsinski, 2005:490).

Thus prior findings suggest that between-word cohesion influences location of interruption and speakers tend to restart interrupted cohesive units, whether the cohesion is caused by syntactic constituency or probability of co-occurrence. In the present study we will examine whether location of the interruption is also sensitive to within-word cohesion and, more specifically, whether high-frequency words are more cohesive than low-frequency words.

### **The Possible Roles of Relative Frequency**

Surprisingly, there has been only one study looking at word frequency as an influence on disfluency location. Biber et al. (1999:1059) observe that the indefinite article is less prone to being repeated than the definite article and propose that “perhaps, all other things being equal, the higher a word’s frequency, the more likely it is to form repeats... It is easy for the speaker to utter a very frequent word, without having a clear plan of what words will follow it. Hence, such a word precedes a natural hesitation point in the utterance”. Biber et al. support the hypothesis by pointing out that *an* is repeated very rarely since before choosing *an* the speaker must at least decide on a vowel-initial word to follow it. Otherwise, the speaker would choose the much more frequent variant *a*. Consequently, the sequence *a an* is much more frequent than the sequence *an a*. In addition, the authors find that frequent subject+verb contractions, those that involve ‘s, ‘re, ‘m and ‘ll, are more likely to be repeated, per number of tokens of the contraction in the corpus, than less frequent contractions involving ‘ve and ‘d (Biber et al., 1999: 1061-2).

Biber et al.’s (1999) hypothesis provides a possible prediction for when a to-be-replaced word’s production will be interrupted. The hypothesis is that a high-frequency word is likely to come to mind faster than a low-frequency word. Thus, if the replaced word is frequent and the replacement word is rare, the replaced word will come to mind long before the replacement word. Thus, the speaker will have enough time to produce the replaced word in its entirety before s/he becomes aware of the more appropriate alternative. On the other hand, if the replacement word is frequent relative to the replaced word, the appropriate replacement is likely to come to mind soon after the speaker starts to utter the less appropriate word, leading the production of the replaced word to be aborted before the entire word is produced. This theory predicts that, other things being equal, interrupted words should be replaced by high-frequency words while uninterrupted words should be replaced by low-frequency words. Thus in the present study, we examine both frequency of the replaced word and frequency of the replacement word as predictors of whether or not the replaced word is interrupted.

In addition, if a high-frequency word comes to mind faster than a low-frequency word, the case in which a frequent inappropriate word is replaced by a rare but more appropriate word should be more common than the case in which a rare word is replaced by a word that is both more appropriate and more frequent. Thus, the replaced word should tend to be more frequent than the replacement word. However, studies of semantic substitution errors have failed to find a difference between the erroneous word (‘the intrusion’) and the correct target (DelViso et al., 1991; Harley & MacAndrew, 2001; Hotopf, 1980; Silverberg, 1998). Garrett (2001) notes that this negative result is inconsistent with existing models of word production as well as experimental data from picture naming, which show that pictures with high-frequency names are faster than pictures with low-frequency names (e.g., Jescheniak & Levelt, 1994;

Oldfield & Wingfield, 1965).<sup>5</sup> In contrast to studies of semantic substitution errors, the present sample shows a small but reliable difference in frequency between replaced words and replacement words in the expected direction, closing the gap between naturalistic and experimental data observed by Garrett (2001) and supporting the role of token frequency in facilitating lexical access in production.

### **Main Interruption Rule and Error Detection**

An assumption made by the model in the preceding section is that interruption is triggered by awareness of an alternative, rather than recognition of the inappropriateness of the word being produced. Alternatively, interruption could be triggered by detection of inappropriateness and the search for an alternative could be initiated by detection of inappropriateness. Under this hypothesis, the location of the interruption would be independent of how fast the alternative is accessed. Rather, a word would be likely to be interrupted if its inappropriateness is detected early relative to when the production of the word is initiated.

Levelt (1983, 1989) proposes that speakers interrupt production as soon as they detect inappropriateness of the word being produced (what he calls the **Main Interruption Rule**). Under the Main Interruption Rule, words may be interrupted if their inappropriateness is detected quickly. The speed of detection could plausibly depend on the severity of the error. Thus, a word that is merely inappropriate may be less likely to be interrupted than a word that is an outright speech error, as found by Levelt (1983).<sup>6</sup> There is some evidence that low-frequency words are more likely to be involved in speech errors (e.g., Harley and MacAndrew, 2001). If high-frequency words are less likely to be uttered in error and more likely to be merely inappropriate than low-frequency words, error detection may be slower in high-frequency words, making high-frequency words less likely to be interrupted than low-frequency words. We will return to this possibility in the analysis section.

### **Experimental Studies of Interruptibility in Language Production**

There have been three previous studies that specifically examined how easy it is to interrupt language production and the factors influencing interruptibility (Ladefoged et al., 1973; Logan, 1982; Slevc & Ferreira, 2006). In all of these studies, on a small proportion of trials, the subject was presented with a stop signal, which indicated to the subject that they should stop production.

While none of these studies were specifically designed to test for frequency effects, Logan (1982, Experiment 3) observed that if the typists were told to stop typing immediately before they started typing the word ‘the’, they tended not to stop until after producing ‘the’, producing 2.72 letters on average. The same subjects produced fewer than 2 letters on average if the stop signal came before a content word (verb or noun). Logan showed that while the word ‘the’ was typed faster than other words, the time it took subjects to stop typing ‘the’ was longer than the time it took them to stop typing content words. He attributed the effect to word frequency, noting that ‘the’ is the most frequent word in English. The present study extends this finding by investigating a much larger range of words and word frequencies in naturalistic speech production.

---

<sup>5</sup> The lexical locus of this effect was confirmed by its disappearance in picture recognition (Jescheniak and Levelt 1994, experiment 2).

<sup>6</sup> Levelt himself (1989:481) seems to reject this possibility, writing “there is no reason to assume that the detection of error occurs more frequently within the troublesome word than the detection of inappropriateness”, suggesting instead that interruption is used by the speaker to tell the listener whether the replaced word is a speech error.

## Method

### The Corpus

For this study we collected all tokens of replacement repair in the Switchboard corpus (Godfrey et al., 1992) that satisfied our inclusion criteria. The Switchboard corpus is a collection of telephone conversations between native American English speakers on predetermined topics that are chosen by the participants from a fixed set of alternatives with no knowledge of the identity of their interlocutor-to-be. The version of Switchboard annotated for disfluencies contains about two million words. The corpus is annotated with a special symbol ('+') which marks the locations of repairs. Sound recordings of the conversations are available online from the LDC (<https://online.ldc.upenn.edu/search/>). To be included in the present sample, a token of repair had to be coded as one in the corpus. In addition, the author listened to the coded tokens of repair and excluded a number of cases based on the exclusion criteria outlined in the next section.

### Exclusions

In the present paper, we concentrate on semantically motivated replacement repair. Thus instances of repair which involve word insertion as in (6) or (7), word deletion, or reordering as in (8), as opposed to replacement were excluded.

- (6) It does give you a **good**, a *real good* workout.
- (7) Just to see whether or not we're **falling**, you know, *getting ahead*, **falling** behind or staying even or what.
- (8) They ought to, you know, go out of the way, I think, a little bit more to, to *help you get*, **help get you** rehabilitated

Since it is difficult to guess the identity of an interrupted replaced word when it is not semantically related to the replacement word, uninterrupted replaced words were excluded as well if they were not semantically related to the replacement. Thus, the example in (9) was excluded from the sample.

- (9) I went to the bike **shock**, I mean, the bike *shop*.

The example in (9) would be excluded from the sample for another reason as well. In (9), the replaced word (*shock*) and the replacement word (*shop*) share beginnings. Therefore, if the replaced word were interrupted, it would be impossible to tell that the sentence involves replacement rather than repetition. Thus, all cases in which the replaced word and the replacement word share beginnings were excluded from the sample if they shared more than one segment. Repairs involving words shorter than three segments or longer than eight segments were excluded because there were very few such words in the sample.

In addition, instances of repair in which the replaced consisted of more than one word were excluded. These include cases of the type shown in (10), where *turned [out]* is abandoned in favor of *was*, as well as cases in which multiple words that are part of the replaced surface as in (11). Contractions like *can't* or *don't* and *going to* in the sense of *will* were considered single words and included in the sample.

- (10) It **turned**, it *was* okay.
- (11) The court systems need to be **more accurate** in, in, *stiffer* in their penalties.

Cases in which the replaced was a function word that was incompatible with what followed the replacement, as in (12) where *has* appears to be replaced by *is*, were also eliminated because it is likely

that in these cases repair is motivated by a desire to replace not the function word itself but some word downstream in the planning sequence or the syntactic construction itself (cf. Stemberger, 1984).

(12) **Has**, *is* this guy a convicted felon?

Uninterrupted replacement repairs included both cases in which the flow of speech was interrupted immediately after the replaced word and those in which it was interrupted later. Thus, cases like (13) were included in the sample.

(13) I haven't **had** a chance, I haven't *got* a chance to look at them yet.

Finally, there is a thin line between replacement repair and certain grammaticalized constructions, which should not be included into a sample of repairs because they disallow interruption. One such construction is the clarification construction in which the 'replacement' is a hyponym of the replaced. Thus one can argue that the example in (14) does not involve repair but rather clarification. However, example (15) in which the replaced word is interrupted, cannot be interpreted in this way. Thus, the speaker may prefer to say (14) instead of (15) regardless of the frequency of *same*. Thus, cases in which the replacement is a hyponym of the replaced were excluded from the sample.

(14) But, no, no real association with TI other than being in the **same** industry, the *electronics* industry.

(15) But, no, no real association with TI other than being in the **sa-**, the *electronics* industry.

The mirror image of the clarification construction illustrated in (14) is presented by subject topicalization in which the 'replaced' is a hyponym of the replaced. An example is presented in (16). To avoid inadvertently including such cases into the sample, all examples in which the replacement is a pronoun, the replaced is a noun phrase, and the two can be coreferential were excluded from the sample.

(16) **My husband and I**, *we* just sit there and cackle.

Another potentially grammaticalized case excluded from the sample is the use of interruption following subject+*just* followed by repetition of the same subject as in (17). Such cases are quite common, although more commonly *just* is either repeated or omitted and may involve an interruption that is preplanned for emphatic purposes rather than generated online when a decision to replace a word is made.

(17) He **just**... He *simply* doesn't care anymore.

Another case in which repair can be confused with a grammatical construction if the replaced word is not interrupted is when the 'replaced' and the 'replacement' are numbers and the second number is larger than the first (in terms of absolute value, as (19) shows). Thus, repairs involving numbers were included only if the second number was closer to zero than the first.

(18) It's taken them **ten**, *fifteen* minutes at a time.

(19) When it's minus **twenty-five**, minus *thirty* degrees...

(20) When you're **twenty**, *thirty* years old...

(21) He was there in nineteen eighty **four**, eighty *five*.

Finally, repairs are important to distinguish from lists. A specific problem is presented by lists of near-synonyms in which the following synonym is 'more intense' than the preceding one, e.g., *big giant trees* or (possibly) the example in (19). In these cases, the second word is not intended to replace the first

word, hence interruption is not an option. In addition, cases in which the speaker can't decide on the correct word and plans to indicate his lack of certainty by using a disjunction in advance are potentially problematic (a possible example is shown in (22)).

(22) He's a computer **programmer**, or a computer *engineer*.

Fortunately, lists that do not have a conjunction usually have more than two elements and were excluded on the basis of this criterion. In addition, both listing constructions and disjunctions can be identified by intonation. The presence of emphasis on the replaced list intonation, or the absence of interruption was sufficient for exclusion. As Ladefoged et al. (1973) observed in their study of experimentally elicited interruption, interrupted words almost invariably end with a glottal stop or at least significant glottalization, while uninterrupted words do not. Thus, glottalization is a very reliable cue for whether the word was interrupted. Nonetheless, 31 tokens were excluded from the study because there was disagreement between the present author and the corpus coders on whether or not the word was interrupted or because the present author was not certain about the status of the word.

### Exclusions Specific to Particular Analyses

While cases in which the replacement consisted of more than one word, as in (3) where *watch* is replaced by *look at* or (23) where *had* is replaced by *came out of*, were included in the complete sample, they were excluded for the purposes of comparing the frequency of the replaced word to the frequency of the replacement word both in terms of their absolute values and as predictors of interruption. Comparing word frequencies to a mix of word and phrase frequencies would be unfair because the frequency of a phrase is on average lower than the frequency of a word just because a phrase contains multiple words.

(23) I **had** a, I *came out of* a thirty-one hundred square foot two story house.

In order to assess whether low-frequency words are more likely to be uttered in error, rather than being merely inappropriate, we need to determine whether a given repair involves an error. Determining whether a repair involves an error in natural conversation is quite difficult and it is not clear that the distinction can be reliably made in all cases. Moreover, a large proportion of cases are similar to example (24) where what the speaker may consider a speech error, the listener, who does not know anything about the speaker's family, would surely not. Thus, the analysis will be restricted to unambiguous cases only.

(24) My **parents**, my *mother* is trying to let my grandmother stay in her house.

Levelt (1983:63), writes that "in an appropriateness repair... the reparandum is correct but needs some qualification". This suggests that a hyponymy relation is involved. Such cases were excluded from the present sample. In addition, the example shown in Levelt(1989:481) suggest that repairs of suboptimal choices can also involve synonymy as shown in (25).

(25) To the left of it a **blanc**, or a *white* crossing point.

While it is not clear whether this example would be included in the present sample because the replaced and the replacement are conjoined with *or* and it is not indicated that the example involved hesitation, the sample does include a number of cases in which the replaced and the replacement are synonymous, as in (26)-(27).

(26) I don't have the expertise to just hurry up and do it like **some**, a professional would.

(27) That's my **private**, you know, my *own* home.

These cases can be compared to tokens in which the replaced and the replacement are incompatible because of having demonstrably different referents as in (28)-(29). A common special case is the replacement of a quantifier by a quantifier with a different range of possible values as when *most* is replaced by *all*, *few* is replaced by *most*, *eleven* is replaced by *twenty-one*, *quite* being replaced by *not really*. On the other hand, cases in which the replaced and the replacement are quantifiers whose ranges of values are similar, as when *several* is replaced by *a few*, can be considered repairs not involving a speech error.

- (28) A sixty-seven **Chev-**, uh, *Mustang*  
 (29) You may be able to take **care**, take *advantage* of that.

An additional class of repairs that can be said to involve speech errors are repairs in which the replaced word does not fit the preceding context as in (30)-(31).

- (30) The person who is the line **own-**, the line *manager*.  
 (31) I was watching the **ra-**, the *TV* today.

Finally, repairs in which one form of a verb is replaced by a different form of the same verb, such as *is* being replaced by *was* can be considered repairs involving speech errors. This does not include cases like *was* being replaced by *has been* in which the two verb forms can have the same referent. Such cases were not included in the analysis.

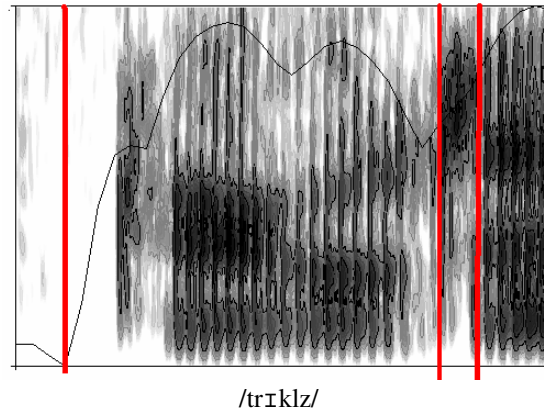
### Measuring Frequency, Duration, and Number of Segments

For each instance of repair included in the sample (N=1749), the duration of what remained of the replaced word (the remainder) and the duration of the replacement word were measured. In order to examine the extent to which any possible effects of frequency are mediated by the effect of frequency on duration (frequent words are shorter), I estimated the length the interrupted word would have if it were not interrupted. Several estimates were obtained. For each of these estimates, the duration of the word did not include the word-final segment. This is because one purpose for which we need estimates of word duration is to compare the durations of interrupted and uninterrupted words. Since a word may not be coded as interrupted if its final segment was perceived by the coders, the final segment is not a possible location for interruption, and the status of the preceding transition is questionable since it can contain strong cues to the final segment's identity.

First, a very crude estimate of word duration was obtained by multiplying the duration of the remainder by the ratio of the number of segments in the remainder to the number of segments in the complete word. Second, the duration of the complete word produced in isolation from text by an adult female native speaker of American English was obtained from LDC's American English Spoken Lexicon (<http://www ldc.upenn.edu/cgi-bin/aesl/aesl>). Finally, ten samples of each of the replaced words were obtained from the Switchboard Corpus and their durations were measured. When ten tokens of the word were not available, all available tokens were used. When more than ten tokens were available, the ten samples used were randomly selected. This last measure proved to be the best of the duration measures in terms of predicting whether or not the word would be interrupted. Hence, the comparisons between frequency and duration as predictors of interruption reported below use this measure.

Durations were measured by hand in Praat. The principal difficulty in measuring duration came from cases in which the to-be-measured boundary fell between two stops or a stop and a pause. When a word began with a stop preceded by another unreleased stop or silence, the beginning of the word was taken to be the point at which the intensity track starts to increase sharply from the floor as shown in

Figure 1 for the word *trickles*. In the case of a stop-final word, the midpoint of the preceding segment was taken to be the end of the word.



**Figure 1.** Measurement of duration of stop-initial words (spectrogram with a superimposed intensity track). Word boundaries are shown by the thick lines. The rightmost line shows the actual end of the word while the next rightmost line shows the end of the word as measured for the purposes of this study.

For the purposes of estimating word length in numbers of segments, affricates, diphthongs, syllabic nasals and liquids and /3r/ were coded as single segments. This decision was made because cases in which a diphthong was interrupted (e.g., [ha- haʊsɪz], [θɜrzdɛ- θɜrzdɛɪ]) and cases in which the schwa was produced without the following sonorant (e.g., [mɑðə- mɑðə], [ivə- ivən]) were exceedingly rare, and there were no cases in which an affricate was interrupted.

Word frequency was operationalized as frequency of occurrence within the Switchboard Corpus, the corpus under analysis in the present study. Since Switchboard consists of conversations on a limited range of topics, frequencies within the corpus may not correlate very well with frequencies elsewhere in the language. Since recent repetitions are likely to be more important for present behavior than earlier repetitions, and since the production of a word may be more automatic when it is related to the topic of conversation and therefore somewhat predictable, frequency within the corpus under analysis is still arguably a more appropriate measure than frequency within some other corpus (e.g., Francis and Kuçera 1982). Surface frequency rather than base frequency was used. That is, frequency was not aggregated across different inflectional forms of a particular word. This decision is based on a regression analysis of the effect of frequency on whether or not a to-be-repeated word is interrupted, which showed that surface frequency was a better predictor of interruption than base frequency.

For the purposes of analysis, frequency was logarithmically scaled since ease of lexical access in both perception (Howes & Solomon, 1951) and production (Oldfield & Wingfield, 1965) is correlated with log frequency better than with raw frequency. The basic idea behind the log transform is that the difference in frequency between a word that occurs only once in the corpus and a word that occurs ten times is much more psychologically significant than the difference between a word that occurs 1000 times in the corpus and one that occurs 1010 times. For 3-segment and 1-syllable words considered separately the distribution of log frequencies is skewed, violating the assumptions of standard statistical tests. For

this reason, frequencies were converted to ranks for the purposes of statistical tests involving the subsamples of three-segment and one-syllable words.<sup>7</sup>

## Analysis

In this section, we first establish that interrupted words that are interrupted and words produced completely do in fact differ in token frequency, that longer words are more likely to be interrupted than shorter words, and that the frequency difference remains when number of segments is controlled. We then examine whether the speed of accessing the replacement can account for the results and establish that interrupted words do not tend to be replaced by high-frequency words while words produced completely are followed by low-frequency words. In the following section we confront the issue that the identity of the replaced word needed to be guessed and show that interrupted replaced words, which I guessed, are as frequent *relative to their replacements* as uninterrupted words, for which no guessing was involved. Therefore, my guesses are argued not to be biased in favor of the hypothesis in that, even if wrong, they tend to produce words that are as frequent as the interrupted words intended by the speakers. We then address the possibility that errors that occur in low-frequency words are more severe and thus easier to detect. Finally, we examine the interaction of frequency, interruptibility and duration, arguing that high frequency of use does not just shorten words but also makes interruption dispreferred.

### Frequency and Number of Segments

Figure 2 shows that the longer the replaced word, in terms of number of segments, the more likely it is to be interrupted. The relationship between number of segments and likelihood of interruption is well approximated by a logarithmic curve. Figure 2 also shows that words longer than four segments are more likely to be interrupted than produced completely. In addition, it should be born in mind that replacements involving only one segment and replacements in which the identity of the replaced word could not be guessed are not included in the sample. As a result, Figure 2 is likely to underestimate the true likelihood of interruption in replacement repair.

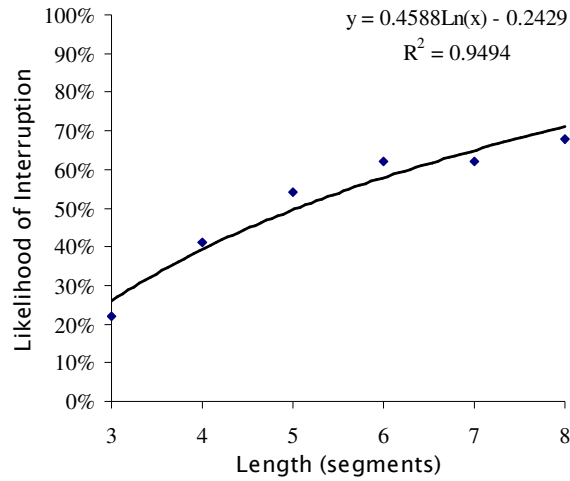
Figure 2 indicates that it is not the case that all words are created equal in terms of the interaction of the Continuity Hypothesis (Clark and Wasow, 1998) with the Main Interruption Rule (Levelt, 1983, 1989). While in general, words are produced completely more often than they are interrupted in the present sample (61% of all words in the sample are not interrupted), Figure 3 suggests that this is an artifact of the fact that there are more short words than long words in the English lexicon. However, between-word transitions can still be privileged locations of interruption relative to word-internal segment-to-segment transitions. An eight-segment word maximally contains seven possible word-internal locations for interruption and one between-word location. Thus, if 40% of all interruptions involving an eight-segment replaced word occur in the between-word location, the between-word transition is privileged relative to the word-internal transitions as a location for interruption, as shown in Figure 3.

Figure 3 confirms that the data in Figure 2 do not contradict the Continuity hypothesis. As predicted by the Continuity hypothesis, for any word length, the between-word transition is a significantly more common location for interruption than any one of the within-word transitions according to the chi-square test (the closest contender among between-word transitions is the location after the third segment in eight-segment words that hosts 18 interruptions relative to 27 cases in which an eight-segment word is not interrupted; the difference is significant,  $\chi^2(1)=7.2$ ,  $p<.01$ ). On the other hand, since many interrupted cases of repair are not included in the sample because the identity of the interrupted word could not be

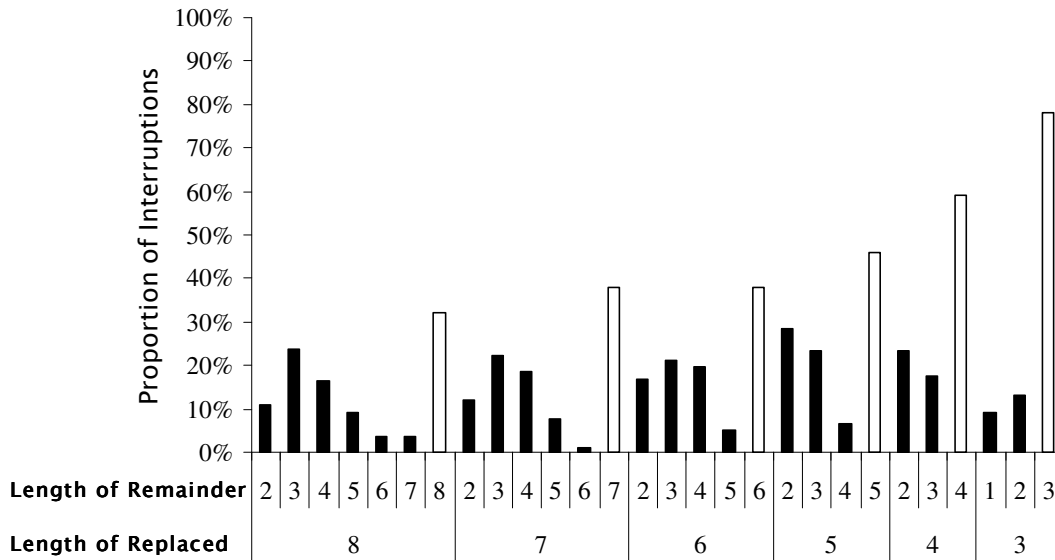
---

<sup>7</sup> Rank conversion of a set of numbers (e.g., frequencies) involves arranging the numbers from the highest to the lowest and replacing each number with its position in the sequence. For instance, if we have a sample of words that have frequencies of 1000, 2, 35, and 99, the corresponding ranks are 1, 4, 3, and 2 respectively. This is a standard way to deal with non-normal data.

guessed, this result does not provide strong evidence for the Continuity hypothesis. For that, we will have to turn to frequency effects.



**Figure 2.** When a speaker intends to replace a word, s/he is more likely to interrupt it if it is long than if it is short.<sup>8</sup>

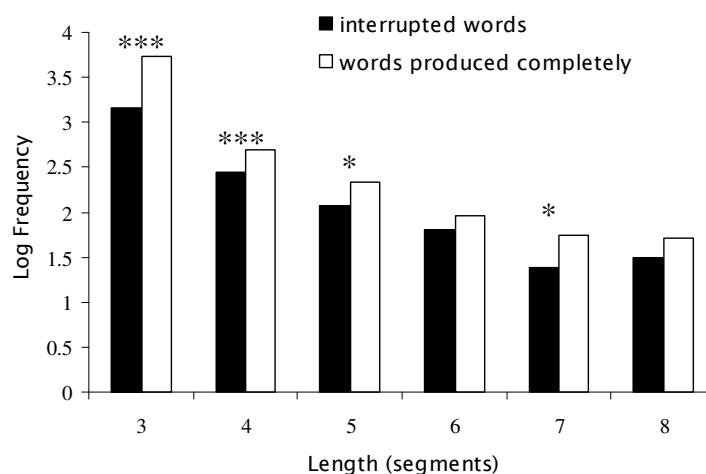


**Figure 3.** Interruptions are more likely to occur in a between-segment transition that spans a word boundary than in any between-segment transition within a word. The proportions shown are out of all interruptions involving replaced words of a given length. Thus, percentages within a bin defined by length of the replaced sum to 1.

High-frequency words tend to have fewer segments and Figures 2-3 show that words that have fewer segments are less likely to be interrupted. Therefore, for an effect of frequency on interruptibility to be established, it needs to be shown that it holds when number of segments is controlled. This is shown in

<sup>8</sup> Grouping the words by number of syllables rather than number of segments produces the same result.

Figure 4. For each word length, replaced words that are interrupted tend to be lower in frequency than words that are produced completely. The difference is statistically significant overall (in a multiple linear regression that also included log number of segments, interruption was a significant predictor of frequency,  $t(1746)=9.934$ ,  $p<.0005$ ; frequency is a significant predictor of interruption when frequency and length are entered into a binomial logistic regression as covariates,  $p<.001$ ), as well as for three-, four-, five- and seven-segment words considered separately (for 3-segment words,  $t(798)=7.821$ ,  $p<.0005$ <sup>9</sup>; for 4-segment words,  $t(406)=4.092$ ,  $p<.0005$ ; for 5-segment words,  $t(190)=2.051$ ,  $p=.042$ ; for seven-segment words,  $t(131)=2.131$ ,  $p=.035$ ). It is not significant for six-segment and eight-segment words.



**Figure 4.** Words that are interrupted tend to be less frequent than words produced completely. One star indicates significance at the .05 level in a two-tailed t-test. Three stars indicate significance at the .005 level.

The data in Figure 4 support the hypothesis that high-frequency words are more cohesive than low-frequency words and their production is more automatized than the production of low-frequency words. However, alternative explanations are possible. The next section will consider an explanation based on speed of accessing the replacement, the following section explores possible observer bias, the one after that considers duration, and the one after that error detectability.

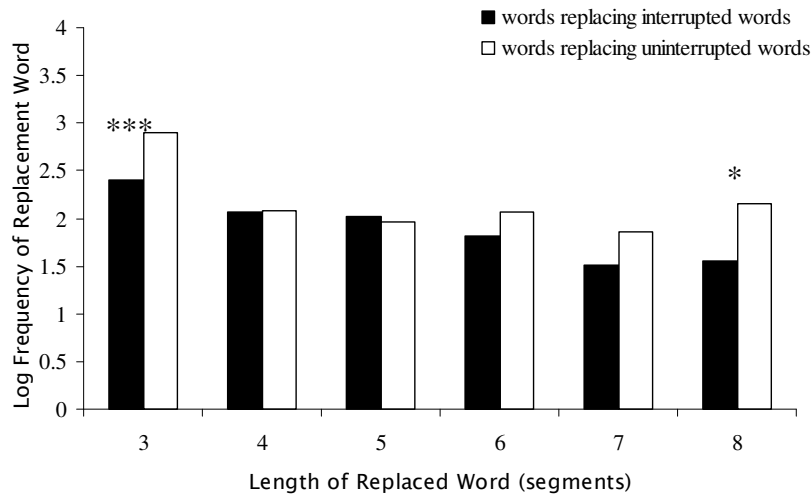
### Frequency of the Replaced vs. Frequency of the Replacing

A possible alternative explanation is suggested by Biber et al.'s (1999) account of why frequent words are more likely to be repeated than rare words. It is possible that words that replace interrupted words are more frequent than words that replace uninterrupted words. If this were the case, interrupted words would be interrupted because the more appropriate alternative would come to mind more quickly. Assuming that the decision to interrupt production in single-word replacement repair is caused by activation of a more appropriate word, this decision would then be made earlier when the replacement word is frequent. And if the replaced word is rare and thus accessed slowly, the decision to replace the word would be made shortly after it accessed, giving speaker more opportunities to interrupt its production. Thus, this hypothesis predicts that interrupted words should be rarer *relative to their*

<sup>9</sup> For this analysis, frequencies were converted to frequency ranks as the distribution of log frequencies was highly skewed for three-segment words.

*replacements* than uninterrupted words without necessarily predicting a difference in absolute token frequency between interrupted and uninterrupted words.

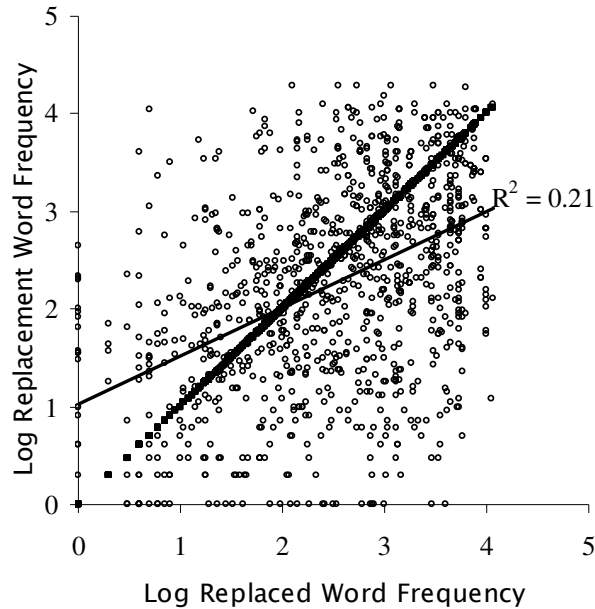
However, the data are inconsistent with this prediction. Not only is there an absolute frequency effect in the data (as shown by Figure 4) but there is also no relative frequency effect. In order to derive estimates of relative frequency, the (log) frequency of each replaced word was divided by the sum of (log) frequencies of the replaced word and the corresponding replacement. Then mean relative frequency of interrupted replaced words was compared to mean relative frequency of uninterrupted replaced words. The mean relative frequency of interrupted words was .54 while the mean relative frequency of uninterrupted words was .53. This non-significant difference ( $t(1029) < 1, p = .4$ ) is in the opposite direction from the one predicted by the hypothesis.<sup>10</sup> Figure 5 shows that words that replace interrupted words tend to be less, rather than more, frequent than words that replace words that are produced completely.



**Figure 5.** Words that replace interrupted words tend to be less frequent than words that replace words that are produced completely. One star indicates significance at the .05 level in a two-tailed t-test. Three stars indicate significance at the .005 level.

The reason for this result can be inferred from the data in Figure 6, which shows that the frequency of the replacement is positively correlated with the frequency of the replaced. Thus, the replaced and the replacement tend to be of similar frequency. This finding has also been observed in studies of lexical substitution errors (DeViso et al., 1991; Harley and MacAndrew, 2001; Hotopf, 1980; Silverberg, 1998). The correlation is very similar in magnitude to that obtained by Harley and MacAndrew (2001) in their study of lexical substitution errors:  $r = .44$  in the present study, compared with  $r = .4$  in Harley and MacAndrew (2001).

<sup>10</sup> For the purposes of the analyses reported in this section, multiple-word replacements and replacement words shorter than 3 or longer than 9 segments were eliminated to make the sample of replacements comparable to the sample of replaced words. Because of this, the sample only contains 1030 tokens.



**Figure 6.** The relationship between the frequency of the replaced and the frequency of the replacement. A linear regression fit is shown as well as the diagonal for which the frequency of the replaced and the frequency of the replacement are equal. The data show a significant positive correlation between the frequency of the replaced and the frequency of the replacement. The fact that fewer points are above the diagonal than below indicates that the replacement tends to be less frequent than the replaced.

However, Figure 6 also shows that, unlike in studies of semantic substitutions, replaced words tend to be more frequent than the replacement words in the present sample. The replaced was more frequent than the replacing in 593 cases while the replacement was more frequent than the replaced in 428 cases. Thus, the replaced was more frequent than the replacement in 58% of the cases in which the two differed in frequency. While the effect is small, it is significant in both the chi-square test and the paired samples t-test ( $\chi^2(1)=26.66, p<.0005; t(1029)=7.307, p<.0005$ ).

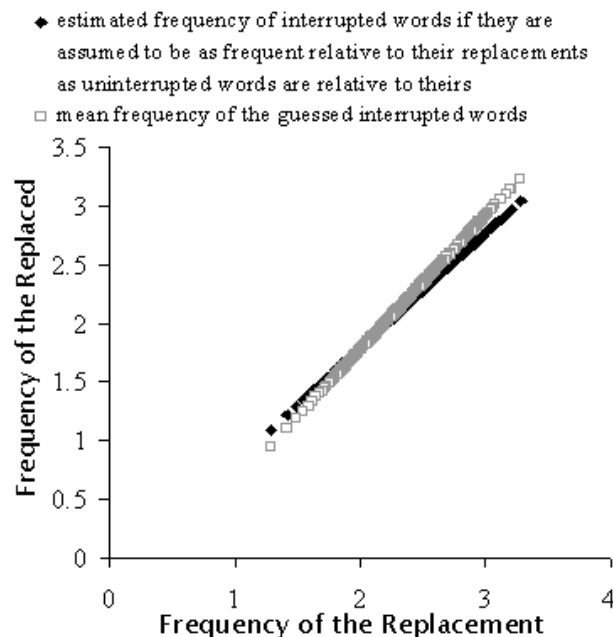
In order to make the data even more comparable to data obtained in studies of semantic substitution errors, all cases in which the replaced was interrupted were then eliminated from the sample, leaving 489 tokens. The frequency asymmetry was still observed. If anything, it became stronger: mean frequency of the replaced was 436 words/million while mean frequency of the replacement was 193 words/million,  $t(488)=8.058, p<.0005$ ; the replaced was more frequent than the replacement in 63% of the cases,  $\chi^2(1)=31.51, p<.0005$ . The correlation between the frequency of the replaced and the frequency of the replacement was present as well ( $r=.38$ ). Thus, the finding that the replaced tends to be more frequent than the replacement cannot be due to inclusion of interrupted replaced words in the present study and their exclusion from previous studies.

While Hotopf (1980: 100) and DelViso et al.'s (1991) report results that are in the same direction (in Hotopf, 1980, 56% of intrusions are more frequent than the corresponding targets; 53% in DelViso et al., 1991), albeit non-significant, Harley and MacAndrew (2001) do not. Furthermore, Harley and MacAndrew's (2001) sample is even larger than the present one ( $N=783$  for Harley and MacAndrew

2001<sup>11</sup>, vs. N=489 here). Furthermore, mean word frequencies in the full sample of repairs (199 for the replaced and 102 for the replacement) are similar to those in Harley and MacAndrew (2001) (153 and 165.5 respectively). It is possible that differences between methods of frequency estimation can account for the discrepancy between the present study and Harley and MacAndrew (2001). Frequencies used in Harley and MacAndrew (2001) are based on the written Brown Corpus (Francis and Kuçera, 1982) and the ones used here are based on the spoken Switchboard corpus (Godfrey et al., 1992). In addition to being spoken, the Switchboard corpus also has the advantage of being larger than the Brown corpus, more recent (the texts used in the Brown corpus were published in 1961), and being the same corpus as the one from which the disfluency tokens are drawn.

### Coder Bias

Identification of an interrupted replaced word necessarily involves guessing whereas identification of a word that has been produced completely does not. Thus a possible explanation for why interrupted words tend to be of lower frequency than uninterrupted words is that my guesses are biased in favor of the hypothesis. That is, it is possible that I tend to come up with words that are lower in frequency than the words the speaker intended to produce. As the results of the previous section show, the frequency of the replaced word is correlated with the frequency of the replacement word. We can use this finding to assess the hypothesis of observer bias. If the frequency of the interrupted replaced words is lower relative to the corresponding replacement words than the frequency of the uninterrupted replaced words is relative to their corresponding replacements, the hypothesis of observer bias would be confirmed.



**Figure 7.** Frequencies of replaced interrupted words can be objectively estimated from the frequencies of the corresponding replacement words based on the relationship between the frequencies of replaced words and replacement words observed with uninterrupted tokens. These estimated frequencies of interrupted replaced words are then compared to the frequencies of guessed interrupted replaced words. If guesses are biased in favor of low-frequency words, the frequencies of guessed words would be lower than expected (the gray line would be below the black line). The present figure shows that this is not the case.

<sup>11</sup> Inferred from the df of the t-test comparing semantic targets and intrusions (Harley & MacAndrew 2001: 408).

First, the strength of the correlation between the frequency of the replaced and the frequency of the replacement does not depend on whether the replaced word is interrupted ( $r=.39$  when the word is interrupted vs.  $.38$  when it is not). More importantly, as Figure 7 shows, the replaced word, if anything, tends to be more frequent relative to the corresponding replacement when I had to guess its identity than when I did not. In other words, frequencies of guessed words are not lower than expected based on the frequencies of the corresponding replacement words, contradicting the hypothesis of observer bias in favor of low-frequency words.

The average frequency for interrupted words used in the sample (and guessed by me) was 47.5 words/million while the average estimated frequency based on the relationship between the frequencies of replaced uninterrupted words and the corresponding replacements was 46.5 word/million. Thus, the hypothesis that the difference in frequency between replaced and replacement words are due to observer bias is disconfirmed.

### Erroneous vs. Suboptimal Choices and the Main Interruption Rule

Replacement repair does not always involve an outright speech error. Levelt (1983, 1989) proposed that repairs involving speech errors are very different from repairs that involve lexical choices that are considered suboptimal but not erroneous by the speaker. Levelt (1989: 481) proposes that “words that are not errors themselves tend to be completed before interruption... By interrupting a word, the speaker signals to the addressee that the word is an error. If a word is completed, the speaker intends the listener to interpret it as correctly delivered.” Levelt (1983: 63) shows that in his corpus, 32% of immediate repairs of erroneously uttered words (91/284) involve interrupting the word, while only 11% of immediate repairs of suboptimal lexical choices involve interruption of the repaired (20/175). Furthermore, there is evidence that semantic substitution errors are more likely to involve low-frequency words than high-frequency words (Harley & MacAndrew, 2001). If high-frequency replaced words are more likely to be merely inappropriate rather than erroneous low-frequency replaced words, the frequency effect observed in the present study could be ascribed to the error severity effect observed by Levelt. (Of course, this argument cuts both ways. One could also argue that the error severity effect is a frequency effect in disguise.)

For maximum coding reliability, only instances of repair in which the replaced was not interrupted were analyzed. There was no tendency for repairs involving speech errors to involve less frequent words than repairs involving suboptimal lexical choices. The results are shown in Table 1. No differences in frequency between erroneous and suboptimal words are significant. Thus, we can reject the hypothesis that high-frequency words are less likely to be interrupted because they are less likely to be uttered in error for the present sample.

Length	Erroneous Words	Suboptimal Words
3	3.42 N=107	3.35 N=112
4	2.62 N=41	2.63 N=67
5	2.26 N=8	2.34 N=18
6	2.05 N=14	2.00 N=13

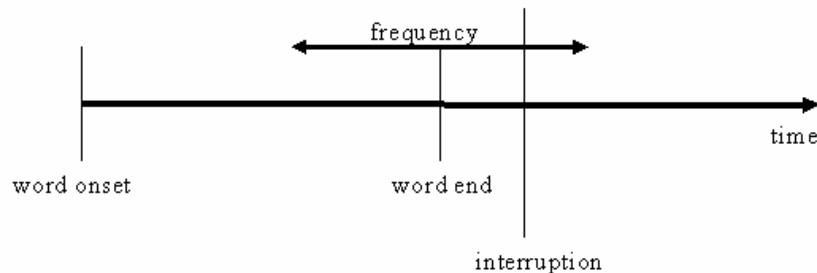
**Table 1.** Log frequencies of uninterrupted replaced erroneous vs. suboptimal words: erroneous words do not tend to be less frequent.

In addition to the problem of accounting for the present data, the hypothesis that high-frequency words are less likely to be interrupted than low-frequency words because high-frequency words are less likely to be produced as errors runs into problems with Logan's (1982) data. In his study, replacement was triggered by an external stop signal, rather than erroneous or inappropriate production. Nonetheless, a frequency effect was present.

A way to maintain the Main Interruption Rule (Levelt, 1983, 1989), which states that the speaker interrupts speech production as soon as the occasion for repair is detected, in the face of the present data would be to say that speakers are slower to detect that a high-frequency word is wrong or inappropriate than they are to detect the incorrectness of a low-frequency word. However, this hypothesis cannot account for Logan's (1982) experimental data where there is no error to be detected.<sup>12</sup> To account for why 'the' is typed completely after the stop signal is presented while less frequent words are truncated without invoking a preference to maintain constituent continuity would be to say that the detection of the stop signal is slowed down when a high-frequency word is being produced. It is not clear why this should be the case. If anything, production of a high-frequency word should be less taxing and demand fewer cognitive resources than the production of a low-frequency word, leaving more cognitive resources free to be used in perceiving the stop signal. Thus, if anything, we would predict the perception of the stop signal to be faster while a high-frequency word is being produced than while a low-frequency word is under construction.

### Frequency and Duration

There is a strong negative correlation between word frequency and word duration ( $r = -.72$  in the present sample), which remains even when number of segments is controlled (in the present sample,  $r = -.63$  for three-segment,  $-.55$  for four-segment,  $-.53$  for six-segment,  $-.35$  for seven-segment, and  $-.65$  for eight-segment uninterrupted replaced words; no correlation is observed for five-segment words,  $r = -.04$ ). Thus, frequent words tend to be shorter than rare words even when number of segments is controlled (as previously found in corpus studies by Gregory et al., 2000, and Jurafsky et al., 2001). This finding is predicted by the hypothesis that high frequency leads to automatization of production but it suggests that the effect of frequency on interruptibility may be accounted for by the effect of frequency on duration.

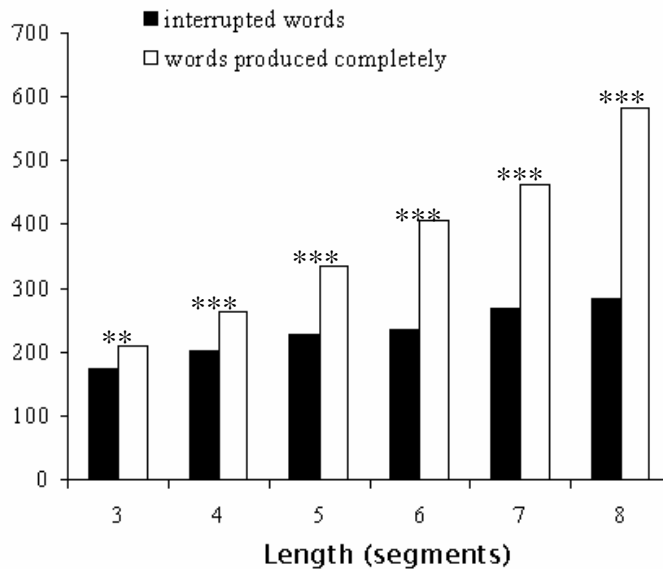


**Figure 8.** A model in which the only variable affected by frequency is word duration. Time since word onset is indicated by the thick line. Vertical lines mark important points in time, such as the end of the word and the location of interruption. The arrows attached to a vertical line indicate the extent to which variation in frequency can influence the location of the vertical line. In this model, the only point in time whose location is influenced by frequency is the end of the word, which can fall after or before the fixed location of interruption.

<sup>12</sup> One could claim that the 'error' being detected is the fact that production is still continuing. Then one could say that detection of this fact is more difficult when the word being produced is more frequent. However, this would mean that the production of high-frequency words is less cognitively penetrable than the production of low-frequency words, which is precisely the claim of the automaticity hypothesis.

The simplest model of the effect of frequency on interruptibility is that there is no effect. How long it takes a speaker to reach and carry out the decision to interrupt a word is independent of frequency. Rather, all that frequency influences is word duration. As Figure 8 shows, the location of interruption relative to the onset of the word is fixed in this model. The only variable affected by frequency is the duration of the word. When frequency is low enough, the word becomes so long that the decision to interrupt speech is carried out before the word is produced completely.

Under this model, words are interrupted only if they are sufficiently long. Therefore, there should be no difference in duration between the remainders of interrupted replaced words and uninterrupted replaced words. This is not the case in the data. Overall, remainders of interrupted words (mean duration = 217 ms) are shorter than uninterrupted replaced words (mean duration = 316 ms):  $t(1136)=15.97$ ,  $p<.001$ . Figure 9 shows the results broken down by the length of the replaced in segments. This result indicates that interruption comes earlier in time, relative to the beginning of the to-be-replaced word, when the word is interrupted than when it is not, contrary to the predictions of the model in Figure 8. Thus, there is something about uninterrupted words that delays interruption when these words are produced. This is consistent with Logan's (1982) results regarding the very frequent word 'the': while typers took less time to type 'the' than other words, the hypothesis that the difference in typing speed accounted for the result was ruled out because the time it took typists to stop while producing 'the' was longer than the time it took them to stop while producing other words.

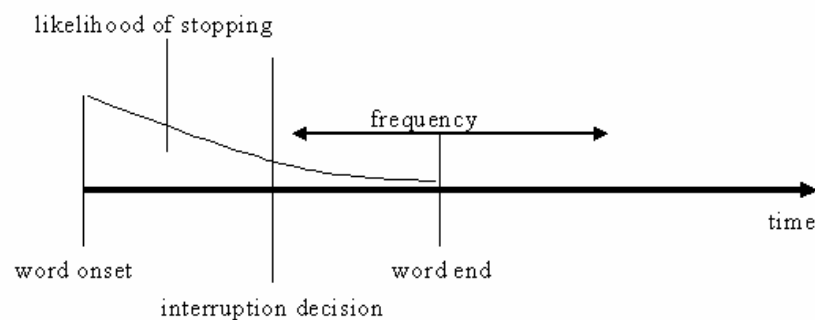


**Figure 9.** Remainders of interrupted words are shorter than uninterrupted replaced words.

An objection that could be raised to our interpretation of the data in Figure 9 is that the coder could sometimes erroneously code words that are interrupted very late during their production as uninterrupted. This is presumably not a problem with Logan's (1982) data because typing involves a discrete response while speech is continuous and involves extensive co-articulation (e.g., Coleman, 2003; Öhman, 1966), which means that the cues for the final segment can be present much earlier in the word. Furthermore, there may be more coarticulation in high-frequency words than in low-frequency words (Yun, 2006). This is in fact suggested by the data in Figure 3 where the probability of interrupting the word drops off just before the word is completed.

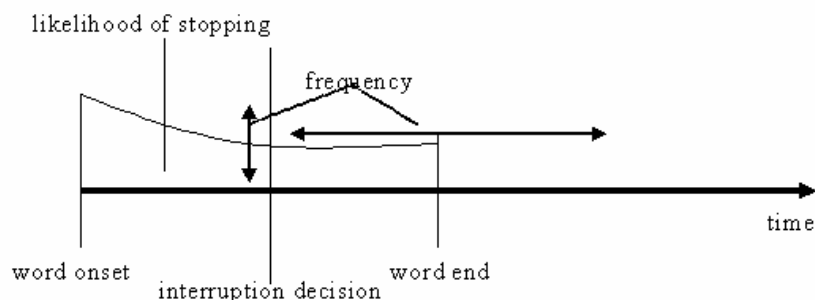
However, the data in Figure 3 can be interpreted in multiple ways, including misperception, a tradeoff between the speaker's desire not to interrupt the word and the desire to interrupt as soon as possible, and generally early detection of errors with interruption being sometimes delayed until the end of the word. In addition, within-word interruption is reliably accompanied by a particular cue, the presence of glottalization (Ladefoged et al., 1973). There was very little disagreement between the present author and the Switchboard corpus coders (only 31 words were eliminated based on this criterion). In addition, the word lengths for uninterrupted words used in the present study do not include the word-final segment. It is highly unlikely that the coders would have coded a word as uninterrupted without perceiving the final segment. Finally, the correlation between word duration and frequency is negative. Therefore, if we were to project the durations of the remainders of interrupted words from the relationship between frequency and duration found in uninterrupted words, we would expect durations of remainders of interrupted words to be longer than the durations of uninterrupted words because the frequencies of interrupted words are lower than frequencies of uninterrupted words.

The data presented so far are sufficient to reject the simple model in Figure 8. The differences in duration between the remainders of interrupted and uninterrupted words are too great to be ascribed to differences in duration between the corresponding complete words. However, the data presented thus far and Logan's (1982) results for typing 'the' are consistent with the model is shown in Figure 11. This model relies on the assumption that the closer the speaker is to the end of the word when s/he reaches the decision that the word is to be replaced, the less likely s/he will be to stop immediately. One can think of the speaker as choosing the better of two evils: to stop immediately, interrupting a cohesive constituent, or to continue producing material that will need to be replaced. In other words, the speaker can be thought of as choosing between violating the Continuity hypothesis (Clark & Wasow, 1998) vs. violating the Main Interruption Rule (Levelt, 1983, 1989). The smaller the amount of material that remains to be produced to avoid interrupting the word, the more likely the speaker is to choose producing the word to the end. Since frequency influences word duration, the amount of material that needs to be produced to complete the word will be smaller in a high-frequency word than in a low-frequency word.



**Figure 11.** A model in which likelihood of interrupting the speech stream immediately is lower if the amount of material that remains to be produced or time that it takes to complete the word is small. The likelihood of stopping immediately is shown by the height of the curved line. The higher the curve at a certain point in time, the higher the likelihood that the word will be interrupted immediately if the decision to interrupt is made at that point in time. In this model, the closer a speaker is to the end of the word, the less likely s/he is to interrupt speech production immediately. Word duration is influenced by frequency, so a speaker is more likely to be close to the end of the word when deciding to interrupt speech production if the word is frequent than if it is rare. Thus, in a frequent word, the interruption decision is likely to occur at a point when likelihood of stopping immediately is low.

An alternative model is presented in Figure 12. Here, the speaker's reluctance to interrupt a word is simply greater if the word is frequent than if the word is rare, regardless of how much linguistic material remains to be produced and how much time it would take to complete the word. There may be an effect of duration but frequency has an effect on likelihood of stopping that is independent of duration.



**Figure 12.** A model in which interruption is dispreferred in frequent words. Frequency in this model influences both the duration of the word and the likelihood of stopping immediately if the interruption decision is reached during word production as indicated by the arrows being attached to the curve indicating likelihood of stopping immediately (the curve is not crucial for this model and could be replaced by a horizontal line).

The difference between the two models lies in whether frequency has any effect on interruption when duration is controlled (both models can account for an independent effect of duration since duration is uncontroversially influenced by factors other than frequency such as speaking rate and number of segments). In order to examine this issue, a binomial logistic regression was conducted. Duration, frequency, and number of segments were entered into the analysis as covariates. Number of segments was subsequently excluded because it was not statistically significant as a separate predictor. Thus, the analyses presented below included only duration and frequency as covariates. Both were significant at the .0001 level on the full sample. The sample was then split by number of syllables so that monosyllabic and multisyllabic words were submitted to the regression analysis separately. Both frequency and duration were significant in both analyses. Frequency was significant with  $p=.001$  for multisyllabic and  $p<.0001$  for monosyllabic words. Duration was significant with  $p=.014$  for multisyllabic and  $p=.01$  for monosyllabic words,  $N=717$  for multisyllabic words,  $N=1032$  for monosyllabic words. Thus we can tentatively conclude that frequency has some effect on interruptibility that is not mediated by the effect of frequency on duration.<sup>13</sup>

## Conclusion

When a speaker intends to replace a word s/he has started producing, s/he has the choice of stopping immediately, obeying Levelt's (1983) Main Interruption Rule, or delaying interruption until the word is completed, obeying Clark and Wasow's (1998) Continuity Hypothesis. The present study has argued that the speaker's choice is influenced by word duration and word frequency. Speakers prefer not

<sup>13</sup> A necessary caveat for this conclusion is that our estimates of frequency and duration are imperfect. The full model achieved only 61% accuracy in predicting whether the word was broken when the word was multisyllabic and 75% accuracy when the word was monosyllabic, suggesting that there is much room for improvement in modeling interruptibility. Perhaps, frequency would not account for any variance that duration does not account for as well if our estimate of duration were better. However, the fact that including number of segments or number of syllables as an additional predictor does not reduce the significance of frequency suggests that this is unlikely.

to interrupt high-frequency words. This effect provides novel empirical support for the hypothesis that the production of high-frequency words is more automatic, being both faster and less susceptible to conscious control than the production of low-frequency words (Bybee, 2002; Logan, 1982). Thus Bybee's (2002) hypothesis that reductive sound change starts with high-frequency words because the production of such words is more automatic is at least psychologically plausible. In addition, the present study found that speakers tend to replace suboptimal lexical choices by less frequent but more appropriate words, supporting the idea that high-frequency words are accessed faster than low-frequency words (e.g., Jescheniak and Levelt, 1994). Thus, frequent words are easier to access, faster to produce, and harder to interrupt than rare words.

## References

- Anderson, J.R. (2000). *Cognitive psychology and its implications*. New York: Worth.
- Beattie, G.F., & Butterworth, B.L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, 22, 201-11.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.
- Boomer, D. S. (1965). Hesitation and grammatical encoding. *Language and Speech*, 8, 148-58.
- Bybee, J.L. (2002). Word frequency and context of use in the lexical diffusion of phonemically conditioned sound change. *Language Variation and Change*, 14, 261-90.
- Bybee, J.L. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.
- Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1430-50.
- Clark, H.H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, 37, 201-42.
- Coleman, J. (2003). Discovering the acoustic correlates of phonological contrasts. *Journal of Phonetics*, 31, 351-72.
- DelViso, S., Igoa, J.M., & Garcia-Albea, J.E. (1991). On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research*, 20, 161-85.
- DuBois, J.W. (1974). Syntax in mid-sentence. *Berkeley Studies in Syntax and Semantics*, 1, III:1-25.
- Fowler, C. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech*, 28, 47-56.
- Fowler, C.A., & Housum, J. (1987). Talkers' signaling of "old" and "new" words in speech and listeners' use of this distinction. *Journal of Memory and Language*, 26, 489-504.
- Fox, B.A., Hayashi, M., & Jasperson, R. (1996). Resources and repair: A cross-linguistic study of syntax and repair. In E. Ochs, E. Schegloff, & S. Thompson (eds.), *Interaction and grammar* (pp. 185-237). Cambridge, UK: Cambridge University Press.
- Fox, B.A., & Jasperson, R. (1995). A syntactic exploration of repair in English conversation. In P.W. Davis (ed.), *Alternative linguistics: Descriptive and theoretical modes* (pp. 77-134). Amsterdam: John Benjamins.
- Francis, W.N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston, MA: Houghton Mifflin.
- Garrett, M. (2001). Now you see it, now you don't: Frequency effects in language production. In E. Dupoux (ed.), *Language, brain and cognitive development* (pp. 227-40). Cambridge, MA: MIT Press.
- Godfrey, J.J., Holliman, E.C., & McDaniel, J. (1992). SWITCHBOARD: Telephone Speech Corpus for Research and Development. *IEEE ICASSP*, 1517-20.
- Goldman-Eisler, F. (1958). Speech production and predictability of words in context. *Quarterly Journal of Experimental Psychology*, 10, 96-106.

- Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. San Diego: Academic Press.
- Gregory, M., Raymond, W.D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (2000). The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society*, 35, 151-66.
- Harley, T.A., & MacAndrew, S.B.G. (2001). Constraints upon word substitution speech errors. *Journal of Psycholinguistic Research*, 30, 395-18.
- Hooper, J. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In W. Christie (ed.), *Current progress in historical linguistics* (pp. 96-105). Amsterdam: North Holland.
- Hotopf, W. (1980). Semantic similarity as a factor in whole-word slips of the tongue. In V. Fromkin (ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 97-110). New York: Academic Press.
- Howes, D.H., & Solomon, R.L. (1951). Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, 41, 401-10.
- Jescheniak, J.D., & Levelt, W.J.M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 824-843.
- Jescheniak, J.D., Meyer, A.S., & Levelt, W.J.M. (2003). Specific word frequency is not all that counts in speech production: Comments on Caramazza, Costa et al. (2001) and new experimental data. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 432-438.
- Jurafsky, D., Bell, A., & Girand, C. (2002). The role of lemma in form variation. In C. Gussenhoven & N. Warner (eds.), *Papers in Laboratory Phonology VII* (pp.1-34). Berlin/New York: Mouton de Gruyter.
- Jurafsky, D., Bell, A., Gregory, M., & Raymond, W.D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J.L. Bybee & P. Hopper (eds.), *Frequency and the emergence of linguistic structure* (pp. 229-54). Amsterdam: John Benjamins.
- Kapatsinski, V.M. (2005). Measuring the relationship of structure to use: Determinants of the extent of recycle in repetition repair. *Berkeley Linguistics Society*, 30, 481-92.
- Ladefoged, P., Silverstein, R., & Papçun, G. (1973). Interruptibility of speech. *Journal of the Acoustical Society of America*, 54, 1105-1108.
- 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.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W.J. Hardcastle & A. Marchal (eds.), *Speech production and speech modeling* (pp. 403-39). Dordrecht: Kluwer.
- Logan, G.D. (1982). On the ability to inhibit complex movements: A stop-signal study of typewriting. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 778-792.
- Maclay, H., & Osgood, C.E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, 15, 19-44.
- Mowrey, R., & Pagliuca, W. (1995). The reductive character of articulatory evolution. *Rivista di Linguistica*, 7, 37-124.
- Nooteboom, S. (1980). Speaking and unspeaking: detection and correction of phonological and lexical errors in spontaneous speech. In V. Fromkin (ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 87-95). San Diego: Academic Press.
- Öhman, S.E.G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 39, 151-168.
- Oldfield, R.C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17, 273-281.
- Pagliuca, W., & Mowrey, R. (1987). Articulatory evolution. In A.G. Ramat, O. Carruba, & G. Bernini (eds.), *Papers from the VIIth International Conference on Historical Linguistics* (pp. 459-72). Amsterdam: John Benjamins.

- Phillips, B.S. (2001). Lexical diffusion, lexical frequency, and lexical analysis. In J.L. Bybee & P. Hopper (eds.), *Frequency and the emergence of linguistic structure* (pp. 123-36). Amsterdam: John Benjamins.
- Slevc, L.R., & Ferreira, V.S. (2006). Halting in single word production: A test of the perceptual loop theory of speech monitoring. *Journal of Memory and Language*, *54*, 515-540.
- Stemberger, J.P. (1984). Structural errors in normal and agrammatic speech. *Cognitive Neuropsychology*, *1*, 281-313.
- Tannenbaum, P.H., Williams, F., & Hillier, C.S. (1965). Word predictability in the environments of hesitation. *Journal of Verbal Learning and Verbal Behavior*, *4*, 134-40.
- Yun, G. (2006). The effects of lexical frequency and stress on coarticulation. Paper presented at Berkeley Linguistics Society Meeting 32, Berkeley, CA.

