Alice's adventures in *um*-derland: Psycholinguistic sources of variation in disfluency production

Scott H. Fraundorf[a] and **Duane G. Watson**[a]

[a]Department of Psychology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Abstract

This study tests the hypothesis that three common types of disfluency (fillers, silent pauses, and repeated words) reflect variance in what strategies are available to the production system for responding to difficulty in language production. Participants' speech in a storytelling paradigm was coded for the three disfluency types. Repeats occurred most often when difficult material was already being produced and could be repeated, but fillers and silent pauses occurred most when difficult material was still being planned. Fillers were associated only with conceptual difficulties, consistent with the proposal that they reflect a communicative signal whereas silent pauses and repeats were also related to lexical and phonological difficulties. These differences are discussed in terms of different strategies available to the language production system.

Keywords

disfluency; language production; discourse

Speech is fraught with interruptions. A speaker telling a story might well produce an utterance such as (1), which contains several *disfluencies*, or interruptions and gaps in the fluent speech stream.

> **(1)** Um and then as she's looking at the baby she th she thinks to herself that she should take it with her.

Disfluencies take many forms and can appear in many places in speech. Example (1) contains what are generally considered two distinct types of disfluency (e.g., Maclay & Osgood, 1959). First, before *and then*, the speaker produces *um*, delaying production of the utterance. Then, while producing *she thinks*, the speaker interrupts the utterance only to eventually repeat *she thinks*. These disfluencies differ in several ways. For instance, one occurs before the utterance has started while the other interrupts it halfway through; the repeated *she thinks* interrupts production of a word, but *um* does not.

Theories of language production must thus account for why interruptions in speech take so many forms. In the present work, we propose a psycholinguistic explanation: Because

disfluencies often[1] result from problems in language production (Clark, 1996), the existence of multiple disfluency types may stem from differences in the strategies made available by the state of the production system at the time of the problem. Indeed, speakers' production of multiple kinds of disfluencies suggests a language production system with multiple options for responding to delays and errors.

Disfluency Types

Disfluencies in human speech take many forms. While precise taxonomies vary, most resemble the four categories used by Maclay and Osgood (1959). *Fillers* (or *filled pauses*), such as the one appearing in (2), are verbal interruptions that do not relate to the proposition of the main message—in English, most commonly *uh* and *um*, but also *er* and *ah*. (In the examples below, specific instances of disfluency are in boldface.) *Silent pauses* (or *unfilled pauses*), as in (3), are periods of silence longer than the pauses in an equivalent fluent utterance. *Repeats*, as in (4), are unmodified repetitions of a word, a part of a word, or a string of words. Finally, *repairs* (or *false starts*) are self-corrections or revisions of material already spoken[2], as in (5). (Note that repairs sometimes, but not always, contain fillers.) It is generally agreed that repairs are used when the speaker previously uttered something erroneous (although theories differ about the timing of repair onset relative to error detection; e.g., Levelt, 1983; Tydgat, Stevens, Hartsuiker, & Pickering, 2011). What are less clear are the circumstances that produce fillers, silent pauses, and repeats and the reasons that these three separate types of disfluency exist.

(2)   She grabs the fan and **uh** one pair of gloves.

(3)   She notices … a small … box that says "EAT ME."

(4)   Alice doesn't think **that cats that cats** can grin.

(5)   (5) And they sent Bill the lizard down the chimney **to find her er to see what was going on**.

One hypothesis (Bock, 1996) is that these disfluencies represent different underlying production problems or different strategies for correcting problems. The availability of multiple strategies for correcting problems and restoring fluent speech may be necessary for a production system that is tasked with not only producing correct speech but doing so quickly and efficiently (e.g., Bock, 1995). Testing how speakers' disfluencies vary as a function of the difficulties they face not only illuminates how language production can go wrong but also informs theories of how language production typically proceeds successfully.

However, little work has directly tested the hypothesis that differences in disfluencies reflect differences in what strategies are available to the production system for responding to disruptions. In general, the frequencies with which a speaker uses different types of

---

[1]Disfluencies may not always reflect planning problems. For example, speakers may sometimes deliberately use fillers like *uh* to introduce a dispreferred response (Schegloff, 2010).

[2]The disfluency literature is presently divided over the use of the term *repair*. Some authors, following Maclay & Osgood (1959), use the term to refer only to corrections of material already spoken. Others, such as Levelt (1983), term essentially all disfluencies *repairs* under the assumption that other disfluencies such as fillers and silent pauses represent covert repairs of speech still being planned. For the present study, we reserve the term *repair* to refer to overt corrections of material already spoken.

disfluencies correlate only weakly (Maclay & Osgood, 1959), so it is plausible that they reflect different processes in production. Moreover, particular disfluency types appear to be differentially influenced by task manipulations or speaker characteristics. For example, Schnadt and Corley (2006) found that the frequency of disfluent *prolongations*, such as pronouncing *the* as *thiy*, was reduced by lexical and visual accessibility of referents in a task, but not that of fillers or repairs. Similarly, Hartsuiker and Notebaert (2011) found that the number of names available for a picture increased the rate of repeats but not of a category combining fillers and silent pauses. Speakers with autism spectrum disorders (Lake, Humphreys, & Cardy, 2011) or attention deficit/hyperactive disorder (Engelhardt, Corley, Nigg, & F. Ferreira, 2010) also produce different disfluencies than control speakers.

These studies have established that various disfluency types are differentially used across speakers and tasks. In the present study, we provide evidence that the use of different disfluency types *within* a speaker and *within* a task reflect at least two differences constraining which strategies are available to responding to production problems. First, repeats can be used when a speaker is already in the process of articulating and has material available to repeat, but fillers and silent pauses are required when a speaker is between units of speech such as utterances or grammatical phrases. Second, whether speakers can quickly initiate a new conceptual plan may influence whether speakers can produce a filler or must pause silently. We detail these differences below.

Availability of Repetitions

One way that speakers may respond to delays or errors in speech production is to restart the problem utterance. Blackmer and Mitton (1991) propose, based on the short latency with which repeats can be initiated, that restarting an utterance can be performed relatively automatically so long as no additional conceptual planning is required. Clark and Wasow (1998) further argue that the ability to restart an utterance in this way leads to a deliberate *commit-and-repair* strategy: speakers begin to articulate a partially planned utterance, and if delays in planning prevent its initial fluent completion, they repeat the beginning so that the entire utterance can be presented fluently.

Although speakers might prefer to present a complete fluent utterance when possible, they may not always have recently articulated material available to repeat. Most theories of language production posit that speech is planned or prepared in units greater than a single word, at least in early stages of the production process (e.g., Bock & Cutting, 1992; Garrett, 1980; M. Smith & Wheeldon, 2004, but see Brown-Schmidt & Tanenhaus, 2006). Some problems or delays with a unit of speech may occur while it is still being planned and articulation has not begun. If there is no recently produced material available to repeat, speakers may produce other disfluencies, such as fillers and silent pauses, that do not require the availability of lexical or syntactic material.

The hypothesis that fillers and silent pauses are used when problem material is still being planned is supported by past evidence that both fillers (Clark & Fox Tree, 2002; Swerts, 1998) and silent pauses (Butterworth, 1975; 1980; Lake et al., 2011) occur more frequently at syntactic, semantic, or prosodic boundaries. These are places at which speakers have

finished one unit of speech but may encounter difficulty in planning what follows (e.g., Butterworth, 1975, 1980; Clark & Fox Tree, 2002; F. Ferreira, 2007; V. L. Smith & Clark, 1993).

Thus, we hypothesized that repeats tend to be used when a planning problem is detected *after* starting to articulate of the problem segment, while fillers and silent pauses are used when the problem occurs while that segment is still being initially planned, *before* articulation. This difference is suggested by the evidence reviewed above, but little work has directly compared disfluency types against each other.

It should be noted that this is not the only possible account of when repeats, fillers, and silent pauses are produced. For example, Maclay and Osgood (1959) propose that speakers produce fillers after they have already become disfluent as a way to avoid losing the conversational floor; this hypothesis does not predict that fillers need always be tightly linked to difficult *upcoming* material. Moreover, Levelt (1983) raises the possibility that at least some repeats occur not because of planning problems but because the repair process was erroneously initiated in response to an acceptable utterance, causing the utterance to be reproduced without change. These competing accounts can be assessed by testing when in speech each disfluency type occurs.

Availability of Message Planning

In the account described above, both fillers and silent pauses are used when speakers are still planning difficult material. What, then, accounts for why speakers sometimes produce a filler and sometime pause silently? We hypothesize that this contrast reflects variation not in the availability of prior material to repeat, but in a different (although potentially correlated) factor: how quickly a new plan can be initiated at the message level (Blackmer & Mitton, 1991).

Most theories of language production posit at least three cascaded levels within the production system: a message level representing preverbal meaning, a grammatical level at which lexical items and morphemes are selected and assembled into morphosyntactic structures, and a phonological level at which individual words' phonology and the utterance's overall prosody are encoded. (For review, see Bock, 1995; Griffin & V. F. Ferreira, 2006.) It is likely that errors and delays may occur at all three levels. Difficulty might occur at the message level if speakers have trouble deciding the message they wish to convey, whereas difficulty at the grammatical and phonological levels may involve delays or errors in selecting lexical items, syntactic constructions, or phonological forms.

We hypothesized that speakers most commonly use fillers when they are still planning their next message. One view is that speakers produce fillers as deliberate communicative signals to indicate their difficulty to their addressees (Clark & Fox Tree, 2002), although this hypothesis has been disputed (Corley & Stewart, 2008). If fillers are communicative signals, they should require speakers to implement a new message-level plan. This account thus predicts that speakers should most commonly produce fillers when they are not already committed to a message-level plan and can easily adopt a new one—namely, when the disruption in planning was itself at the message level. By contrast, when speakers have

already decided on a particular message and are simply delayed in lexical or phonological retrieval, producing a filler requires the existing message to be halted and a new communicative intention started. Although theories of language production differ as to whether activation flows directly back from lower stages to the message level or requires a separate monitoring process (for review, see Griffin & V. F. Ferreira, 2006), in all accounts this would be more time consuming than using information already at the message level.

Some existing evidence suggests that fillers are indeed particularly common in circumstances in which speakers are engaged in message–level planning. Fillers are more common at stronger discourse boundaries than weaker ones, possibly because more planning is required to determine the next message (Butterworth, 1975; Swerts, 1998). Fillers are used less frequently in those college lectures that use more facts and deductive methodology (Schachter, Christenfeld, Ravina, & Bilous, 1991); it is likely that these lectures minimize the uncertainty speakers encounter in planning their messages. Conversely, speakers produce more fillers when answering questions about which they are less certain (V. L. Smith & Clark, 1993). These findings all suggest that fillers frequently arise when speakers encounter difficulty in message selection and planning because of a new or difficult topic. Such circumstances would allow a new, different message-level plan—signaling difficulty with a filler—to be easily initiated.

By comparison, silent pauses do not require speakers to plan any overt signal of trouble. They require only the cessation of speech. Thus, problems at lower levels of production— the grammatical and phonological levels—may be more likely to give rise to silent pauses than to fillers because only silent pauses need not be filtered through a new communicative plan at the message level.

Evidence suggests that even grammatical and phonological difficulties, such as delays in lexical or phonological retrieval, can give rise to silent pauses. Maclay and Osgood (1959) observed that, while fillers usually occurred between phrase boundaries, silent pauses tended to occur *within* phrases. Because the unit of message planning has frequently been argued to be at least an entire phrase (e.g., Bock & Cutting, 1992; Garrett, 1988; but see Brown-Schmidt & Tanenhaus, 2006), disfluencies within a phrase may be attributed more to problems associated with grammatical and phonological planning than to message planning. In addition, patients with jargon aphasia, who use neologisms when they cannot retrieve desired lexical items, produce neologisms more frequently after silent pauses than elsewhere, suggesting a link between silent pauses and lexical retrieval difficulties (Butterworth, 1980). Nevertheless, silent pauses may be influenced by conceptual difficulties as well. Even when controlling lexical frequency, silent pauses are more common before nouns representing abstract concepts than concrete ones (Reynolds & Paivio, 1968) and in semantically anomalous utterances as compared to typical utterances with the same syntactic structure (e.g., *Duane bites dog* versus *Dog bites Duane*; Butterworth, 1980).

Finally, if speakers prefer to produce repeats whenever material is available to repeat, as hypothesized above, then we would not expect their distribution to be conditioned on

message-level planning. Repeats could arise regardless of the level of production at which the disruption occurred.

Present Study

We have proposed two hypotheses about how the state of the production system constrains what strategies are available for responding to production delays or difficulties. First, we hypothesized that repeats are used when the problem is detected after the problem segment is already being articulated and material is available to repeat. Fillers and silent pauses must be used when a delay or difficulty is detected before articulation of a segment has begun and there is no recent material available to repeat. Second, we hypothesized that fillers require a new message-level plan and can be most easily initiated when speakers have not already committed to a message-level plan, such as when they are still planning the next plot element of a story. Silent pauses, by contrast, do not require such a plan and can result from problems at the grammatical and phonological levels as well as the message level.

To test these hypotheses, we sought a paradigm that balanced naturalistic speech with experimental control (for further discussion, see, e.g., Brown-Schmidt & Tanenhaus, 2008; Jaeger, Furth, & Hilliard, 2012). Unconstrained language production, as opposed to highly scripted or restricted laboratory speech, more closely resembles natural language production (Clark, 1996). However, to assess potential message-level influences on disfluency, we also wanted potential conceptual difficulty matched across participants, which was unlikely if participants could freely choose their topic of speech. Thus, we adopted a storytelling paradigm in which participants read passages from *Alice's Adventures in Wonderland* (Carroll, 1865) and retold them in their own words. Each passage included a set of fourteen plot points, each of which was either a single action or two related actions crucial to the plot of the passage, such as *Alice finds a cake marked "EAT ME" and eats it*. Participants were instructed to include these plot points in their retelling. This paradigm presents similar sources of conceptual difficulty to all participants while still eliciting relatively unconstrained, natural speech.

We had hypothesized that fillers and silent pauses are preferentially used at points at which speakers are between units of speech whereas repeats are used when speakers are articulating material that already been planned. To test this hypothesis, we examined whether each of these disfluencies was more or less prevalent at three points at which speakers were likely to be planning new material they had not yet begun to articulate: immediately before new utterances, immediately before new grammatical phrases, and immediately before initiating a repair.

We had also hypothesized that fillers require a new message-level plan and could be most easily produced when the disruption itself occurred in message-level planning. By contrast, we hypothesized that silent pauses did not require such a plan and could easily arise from difficulties at all levels. To test this hypothesis, we investigated the relation of each disfluency type to difficulty at the message level and at later levels. To index difficulty at the message level, we used the beginning of new plot points. To indicate planning at later levels, we examined lexical frequency, which models of language production have shown to

influence both[3] grammatical and phonological retrieval (Kittredge, Dell, Verkulien, & Schwartz, 2008). We also tested whether these lexical effects were modulated by prior mention of the word, which may facilitate subsequent access (Bell, Brenier, Gregory, Girand, & Jurafsky, 2009). As a further indicator of grammatical-level difficuly, we included an interaction between lexical frequency and lexical class (content or function), since it has been argued (e.g., Bell et al., 2009; Garrett, 1980; Griffin & V. F. Ferreira, 2006) that function word placement is controlled by more automatic processes that are consequently less sensitive to frequency (Bell et al., 2009).

Method

Participants

Fifteen University of Illinois undergraduate students participated in partial fulfillment of a course requirement. All were between the ages of 18 and 22 and were native speakers of English.

Materials

Participants read three passages, each approximately 2000 words in length, excerpted from *Alice's Adventures in Wonderland* (Carroll, 1865). Each passage represented a distinct incident in the plot that involved a number of discrete actions and had a specific beginning and end. The list of plot points for each passage was printed in bullet-point format on a separate sheet. The Appendix presents these lists.

Procedure

Each participant read and retold all three stories, presented in randomized order for each participant. The experimenter instructed participants to read at their preferred speed and not to worry about memorizing the story because they would receive a list of the plot points to use in their retelling. After they had read one printed passage, it was taken away and the list of plot points was presented. The experimenter instructed participants to include all of the plot points when retelling the story. The retellings were recorded using a Marantz PMD670 Professional digital recorder. Participants could consult the plot points while speaking but had to retell the story in their own words. No time limit was imposed; the recordings continued until speakers indicated they were finished. They then read and retold the next passage.

Transcription and Coding

Two transcribers coded the presence or absence of each of three types of disfluency between every pair of adjacent words in the transcript. *Fillers* included *uh*, *um*, *ah*, or *er*. Fillers have sometimes been observed to contrast in the length of the pause they precede (Clark & Fox Tree, 2002, but see Corley & Stewart, 2008), but in the present study we were predicting the

---

[3] Although work continues to investigate what proportion of the effect of lexical frequency on planning should be attributed to the grammatical level versus the phonological level (see Kittredge et al., 2010, and references therein), this issue does not bear directly on the present hypotheses, which concerned differences between planning at the message level and at either of the later levels. What is important for the present study is that the lexical frequency influences some level after the message level, not *which* of those levels it influences.

presence or absence of disfluency rather than its duration, so we combined all fillers into a single category. In five cases (less than 1% of the total number of fillers), two fillers occurred in a row between words. These cases were infrequent enough that we could not analyze them separately, so we coded as a binary variable the presence of at least one filler versus the absence of fillers. *Repeats* were the initiation of repetition without modification of the same word, part of word, or string of words that had been spoken immediately prior. *Silent pauses* were gaps in the fluent speech stream. Because the length of the pauses licensed in fluent speech varies as a function of the syntactic and semantic context, pause duration is not a direct index of fluency or disfluency (F. Ferreira, 2007). Consequently, rather than coding all pauses over a certain length as disfluent, silent pauses were coded based on the transcribers' subjective perception of a disfluent gap. The transcribers were explicitly instructed to consider the speaker's typical speech rate and the surrounding prosodic context when judging whether a pause was disfluent. To assess the reliability of this procedure, we calculated the reliability between the two transcribers. Agreement was almost perfect ($\kappa = .96$) using the criterion of Landis and Koch (1977). Only silent pauses coded by both transcribers were included. There were almost no disagreements about the other disfluency types or about the fluent words; those disagreements were resolved by discussion.

The transcripts were also coded for potential sources of difficulty in language production. Log_{10}-transformed lexical frequency was obtained from the SUBTLEX$_{US}$ corpus (Brysbaert & New, 2009), which in comparisons to other corpora has best predicted psycholinguistic outcome measures such as performance in lexical decision and naming tasks (Brysbaert & New, 2009). Other work (e.g., Brysbaert & New, 2009) has found that the lemma stem frequency (e.g., *camel* for *camels*) does not predict behavior appreciably better than inflected form frequency, so we used the latter. Each word's lexical class was coded as either *function word* or *content word*. Function words were prepositions, determiners, auxiliary verbs, pronouns, quantifiers, verb particles, and conjunctions; content words were nouns, verbs, adverbs, adjectives, and discourse markers. The beginning of plot points were scored as the first phrase introducing a fact from the printed bullet-point list. Discrepancies between the two transcribers in the plot point beginnings were again resolved by discussion.

Both transcribers also scored the beginning and ends of utterances, defined as a subject and predicate together separated by a discernable prosodic break. Agreement was almost perfect ($\kappa = .99$); where the transcribers disagreed, only an utterance onset coded by both was included. Onsets of grammatical phrases were first coded using the Illinois Chunker (Rizzolo, 2010) and adjusted by hand by the first author as needed. Finally, the beginnings of *repairs* were coded as mid-utterance alterations of material already produced, including abandonment of the entire utterance; discrepancies between the two transcribers in locating the repairs were resolved by discussion. Words and disfluencies inside a repair were included as part of the transcript.

Data points were excluded if lexical frequency information was missing (fewer than 0.5% of observations), leaving 22,801 transitions between words.

Results

One participant did read and retell the third passage due to session time constraints. For this participant, we simply included the two completed passages.

Table 1 presents means and standard deviations across subjects of the frequencies of each of the three types of disfluencies of interest and of repairs.

Analytic Strategy

Each transition between words presents speakers with a different set of constraints on the production system—such as an utterance boundary or a low-frequency word—and provides another opportunity for the speech stream to be interrupted by a disfluency. A desirable analytic strategy, then, would be to model whether each new word was preceded by a disfluency as a function of factors influencing the state of the production system at that point. We thus adopted multilevel logit models (Baayen, Davidson, & Bates, 2008; Jaeger, 2008), in which the unit of analysis is a single word rather than averages over trials or participants. The dependent measure in our models was whether each word was or was not preceded[4] by a particular type of disfluency (such as a filler), as measured by the *logit*, or log odds, of that disfluency. These odds were modeled as a simultaneous function of multiple variables representing (a) points at which speakers were likely to be planning new material, (b) difficulty at the message level, and (c) difficulty at later, non-message levels.

One concern in modeling the odds of disfluency at each possible location is the potential non-independence of observations: Having just been disfluent may make a speaker more or less likely to be disfluent again (Shriberg, 1994). These relationships can be controlled by including behavior on previous observations in the model (Baayen & Milin, 2010). Thus, the models incorporated as predictors the presence or absence before the previous word of each type of disfluency; these variables were included to control their influence on the dependent variable rather than for hypothesis testing. To further reduce dependence between disfluencies, we also excluded from the analysis words that were themselves part of a disfluent repetition.

Multiple coding systems for predictor variables are available for multilevel models depending on the hypotheses under investigation. We mean-centred most predictors to obtain estimates equivalent to the main effects in an ANOVA analysis. However, for lexical class and prior mention, our interest was in separately testing effects among content and among function words and among first versus repeated mentions. This was accomplished using a dummy coding system, under which the simple main effect of lexical frequency represents its effect within function words and first mentions, and the interaction of lexical frequency with lexical class represents whether the effect of lexical frequency differed for content words.

---

[4]In an additional analysis, we also considered whether the properties of a word influenced whether a disfluency *followed* it. Disfluencies were less influenced by what preceded them than by followed them, and the relationship that did exist were consistent with our other results: silent pauses but not fillers were influenced by lexical properties (lexical class and previous mention), and repeats were more common after a new plot point had already been initiated rather than before.

Multilevel models can also incorporate *random effects*, effects for which the particular levels observed were sampled from a larger population; these included the speaker and the upcoming word. Between-participant variability in a particular effect (e.g., variability between participants in how strongly lexical frequency affects their rate of disfluency) can be modeled with a *random slope* of that variable by subjects. However, estimations of this variability is obtained through an iterative fitting process, which may fail to converge on an estimate when attempting to fit too many random slopes simultaneously (Freeman, Heathcote, Chalmers, & Hockley, 2010). Consequently, we performed likelihood-ratio tests to assess whether each model's fit was significantly improved by random slopes (and their correlations) for each set of variables—points of planning, message-level difficulty, grammatical- and phonological-level difficulty, and prior disfluency—and retained those slopes that contributed to the fit of the model. (This process could not converge on a stable estimate of the parameter for the complex three-way interaction of lexical frequency, lexical class, and prior mention. We thus excluded this interaction, which never approached significance in the fixed effects and was not of primary interest, from the random effects structure.)

It would in principle also be possible to take a within-item measure of the influence of each variable, and capture any variability across items in those measures with a random slope by words. However, many of the variables of interest were constant properties of the lexical items, such as lexical frequency, for which no within-item comparison, and thus no random slope, is appropriate. (It is not logically or statistically possible to compare high lexical frequency versus low lexical frequency within just the word *rabbit*.) Within-item measures were also generally not possible for the remaining variables because typically not all levels of a variable were represented within each item. For instance, many of the words used by speakers *never* appeared as the first word in an utterance, so it was not possible to measure the effect of utterance boundaries separately within each word.

All models were fit via Laplace estimation in the R environment for statistical computing using the *lmer()* function of the *lme* package (Bates, Maechler, & Bolker, 2011).

Analysis Results

Although we had specific hypotheses, described above, about how each disfluency type related to the strategies available to the production system at any given point, we included in each model the full set of predictor variables to detect any unexpected effects we had not predicted. This approach also ensured that any effect of an included variable was not merely due to a confound with some other, excluded variable. We report the full set of parameter estimates in the tables, but in the text we focus on those variables relevant to our hypotheses. Parameter estimates from logit models are expressed as log odds; to facilitate interpretation, the text presents these estimates back-transformed into odds ratios[5]. All reported effects are reliable at the $\alpha = .05$, two-tailed, level unless otherwise noted.

---

[5]Confidence intervals are symmetric around the log odds but become asymmetric when parameters are back-transformed into odds ratios.

Fillers

For fillers, the maximal set of random effects justified by the data included random slopes for points of planning, $\chi^2_{(9)} = 81.40$, $p < .001$, and for message-level planning, $\chi^2_{(14)} = 85.29$, $p < .001$; other slopes did not further improve the model, all ps > .60. Table 2 displays fixed effect parameter estimates from the final model of when speakers did and did not produce fillers.

We had hypothesized that fillers are used when speakers encounter difficulty in planning upcoming material they are not yet articulating. The evidence was consistent with the predictions of this hypothesis. Fillers were more prevalent at all three types of locations at which speakers were likely to be planning upcoming material: all other things being equal, the odds of producing a filler were 3.02 times greater (95% CI: [2.25, 4.04]) before beginning a new phrase, 11.37 times greater (95% CI: [4.07 31.77]) before beginning a new utterance, and 22.79 times greater (95% CI: [10.19, 50.95]) before initiating a repair.

We had further hypothesized that fillers required a new message-level plan and would most commonly arise when the difficulty itself arose at the message level. This hypothesis was also supported: fillers were especially prevalent when participants had to plan a new plot element, with the odds of a filler being 3.45 times greater (95% CI: [2.25, 5.30]) immediately before a new plot point. By contrast, filler production showed little influence of any grammatical or phonological factors. There was a marginal effect whereby the odds of a filler were 2.00 times (95% CI: [0.97, 4.14]) greater before a previously unmentioned word, but this effect did not reach conventional levels of significance ($z = 1.88$, $p = .06$), and no other grammatical or phonological variables, not even lexical frequency, approached significance (all ps > .15).

Silent pauses

For silent pauses, the maximal set of random effects justified by the data included random slopes only for points of planning, $\chi^2_{(9)} = 25.88$, $p < .01$; no other slopes improved the model, all ps > .80. Parameter estimates for the model of silent pause production are displayed in Table 3.

We had hypothesized that silent pauses, like fillers, are used when the problem material is still being planned and articulation has not begun. This hypothesis was supported, although the points of planning had a smaller influence on silent pauses than on fillers. The odds of a silent pause were 1.43 times greater before initiating a new phrase (95% CI: [1.01, 2.02]), 2.19 times greater before initiating a new utterance (95% CI: [1.06, 4.54]), and 6.60 times greater before initiating a repair (95% CI: [4.29, 10.16]).

Because pausing silently does not require a message-level communicative intention, we also predicted that silent pauses would not be conditioned only on message-level planning and could also stem from grammatical or phonological difficulties. This prediction was supported. Like fillers, silent pauses were more common before a new plot point, with the odds of a silent pause being 2.18 times greater (95% CI: [1.37, 3.45]) there. However, unlike fillers, silent pauses also showed sensitivity to several factors expected to influence grammatical- or phonological-level planning. The odds of a silent pause were 2.90 times

greater (95% CI: [1.02, 8.24]) before a content word than a function word. More importantly, lexical frequency influenced the distribution of silent pauses, although it interacted with lexical class: For content words, silent pauses were less common the more frequent the next word was, with every 1-unit increase in log word frequency decreasing the odds of a silent pause to just 0.03 times (95% CI: [0.01, 0.71]) what they would otherwise be. This pattern was reversed for function words, with every 1-unit increase in log function word frequency *increasing* the odds of a silent pause by 28.56 times (95% CI: [1.93, 423.44]). This increase in disfluency for high-frequency function words was unexpected, so we explored it further (see Figure 1). The effect was driven by some low-frequency words that were *never* observed to be preceded by a silent pause—perhaps because there were simply too few tokens of them. Among those function words sometimes preceded by a silent pause, the relation held as predicted: greater frequency predicted fewer disfluencies.

Repeats

Across the dataset, there were *no* cases of a repeat occurring immediately before a repair. Because we could not quantify how much more or less likely repeats were in this location, we eliminated this variable from the model.

Parameter estimates for the remaining variables are displayed in Table 4. The model of repeat production was not reliably improved by any random slopes; all ps > .10.

Our main hypothesis concerning repeats was that they are used when production difficulties are detected while a unit of speech is already being articulated and there is material readily available to repeat; consequently, they should be uncommon at junctures where speech is still being planned. This hypothesis was supported: unlike fillers and silent pauses, repeats were *less* common at points associated with initiating new material. The odds of a repeat before a new utterance were just 0.07 times (95% CI: [0.01, 0.57]) what they were elsewhere, and at a phrase boundary they were 0.05 times (95% CI: [0.03, 0.08]) what they were elsewhere. Furthermore, as noted previously, there were *no* cases of repeats initiated immediately before beginning a repair, a particularly striking absence of repeats at points of planning upcoming material.

As repeats may be a preferential strategy whenever material is available to repeat, we did not hypothesize them to be associated with any particular level of production. Indeed, no such pattern was observed; greater use of repeats was predicted by neither message nor non-message difficulty. In fact, the odds of a repeat immediately before a new plot point at the message level decreased to 0.12 times (95% CI: [0.03, 0.54]) what they were elsewhere; this may reflect the fact that new plot points also constitute points at which new material was still being planned and hence at which a repeat would be unlikely.

Discussion

The present study investigated differences between three common types of disfluency: fillers, silent pauses, and repeats. Based on the proposal that disfluency reflects a disruption in an ideal delivery (Clark, 1996), it was hypothesized that the existence of these three separate types of disfluencies reflect what strategies for responding to the disruption are

made available by the state of the production system. In particular, we hypothesized that repeats are used when there is recently articulated material available to repeat, but fillers and silent pauses are used when speakers are between units of speech and articulation has not yet begun. We further hypothesized that fillers, unlike silent pauses, require new message planning and thus are most available when speakers had not yet committed to a message-level plan.

The distribution of disfluencies in a storytelling task was consistent with both hypotheses. Table 5 summarizes these distributional differences. Fillers and silent pauses occurred most commonly where an upcoming unit of speech was still being planned and not yet articulated: the beginnings of utterances and grammatical phrases and before repairs were initiated. But it was precisely these points at which repeats were less frequent; repeats were used after articulation of a unit of speech had already begun and could easily be repeated.

Although both fillers and silent pauses occurred at points of speech planning, they differed in the problems with which they were associated. The distribution of fillers suggested they were most commonly used when speakers were still planning their next message: fillers occurred most before new plot elements, which were expected to introduce difficulty at the message level, but were not reliably affected by variables that exert their influence at later levels of production. Silent pauses, by contrast, were also affected by variables expected to influence grammatical or phonological planning: lexical frequency and lexical class.

This differing pattern of influences is not simply due to differences in statistical power. The predictors of fillers and silent pauses differed not only in their statistical significance but in the magnitude of their influences: the magnitude of lexical class × frequency interaction, for instance, was −0.84 logits for fillers but −3.53 for silent pauses, a difference of over four times. Furthermore, Table 1 indicates that fillers were actually more common than silent pauses. Both of these patterns indicate that the results cannot simply be attributed to having an insufficiently large sample to detect a similar lexical influence on filler production.

Disfluency Types as Differing Strategies

The present study demonstrated that different types of disfluency occur in different places in speech. But why do different disfluency types exist to begin with, and why should they have different distributions? We propose that the processes and representations involved in language production constrain which options for responding to delays or errors in planning are available to the production system at any given point in time.

In particular, the availability of recently produced material may determine whether a repeat is used. When speakers have some recently articulated material available, they may prefer to respond to production delays by simply repeating what was recently uttered. Blackmer and Mitton (1991) provide evidence that the articulator can repeat recently material relatively automatically, and Clark and Wasow (1998) argue that such repetitions are the preferred response to production difficulties because they eventually result in fluent delivery of the intended utterance. However, if speakers have already stopped to plan a new utterance or clause, then there may no longer be recent material in the articulatory buffer to repeat.

When a repeat is infeasible, two alternatives are fillers and silent pauses. Which of these alternatives is used may depend on how quickly a new conceptual plan can be initiated: if fillers constitute a message-level communicative signal (Clark & Fox Tree, 2002; but see Corley & Stewart, 2008), they should require a new message-level plan. Given most theories of language production (Bock, 1995; Griffin & V. F. Ferreira, 2006), a revision of the message should be initiated more quickly when the difficulty occurs at the message level itself rather than at a later level. In some theories, this is because information about trouble at the grammatical and phonological levels requires a separate monitoring process to reach the message level; this time-consuming (Nozari & Dell, 2009) monitoring process would delay initiation of a filler in response to a grammatical or phonological problem relative to a message problem. Other theories of production posit bidirectional information flow between levels, but even in these theories, initiation of a filler at the message level should take longer when information about the difficulty must propagate back from the grammatical and phonological levels than when the difficulty occurred on the message level itself. Consequently, difficulties in conceptual planning can result in the quick initiation of a filler, but difficulties in grammatical and phonological planning require more time to inform production of a filler, by which time the problem might already be resolved.

Thus, one explanation of the existence of multiple disfluency types is that they represent different strategies required by differences in the state of the production system at the time of disruption. When possible, speakers repeat articulated material to maintain fluent delivery. When a recent portion of an utterance is not available to repeat, speakers must instead halt the utterance. Whether this interruption is filled with an *uh* or *um* then depends on how quickly a message-level plan can be initiated.

Note that in this account, silent pauses are essentially a dispreferred production used when other strategies are unavailable. This accords with the proposal that speakers use other types of disfluency in part to *avoid* pausing silently; a silent speaker may lose the conversational floor (Maclay & Osgood, 1959) or be perceived as insufficiently knowledgeable (V. L. Smith & Clark, 1993),

Conceptual and Lexical Difficulty in Language Production

The message-level effects reported here are particularly important because it has been unclear whether disfluencies reflect difficulties in message planning or in grammatical and phonological retrieval. For instance, although fillers are sometimes associated with conceptual difficulties or boundaries (Butterworth 1975); V. L. Smith & Clark, 1993; Swerts, 1998), prior work has also discussed them in terms of lexical access. Fillers are more common before sentences ending with low cloze probability (Corley, MacGregor, & Donaldson, 2007) or low transitional probability (Cook, 1969), and these effects have often been attributed to lexical retrieval because improbable or unexpected lexical items are assumed to be more difficult to retrieve (e.g., Bell et al, 2009; Griffin & V. F. Ferreira, 2006). However, uncommon or unpredictable transitions generally differ not only in the lexical items used but also in the message the speaker intended to convey. Thus, it is unclear whether these effects should be attributed to message-level variables such as plausibility or to retrieval of lexical forms at the grammatical or phonological level.

In the present study, fillers were substantially more prevalent before major plot transitions, but they were not reliably affected by lexical variables such as lexical frequency, lexical class, or previous mention. This pattern suggests that fillers may be more tightly linked to planning at the message level than at the grammatical or phonological level, for reasons described above. It is possible that some of the past influences on filler production assumed to be lexical in nature actually resulted from conceptual properties correlated with lexical frequency, such as imageability or conceptual familiarity. Nevertheless, grammatical and phonological properties clearly play a role in some disfluencies: silent pauses were more common before content words, especially infrequent ones.

Fillers in Language Production

One account (Clark & Fox Tree, 2002) of fillers has been that speakers use them to communicate to their interlocutors that they are encountering delays in planning. Fillers differ in their form across languages, and this has been interpreted (Clark & Fox Tree, 2002) as indicating that they are conventionalized words. Moreover, differences between fillers (*uh* and *um*) sometimes predict the length of upcoming delays, which has been taken as evidence that speakers make a choice between them (Clark & Fox Tree, 2002). However, relations between filler form and pause time have not always been replicated, and it has been argued that evidence for the filler-as-signal hypothesis is presently insufficient (Corley & Stewart, 2008).

In the present study, fillers were only reliably affected by message-level difficulties, not grammatical- or phonological-level ones. This pattern can be explained if fillers indeed require a message-level communicative intention: it is easier to initiate this signal when the delay or difficulty itself is in message-level planning. Thus, the present data are consistent with the filler-as-signal hypothesis.

Limitations

The present study provides evidence that differences in the availability of material to repeat and in the level of production at which a problem occurred contribute to the diversity of disfluencies. But, it is likely that other variables that we did not examine also influence which disfluencies speakers produce. For instance, the severity of the problem may determine which particular filler is produced: Clark and Fox Tree (2002) observe that *uh* and *um* contrast in the length of pause that follows them (but see Corley & Stewart, 2008).

Another potential limitation of the study is that participants had to base their productions around the provided plot points, which may not have captured participants' own perceptions of what constituted important events in each story. However, other potential organizations of the material are unlikely to bear on the plot points' influence on disfluency: participants were explicitly instructed to base their retelling on the listed points and had ample time to organize their narration before beginning. The high rate at which the plot points were incorporated into participants' recordings ($M = 13.3$ of a possible 14 per story) suggests that participants were able to base their retelling on them.

A related concern is that participants' disfluencies might reflect task constraints introduced by the printed list of points. For example, because disfluencies can be produced between elements of a list (Fox Tree, 2006), the increase in disfluencies before new plot points might simply reflect participants reading the plot points as a list. However, between one plot point and the next, participants produced an average of 21 words (SD = 15 participants) that elaborated on the plot point or mentioned events not on the printed list. This provides evidence that participants were using the printed list as a guide for their own elaborated retelling of the story, as intended, and not simply to read as a list. Moreover, the rate of disfluency in the present task (M = 4.76 per 100 words, excluding silent pauses) accords fairly well with prior estimates from conversational speech (Fox Tree 1995; Shriberg, 2001) of about 6 disfluencies per 100 words, excluding silent pauses; this similarity suggests the task elicited production that was fairly representative of natural speech.

Similarly, disfluencies could have resulted before new plot points from participants slowing production to consult the printed list. However, introducing a new plot point did not simply produce a uniform increase in disfluencies—indeed, repeats were reliably *less* common before a plot point from the list. Rather, plot points particularly influenced the distribution of fillers, the disfluency type that, based on the literature, we had hypothesized would require message-level planning. This selective relation makes it more likely that the disfluencies associated with new plot points truly reflected message-planning demands. Nevertheless, it would be informative to examine whether similar relations are observed in a task that does not involve a printed list of plot points.

Conclusion

Disfluency is not a unitary phenomenon. Fluent speech is subject to multiple types of disruptions, which can represent different problems and different responses from the production system. Different disfluencies occur depending on both whether recently articulated material is available to repeat and whether a new communicative plan can be quickly initiated. Speakers use disfluent repetitions when they are already articulating a segment of speech and can easily repeat it. By contrast, speakers typically use fillers and silent pauses when they are still planning the start of a segment and have less recently articulated material to repeat. Fillers, in particular, may require a new message-level plan and are thus most commonly used when speakers are already engaged in conceptual planning. Silent pauses do not require a communicative intention and can frequently arise from difficulty in grammatical or phonological planning.

These findings concur with the observation that, in comprehension, listeners respond differently to hearing different types of disfluencies or interruptions (e.g., Arnold & Tanenhaus, 2011; Barr & Seyfeddinipurr, 2010; Brennan & Schober, 2001; Fraundorf & Watson, 2011; but see Corley & Hartsuiker, 2011). In fact, one possible reason for these differences in comprehension is that different disfluency types represent different production problems and consequently give listeners different expectations about how the speech stream will unfold.

More broadly, the production of different disfluencies given different underlying difficulties speaks to the flexibility of the production system. Bock (1995) argues that one striking

feature of the language production system is that it not only accomplishes the complex task of speaking but does so quickly and with relatively few errors. The present results suggest one reason for this efficiency: When a difficulty *does* occur, the production system is equipped with multiple strategies to quickly restore fluent production depending on what problem occurred and when it was detected.

Acknowledgments

References

Arnold, JE.; Tanenhaus, MK. Disfluency effects in comprehension: How new information can become accessible. In: Gibson, E.; Pearlmutter, N., editors. The processing and acquisition of reference. Cambridge, MA: MIT Press; 2011. p. 197-217.

Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language. 2008; 59:390–412.

Baayen RH, Milin P. Analyzing reaction times. International Journal of Psychological Research. 2010; 3:12–28.

Barr D, Seyfeddinipurr M. The role of fillers in listener attributions for speaker disfluency. Language and Cognitive Processes. 2010; 25:441–455.

Bates D, Maechler M, Bolker B. *lme*4: Linear mixed-effects models using S4 classes (version 0.999375-2) [Computer software]. 2011 Retrieved from http://lme4.r-forge.r-project.org.

Bell A, Brenier JM, Gregory M, Girand C, Jurafsky D. Predictability effects on durations of content and function words in conversational English. Journal of Memory and Language. 2009; 60:92–111.

Blackmer ER, Mitton JL. Theories of monitoring and the timing of repairs in spontaneous speech. Cognition. 1991; 39:173–194. [PubMed: 1841032]

Bock, K. Sentence production: From mind to mouth. In: Miller, JL.; Eimas, PD., editors. Handbook of perception and cognition. Vol. 11 Speech, language, and communication. Orlando, FL: Academic Press; 1995. p. 181-216.

Bock K. Language production: Methods and methodologies. Psychonomic Bulletin & Review. 1996; 3:395–421. [PubMed: 24213975]

Bock K, Cutting JC. Regulating mental energy: Performance units in language production. Journal of Memory and Language. 1992; 31:99–127.

Brennan SE, Schober MF. How listeners compensate for disfluencies in spontaneous speech. Journal of Memory and Language. 2001; 44:274–296.

Brown-Schmidt S, Tanenhaus MK. Watching the eyes when talking about size: An investigation of message formulation and utterance planning. Journal of Memory and Language. 2006; 54:592–609.

Brown-Schmidt S, Tanenhaus MK. Real-time investigation of referential domains in unscripted conversation: A targeted language games approach. Cognitive Science. 2008; 32:643–684. [PubMed: 19890480]

Brysbaert M, New B. Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior Research Methods. 2009; 41:977–990. [PubMed: 19897807]

Butterworth B. Hesitation and semantic planning in speech. Journal of Psycholinguistic Research. 1975; 4:75–87.

Butterworth, B. Evidence from pauses in speech. In: Butterworth, B., editor. Language Production, vol. 1: Speech and talk . London: Academic Press; 1980. p. 155-176.

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

Clark, HH. Using language. Cambridge: Cambridge University Press; 1996.

Clark HH, Fox Tree JE. Using *uh* and *um* in spontaneous speaking. Cognition. 2002; 84:73–11. [PubMed: 12062148]

Clark HH, Wasow T. Repeating words in spontaneous speech. Cognitive Psychology. 1998; 37:201– 242. [PubMed: 9892548]

Cook M. Transition probabilities and the incidence of filled pauses. Psychonomic Science. 1969; 16:191–192.

Corley M, Hartsuiker RJ. Why*um*helps auditory word recognition: The temporal delay hypothesis. PLoS ONE. 2011; 6:e19792. [PubMed: 21611164]

Corley M, MacGregor LJ, Donaldson DI. It's the way that you er say it: Hesitations in speech affect language comprehension. Cognition. 2007; 105:658–668. [PubMed: 17173887]

Corley M, Stewart OW. Hesitation disfluencies in spontaneous speech: The meaning of *um*. Language and Linguistics Compass. 2008; 2:589–602.

Engelhardt PE, Corley M, Nigg JT, Ferreira F. The role of inhibition in the production of disfluencies. Memory & Cognition. 2010; 38:617–628. [PubMed: 20551341]

Ferreira F. Prosody and performance in language production. Language and Cognitive Processes. 2007; 22:1151–1177.

Fox Tree JE. Placing *like* in telling stories. Discourse Studies. 2006; 8:723–743.

Fraundorf SH, Watson DG. The disfluent discourse: The effect of fillers on recall. Journal of Memory and Language. 2011; 65:161–175. [PubMed: 21765590]

Freeman E, Heathcote A, Chalmers K, Hockley W. Item effects in recognition memory for words. Journal of Memory and Language. 2010; 62:1–18.

Garrett, MF. Levels of processing in sentence production. In: Butterworth, B., editor. Language production. Vol. Vol. 1. London: Academic Press; 1980.

Garrett, MF. Processes in language production. . In: Newmeye, FJ., editor. Linguistics: The Cambridge Survey: Vol 3. Language: Psychological and biological aspects. Cambridge, UK: Cambridge University Press; 1988. p. 69-96.

Griffin, Z.; Ferreira, VF. Properties of spoken language production. In: Traxler, MJ.; Gernsbacher, MA., editors. Handbook of Psycholinguistics. 2nd ed.. London: Elsevier; 2006. p. 21-59.

Hartsuiker RJ, Notebaert L. Lexical access problems lead to disfluencies in speech. Experimental Psychology. 2011; 57:169–177. [PubMed: 20178924]

Jaeger TF. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. Journal of Memory and Language. 2008; 59:434–446. [PubMed: 19884961]

Jaeger TF, Furth K, Hilliard C. Incremental phonological planning during unscripted sentence production. Frontiers in Psychology. 2012; 3:1–22. [PubMed: 22279440]

Kittredge AK, Dell GS, Verkulien J, Schwartz MF. Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. Cognitive Neuropsychology. 2008; 25:463–492. [PubMed: 18704797]

Lake J, Humphreys K, Cardy S. Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism. Psychonomic Bulletin & Review. 2011; 18:135–140. [PubMed: 21327345]

Landis JR, Koch G. The measurement of observer agreement for categorical data. Biometrics. 1977; 33:159–174. [PubMed: 843571]

Levelt WJM. Monitoring and self-repair in speech. Cognition,. 1983; 14:41–104. [PubMed: 6685011]

Maclay H, Osgood CE. Hesitation phenomena in spontaneous speech. Word. 1959; 14:19–44.

Nozari N, Dell GS. More on lexical bias: How efficient can a "lexical editor" be? Journal of Memory and Language. 2009; 60:291–307. [PubMed: 20126302]

Reynolds A, Paivio A. Cognitive and emotional determinants of speech. Canadian Journal of Psychology. 1968; 22:164–175. [PubMed: 5658926]

Rizzolo N. Illinois Chunker [Computer software]. 2010 Retrieved from http://cogcomp.cs.illinois.edu/page/software_view/13.

Schachter S, Christenfeld N, Ravina B, Bilous F. Speech disfluency and the structure of knowledge. Journal of Personality and Social Psychology. 1991; 60:362–367.

Schegloff EA. Some other "uh(m)"""s. Discourse Processes. 2011; 47:130–174.

Schnadt, MJ.; Corley, M. Proceedings of the twenty-eighth meeting of the cognitive science society [CD-ROM]. Mahwah, NJ: Lawrence Erlbaum Associates; 2006. The influence of lexical, conceptual and planning based factors on disfluency production.

Shriberg, E. Preliminaries to a theory of speech disfluencies [Unpublished doctoral dissertation]. University of California at Berkeley; 1994.

Shriberg E. To 'errrr' is human: ecology and acoustic of speech disfluencies. Journal of the International Phonetic Association. 2001; 31:153–169.

Smith M, Wheeldon L. Horizontal information flow in spoken sentence production. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2004; 30:675–686.

Smith VL, Clark HH. On the course of answering questions. Journal of Memory and Language. 1993; 32:25–38.

Swerts M. Filled pauses as markers of discourse structure. Journal of Pragmatics. 1998; 30:485–496.

Tydgat I, Stevens M, Hartsuiker RJ, Pickering MI. Deciding where to stop speaking. Journal of Memory and Language. 2011; 64:359–380.

Appendix

List of plot points provided to participants.

Passage #1

1. Alice finds a golden key that opens a tiny door leading into a garden.

2. She wants to go into the garden, but is too tall to fit through the door.

3. Alice returns to the table where she found the golden key and discovers a bottle labeled "DRINK ME."

4. Alice drinks the bottle and shrinks down so she is only ten inches tall.

5. Alice tries to go through the door to the garden, but she left the key on the table and is now too short to reach it.

6. Alice cries.

7. Alice finds a cake marked "EAT ME" and eats it.

8. The cake makes Alice grow until she is too tall to fit through the door.

9. Alice cries so much her tears form a pool.

10. The White Rabbit runs by and drops his fan and gloves. Alice picks them up.

11. Alice feels like she's not herself and compares herself with other girls she knows to make sure she's not one of them.

12. Alice recites a poem about a crocodile.

13. Alice concludes that she must be Mabel, not Alice.

14. Alice puts on one of the gloves and realizes that holding the fan is making her shrink. She drops the fan.

Passage #2

1. Alice sees the White Rabbit looking for a fan and a pair of gloves.

2. The great hall Alice had been in has vanished completely.

3. The rabbit calls Alice "Mary Ann" and orders her to find his gloves and fan.

4. Alice runs away.

5. Alice comes to the White Rabbit's house and goes inside to look for the fan and gloves.

6. Alice finds the fan and gloves on a table.

7. Alice finds a bottle with the fan and gloves and drinks it.

8. After drinking the bottle, Alice keeps growing and growing and becomes so tall that she has to lie down to fit inside the house.

9. Alice talks to herself.

10. The White Rabbit returns home but can't get the door open, and so he tries to climb in through the window.

11. One of the White Rabbit's servants tries to climb down the chimney.

12. The White Rabbit throws pebbles through the window at Alice.

13. Alice finds and eats a cake that makes her tiny.

14. Alice runs off into the wood.

Passage #3

1. Alice is looking at the Duchess's house.

2. A footman who looks like a fish comes out of the woods and knocks at the door.

3. The door is answered by a footman who looks like a frog.

4. The first footman delivers an invitation to the Duchess to play croquet with the Queen.

5. The two footmen bow to each other, causing their curls to get tangled.

6. Alice starts laughing and has to hide in the forest.

7. Alice knocks on the door, but the footman tells her she can't get inside because he's outside and can't let her in.

8. Alice opens the door anyway and goes inside.

9. Inside, the Duchess is nursing a baby and the cook is cooking soup with too much pepper, causing everyone to sneeze.

10. Alice sees the Cheshire Cat sitting on the hearth and grinning

11. The Cook hurls plates and other items at the Duchess and the baby, but the Duchess doesn't mind.

12. The Duchess sings to the baby while tossing it up and down.

13. The Duchess throws the baby to Alice and leaves.

14. Alice takes the baby outside and realizes it's turned into a pig. She releases it into the wood.
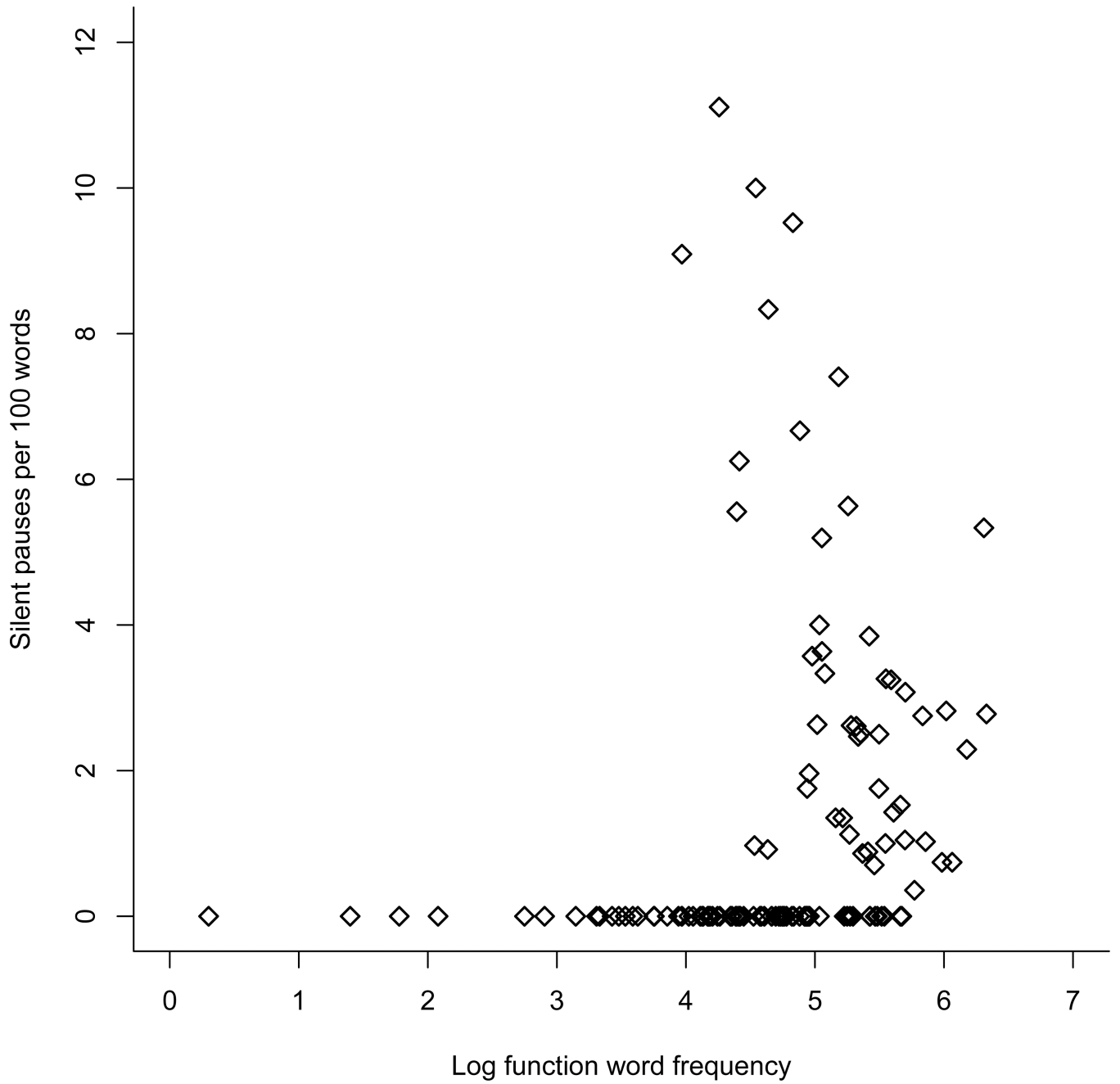
Figure 1.
Rate of silent pause production as a function of log lexical frequency.

Table 1

Rate of disfluency per 100 words and total count in sample by disfluency type.

Type	*M*	*SD*	Total Count
Filler	2.41	1.88	613
Silent Pause	1.73	0.77	367
Repeat	0.99	1.25	236
Repair	1.36	1.00	337
Total	6.49	3.65	1553

Note. SD = standard deviation across subjects.

Table 2

Fixed effect estimates for multilevel logit model of filler production (N = 22,801, log-likelihood: −2198).

Fixed effect	Coefficient	SE	Wald z	p
Intercept (base rate)	−5.41	0.50	−10.78	<.001
Points of planning				
Beginning of utterance	2.43	0.52	4.64	<.001
Beginning of phrase	1.10	0.15	7.43	<.001
Beginning of repair	3.13	0.41	7.62	<.001
Message-level difficulty				
New plot point	1.24	0.22	5.67	<.001
Grammatical- or phonological-level difficulty				
Lexical frequency (function words)	0.20	0.79	0.25	.80
First mention of word (function words)	0.70	0.37	1.88	.06
Lexical frequency × first mention (function words)	−0.84	1.07	−0.79	.43
Content word	0.37	0.31	1.17	.24
Content word × lexical frequency	−0.84	1.07	−0.79	.43
Content word × first mention	−0.17	0.40	−0.42	.67
Content word × lexical frequency × first mention	0.03	1.44	0.02	.98
Control variables				
Filler before previous word	−1.26	0.35	−3.58	<.001
Silent pause before previous word	1.00	0.29	3.40	<.001
Repeat before previous word	−0.50	0.47	−1.05	.29

Note. SE = standard error.

Table 3

Fixed effect estimates for multilevel logit model of silent pause production (N = 22,801, log-likelihood: −1744).

Fixed effect	Coefficient	*SE*	Wald *z*	*p*
Intercept (base rate)	−5.67	0.52	−10.98	<.001
Points of planning				
Beginning of utterance	0.79	0.37	2.12	<.05
Beginning of phrase	0.36	0.18	2.04	<.04
Beginning of repair	1.89	0.22	8.60	<.001
Message-level difficulty				
New plot point	0.78	0.24	3.31	<.001
Grammatical- or phonological-level difficulty				
Lexical frequency (function words)	3.35	1.38	2.44	<.05
First mention of word (function words)	0.73	0.76	0.97	.33
Lexical frequency × first mention (function words)	−0.57	0.78	−0.73	.46
Content word	1.07	0.53	2.00	<.05
Content word × lexical frequency	−3.53	1.63	−2.17	<.05
Content word × first mention	−1.32	0.78	−0.73	.46
Content word × lexical frequency × first mention	0.62	2.50	0.25	.80
Control variables				
Filler before previous word	−0.03	0.29	−0.11	.91
Silent pause before previous word	1.70	0.22	7.54	<.001
Repeat before previous word	0.16	0.48	0.34	.73

Note. SE = standard error.

Table 4

Fixed effect estimates for multilevel logit model of repeat production (N = 22,801, log-likelihood: −1094).

Fixed effect	Coefficient	*SE*	Wald *z*	*p*
Intercept (base rate)	−12.05	2.68	−4.50	<.001
Points of planning				
Beginning of utterance	−2.61	1.05	−2.50	<.05
Beginning of phrase	−3.01	0.23	−12.92	<.001
Beginning of repair	—	—	—	—
Message-level difficulty				
New plot point	−2.09	0.75	−2.79	<.01
Grammatical- or phonological-level difficulty				
Lexical frequency (function words)	7.71	7.72	1.00	.32
First mention of word (function words)	−0.91	1.41	−0.64	.52
Lexical frequency × first mention (function words)	−0.72	4.06	−0.18	.86
Content word	−0.73	2.82	−0.26	.80
Content word × lexical frequency	−10.73	9.35	−1.15	.25
Content word × first mention	1.44	1.49	0.97	.33
Content word × lexical frequency × first mention	3.67	4.78	0.77	.44
Control variables				
Filler before previous word	0.23	0.39	0.60	.55
Silent pause before previous word	−0.02	0.57	−0.03	.98
Repeat before previous word	0.58	0.52	1.12	.26

Note. SE = standard error.

Table 5

Comparison of fixed effect point estimates for influences on production of fillers, silent pauses, and repeated words.

Fixed effect	Fillers	Silent Pauses	Repeats
Intercept (base rate)	−5.41***	−5.67***	−12.05***
Points of planning			
Beginning of utterance	2.43***	0.78*	−2.61*
Beginning of phrase	1.10***	0.36*	−3.01***
Beginning of repair	3.13***	1.89***	none
Message-level difficulty			
New plot point	1.24***	0.78***	−2.09**
Grammatical- or phonological-level difficulty			
Lexical frequency (function words)	0.20	3.35*	7.71
First mention of word (function words)	0.70	0.72	−0.91
Lexical frequency × first mention (function words)	−1.56	−1.32	−0.72
Content word	0.37	3.35*	−0.73
Content word × lexical frequency	−0.84	−3.53*	−10.73
Content word × first mention	−0.17	−0.57	1.44
Content word × lexical frequency × first mention	0.03	0.62	3.67
Control variables			
Filler before previous word	−1.27***	−0.03	0.23
Silent pause before previous word	1.00***	1.67***	−0.02
Repeat before previous word	−0.50	0.16	0.58

Note.

*
$p < .05$.

**
$p < .01$.

***
$p < .001$.