Language and Speech
http://las.sagepub.com/

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

Vsevolod Kapatsinski

Language and Speech 2010 53: 71
DOI: 10.1177/0023830909351220

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

Vsevolod Kapatsinski

University of Oregon, U.S.A.

Key words

automaticity

disfluencies

repair

speech production

word frequency

Abstract

In spontaneous speech, speakers sometimes replace a word they have just produced or started producing by another word. The present study reports that in these replacement repairs, low-frequency replaced words are more likely to be interrupted prior to completion than high-frequency words, providing 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). Frequency appears to have an effect on interruptibility even when word duration is statistically controlled. In addition, the frequency of the replaced word is positively correlated with the frequency of the word it is replaced by, supporting the hypothesis that high-frequency words are easier to access in production (Kittredge, Dell, Verkuilen, & Schwartz, 2008): the more frequent the target, the more frequent an inappropriate word needs to be to overcome the target and be uttered, only to be replaced.

1 Introduction

Shiffrin and Schneider (1977) define a process as being automatic if it "nearly always becomes active in response to a particular input" (p.155) being "largely beyond subject control" and "difficult to suppress or alter" (p.156). Ladefoged, Silverstein, & Papçun (1973, p.1105) hypothesized that if speech production is a sequence of automatized

Acknowledgements: I would like to thank the NIH for financial support through Training Grant DC-00012 and Research Grant DC-00111, as well as Adam Albright, Tessa Bent, Joan Bybee, Ken DeJong, Barbara Fox, Dan Jurafsky, David Pisoni, Rena Torres-Cacoullos, and the audiences at HDLS VII and the Workshop on Gradience and Frequency Effects in Phonology for helpful comments.

Address for correspondence. Vsevolod Kapatsinski, Department for Linguistics, 1290 university of Oregon, Eugene, OR 97403 U.S.A.; <e-mail: vkapatsi@uoregon.edu>

Language and Speech 0023-8309; Vol 53(1): 71–105; 351220; DOI:10.1177/0023830909351220
http://las.sagepub.com

actions, similar to the stroke of a tennis player or golfer, then "there are some moments in the stream of speech when a speaker would find it more difficult to interrupt himself than at other moments" since "any tennis player or golfer knows how difficult it is to interrupt himself in the middle of a stroke." In particular, if words are units of speech production (e.g., Bybee, 2002; Dell, 1986; Jescheniak & Levelt, 1994; Oldfield & Wingfield, 1965), one may expect it to be more difficult to interrupt speech production within a word than between two words.

To test this hypothesis, Ladefoged et al. (1973) asked subjects to say the sentence "Ed had edited Id" repeatedly. On some trials, as the subjects were producing the sentence, they were asked to stop and say [ps] instead. Contrary to the hypothesis, stopping was equally fast in all locations within the sentence. Similarly, Logan (1982) found that skilled typists were able to stop typing very quickly in the middle of a word and did not tend to delay stopping until the end of the word: the number of letters typed following the signal indicating that they should stop did not depend on the number of letters remaining until the end of the current word. These findings suggested that the word is not a unit whose production is automatized in the traditional sense of the term in either speech or typing (Logan, 1982).

However, not all words are created equal. Bybee (2002) and Logan (1982) suggest that the production of high-frequency words and phrases is more automatic than the production of low-frequency words and phrases because 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. Linguistic evidence for a link between cohesion and frequency of use has come from studies showing that high-frequency words are more likely to undergo reductive sound change (Bybee, 2002; Hooper, 1976). For instance, Hooper (1976) discusses vowel reduction in English, which tends not to affect the second vowel in the low-frequency word "mammary" but does affect the second vowel in the more frequent "memory" and has proceeded to completion in the very frequent "every." Bybee (2001, pp.79–83) and Phillips (2001) suggest that reductive sound changes that involve lexical diffusion from high-frequency to low-frequency words can be explained by a repetition-driven increase in gestural compression (see Browman & Goldstein, 1992, and Mowrey & Pagliuca, 1995, for an in-depth discussion of how gestural compression can lead to different types of sound changes).

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 of sound change, 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, p.269; Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 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 & 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, p.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 (although cf. Bard et al., 2000, for an argument against the view that speakers model the knowledge of their listeners).

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 influence contextual predictability, which serves as a constraint on how much reduction the speaker thinks s/he can get away with. Thus, the finding that reductive sound change starts with high-frequency words does not provide unambiguous evidence for the hypothesis that the production of high-frequency words is more automatic than the production of low-frequency words. In the present article, we test this hypothesis by examining whether the production of a frequent word is less likely to be interrupted than the production of a rare word when the word is uttered in error.

While neither of the experimental studies of word interruptibility was specifically designed to test for frequency effects, Logan (1982, Experiment 3) observed that if 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 two 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. In naturalistic speech production, interrupting speech as soon as possible is not a top priority for the speaker (e.g., Jasperson, 1998; Jefferson, 1974), unlike in stop-signal experiments (Ladefoged et al., 1973; Logan, 1982). Therefore, whereas interruption may not be delayed until the end of a word in stop-signal experiments unless the word is extremely frequent, especially if the word is produced by typing, even lower frequency values may delay interruption when interrupting quickly is not as important for the speaker. In effect, having an explicit goal to stop as soon as possible, as in the stop-signal experiments, may raise the minimum frequency that words must have in order to resist interruption. The use of typing may have a similar effect: since typing is not as automatic as speech, the minimum frequency of use needed for resisting interruption may be higher in typing. Thus in the present article I analyze a corpus of conversations among native English speakers (Switchboard: Godfrey et al., 1992). 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-replaced word if the word is frequent than if it is infrequent.

Interruption of the flow of speech for the purpose of replacing a word that has already been produced or whose production has begun is termed *replacement repair.* 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, for example, *wa* in (3), the *remainder*, reserving the term *replaced word* for the inferred complete lexical item, for example, *watch* in (3). The replaced word will be said to be *interrupted* if it is marked as such by the Switchboard Corpus (Meteer & Taylor, 1995) and subsequently auditorily confirmed to be recognizably incomplete (Fox et al., in press) by the author. Examples in (1)–(4) show that the speaker has a choice of producing the replaced word completely or interrupting its production. The present article 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 (see also Jasperson, 1998, for evidence that semantic replacements are distinct from phonological ones), which eliminated approximately 12% of the cases in which the replaced word was not interrupted.[1]

(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.

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 the amount of material that 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), Cook

1 For the interrupted replaced words, it is impossible to estimate whether a word is excluded because there is no semantic relationship or because the word's identity cannot be guessed.

(1969), Goldman-Eisler (1958, 1968) and Tannenbaum, Williams, & Hillier et al. (1965) demonstrated that hesitations tend to occur in between-word transitions of maximum uncertainty, as indicated by low transitional or Cloze probability.[2] 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 (Clark & Wasow, 1998; DuBois, 1974; Fox & Jasperson, 1995; Kapatsinski, 2005; Levelt, 1983; Maclay & Osgood, 1959; Nooteboom, 1980). Definitions of major constituent boundaries differ somewhat across studies, with most researchers taking such boundaries to include clause, direct object, and oblique (prepositional phrase) boundaries (Clark & Wasow, 1998; Fox & Jasperson, 1995; Kapatsinski, 2005; Maclay & Osgood, 1959).[3] Based on this work, Clark and Wasow (1998, p.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; in Fox et al.'s, in press, sample, 56% of repetitions involve a single word). The likely reason, according to the Continuity Hypothesis, is that the speaker wants to produce the entire prepositional phrase without interruption.[4] 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, Hayashi, & Jasperson, 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 (Clark & Wasow, 1998).[5]

2 Cloze probability is the experimentally derived probability of a listener guessing the word given the rest of the sentence.

3 The status of the subject–verb boundary is questionable (Fox & 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.

4 Alternatively, speakers may have difficulty initiating production from the middle of a cohesive unit (Clark & Wasow, 1998).

5 I have been able to find only one example of the latter in Switchboard.

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 (i.e., if the probability of the word preceding the constituent boundary given the following word is high). 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 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 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, p.490).

Thus prior findings support what I shall call the *Gradient Continuity Hypothesis*: speakers prefer to produce *cohesive sequences* with a continuous delivery, whether the cohesion is caused by linguistic constituency or probability of co-occurrence. In the present study we will examine the hypothesis that high-frequency words are more cohesive (and therefore less interruptible) than low-frequency words.

2 Analysis I: Frequency and number of segments

Both versions of the Continuity Hypothesis predict that interrupting the flow of speech within a word should be dispreferred relative to interrupting the flow of speech between words (Clark & Wasow, 1998). Thus the probability of interrupting the flow of speech at any point within a word should be lower than the probability of interrupting the flow of speech between words.

On the other hand, Levelt (1983) proposes the *Main Interruption Rule*, which claims that the speaker interrupts the flow of speech as soon as the present word is detected to be an error. In the present article, I will test a gradient version of the rule, which militates against producing erroneous material: the more material is produced between error detection and the interruption of speech, the greater the violation. The Continuity Hypothesis and the Main Interruption Rule are inherently in conflict. If speakers obey both, a difference between short words and long words is predicted. It appears safe to assume that the distance between the point in time at which the inappropriateness of a long word can be detected and the point in time at which the word is complete is larger than the distance between the error detection point and the word completion point in a short word (Brédart, 1991). A speaker who has detected that a word s/he is producing is inappropriate faces a choice: continue producing the word to completion, satisfying the Continuity Hypothesis, or interrupt immediately, satisfying the Main Interruption Rule. When the word is long, the amount of material that remains to be produced to satisfy the Continuity Principle is large, so immediate interruption is likely to win out. When the word is short, the amount of material is small, so the word is likely to be produced completely. Thus, long words are predicted to be more likely to be interrupted than short words (for tests of this hypothesis in other corpora, see also Brédart, 1991, and Fox et al., in press).

In addition to the word length effect and the overall dispreference for word-internal interruption, the Gradient Continuity Hypothesis predicts that high-frequency words should be less likely to be interrupted than low-frequency words. The reason for this prediction is that the production of high-frequency words is hypothesized to be more automatic and thus less subject to monitoring and/or interference than the production of low-frequency words. Thus, the analysis in this section tests whether interrupted replaced words tend to have lower token frequency than uninterrupted replaced words even when length is controlled.

2.1 Methods

2.1.1 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.[6] 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. They are reminded of the topic of conversation immediately prior to its beginning. The version of Switchboard annotated for disfluencies contains about 1,400,000 words (Fillmore, Jurafsky, Ide, & Macleod, 1998). The corpus is annotated with a special symbol ('+') that marks the locations of disfluencies, including 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 (see Meteer & Taylor, 1995, and Shriberg, 1994, for details on the coding procedures). 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.

2.1.2 Exclusions

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

6 Repetition repairs were excluded from the present report. While these repairs exhibit the same pattern as replacement repairs discussed here (Kapatsinski, 2007), the motivation for interrupting a to-be-repeated word is unclear, given that the main function of repetition repair is assumed to be buying time (Fox et al., in press; Rispoli, Hadley, & Holt, 2008; Schegloff, 1979). Since we do not have a compelling reason to believe that the speaker wants to interrupt a to-be-repeated word as soon as possible by default, it is not clear that we can make claims about whether interruption of a to-be-repeated word is delayed by high frequency.

(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

Uninterrupted replaced words were excluded as well if they were not semantically related to the replacement. First, the identity of an interrupted word is very difficult to guess if it is not semantically related to its replacement. Second, a speaker who produces a phonologically motivated repair may never have the replaced word in mind during the production process at all. Thus, the speaker of (9) may never have selected the word "shock" for production. Instead, the phoneme /k/ may have been erroneously selected. If this is the case, then the frequency and duration of "shock" are irrelevant the production process. 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 and the example would be excluded. 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 replaced words shorter than three segments or longer than eight segments were excluded because there were very few such words in the sample. Length was defined in terms of segments for the purposes of this exclusion and the analyses presented below. A segment-based length measure appears appropriate because there is no effect of syllable structure on interruption location (Levelt, 1983; Nooteboom, 1980). In the present dataset, the interruption rate was 27% between syllables, within the onset of a syllable, or at the onset–nucleus boundary, and 32% at the nucleus–coda boundary. Interruption within the coda was unlikely (17%), but this is likely to be an effect of proximity to the final boundary of the word, rather than syllable structure.

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 (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 (Clark & Wasow, 1998; Fox et al., in press), which should not be included in a sample of repairs because they disallow word-internal interruption. That is, if the word were interrupted, the utterance would not be an example of the grammatical construction, which may make it dispreferred relative to the version that does not feature word-internal interruption and instantiates a grammaticalized construction. 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 left dislocation, in which the "replaced" is a hyponym of the replacement. 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 (see Clark & Wasow, 1998, Fox et al., in press, and Jefferson, 1974, for further discussion of preplanning of interruption in repair).

(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, for example, *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/her 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 cues were sufficient for exclusion. As observed by Ladefoged et al. (1973) and Jasperson (1998, chapter 7), 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. The number of instances of repair retained in the sample after exclusions was 1794.

2.1.3 Defining the dependent variable

The dependent variable in the present study is whether the replaced word is interrupted or produced completely. While every token of repair features an interruption in the flow of speech, this interruption can follow the replaced word, in which case the replaced word will be said to be *uninterrupted* or it can leave the replaced word incomplete, or *interrupted*. Words were considered to be interrupted if the author could not detect an acoustic cue to the final segment of the word in the signal, detected glottalization, and the Switchboard coders have identified the word as interrupted (coded with a dash, Meteer & Taylor, 1995). Thirty-one 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. For the present analyses, all word-internal locations were considered equally internal. This is motivated by the fact that I did not find any significant differences among word-internal locations. Words interrupted early in their production (in terms of number of segments that have elapsed since the beginning of the word prior to interruption) did not differ significantly from words interrupted late in terms of frequency of use. Furthermore, there were no effects of word-internal syllabic constituency. All locations within the syllable were as likely to host an interruption as the between-syllable location (with a constant interruption rate of 27%–32%). Thus, all word-internal locations of interruption were treated identically in the present analyses.

2.1.4 Measuring frequency and length

For the purposes of estimating word length in terms of segments, affricates, diphthongs, syllabic nasals and liquids and /ɜr/ were coded as single segments. This decision was made because cases in which a diphthong was interrupted (e.g., ‘ho- houses’ [ha- haʊsɪz], ‘Thursda- Thursday’ [θɜrzdɛ- θɜrzdɛɪ]) and cases in which the schwa was produced without the following sonorant (e.g., ‘mothe- mother’ [mɑðə- mɑðɚ], ‘eve-even’ [ivə- ivən]) were exceedingly rare, and there were no cases in which an affricate was reported as 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. As recent repetitions are likely to be more important for present behavior than earlier repetitions, and as the production of a word may be more automatic when it is related to the topic of conversation and therefore somewhat predictable (Fowler, 1988; Gregory et al., 2000), frequency within the corpus under analysis is arguably a more appropriate measure than frequency within some other corpus (e.g., Francis & Kučera, 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 (not discussed in this article) 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.

Frequency was logarithmically scaled because ease of lexical access in both perception (Howes & Solomon, 1951) and production (Oldfield & Wingfield, 1965) is linearly correlated with log frequency better than with raw frequency.[7] For three-segment and one-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.

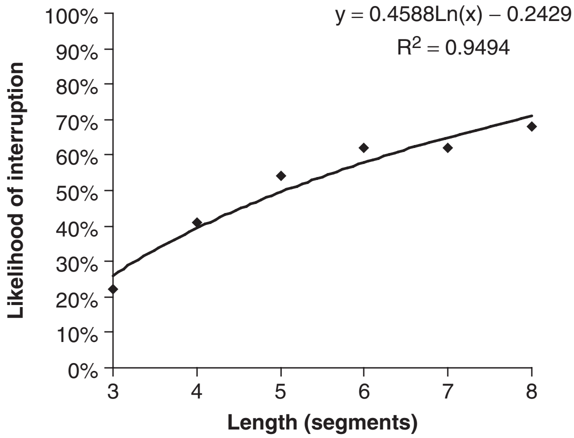
2.2 Results and discussion

Figure 1 shows that the longer the replaced word, in terms of number of segments, the more likely it is to be interrupted (see also Brédart, 1991; Fox et al., in press). The relationship between number of segments in a word and likelihood of word-internal interruption is well approximated by a logarithmic curve. Figure 1 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

7 The basic idea is that the difference in frequency between a word that occurs only once in the corpus and a word that occurs ten times is psychologically much more significant than the difference between a word that occurs 1000 times in the corpus and one that occurs 1010 times.

Figure 1

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[8]



guessed are not included in the sample. As a result, Figure 1 is likely to underestimate the true likelihood of word-internal interruption in replacement repair.

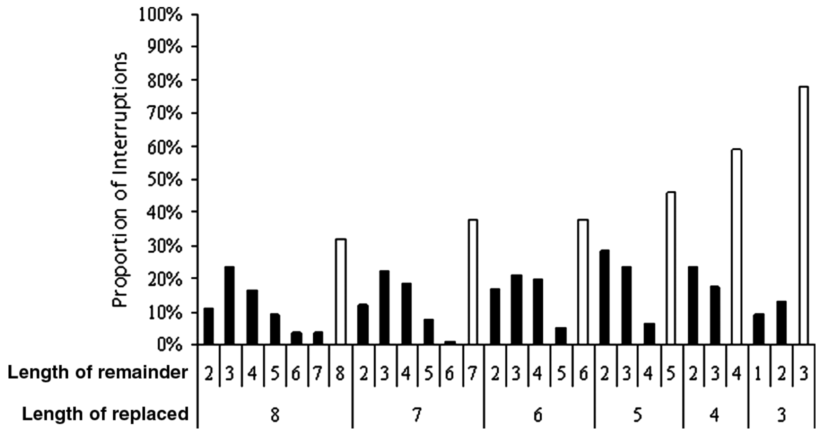
In general, to-be-replaced words are produced completely more often than they are interrupted in the present sample (61% of all words in the sample are not interrupted), a result that has also been observed by Boomer (1965), Clark and Wasow (1998), Levelt (1983), and Maclay and Osgood (1959). However, Figure 1 suggests that this is an artifact of the fact that there are more short words than long words in the English lexicon: if the English lexicon consisted mostly of words that are five or more segments long, most to-be-replaced words would be interrupted prior to completion (Fox et al., in press, report that this is indeed the case in languages where most words are long, such as Finnish). 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 at least seven possible word-internal locations for interruption and one between-word location. Thus, if an eight-segment replaced word is interrupted 68% of the times when it is replaced, as shown in Figure 2, the between-word transition is privileged relative to any one of the word-internal transitions as a location for interruption.

Figure 2 confirms that the data in Figure 1 do not contradict the Continuity Hypothesis (contra Brédart, 1991). As predicted by the Continuity Hypothesis, for any word length, the between-word transition is a significantly more common location for the onset of repair (interruption in the flow of speech) than any one of the

8 Grouping the words by number of syllables rather than number of segments produces the same result. However, a segmental measure is favored by the fact that there was no effect of syllable structure on interruption location: interruption was not dispreferred in syllable-internal positions while segment-internal interruption is impossible given Switchboard's orthographic coding and never observed auditorily even in diphthongs and affricates.

Figure 2

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

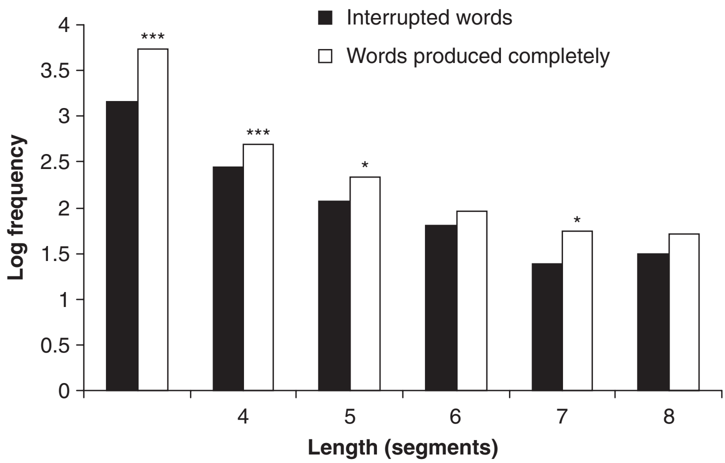


within-word transitions according to the chi-square test (the closest contender among between-word transitions is the location after the third segment in eight-segment words that hosts 18 interruptions relative to 27 cases in which an eight-segment word is not interrupted; the difference is significant, $\chi^2(1) = 7.2$, $p < .01$). Additional supporting evidence is provided by Fox et al. (in press) who find that in their sample English is the language that has the weakest preference for completing to-be-replaced words prior to interrupting the flow of speech. Finally, the Switchboard Corpus consists mostly of expression of opinion on, sometimes, relatively unfamiliar topics and agreement or disagreement with those opinions. This context may make speakers particularly careful about wording (B. Fox, personal communication) and thus particularly likely to monitor their speech for inappropriate words and catch them prior to completion. On the other hand, as many cases of word-internal interruption are not included in the sample because the identity of the interrupted word could not be guessed, while most cases of between-word interruption are included, this result does not provide strong evidence for the Continuity Hypothesis. For that, we will have to turn to frequency effects.

High-frequency words tend to have fewer segments and Figures 1 and 2 show that words that have fewer segments are less likely to be interrupted. Therefore, for an effect of frequency on word interruptibility to be established, it needs to be shown that it holds when number of segments is controlled. This is shown in Figure 3. 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, word interruption was a significant predictor of frequency, $t(1746) = 9.934$, $p < .0005$; frequency is a significant predictor of word interruption when frequency and length are entered into a binomial logistic regression as covariates, $p < .001$), as well as for

Figure 3

Words that are interrupted tend to be less frequent than words produced completely (* $p < .05$, *** $p < .0005$, two-tailed)



three-, four-, five-, and seven-segment words considered separately (for three-segment words, $t(798) = 7.821$, $p < .0005$;[9] for four-segment words, $t(406) = 4.092$, $p < .0005$; for five-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.

The data in Figure 3 indicate that even when length (in segments) is controlled, high-frequency words are less likely to be interrupted than low-frequency words. Assuming that the speaker interrupts the flow of speech as soon as possible once s/he decides to replace a word under construction (Levelt, 1983), the interruption is delayed until the end of the word when the word is frequent, supporting the hypothesis that the production of high-frequency words is more automatic than the production of low-frequency words. However, alternative explanations are possible. The following sections will in turn address explanations based on speed of accessing the replacement word, observer bias, error detectability, and the effect of frequency on word duration.

3 Analysis II: Frequency of the replaced versus frequency of the replacement

The finding that the replaced words that are interrupted tend to be lower in token frequency than the replaced words that are not interrupted is consistent with the hypothesis that high frequency makes the production of a word more automatic, and hence more difficult to interrupt. An alternative explanation may lie in how fast the speaker can access the replacement word relative to initiating the production of the word that the replacement word eventually replaces.

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

Biber , Johansson, Leech, Conrad, and Finegan (1999, p.1059) and Clark and Wasow (1998) observed that the indefinite article is less prone to being repeated than the definite article and proposed, in the words of Biber et al., that "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. (1999) and Clark and Wasow (1998) support the hypothesis by pointing out that *an* is repeated very rarely because, 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* (Biber et al., 1999). In addition (Biber et al., 1999, pp.1061–1062) 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*.

The hypothesis is that a high-frequency word is likely to come to mind faster than a low-frequency word (a standard assumption in models of speech production, e.g., Dell, 1986; Jescheniak & Levelt, 1994; Oldfield & Wingfield, 1965) provides a possible prediction for when a to-be-replaced word's production will be interrupted. 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.[10] Thus, this hypothesis predicts that interrupted words should be rarer *relative to their replacements* than uninterrupted words without necessarily predicting a difference in absolute token frequency between interrupted and uninterrupted words.

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.

3.1 Methods

The sample of repairs described in the previous section was reduced for the purposes of comparing the frequency of the replaced word to the frequency of the replacement word. To make the sample of replacement words comparable to the sample of replaced words, replacement words shorter than three or longer than eight segments were eliminated. In addition, 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

10 A reviewer asks why a speaker would start producing the replaced word at all if it is less frequent and less appropriate. A likely explanation appears to be priming. Other things being equal, a frequent word should be activated more easily than an infrequent word, but an infrequent (and, perhaps, inappropriate) word may be activated more than a frequent (and appropriate) word in context.

came out of, were included in Analysis I but excluded in Analysis II. 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. Thus the total sample size in Analysis II is 1030.

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

3.2 Results and discussion

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, a non-significant difference, $t(1029) < 1$, $p = .4$. These results are inconsistent with the hypothesis that frequent to-be-replaced words are less likely to be interrupted than rare to-be-replaced words because frequent to-be-replaced words but not rare to-be-replaced words are likely to be more frequent than the corresponding replacements. Whether a replaced word is interrupted appears to be independent of the frequency of the replaced word relative to the frequency of the corresponding replacement word.

While interrupted words tend to be lower in token frequency than uninterrupted words, words that replace interrupted words also tend to be less frequent than words that replace words that are produced completely (Figure 4), indicating that the speed

Figure 4

Words that replace interrupted words tend to be less frequent than words that replace words that are produced completely

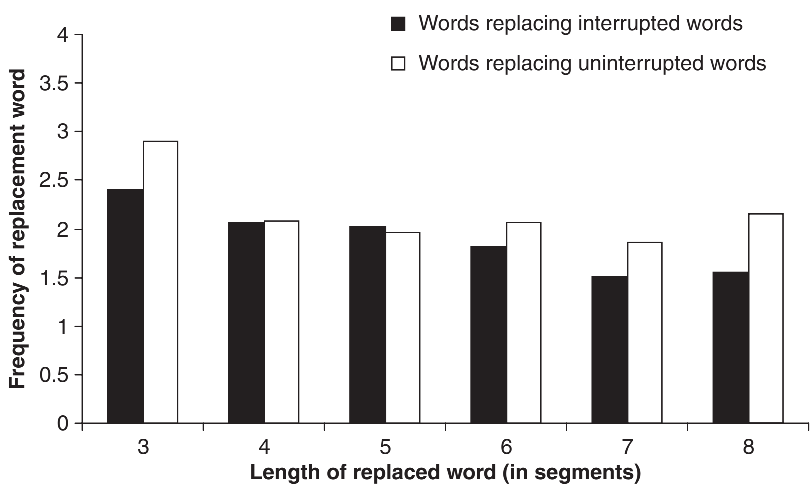
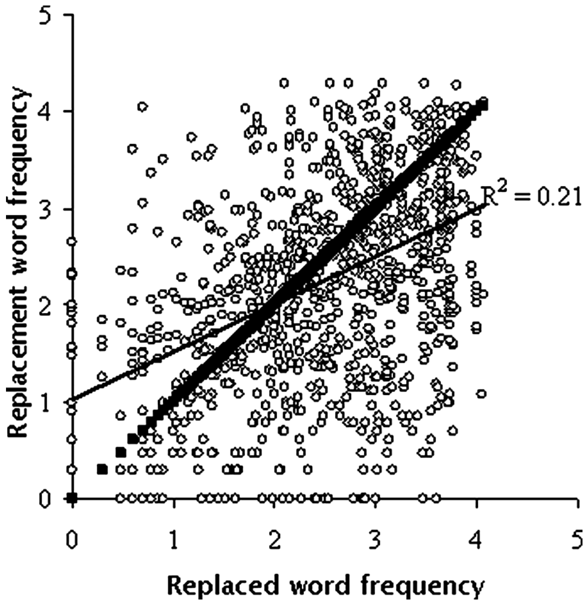


Figure 5

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



of accessing the replacement word is unlikely to account for why frequent replaced words are less likely to be interrupted than rare replaced words.

The reason for the finding that words that replace interrupted words tend to be less frequent than words that replace uninterrupted words (Figure 4) can be inferred from the data in Figure 5, which shows that the frequency of the replacement is positively correlated with the frequency of the replaced. Thus, as the frequency of the replaced increases, so does the frequency of the replacement. This finding has also been observed in studies of lexical substitution errors (DelViso, Igoa, & Garcia-Albea, 1991; Harley & MacAndrew, 2001; Hotopf, 1980; Kittredge et al., 2008; 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). Kittredge et al. (2008) point out that the positive correlation between the frequency of the replaced and the frequency of the replacement provides support for the hypothesis that frequent words can be accessed more quickly than rare words in speech production (Biber et al., 1999; Clark & Wasow, 1998; Dell, 1986; Jescheniak & Levelt, 1994). The more frequent the target, the more frequent a competitor needs to be to become activated before the target and thus be erroneously uttered in production.

4 Analysis III: Observer 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 Analysis II 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 by estimating the expected frequencies of the interrupted replaced words given the frequencies of the corresponding replacement words by correlating frequencies of replaced uninterrupted words with the corresponding replacement words. If the frequency of 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 in favor of the hypothesis that high frequency decreases interruptibility would be confirmed.

4.1 Results and discussion

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 6 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. 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 interrupted and uninterrupted replaced words is due to observer bias is disconfirmed.

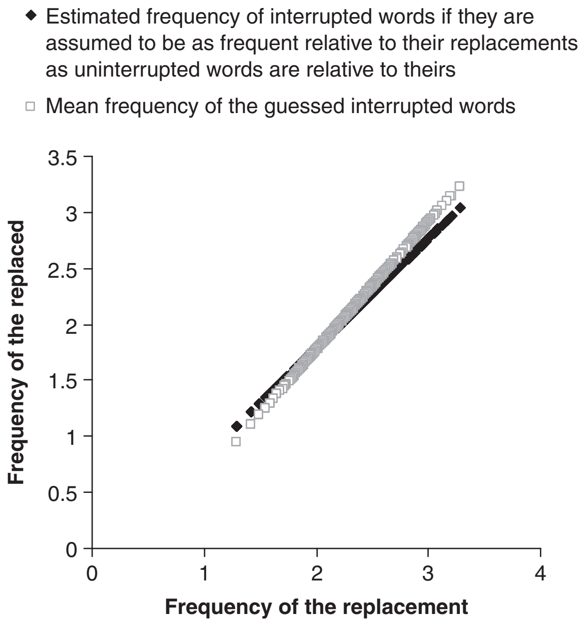
5 Analysis IV: Error detection

An assumption made by the model in Analysis II is that interruption is triggered by awareness of an alternative, rather than recognition of the inappropriateness of the word being produced. Alternatively, interruption and the search for an alternative could be triggered by detection of inappropriateness. Under the latter 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 that the word being produced is erroneous (what he calls the *Main Interruption Rule*). Under the Main Interruption Rule, words may be interrupted if they are detected to be errors quickly. The speed of detection could plausibly depend on the severity of the error. Thus, a word that is merely inappropriate and simply "needs

Figure 6

Estimated frequencies of interrupted replaced words compared to the frequencies of guessed interrupted replaced words. If guesses were 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), which does not appear to be the case



some qualification" (Levelt, 1983, p.63) may be less likely to be interrupted than a word that is an outright speech error, as found by Levelt (1983).[11] There is some evidence that low-frequency words are more likely to be involved in semantic speech errors (e.g., Harley & MacAndrew, 2001). If high-frequency words are less likely to be errors 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. Thus, in the present analysis, I test whether high-frequency words are more likely to be merely inappropriate (and less likely to be speech errors) than low-frequency words.

5.1 Methods

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

11 Levelt himself (1989, p.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 word-internal interruption is used by the speaker to tell the listener whether the replaced word is a speech error.

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. For maximum coding reliability, only instances of repair in which the replaced was not interrupted were analyzed.

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

Levelt (1983, p.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, p.481) suggests that appropriateness repairs 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), where the replaced clearly would not be considered a speech error if left uncorrected.

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

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

These cases can be compared to tokens in which the replaced and the replacement are incompatible because there are aspects of their meanings that are demonstrably different as in (28)–(29). The replaced in (28)–(29) is severe enough to be considered a speech error, a semantic intrusion. 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*, and *quite* is replaced by *not really*. On the other hand, cases in which the replaced and the replacement are quantifiers whose ranges of possible values overlap, as when *several* is replaced by *a few*, can be considered repairs not involving a speech error but rather mere inappropriateness.

(28) A sixty-seven **Chev-**, uh, *Mustang*. (Error)

(29) You may be able to take **care**, take *advantage* of that. (Error)

An additional class of repairs that can be said to involve 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*. (Error)

(31) I was watching the **ra-**, the *TV* today. (Error)

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

Table 1

Log frequencies of uninterrupted replaced erroneous vs. merely inappropriate words: erroneous words do not tend to be less frequent

Length	*Erroneous words*	*Inappropriate 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 $N = 13$

speech errors. This does not include cases such as *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.

5.2 Results and discussion

There was no tendency for repairs involving speech errors to involve less frequent words than repairs involving mere inappropriateness. The results are shown in Table 1. No differences in frequency between erroneous and merely inappropriate 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.

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 without acknowledging the Continuity Hypothesis, 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. 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.

In addition, the Main Interruption Rule predicts that speech should be interrupted prior to planning of the replacement: interruption occurs as soon as an error is detected and before an alternative formulation has been planned. Blackmer and Mitton (1991) point out that this hypothesis predicts that there should be no repairs in which there is no pause between the replaced and the replacement. Contrary to the Main Interruption Rule, Blackmer and Mitton (1991) find such "impossible" repairs to be fairly common, forming 19% of covert repairs in their data. Some of the examples involve clear instances of speech error correction (e.g., "willfiddily" being replaced by "willfully") rather than correction of mere inappropriateness. Furthermore, the average delay between interruption and replacement is lower for error correction repairs than for appropriateness correction repairs (119 ms vs. 304–402 ms in Blackmer & Mitton, 1991, p. 197), contradicting Levelt's proposal that error correction repairs involve immediate interruption while appropriateness repairs frequently do not, which would predict that error correction repairs would require more planning time following interruption. This failing of the Main Interruption Rule appears to be avoided only if the rule is combined with the Continuity Hypothesis.

6 Analysis V: The effects of frequency on duration and interruptibility

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 infrequent words even when number of segments is controlled (as previously found in corpus studies by Gregory et al., 2000, and Jurafsky, Bell, Gregory, & Raymond, 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.

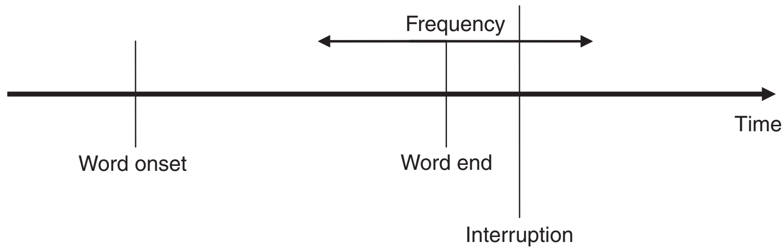
In this section, I consider three possible models of the relationship between frequency, duration, and interruptibility. The simplest model of the effect of frequency on interruptibility is that there is no effect. How long it takes a speaker to reach and carry out the decision to interrupt a word is independent of frequency. Rather, all that frequency influences is word duration. As Figure 7 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 (long enough to exceed the fixed time required for interruption). Therefore, the remainders of interrupted replaced words should be longer than uninterrupted replaced words.

An alternative model is shown in Figure 8. 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 lesser of two evils: to stop immediately,

Figure 7

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



interrupting a cohesive sequence, 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. As 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.

The final model is presented in Figure 9. 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. The difference between the model in Figure 8 and the model in Figure 9 lies in whether frequency has any effect on interruption when duration is controlled (both models can account for an independent effect of duration because duration is uncontroversially influenced by factors other than frequency such as speaking rate and number of segments).

The simple model in Figure 7 is easy to test: all we need to do is compare the durations of the remainders of interrupted replaced words and the durations of uninterrupted replaced words. Both can be measured directly. Testing the difference between the model in Figure 8 and the model in Figure 9 is more difficult. Both models require an estimate of what the duration of an interrupted replaced word would be if it were produced completely. The model in Figure 8 states that the estimated total duration of the replaced word will account for all the variance in interruptibility that frequency accounts for, while the model in Figure 9 predicts that frequency should account for additional variance in interruptibility. However, since there is no perfect estimate of what duration a word would have if it were not interrupted, the model in Figure 8 can never be definitively rejected unless its assumption that a word is less likely to be interrupted when little of it remains to be said is shown to be false.

Figure 8

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

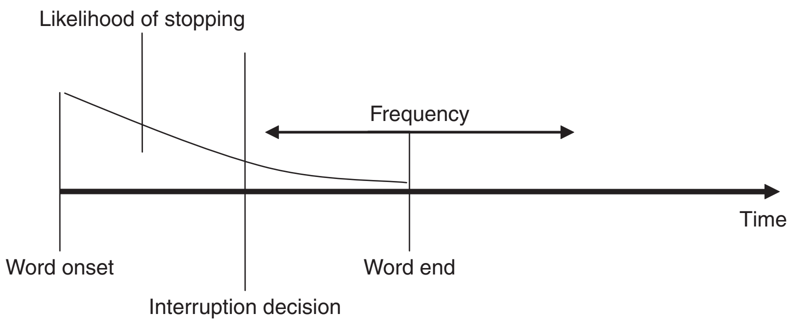
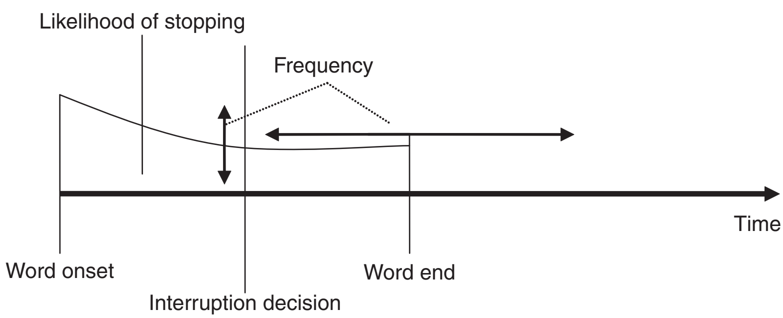


Figure 9

A model in which word-internal 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)



6.1 Methods

For each instance of repair included in the complete sample used for Analysis I (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. 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. As 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 because it can contain strong cues to the final segment's identity.

Ten samples of each of the interrupted 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 (this happened for 74 repair tokens). When more than 10 tokens were available, the 10 samples used were selected by taking the first 10 samples from Switchboard that were generated by the corpus search engine at LDC online (https://online.ldc.upenn.edu/search/) that satisfied the following three criteria: (1) the sampled word was preceded by the same word as the interrupted replaced word; (2) the sampled word belonged to the same syntactic category as the interrupted replaced word; and (3) the sampled word was in the same relation to the following clause boundary as the interrupted replaced word (adjacent or non-adjacent). The search procedure involved searching for the two-word sequence first. If fewer than 10 tokens that satisfied the other two criteria for inclusion were found, the search was expanded to include all instances of the replaced word regardless of preceding context.

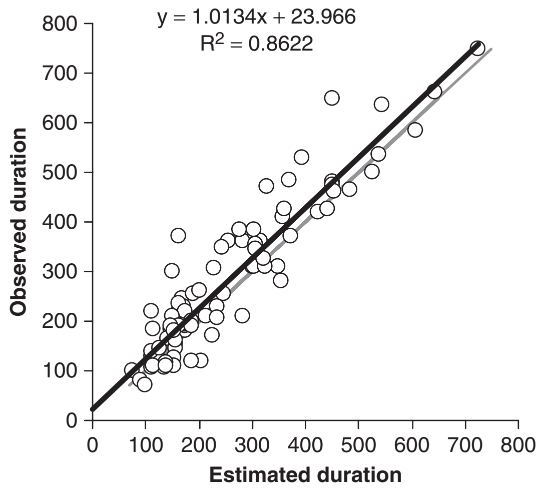
In order to assess the validity of duration measurements, I collected a sample of 100 uninterrupted replaced words and estimated their durations in the same way as for the interrupted replaced words. The sample was collected by random sampling of the uninterrupted replaced word types followed by random sampling within each type so that only one token of each word was included in the sample. Ten tokens could be collected for 93/100 words. For the remaining words, the estimate was based on one to nine tokens. The estimated durations are compared to the observed durations in Figure 10. As shown in Figure 10, estimated duration is a very good predictor of observed duration, predicting 86% of the variance. Estimated duration is also slightly lower than observed duration, suggesting that replaced words tend to be lengthened. Thus the use of estimated duration for interrupted words may slightly inflate their durations relative to uninterrupted words. This bias is in favor of the hypothesis that words are interrupted just because they are long rather than because speakers prefer not to interrupt high-frequency words.[12]

Durations were measured by hand in Praat (Boersma & Weenink, 2005). 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

12 The measure assessed in Figure 10 proved to be better at predicting whether a to-be-replaced word would be interrupted than two alternative measures: multiplying the duration of the remainder of the actual replaced word by the ratio of the number of segments in the remainder to the number of segments in the complete replaced word or the duration of the replaced word produced in isolation from text by an adult female native speaker of American English (obtained from LDC's American English Spoken Lexicon, http://www.ldc.upenn.edu/cgi-bin/aesl/aesl).

Figure 10

Estimated vs. observed durations of 100 uninterrupted replaced words. The light line indicates the diagonal at which estimated and observed durations are equal. The dark line is a linear regression fit, the equation and R^2 values for which are shown on the graph



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 11 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. Durations were rounded to the nearest 10 ms.

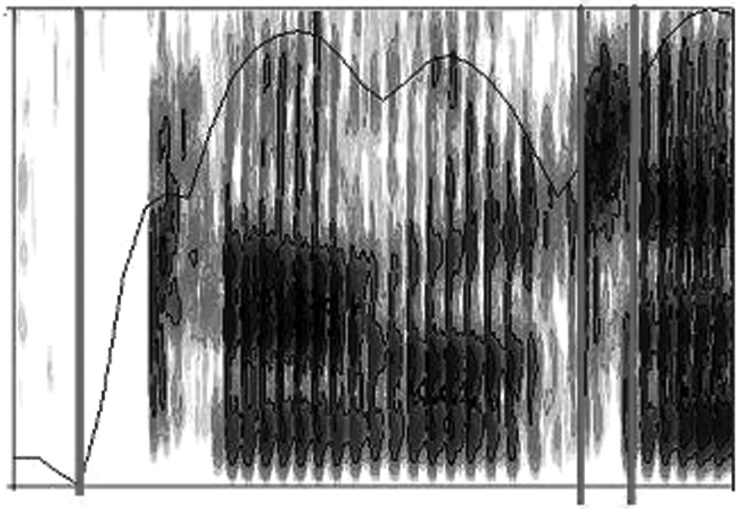
6.2 Results and discussion

The model in Figure 7, which states that words are interrupted if they are "too long" and all that frequency influences is word duration, predicts that the remainders of interrupted words should be longer than 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): $t(1136) = 15.97$, $p < .001$.[13] Figure 12 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 prior to completion than when it is not, contrary to the predictions of the model in Figure 7. 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 typists 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.

13 Note that this result does not depend on an estimate of the duration of the replaced word. Rather it is based purely on the observed duration of the remainder, which is present in the signal.

Figure 11

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

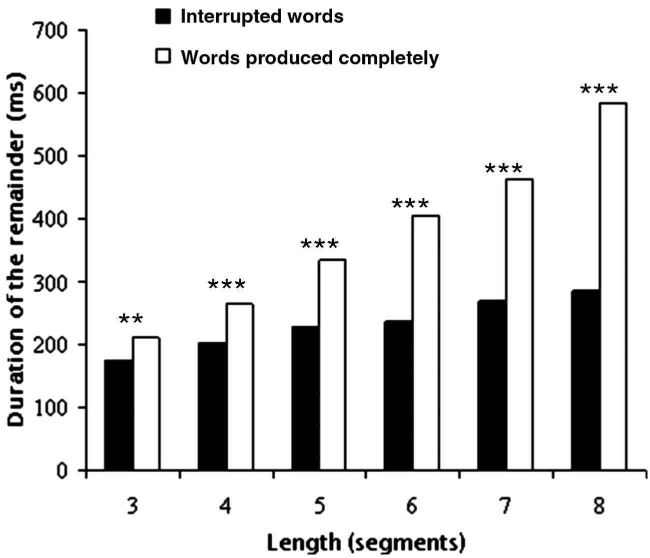


An objection that could be raised to our interpretation of the data in Figure 12 is that the coder could sometimes erroneously code words that are interrupted very late during their production as uninterrupted, inflating the difference in duration between the remainders of interrupted words and uninterrupted words. 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 2 where the probability of interrupting the word drops off just before the word is completed (also observed by Fox et al., in press).

However, the data in Figure 2 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 (Jasperson, 1998; Ladefoged et al., 1973). There was very little disagreement between the present author and the Switchboard Corpus coders on whether a given token featured a word-internal interruption or a completely produced replaced word (only 31 words were eliminated based on this criterion). 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

Figure 12

Remainders of interrupted words are shorter than uninterrupted replaced words (* $p < .01$, *** $p < .0005$, two-tailed)



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 in Figure 12 are sufficient to reject the simple model in Figure 7. The difference between the durations of the uninterrupted replaced words and the durations of the remainders of interrupted replaced words is significant at every word length and is in the opposite direction from what is predicted by the model in Figure 7. However, the data presented thus far and Logan's (1982) results for typing "the" are consistent with the model shown in Figure 8.

In order to test the difference between the model in Figure 8 and the model in Figure 9, a binomial logistic regression was conducted. Duration, frequency, and number of segments were entered into the analysis as covariates. The model in Figure 8 states that frequency influences word duration and that, if the inappropriateness of the replaced word is detected close to its end, the speaker prefers not to interrupt the word because finishing the word would satisfy the Continuity Hypothesis without violating the Main Interruption Rule too much. When the word is long it is detected to be inappropriate when much of the word still remains to be produced. As a result, completing the word would greatly violate the Main Interruption Rule, making the speaker prefer to interrupt the word. This model is different from the model in Figure 7 because here frequency delays interruption relative to the detection of inappropriateness, while in Figure 7 interruption is not delayed by frequency. However, the model in Figure 8 differs from the model in Figure 9 in that the latter states that the effect of frequency on interruptibility is only partially mediated by duration, while the model in Figure 8 claims that high frequency delays the interruption through its effect on duration. Thus, the model in Figure 8 predicts no effect of frequency when duration is controlled, while the model in Figure 9 predicts an additional effect of frequency above and beyond duration.

Number of segments was excluded from the model because it was not statistically significant as a separate predictor. Duration and frequency were both 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.

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 interrupted when the word was multisyllabic and 75% accuracy when the word was monosyllabic, suggesting that there is much room for improvement in modeling interruptibility. It is possible that frequency would not account for any variance that duration does not account for as well if our estimate of duration were better. However, Figure 10 suggests that our estimate of duration is quite good, and including number of segments, number of syllables, or some alternative duration estimate (see footnote 12 for alternatives tested) as an additional predictor does not remove the effect of frequency. This result provides some measure of confidence in the evidence supporting the model in Figure 9 over the model in Figure 8.

The crucial question is whether frequency can improve an estimate of total duration of the replaced word over the benchmark estimate obtained by taking 10 samples of the replaced word from utterances in which it is not replaced. If frequency can improve an estimate of total duration of the replaced word, it is possible that the significant improvement in predictability of word interruptibility seen in the logistic regression analysis when frequency is included as a predictor is due to an improvement in the predictability of total duration the replaced word would have if it were not interrupted.

One way to address this issue is to examine whether frequency can help us predict the durations of uninterrupted replaced words when the average duration of 10 tokens of the same word found elsewhere is taken into account.[14] As shown in

14 Thus the present sample contains 100 different types. For comparison, the sample of monosyllabic interrupted replaced words contains 334 different types. Thus the sizes of samples of estimated word lengths and word frequencies are quite comparable.

Figure 10, estimated duration accounted for 86% of variance in observed duration of the uninterrupted replaced words, while log frequency of use accounted for only 56%. Importantly, frequency of use did not improve predictability of observed duration over the estimate obtained by measuring 10 tokens of the same word elsewhere in the corpus (a linear regression with observed duration as the dependent variable shows a significant effect of estimated duration, $p < .000001$, but no significant effect of frequency, $p = .466$, and no significant interaction, $p = .821$). This result strongly suggests that the reason frequency predicts whether a replaced word is interrupted is not because it accounts for variance in word duration that our estimate of duration does not account for. Thus, the influence of frequency on interruptibility does not appear to reduce to its influence on duration and the resulting change in the cost of satisfying the Continuity Hypothesis in terms of the degree of violation of the gradient version of the Main Interruption Rule.

7 General discussion

The main finding of the present study is that frequent to-be-replaced words are less likely to be interrupted than rare to-be-replaced words. This finding provides support for an expanded version of the Continuity Hypothesis, stated by Clark and Wasow (1998) as "speakers prefer to produce constituents with a continuous delivery."

Several recent studies have shown that constituency is not an all-or-none matter and is most straightforwardly interpreted as variation in degree of cohesion: the speech units that belong to a single constituent are more tightly bound together than speech units that are separated by a constituent boundary. One factor that increases cohesion between adjacent speech units is frequency of co-occurrence. For instance, Lee and Goldrick (2008) show that when English and Korean speakers are asked to memorize lists of consonant–vowel–consonant syllables, they sometimes make errors in which a segment from one syllable "moves into" a different syllable. Importantly, pairs of segments that co-occur in the language of the speaker tend to stick together, resisting being erroneously separated through replacement of one of the segments with a segment from a different syllable. Kapatsinski (2005) finds that speakers who have interrupted their flow of speech are unlikely to restart speech production from the middle of a constituent if that constituent is cohesive, as measured by between-word transitional probability. To account for the gradience of constituency, the Continuity Hypothesis can be rephrased as "speakers prefer to produce cohesive sequences without interruption."

If frequency of co-occurrence increases cohesion and therefore constituency, we may expect frequent words to be more cohesive than low-frequency words. That is, high-frequency words are more constituent-like than low-frequency words. If a high-frequency word is more of a constituent than a low-frequency word then the high-frequency word should be less likely to be interrupted, as observed in the present study. There is also evidence that a gradient version of the Continuity Hypothesis interacts with a gradient version of the Main Interruption Rule (Levelt, 1983). While speakers prefer not to interrupt cohesive sequences, they may not want to produce long strings of incorrect material (Levelt, 1983). The interaction of these two principles predicts that speakers should be most unlikely to interrupt speech production when completion of the current constituent would incur little cost in terms of amount of incorrect material that remains to be produced.

The Continuity Hypothesis can have at least two possible interpretations in a model of speech production. First, it may be the case that high-frequency words are harder to interrupt than low-frequency words because their production is automatized and thus relatively impenetrable to outside influences. Cognitive impenetrability is not a necessary consequence of automatization, as shown by Logan (1983), who found that skilled typists can stop typing just as quickly as unskilled typists in response to a stop signal, and by Cohen and Poldrack (2008), who found that training in a motor sequence learning task did not slow stopping while improving the ability of subjects to simultaneously perform a secondary task. However, automatization *often* leads to cognitive impenetrability (Shiffrin & Schneider, 1977). For instance, the typing of "the" in Logan (1982) is both more automatic (as indicated by typing speed) and more difficult to inhibit (as measured by stopping time) than the typing of lower-frequency words. More anecdotally, many people have difficulty inhibiting curse words when they are inappropriate, which seems difficult to explain if automaticity of production does not lead to impaired inhibition. In addition, King and Kutas (1998) and Münte et al. (2001) have provided ERP data showing that, even when grammatical class is controlled, frequent words produce more activation in Broca's area (linked to speech production, e.g., Dronkers, 1996) than rare words in a visual lexical decision task, which does not require word production, suggesting that activation of motor representations is more automatic for frequent words than for rare words. As argued by Bybee (2002), the automaticity hypothesis also provides a natural account of articulatorily-driven sound changes, such as those documented by Hooper (1976).

The other major possible explanation for the Continuity Hypothesis is that it is more difficult to detect the inappropriateness of a high-frequency word than it is to detect the inappropriateness of a low-frequency word. In other words, high-frequency words may be more likely to fly under the radar of an error/appropriateness monitor. This possibility is not ruled out by the present study, although it is difficult to reconcile with the presence of a frequency effect in the stop-signal experiment reported by Logan (1982). In the stop-signal experiment, there is no error to be detected, hence why stopping is more difficult in the most frequent word "the" than in less frequent words is unexplained. 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 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. Nonetheless, an experimental stop-signal study of speech production that explicitly manipulates word frequency would be extremely useful to test the error-monitoring hypothesis.

8 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. The influence of duration is argued to lie in both its influence on the probability of detecting the inappropriateness of a word while it is still being produced (Brédart, 1991) and in that speakers may be more prone to delay interruption until the completion of the word when little of the word remains to be produced. In addition, speakers prefer not 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 a correlation between the frequency of the replaced word and the frequency of its replacement, supporting the idea of a frequency-driven competition for production (e.g., Dell, 1986; Jescheniak & Levelt, 1994; Levelt, 1989; Oldfield & Wingfield, 1965): the more frequent the target word, the more frequent a semantic competitor needs to be in order to be erroneously produced instead of the target (see also Kittredge et al., 2008). Thus, frequent words are easier to access, faster to produce, and harder to interrupt than rare words.

References

BARD, E. G., ANDERSON, A. H., SOTILLO, S., AYLETT, M., DOHERTY-SNEDDON, G., & NEWLANDS, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, **42**, 1–22.

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**, 201–211.

BIBER, D., JOHANSSON, S., LEECH, G., CONRAD, S., & FINEGAN, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.

BLACKMER, E. R., & MITTON, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, **39**, 173–194.

BOERSMA, P., & WEENINK, D. (2005). *Praat: doing phonetics by computer* (Version 4.3) [Computer program]. Retrieved January 27, 2005, from http://www.praat.org/

BOOMER, D. S. (1965). Hesitation and grammatical encoding. *Language and Speech*, **8**, 148–158.

BRÉDART, S. (1991). Word interruption in self-repairing. *Journal of Psycholinguistic Research,* **20**, 123–138.

BROWMAN, C. P., & GOLDSTEIN, L. (1992). Articulatory Phonology: An overview. *Phonetica*, **49**, 155–180.

BYBEE, J. L. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

BYBEE, J. L. (2002). Word frequency and context of use in the lexical diffusion of phonemically conditioned sound change. *Language Variation and Change*, **14**, 261–290.

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

COHEN, J. R., & POLDRACK, R. A. (2008). Automaticity in motor response learning does not impair response inhibition. *Psychonomic Bulletin and Review*, **15**, 108–115.

COLEMAN, J. (2003). Discovering the acoustic correlates of phonological contrasts. *Journal of Phonetics*, **31**, 351–372.

COOK, M. (1969). Transition probabilities and the incidence of filled pauses. *Psychonomic Science*, **16**, 191–192.

DELL, G. S. (1986). A spreading activation theory of retrieval in language production. *Psychological Review,* **93**, 283–321.

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**, 161–185.

DRONKERS, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, **384**, 159–161.

DuBOIS, J. W. (1974). Syntax in mid-sentence. *Berkeley Studies in Syntax and Semantics*, **1**, III.1–III.25.

FILLMORE, C. J., JURAFSKY, D., IDE, N., & MacLEOD, C. (1998). An American national corpus: A proposal. In *Proceedings of the First International Conference on Language Resources and Evaluation*. Paris: European Language Resources Association.

FOWLER, C. A. (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: 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.

FOX, B., WOUK, F., HAYASHI, M., FINCKE, S., TAO, L., SORJONEN, M.-L., et al. (in press). A cross-linguistic investigation of the site of initiation in same turn self repair.

FRANCIS, W. N., & KUČERA, H. (1982). *Frequency results and discussion of English usage: Lexicon and grammar*. Boston, MA: Houghton Mifflin.

GODFREY, J. J., HOLLIMAN, E. C., & McDANIEL, J. (1992). SWITCHBOARD: Telephone Speech Corpus for Research and Development. *IEEE ICASSP*, I517–I520.

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, CA: 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**, 151–166.

HARLEY, T. A., & MacANDREW, S. B. G. (2001). Constraints upon word substitution speech errors. *Journal of Psycholinguistic Research*, **30**, 395–418.

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**, 401–410.

JASPERSON, R. M. (1998). *Repair after cut-off: explorations in the grammar of focused repair of the turn-constructional-unit-so-far*. Unpublished doctoral dissertation, University of Colorado, Boulder, CO.

JEFFERSON, G. (1974). Error correction as an interactional resource. *Language in Society*, **2**, 181–199.

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.

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.229–254). Amsterdam: John Benjamins.

KAPATSINSKI, V. (2005). Measuring the relationship of structure to use: Determinants of the extent of recycle in repetition repair. *Berkeley Linguistics Society*, **30**, 481–492.

KAPATSINSKI, V. (2007, July). *Does high frequency lead to automaticity? A corpus study.* Poster presented at the Workshop on Variation, Gradience, and Frequency in Phonology, Stanford, CA, U.S.A. Retrieved November 2, 2009, from http://www.stanford.edu/dept/linguistics/linginst/nsf-workshop/Kapatsinski_Poster.pdf

KING, J. W., & KUTAS, M. (1998). Neural plasticity in the dynamics of human visual word recognition. *Neuroscience Letters*, **244**, 61–64.

KITTREDGE, A. K., DELL, G. S., VERKUILEN, J., & SCHWARTZ, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, **25**, 463–492.

LADEFOGED, P., SILVERSTEIN, R., & PAPÇUN, G. (1973). Interruptibility of speech. *Journal of the Acoustical Society of America*, **54**, 1105–1108.

LEE, Y., & GOLDRICK, M. (2008). The emergence of sub-syllabic representations. *Journal of Memory and Language,* **59**, 155–168.

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.403–439). 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.

LOGAN, G. D. (1983). Time, information, and the various spans in typewriting. In W. E. Cooper (Ed.), *Cognitive aspects of skilled typewriting* (pp.197–224). New York: Springer.

MACLAY, H., & OSGOOD, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, **15**, 19–44.

METEER, M., & TAYLOR, A. (1995). Dysfluency annotation stylebook for the Switchboard Corpus. Retrieved November 2, 2009, from ftp://ftp.cis.upenn.edu/pub/treebank/swbd/doc/DFLbook.ps

MOWREY, R., & PAGLIUCA, W. (1995). The reductive character of articulatory evolution. *Rivista di Linguistica*, **7**, 37–124.

MÜNTE, T. F., WIERINGA, B. M., WEYERTS, H., SZENTKULI, A., MATZKE, M., & JOHANNES, S. (2001). Differences in brain potentials to open and closed class words: Class and frequency effects. *Neuropsychologia*, **39**, 91–102.

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, CA: 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.

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.123–136). Amsterdam: John Benjamins.

RISPOLI, M., HADLEY, P., & HOLT, J. (2008). Stalls and revisions: A developmental perspective on sentence production. *Journal of Speech, Language, and Hearing Research,* **51**, 953–966.

SCHEGLOFF, E. (1979). The relevance of repair to syntax-for-conversation. In T. Givón (Ed.), *Syntax and Semantics 12: Discourse and Syntax* (pp.261–288). New York: Academic Press.
SHIFFRIN, R. M., & SCHNEIDER, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, **84**, 127–190.
SHRIBERG, E. E. (1994). *Preliminaries to a theory of speech disfluencies*. Unpublished doctoral dissertation, University of California, Berkeley, CA, U.S.A.
SILVERBERG, N. (1998). *How do word meanings connect to word forms?* Unpublished doctoral dissertation, University of Arizona, Tucson, AZ, U.S.A.
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**, 134–140.
YUN, G. (2006, February). *The effects of lexical frequency and stress on coarticulation*. Paper presented at the 32nd Annual Meeting of the Berkeley Linguistics Society, Berkeley, CA, U.S.A.