Published in final edited form as:
J Mem Lang. 2011 August 1; 65(2): 161–175. doi:10.1016/j.jml.2011.03.004.

The disfluent discourse: Effects of filled pauses on recall

Scott H. Fraundorf and **Duane G. Watson**
University of Illinois at Urbana-Champaign

Abstract

We investigated the mechanisms by which fillers, such as *uh* and *um*, affect memory for discourse. Participants listened to and attempted to recall recorded passages adapted from *Alice's Adventures in Wonderland*. The type and location of interruptions were manipulated through digital splicing. In Experiment 1, we tested a processing time account of fillers' effects. While fillers facilitated recall, coughs matched in duration to the fillers impaired recall, suggesting that fillers' benefits cannot be attributed to adding processing time. In Experiment 2, fillers' locations were manipulated based on norming data to be either predictive or non-predictive of upcoming material. Fillers facilitated recall in both cases, inconsistent with an account in which listeners predict upcoming material using past experience with the distribution of fillers. Instead, these results suggest an attentional orienting account in which fillers direct attention to the speech stream but do not always result in specific predictions about upcoming material.

Keywords

disfluency; fillers; discourse; recall; language comprehension

Natural conversation, unlike most laboratory speech, is rife with *disfluencies*, or interruptions in the fluent flow of speech. For instance, speech frequently includes fillers such as *uh* and *um*. Estimates have placed the rate of disfluencies in speech as high as 6 per 100 words (Fox Tree, 1995). What are the consequences of these frequent interruptions for understanding and remembering information from a discourse?

Eye tracking and electrophysiological measures have found that fillers often benefit online comprehension of simple materials (see Corley & Stewart, 2008, for review). However, little work has investigated whether findings from these paradigms generalize to the discourse level and to situations involving later memory. Moreover, the specific mechanisms by which fillers benefit comprehension remain unclear.

In two experiments, we tested whether fillers affect later memory for discourse and examined the mechanisms by which they might do so. We compare three theoretical accounts of how fillers affect processing: fillers may allow participants to predict what will come next, fillers may orient attention to the speech stream, or fillers may allow more time to process the discourse.

Correspondence concerning this article should be addressed to Scott H. Fraundorf, Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL 61820. sfraund2@illinois.edu.

Comprehending Disfluent Speech

Although precise taxonomies of disfluencies vary (e.g. Clark, 1998; Maclay & Osgood, 1959; Shriberg, 1994), all include filled pauses, or *fillers*, verbal interruptions that do not relate to the proposition of the main message. In American English, the most common fillers are *uh* and *um* (Clark & Fox Tree, 2002). Fillers are one of the most frequent types of disfluency, accounting for one-third to over one-half of all disfluencies in several corpora (Shriberg, 1994).

Prior work has found that, although fillers interrupt the fluent delivery of an utterance, they often benefit listeners' online comprehension. Fillers facilitate judgments of whether a word in running speech matches an earlier probe (Fox Tree, 2001) and allow listeners to more quickly respond to an instruction in which the speaker repairs a prior error (Brennan & Schober, 2001). Listeners use the presence of fillers to anticipate that the speaker will refer to a less accessible referent rather than a more accessible one (Arnold, Hudson Kam, & Tanenhaus, 2007; Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Barr & Seyfeddinipur, 2009).

Although these studies have established the benefits of fillers in lexical and referential processing, several outstanding issues remain. First, it is unclear to what degree these effects generalize to later memory. Some evidence that fillers modulate memory comes from Corley, MacGregor, and Donaldson (2007), who found that sentence-final words were more apt to be recognized on a later memory test when they were preceded by a filler than when they were fluent. In general, however, less work has examined what the consequences of fillers and other disfluencies are for the long-term understanding of a connected discourse.

Second, many studies of disfluency have focused on the effects of fillers on identifying individual referents or lexical items. These studies have often used either isolated sentences or discourses with only a few possible referents. Thus, it is unclear whether fillers benefit processing only at lower levels of language comprehension or also at the level of the discourse. In the present work, we address these two issues by examining whether the benefits of fillers generalize to long-term memory for a complex discourse.

Finally, the specific mechanisms by which fillers benefit processing remain uncertain. Below, we review several accounts that have been proposed about how fillers modulate online language comprehension, and discuss how these accounts could be applied to the effect of fillers on long-term memory for a discourse. These hypotheses include predictive processing, attentional orienting, and increased processing time.

Predictive Processing Hypothesis—One reason that fillers and other disfluencies may benefit comprehension is that listeners can use them to predict what they will hear next. Speakers are most apt to produce fillers before material that is less accessible, such as a referent that is new to the discourse or that is difficult to name. Prior experience with this distribution of fillers might allow listeners to use the presence of a filler to predict that the speaker will next refer to a less accessible referent. This type of finding has been obtained in studies using the visual world paradigm; in these studies, eye fixations are recorded as participants follow instructions that refer to referents in simple scenes (e.g. Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Participants' eye movements suggest that when they hear a disfluency, *thee uh*, before a noun phrase, they anticipate that the upcoming noun phrase refers to a new item in the discourse (Arnold, Tanenhaus, Altmann, & Fagnano, 2004) or one that is difficult to name (Arnold, Hudson Kam, & Tanenhaus, 2007). Similar patterns have also been obtained by tracking participants' movements of a computer mouse to select one of several objects (e.g., Barr & Seyfeddinipur, 2009).

Although predictive processing accounts have focused on the ability of listeners to quickly identify particular referents, it is plausible that predictions could have consequences at a broader discourse level. Most theories of discourse propose that discourse comprehension requires processing both at a global level and at the level of local reference (for review, see Graesser, Mills & Zwaan, 1997). Facilitating the local level of processing would allow more time or resources to be devoted to the global level. Furthermore, predictions could also be made at the discourse level itself. Speakers produce fillers more frequently at discourse boundaries (Fraundorf & Watson, 2008; Swerts, 1998), so listeners could potentially use a filler to successfully anticipate a topic change rather than the continuation of an existing topic.

Importantly, however, the predictive processing hypothesis predicts that fillers must lead listeners to make accurate predictions in order to benefit comprehension. When fillers lead to predictions that are later disconfirmed, such as when a filler is followed by a highly accessible referent or the continuation of the same topic, fillers should not benefit, and might even impair, comprehension.

Attentional Orienting Hypothesis—An alternative to the predictive processing hypothesis is that the linguistic nature of fillers orients attention to the speech stream, but does not necessarily result in specific predictions about the nature of upcoming material. As with the predictive processing account, this hypothesis has typically focused on local referential or lexical processing. But attention would likely have consequences for discourse-level comprehension as well. The construction of discourse coherence at a global level is influenced by motivation and goals (Graesser et al., 1997), which would likely be sensitive to changes in attention. Thus, increases in attention in response to fillers might also improve discourse-level comprehension.

In an attentional orienting account, fillers need not be predictive of later material in order to benefit comprehension. For instance, fillers can affect responses even to acoustic events that do not occur in natural speech. Collard, Corley, MacGregor, and Donaldson (2008) tested this hypothesis by examining the effect of fillers on the event-related potentials (ERPs) evoked in response to an oddball auditory effect: the temporary acoustic compression of the speech stream. In general, such oddballs produced large MMN (mismatch negativity) and P300 components, ERP components argued to reflect the orienting and updating of attention. However, a filler before the oddball greatly reduced the magnitude of these effects. Collard et al. interpreted this pattern as indicating that attention was already oriented to the speech stream as a result of the filler. Crucially for the attentional orienting hypothesis, fillers had this effect even though listeners could not predict the unusual, laboratory-only oddball effect based on prior experience about the distribution of fillers.

The attentional orienting hypothesis also predicts that a filler should still benefit comprehension even if it occurs in an atypical location, such as in the middle of a discourse segment or before a highly accessible response. Support for this prediction comes from Corley and Hartsuiker (2003), who asked participants to click on one of several pictures on a computer screen based on an auditory instruction. A filler (*um*) in the instruction facilitated response times to targets of both high lexical frequency (high accessibility) and low lexical frequency (low accessibility). One interpretation of these results is that the fillers did not lead participants to make a specific prediction that the upcoming referent would be the less accessible one, but rather increased general attention to the referring expression.

Processing Time Hypothesis—The two prior accounts have proposed that fillers affect comprehension *qua* fillers, either because they lead to specific predictions or because they increase attention to the speech stream. But a third hypothesis is that interruptions of the

speech stream simply add additional time for comprehension processes, including those at the discourse level, to unfold. This processing time hypothesis predicts that *any* interruption, including a silent pause or environmental noise, should achieve the same effect as a filler, whether or not the interruption is linguistic in nature or diagnostic of the speaker's mental state.

Evidence in support of this hypothesis is mixed. When listeners followed instructions in which the speaker self-corrected an erroneous object description, a silent pause was as effective as a filler in speeding responses to the correction (Brennan & Schober, 2001). Listeners were also equally likely to expect that a referring expression would refer to a complex object when it was preceded by a filler as by a silent pause (Watanabe, Hirose, Den, & Minematsu, 2008). In addition, fillers, environmental noises, and fluent modifying expressions all have similar effects on listeners' interpretation of temporary syntactic ambiguities (Bailey & Ferreira, 2003). However, not all of the results in the literature point to effects of processing time: Barr and Seyfeddinipur (2009) found that fillers directed mouse movements towards a new referent more quickly than did coughs or sniffles at the same point in the utterance. This suggests that fillers had effects beyond the time they spent interrupting the speech stream. Fillers also had larger effects on offline judgments of a speaker's knowledgeability than silent pauses matched in duration (Brennan & Williams, 1995).

The mixed evidence for the processing-time hypothesis may owe in part to the difficulty of determining the appropriate control condition. Although periods of silence might at first appear to present a useful control interruption, silent pauses are themselves a form of disfluency associated with planning difficulties (Maclay & Osgood, 1959) and would also be diagnostic of a speaker's planning difficulties. Other experiments (e.g., Bailey & Ferreira, 2003) have compared fillers to environmental noises such as telephone rings and animal calls. These interruptions present their own challenges: they rely on the assumption that an external noise would plausibly explain the cessation in speech. If in natural production many speakers prefer to talk over such interruptions, then the absence of speech during an interruption might still be interpreted by listeners as a disfluency.

In the present study, we follow Barr and Seyfeddinipur (2009) by using coughs as a control condition for the processing time hypothesis. Coughs are interruptions that are generated by the speaker and that provide a plausible explanation for the cessation of speech, but they should be not be interpreted as related to planning difficulties. Unlike in the Barr and Seyfeddinipur experiment, which investigated online reference resolution, we compare fillers to coughs in their effect on memory for a discourse.

Present Work

The present work aimed to answer two questions: First, do the benefits of fillers in online, local processing generalize to long-term understanding of a discourse? Although it is plausible that disfluencies could also lead to attention to, or predictions about, a more global level of discourse, and that this would improve long-term understanding, this hypothesis has not yet been tested. Second, if fillers *do* benefit memory for a discourse, are these effects attributable to listeners' predictions about upcoming materials, to attentional orienting, or to increased processing time?

We examined the influence of fillers on discourse memory using a storytelling paradigm. Participants listened to short recorded stories, excerpted from *Alice's Adventures in Wonderland* (Carroll, 1865), and then attempted to retell them from memory. The presence or absence of various interruptions was manipulated by splicing them in and out of the recorded stories. This paradigm permits an assessment of the effects of fillers on discourses

that are more complex than those used in many experiments, but still allows precise control over the existence and location of interruptions.

One concern with this paradigm is that if participants found the spliced disfluencies unusual or unnatural, they might devote special attention to them even if they would not do so in natural language comprehension. To ensure that participants found the interruptions plausible, participants were told a cover story that the discourses had been recorded by a participant in a previous study who had to learn the stories and retell them from memory. (Post-experiment debriefing, discussed below, did not find any participants who detected the splicing.)

In Experiment 1, we tested the processing time hypothesis by comparing comprehension of a story containing fillers to a fluent story and to a story containing a non-linguistic interruption—the speaker coughing—matched in duration to the fillers. If fillers benefit comprehension simply because they allow more processing time, a cough of equal duration should equally benefit processing. However, if the association of fillers with language production difficulties is critical to their facilitation of comprehension, then non-linguistic interruptions such as coughs should not produce the same effect as fillers.

In Experiment 2, we then compared the predictive processing and attentional orienting hypotheses by manipulating the location of fillers within a story. According to the predictive processing hypothesis, fillers only benefit comprehension when they allow listeners to predict upcoming material. Thus, fillers at a more likely location—between discourse segments—should have more predictive utility and benefit comprehension more than fillers in unlikely locations—such as within a plot point. By contrast, if fillers simply orient attention to the speech stream, they may facilitate comprehension relative to a fluent story no matter where they are located.

Experiment 1

Experiment 1 assessed the effect of fillers on recall of a discourse. Specifically, we contrasted the processing time hypothesis with the predictive processing and attentional orienting hypotheses by comparing the effects of fillers and coughs.

To verify that fillers were generally associated with production difficulties in these materials, we conducted a production experiment (Fraundorf & Watson, 2008) for norming. Participants in the norming study read passages divided into 14 plot points and then attempted to retell them from memory. Consistent with the idea that fillers are associated with planning difficulties, participants in the production experiment were approximately three times more likely to produce fillers immediately before new plot points than elsewhere in the story.

Then, participants in Experiment 1 listened to recorded passages based on the materials from the norming study and attempted to recall as many details as possible about each passage after listening to it. Passages were presented in one of three conditions: a fluent condition, as in (1a), a *linguistic interruption* condition in which fillers occurred before some of the plot points, as in (1b), and a *non-linguistic interruption* condition in which coughs occurred before some of the plot points, as in (1c).

(1a) Meanwhile, the cook keeps hurling plates and other items at the Duchess and the baby, but the Duchess doesn't even seem to mind.

(1b) Meanwhile, **uh**, the cook keeps hurling plates and other items at the Duchess and the baby, but the Duchess doesn't even seem to mind.

(1c) Meanwhile, ***cough***, the cook keeps hurling plates and other items at the Duchess and the baby, but the Duchess doesn't even seem to mind.

Past work has mostly tested the effects of fillers on isolated materials in which sentences did not relate to one another. However, most theories of discourse (e.g. Kintsch & van Dijk, 1978) propose that individual propositions or details within a discourse are organized together. Presenting the fillers in the context of a discourse, then, permits a test of the degree of locality of fillers' effects: would fillers affect memory only for the specific plot points they precede, or would the effects extend to the entire discourse? To test this, fillers were placed before only 6 of the 14 plot points in the passages, which we term the *manipulated* plot points. If fillers have only local effects on comprehension, they should affect memory for only the 6 manipulated plot points. However, if fillers have broader effects on discourse comprehension and memory, they might affect memory for even the plot points that were not specifically preceded by fillers.

A second reason for preceding only some of the plot points with fillers was to maintain a natural rate of disfluency. Since the present stories averaged 300 words in length, a total of 6 fillers provided a close match to the overall rate of 1.86 fillers per 100 words found in the norming experiment (Fraundorf & Watson, 2008). The use of a natural rate of disfluency increased the plausibility of the cover story that the stories were recorded by a naïve participant.

Method

Participants—In all experiments, the participants were undergraduate students at the University of Illinois at Urbana-Champaign who were native speakers of American English and who participated to fulfill a course requirement or for a cash honorarium. 72 people participated in Experiment 1, none of whom had participated in the norming production experiment.

Materials—Participants listened to three short stories, plus a short practice story, paraphrased from *Alice's Adventures in Wonderland* (Carroll, 1865). Based on prior production work using the passages (Fraundorf & Watson, 2008), each of three main passages was divided into 14 plot points, defined as single events, or two closely related events, that were important to the outcome of the story.

Each story was first recorded from a script by a female research assistant with an Inland North American accent (Labov, Ash, & Boberg, 2006), appropriate for the region. To maintain a natural delivery, the research assistant memorized the story and retold it from memory one plot point at a time, rather than reading directly from the script. The stories were recorded with the prosody of normal conversational speech. The initial version of the passage contained no disfluencies; if the research assistant produced disfluencies when recording a plot point, the plot point was re-recorded until it was produced without disfluencies. This version served as a carrier version for subsequent modification in which different interruptions were spliced into the passage.

Six of the 14 plot points in each passage were then randomly chosen to be the critical plot points that would be modified with interruptions. The same 6 critical plot points were manipulated across both experimental versions of the passages. The practice story never contained any interruptions.

The linguistic interruption version of each passage was created by splicing into the carrier version 3 *uh* fillers (mean duration = 1.93 ms, SD = 0.46 ms) and 3 *um* fillers (mean duration = 2.10 ms, SD = 0.51 ms), taken from a separate recording of each passage by the

same research assistant. Each passage used different filler tokens. The non-linguistic interruption versions were then created by splicing in 6 coughs recorded by the research assistant. The coughs were placed at points that corresponded with the location of the fillers in the linguistic interruption condition. Each cough was matched in duration to the filler token that was at the same location in the linguistic interruption version of the passage. Matching was accomplished by recording multiple coughs until one closely matched in duration to a filler token; brief silence was then added before or after the cough token until it matched the filler's duration in milliseconds. In addition, at the start of the non-linguistic interruption condition, the research assistant coughed once and said "I'm sorry" to introduce the fact that the storyteller was purportedly ill when telling this story.

The complete stories, and the locations of the interruptions in the critical conditions, are available in Appendix A, and the 14 plot points scored for each story are available in Appendix B.

Design—These manipulations resulted in a 2 (interrupted or not) × 3 (interruption type) design. Whether or not a plot point was one of the six designated for interruption was an item-level characteristic and was consistent across all participants. The type of interruption was manipulated within participants: participants heard each passage once, and heard one passage per condition (fluent, linguistic interruption, and non-linguistic interruption). The assignment of the passages to conditions and the order of the conditions were counterbalanced across participants using a Latin Square design.

Procedure—At the beginning of the session, participants were given a cover story that the stories came from a previous experiment in which participants had come into the lab on three days to learn stories from *Alice's Adventures in Wonderland* and then try to retell them. The actual participants were told that the past recordings were now being played to new participants to see how well the original participants had explained the story. The experimenter then pretended to check a list of participant numbers. After checking this list, the experiment told the current participant that the particular story-teller that she or he had been assigned might have been sick on one of the days she was telling the stories, but that the listener should just do the best job he or she could.

Stimuli were presented and responses were collected using MATLAB and Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Before the stories were presented, participants were able to adjust the volume of the computer to ensure that they were able to hear the stories. Participants then heard the short practice passage followed by the three experimental passages. For each passage, participants first listened to the recorded passage and then attempted to verbally recall the passage in as much detail as possible. Participants' responses were recorded on a digital recorder. The computer screen was blank during both the presentation of the recorded passage and the participants' response.

When participants had finished recalling the passage, they made a rating on the computer of how difficult the story was to understand. Ratings were made on a Likert scale from 1 (*very easy*) to 7 (*very hard*). These ratings were not reliably affected by the fillers in either of the experiments reported here and will not be discussed further. Once participants had rated a passage, they advanced to the next passage.

At the end of the session, participants completed a structured debriefing questionnaire to assess whether they had noticed the spliced audio. None of the participants reported noticing the splicing.

Coding—Gist recall was scored for each of the plot points in each passage. Scoring of each passage was done blind to the assignment of passages to conditions for each participant. Each of the participants was scored by one of two judges, one of whom was the first author. To verify the reliability of the coding procedures, the two judges independently scored 20 of the same participants. Reliability between the judges was almost perfect ($\kappa = .85$) according to the criteria of Landis and Koch (1977). Where the two judges disagreed, the first author's ratings were used.

Results

Mean recall of both the manipulated and non-manipulated plot points in each condition is displayed in Figure 1.

Results were analyzed using a multi-level logit model. The model was fit in the R software package (R Development Core Team, 2008) with Laplace estimation using the *lmer()* function of the *lme4* package (Bates, Maechler, & Dai, 2010).

The dependent measure in the model was the log odds of correct recall of each plot point. The predictors included the fixed effects of passage condition (fluent, fillers, or coughs) and plot point type (manipulated or non-manipulated) and their interaction, as well as random intercepts for participants and for plot points. Plot points constituted the "items" in the model because recall of individual plot points was the unit of analysis. Passage condition was coded using mean-centered effects coding to compare each passage-level interruption condition to the grand mean. Plot point type was coded using a mean-centered variable at the item level to code whether or not each plot point was one of the 6 of the 14 plot points that were manipulated with an interruption.

As noted above, only 6 of the 14 plot points in each passage—the manipulated plot points—actually had interruptions preceding them. To the extent that the fillers holistically benefit or impair recall of the *entire* discourse, they would have an effect on recall of every plot point in the passage. This would be represented as a main effect of passage condition on recall, independent of whether or not a particular plot point was preceded by an interruption. Alternately, fillers might selectively benefit just the plot points they precede. This would be represented as an interaction between passage condition (e.g. fillers or coughs) and interruption, such that recall would be only affected by the fillers if the particular plot point was one preceded by a filler. Finally, it is also possible that the plot points selected for manipulation might coincidentally be more difficult to remember. This would be represented in the model as a main effect of plot point type across conditions, even in the fluent story.

In a multi-level model, variability in an effect across participants or across items is reflected in a random slope of that effect by participants or items. However, parameter estimates for multi-level models must be obtained through an iterative fitting process, and models with large numbers of random slopes may fail to converge on a solution (Freeman, Heathcote, Chalmers, & Hockley, 2010). Consequently, we first tested which random slopes contributed significantly to the model. A likelihood-ratio test indicated that the model was improved by a random slope for the effect of fillers across participants, $\chi^2_{(2)} = 27.48$, $p < .001$, and further improved by a random slope for the effect of coughs, $\chi^2_{(3)} = 14.45$, $p < .01$, indicating some variability across participants in the effect of the interruption conditions. All other random slopes made non-significant contributions to the model (all *p*s > .75). Therefore, we excluded those random slopes from the final model in order to obtain a successful fit. The *lme4* model formula used to fit the model is displayed in Appendix C.

Tables 1 and 2 display parameter estimates for the model with the two random slopes for the effect of interruption condition. Both types of interruptions reliably affected performance.

Fillers facilitated recall; the odds of correctly recalling a plot point were 1.57 times greater (95% confidence interval of the difference: [1.01, 2.46]) in the linguistic interruption condition relative to the grand mean. However, the coughs impaired recall; the odds of correctly recalling a plot point were 0.64 times less (95% CI: [0.44, 0.93]) in the non-linguistic interruption condition relative to the grand mean.

There was no reliable interaction between the passage-level condition and whether or not an individual plot point was interrupted. This indicates that the effects of the fillers and coughs applied to the whole recording and were not limited to the particular plot points preceded by a filler or cough.

Because half of the fillers in each story were *uh* and half were *um*, we performed a post-hoc analysis to test whether these fillers had differing effects on recall. We added an additional contrast between plot points designated to be interrupted with *uh* in the linguistic interruption condition and plot points designated to be interrupted with *um*. This change did not improve the fit of the model, $\chi^2_{(3)} = .55$, $p < .91$. Moreover, in the expanded model, the *uh* and *um* distinction had no main effect on recall and did not interact with any other factors. Therefore, there was no evidence that *uh* and *um* differed in their effects on recall in Experiment 1.

Discussion

Experiment 1 examined the effects of fillers on discourse memory in the context of a story recall task. Participants listened to short stories that varied in the types of interruptions they contained. Fillers, an interruption related to language production, facilitated recall of the stories. Conversely, an interruption that was not linguistic in nature—coughing—actually impaired recall of the story. Both interruptions were matched in duration but had opposite effects on recall. This divergence demonstrates that the benefit from the fillers cannot be attributed simply to the increase in total processing time available.

Both the benefit to memory from fillers and the penalty from coughs applied broadly across the entire discourse. The global benefit of fillers may reflect the connected nature of discourse: if a filler facilitates memory for one plot point, that plot point may in turn serve as a cue that facilitates memory for other portions of the discourse as well. We return to this point in the General Discussion.

It is worth noting that mean performance was numerically lower for the plot points designated to be preceded by interruptions. However, this was true even in the fluent conditions, where no actual interruption occurred. Thus, this difference can likely be attributed to the baseline memorability of those particular plot points.

One possible concern with Experiment 1 is that the coughs in the non-linguistic interruption condition might have also been interpreted as related to production processes. Because speakers prefer to minimize the number of interruptions in speech (Clark & Clark, 1977), speakers might strategically postpone coughing until a point of production difficulty when they would be pausing anyway. If speakers distribute their coughs in this way, and listeners are sensitive to this distribution, coughs and fillers could be equally informative about a speaker's planning. However, Experiment 1 provides evidence against this possibility: while the fillers reliably improved memory for the stories, the coughs reliably impaired memory. The opposite ordinality of these effects indicates that listeners did not interpret fillers and coughs as equivalent signals of production difficulty.

These data are also evidence against within-experiment learning about the predictive value of coughs. Participants could have learned that, within the context of this experiment, the

coughs in the recorded story actually did predict the beginning of a new plot point, and could have then used this information to correctly anticipate new plot points. However, this possibility is disconfirmed by the fact that the coughs actually impaired recall.

One final concern with Experiment 1 is whether the difference between fillers and coughs was influenced by the task instructions. The experimenter's initial instructions to participants cautioned the participants about the ill speaker but made no mention of anything related to the filler condition. This may have led participants to treat the cough and filler stories differently, independent of the interruption manipulation. Although this is possible, it cannot explain the difference between the filler and fluent condition, neither of which was mentioned in the instructions. The difference between those two conditions suggests that fillers do benefit memory. If the processing time permitted by coughs also improves memory, but its effects were masked by the subtle difference in instructions, the effect of these instructions would have to be relatively large since performance in the cough condition was lower than in the fluent condition. We think this is unlikely, but future work will need to investigate this possibility.

Thus, Experiment 1 provides evidence that the effects of fillers on memory for a discourse cannot be solely attributed to an increase in processing time. Recall was facilitated by the fillers, but impaired by an interruption that was matched in duration but that was non-linguistic in nature.

The other two hypotheses could both explain the memory benefit. According to the predictive processing hypothesis, fillers benefited performance because they appeared in places where they were most apt to occur in natural production: before new plot points. Consequently, listeners' past experience with the distribution of fillers may have allowed them to anticipate the structure of the discourse, which would then have allowed them to more quickly or more accurately encode incoming information. For example, listeners might have been able to predict that the next propositions in the story would begin a new plot event rather continue an existing event.

In contrast, according to the attentional orienting hypothesis, the fillers more generally oriented attention to the linguistic input, facilitating comprehension even if they did not lead to specific predictions about discourse structure.

Experiment 2 teased apart these explanations by manipulating whether or not the fillers accurately predicted the onset of a new plot point.

Experiment 2

In Experiment 2, we compared the predictions of the predictive processing and attentional orienting hypotheses by splicing fillers into typical or atypical locations. We determined typical locations for fillers for these materials empirically on the basis of the norming production experiment described above. In this experiment, fillers occurred frequently before what we designated as a new plot point, but far less frequently within a plot point. The mean rate of filler use was 5.19 fillers per 100 words in the three words immediately preceding the start of a plot point, but only 1.18 per 100 within a plot point. In other words, fillers were approximately 4.5 times more likely at the start of a plot point than within a plot point, consistent with the idea that fillers tend to precede points of difficulty.

Based on this norming work, we constructed two experimental conditions for Experiment 2. In the *predictive filler condition*, as in (2a), fillers preceded new plot points, a typical location. In the *non-predictive filler condition*, as in (2b), fillers appeared in the middle of

NIH-PA Author Manuscript

NIH-PA Author Manuscript

the corresponding plot point, an atypical location. Finally, a fluent version of the passages was also included as a control.

(2a) Meanwhile, **uh**, the cook keeps hurling plates and other items at the Duchess and the baby, but the Duchess doesn't even seem to mind.

(2b) Meanwhile, the cook keeps hurling plates, **uh**, and other items at the Duchess and the baby, but the Duchess doesn't even seem to mind.

According to the predictive processing hypothesis, listeners use their past experience with the distribution of fillers to make specific predictions about upcoming material. In this account, fillers should only benefit comprehension in the predictive condition, in which they are consistent with the past distribution and allow accurate predictions. In the non-predictive filler condition, the fillers would lead listeners to make inaccurate predictions—for instance, that a new plot event is forthcoming when the current event is actually continuing. These inaccurate predictions would not be expected to benefit comprehension relative to a fluent story, and might even impair it.

By contrast, if fillers simply orient attention to the speech system, they should benefit comprehension in both the predictive and non-predictive conditions relative to the fluent condition. In fact, to the extent that the non-predictive fillers are even more unusual than the predictive fillers, performance might even be superior in the non-predictive filler condition.

It is of course plausible that, in a predictive processing account, predictions made about the upcoming discourse structure could also lead to increased attention as a downstream consequence. The cornerstone of the predictive processing hypothesis, however, is that such modulation in attention would be driven by listeners' experience with the distribution of fillers. Consequently, in a predictive processing account, the effectiveness of this attentional allocation would be affected by whether or not the fillers accurately predicted discourse structure. By contrast, a more general attentional orienting attentional account would permit fillers to benefit processing regardless of their relationship to the general distribution of fillers.

Method

Participants—42 people who had not participated in the previous experiment participated in Experiment 2.

Materials—Experiment 2 used the same three passages as Experiment 1. The fluent version of each passage was the same recording as Experiment 1, and the predictive filler condition was the same as the Experiment 1 filler condition. The non-predictive filler condition was created by moving each filler to a location *within* the plot point that the filler preceded in the predictive condition. The coughs were not used in Experiment 2. To ensure that both types of interruptions sounded natural, fillers were always placed at prosodic boundaries in both conditions. The location within the passages of the fillers in the predictive and non-predictive conditions can be found in Appendix A.

Design—Like Experiment 1, Experiment 2 had a 2 (manipulated or not) × 3 (disfluency type) design. Whether or not a particular plot point was designed for interruption was an item-level characteristic consistent across participants, while disfluency type was manipulated within participants. Three passages were presented to each participant, one in each condition. In Experiment 2, the ordering of passages was randomized across participants rather than counterbalanced, with the constraint that each passage was still used an equal number of times in each condition across participants.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Procedure—Aside from the change in stimuli, the procedure in Experiment 2 was largely the same as in Experiment 1. However, because no condition in Experiment 2 involved coughs, the portion of the cover story related to the ill participant was omitted.

Coding—As in Experiment 1, recall was scored using a gist criterion method while blind to the assignment of passages to conditions across participants. All participants were scored by the same rater (the first author). To verify the reliability of coding procedures, 4 participants were scored by a second rater. Agreement between the two raters on these participants was substantial ($\kappa = .77$). Where the two raters disagreed, the first author's ratings were used.

Results

Mean recall of the manipulated and non-manipulated plot points in each condition is displayed in Figure 2.

Results were analyzed using a multi-level logit model, fit using the same software as in Experiment 1. Log odds of correct recall were modeled as a function of the fixed effects of passage condition (predictive fillers, non-predictive fillers, or fluent), plot point type (manipulated or not), and their interaction, with random intercepts for plot points (items) and participants. Two mean-centered orthogonal contrasts were used to assess the effect of fillers. The first contrast compared the two filler conditions against the fluent condition; the second compared the predictive fillers to the non-predictive fillers.

For the Experiment 2 data, the only random slope that reliably contributed to the model was the random slope by participants of the contrast between predictive and non-predictive fillers, $\chi^2_{(2)} = 25.91$, $p < .001$. The random slope by participants of the contrast between the filler and fluent conditions did not approach significance in Experiment 2, $\chi^2_{(2)} = 1.51$, $p = .47$. The addition of a random slope by items of the effect of fillers trended towards but did not reach significance, $\chi^2_{(2)} = 3.51$, $p = .17$. Because this slope did not reach conventional levels of significance, we report parameter estimates from the model without it, but the pattern of results was the same with or without this slope in the model. All other random slopes did not reliably contribute to the model (all *p*s> .20). The *lme4* model formula appears in Appendix C.

Parameter estimates for the final model are displayed in Tables 3 and 4. Fillers reliably facilitated recall. Odds of correct recall in the two disfluent conditions were 1.38 times (95% CI: [1.05, 1.81]) greater than in the fluent condition. However, contrary to the predictions of the predictive processing hypothesis, recall was not reliably better in the predictive filler condition than the non-predictive condition (Wald $z = 0.51$, $p = .61$).

As before, there was no reliable interaction between the passage-level disfluency effects and whether or not a particular plot point contained a filler. That is, the effect of the fillers applied to the entire discourse and not selectively to the plot points that were disfluent. Once again, performance was numerically lower for the plot points that received manipulations, although this difference was not statistically reliable ($z = -.01$, $p = .99$). Again, however, this was true even in the fluent condition with no interruptions. This provides further evidence that this difference likely reflects baseline memorability rather than an effect of the fillers.

There was again no evidence that *uh* and *um* differed in their effects on recall. Adding the contrast between plot designated for *um* versus those designated for *uh* did not improve the fit of the model, $\chi^2_{(3)} = 1.99$, $p = .57$, and the *uh* and *um* contrast did not improve the fit of the model nor interact with any other factors.

Discussion

Experiment 2 replicated the overall benefit of fillers for memory for a discourse. As in Experiment 1, this benefit applied across the entire story; memory for a plot point was facilitated by having fillers elsewhere in the recording, even if that particular plot point did not have a filler.

In Experiment 2, we pit the predictive processing account of this benefit against an attentional orienting account. If fillers facilitated performance because listeners' past experience allowed them to use fillers to predict the start of a new plot point, fillers should only have been beneficial when those predictions were accurate. Contrary to this prediction, fillers benefited performance equally regardless of the location of the fillers.

These results are most consistent with the attentional orienting hypothesis. Fillers may increase a listener's attention to what a speaker is saying without necessarily resulting in specific predictions about whether there will be, for instance, a topic shift. Consequently, fillers can benefit performance even when they appear in locations that are at odds with the typical distribution of fillers.

One possible concern is that the absence of a predictive processing effect might reflect within-experiment learning. Hearing the non-predictive fillers might have taught participants that fillers were not predictive of an upcoming discourse boundary and so led participants to discontinue using fillers as a predictive cue. This explanation predicts an order effect: Fillers should have more strongly benefited performance before participants had been exposed to the misleading non-predictive condition. To test this prediction, an additional mean-centered contrast was added at the subject level comparing participants who heard the non-predictive condition before the predictive condition with those who heard the opposite ordering. Neither this variable nor any of its interactions with the passage type variables were reliable, nor did adding this variable change the patterns of the other fixed effects or their reliability. Adding the order variable also did not reliably improve the overall fit of the model, $\chi^2_{(3)} = 4.26$, $p = .23$. There was no evidence, then, that the lack of predictive processing could be attributed to within-experiment learning.

General Discussion

In two experiments, we examined both the effects of fillers on participants' ability to correctly recall elements of a short discourse as well as the potential mechanisms underlying those effects.

In Experiment 1, two different types of interruptions, filler and coughs, appeared before new plot points. The fillers facilitated recall of the stories relative to a fluent version. However, coughs—an interruption unrelated to language—*impaired* recall. This divergence provides evidence against the hypothesis that the effects of fillers can be attributed simply to allowing more processing time; the two types of interruptions were matched on duration but yielded opposite effects on comprehension.

In Experiment 2, the location of fillers within the stories was manipulated. In the predictive condition, fillers appeared before the start of new plot points, the location at which they are most apt to occur in production (Fraundorf & Watson, 2008). In the non-predictive condition, fillers appeared within a plot point, a less common location in production. If participants predict upcoming material based on the typical distribution of fillers in speech, fillers should only have benefited performance when they appeared at points that would allow listeners to make accurate predictions. For example, a filler might have led listeners to predict that the next propositions would constitute a new plot event that was not strongly

related to the previous event. This prediction would have been useful when the filler actually did precede a new plot point, but useless or harmful when the filler preceded a continuation of the same plot point. However, no such distinction between predictive and non-predictive fillers was observed.

Together, these results support an attentional account of the effect of fillers on memory for complex discourses. Because fillers originate in the language production system, they may lead listeners to attend more to what the speaker is producing, which can facilitate comprehension and recall of a discourse relative to a fluent version. Although fillers may also sometimes lead listeners to make specific predictions about what is coming next (Arnold et al., 2004, 2007; Barr & Seyfeddinpur, 2009), the present results suggest that this is not always the case. This result is consistent with other findings that fillers can sometimes facilitate language comprehension even when they appear in atypical locations. For instance, fillers speeded response time to an auditory instruction even when they appeared before a high frequency, accessible lexical item (Corley & Hartsuiker, 2003), which is uncharacteristic of the natural distribution of fillers. Fillers also altered responses even to unusual auditory oddballs that do not appear outside of the lab (Collard et al., 2008).

The present study generalizes these past findings by extending them to more complex materials and to the level of discourse processing. Past work has often examined fillers in simple contexts such as single sentences or discourses with restricted referential contexts. Moreover, these studies have typically examined effects of filler on lower-level processes such as referential resolution or lexical identification, and effects at these levels may or may not generalize to the level of global discourse understanding. The present experiments demonstrate that fillers can facilitate comprehension even at the level of a discourse and even when they appear in more complex materials.

The use of a complex discourse also allowed a test of the locality of filler effects. In both experiments, fillers were observed to have a broad benefit on memory: plot points were remembered better when they came from a story with fillers, regardless of whether or not that particular plot point was preceded by or contained a filler. One reason for these non-local effects may be that propositions in a discourse are generally argued to be organized together in larger structures (e.g. Kintsch & van Dijk, 1978). Consequently, if a filler orients attention towards a particular plot point, the benefit to memory for that plot point may also extend to other, related points. For instance, remembering that Alice drank from a mysterious bottle may serve as a cue to remember that, later in the story, she also ate an unusual piece of cake. This effect demonstrates that qualitatively different effects may emerge at the discourse level of language comprehension.

Individual Differences in Discourse Comprehension

While the benefit of fillers was robust across participants and plot points in both of the present experiments, there was nevertheless evidence for some variability between participants in the strength of the effect. In Experiment 1, the magnitude of the filler benefit varied reliably across participants; in Experiment 2, the magnitude of the filler benefit did not vary, but the difference between the predictive and non-predictive fillers did.

Some of this variability may simply reflect sampling error, especially given that there were only 14 observations per participant per condition. However, it is also possible that this variability reflects genuine differences in the extent to which fillers affect individuals' discourse comprehension. Other work has found substantial individual differences in the degree to which language comprehension is sensitive to cues such as lexical distributional statistics (e.g. MacDonald & Christiansen, 2002) or even extra-linguistic information such as referential context (Novick, Thompson-Schill, & Trueswell, 2008). This variation in

sensitivity may reflect differences in factors such as linguistic experience or basic cognitive abilities (MacDonald & Christiansen, 2002). In the case of attentional effects of fillers, for instances, such individual differences might partially reflect individual differences in the efficacy of attentional networks (e.g. Fan, McCandliss, Sommer, Raz, & Posner, 2002).

Because other measures of individual differences were not collected in the present study, such an account remains speculative at this point. However, future work investigating potential individual differences in sensitivity to non-lexical cues such as disfluency could inform theories both of disfluency and of individual differences in language comprehension.

Types of Disfluency

The present study tested the effects of fillers—*uh* and *um*—on recall of a discourse. However, fillers are by no means the only type of disfluency that can occur in the natural production of a discourse. Among other types of disfluencies are silent pauses (e.g., Duez, 1985), repetitions of one or more words (e.g. Clark & Wasow, 1998; Fox Tree, 1995; MacGregor, Corley, & Donaldson, 2009), self-initiated repairs (e.g. Brennan & Schober, 2001; Levelt, 1983), prolonged words (e.g. Schnadt & Corley, 2006), and segmental changes such as pronouncing *the* with a non-reduced vowel rather than a reduced one (Fox Tree & Clark, 1997).

Different types of disfluency have sometimes been observed to have different effects on language processing. For instance, a disfluent repetition before a sentence-final word did not benefit later recognition memory for that word (MacGregor et al., 2009), whereas a filler in the same location did (Corley et al., 2007). And, fillers had a greater influence on judgments of speaker knowledge than did silent pauses (Brennan & Williams, 1995).

Fillers may even differ amongst themselves. For example, Fox Tree (2001) found that *uh* facilitated lexical identification but *um* did not. Fox Tree argued that this difference arose because *um* often precedes a longer delay than *uh* (Clark & Fox Tree, 2002; Smith & Clark, 1993) and listeners may stop attending to the speech stream when expecting a long delay. In the present study, however, *uh* and *um* did not reliably differ in their effects on recall. There are multiple reasons why these fillers might have had equivalent effects in the present study First, there were relatively few data points per participant for each filler type, which reduced the power to detect a difference. Second, the benefit from fillers in the present experiments was consistently observed to apply across a whole passage. Thus, it was less likely that there would be an effect of whether a particular plot point was preceded by *uh* or *um*.

Nevertheless, it is clear that, in general, different types of disfluencies can have different effects. One likely reason that disfluencies can vary in their effect on comprehension is that they stem from different types of problems in production. Because the rates of use of various types of disfluencies correlate only weakly (Maclay & Osgood, 1959), different types of disfluencies have often been argued to reflect different problems in production (Bock, 1996; Fraundorf & Watson, 2008; Schnadt & Corley, 2008) or delays of different severity (Clark & Fox Tree, 2002). For instance, Fraundorf and Watson (2008) found that fillers were especially apt to occur at those clause boundaries that introduced new plot elements into a discourse, whereas silent pauses were equally likely at all clause boundaries. Moreover, fillers were the only type of disfluency whose rate was not predicted by the length of the utterance. Based on this and other evidence, Fraundorf and Watson (2008) argued that fillers were most apt to result from difficulties in conceptual planning rather than in lexical retrieval or grammatical planning. To the extent that fillers are most associated with conceptual planning, they might be more apt than other disfluencies to orient attention to the semantic content of a discourse.

To date, however, differences between disfluency types in their effects on comprehension remain unclear. Moreover, in many experiments, the disfluent condition has consisted of several disfluencies occurring in a row (e.g. a prolongation followed by a filler). To the extent that such disfluencies may co-occur (Bell et al., 2003), this provides for natural-sounding stimuli. However, the effects of specific types of disfluency on comprehension and memory, and possible differences between those types, remain open for future research.

Disfluency and Processing Time

A recurring question in the literature on fillers and other disfluent interruptions is whether their effects can be attributed simply to changes in overall processing time. To date, results have been mixed. The present Experiment 1 tested the processing time hypothesis by comparing fillers to duration-matched coughs and found that, while the fillers facilitated memory, the coughs impaired it. Since the two interruptions were exactly matched in duration, the effects of fillers could not be driven only by the delay they introduced.

In other experiments (e.g. Brennan & Schober, 2001; Watanabe et al., 2008), however, fillers have been observed to have the same effect as interruptions matched in duration. One reason for these discrepancies may be the variety of control conditions used in the literature. Tests of the processing time hypothesis have often relied on a comparison between fillers and silent pauses. However, silent pauses are themselves a type of disfluency (Maclay & Osgood, 1959). Fillers and silent pauses might have had similar effects not because of processing time but because they are both disfluencies, and consequently could both orient attention to the speech stream or both allow similar inferences about the speaker. In the present work, we attempted to avoid this confound by using an interruption unrelated to linguistic planning—coughs—and found that this interruption did not have the same effects as fillers. Barr and Seyfeddinipur (2009) used a similar comparison and also found that fillers were more effective than coughs at directing mouse movements towards new referents.

Some evidence in support of the processing time hypothesis that is not subject to the silent pause confound comes from Bailey and Ferreira (2003), who examined the effect of interruptions on processing of temporary syntactic ambiguities such as (3a), in which *the deer* could be incorrectly interpreted as the object of *hunted*.

(3a) While the man hunted the deer ran into the woods.

(3b) While the man hunted the deer uh uh ran into the woods.

(3c) While the man hunted the uh uh deer ran into the woods.

When an interruption increased the delay between the head of the ambiguous noun phrase (NP) and the main clause verb that resolved the ambiguity, as in (3b), utterances were less likely to be judged as acceptable. (A control condition, (3c), in which the interruption occurred before the head of the ambiguous NP, did not produce this effect.) Moreover, the effect in (3b) was of the same magnitude for fillers as for wholly non-linguistic interruptions such as a telephone ringing or a dog barking. Bailey and Ferreira interpret these results as indicating that the delay of the head NP increases commitment to the incorrect interpretation, and that any interruption that introduces a delay should thus produce the effect.

One interpretation of these results is that whether fillers have effects beyond mere processing time may depend on the task. Bailey and Ferreira's task involved the online resolution of a syntactic ambiguity, whereas the present task involved memory for the semantic content of a discourse. As noted above, Fraundorf and Watson (2008) have argued that fillers are particularly associated with conceptual planning in production. If this is

correct, fillers may be more apt to show unique effects in semantic memory (the present experiment) or reference resolution (Barr & Seyfeddinipur, 2009) rather than in syntactic processing.

Prediction and Disfluency

Experiment 2 tested a predictive processing account of the effects of fillers. In this account, fillers facilitate memory because they allow listeners to make predictions about upcoming material on the basis of their prior experience with the distribution of fillers. However, in Experiment 2, fillers facilitated memory regardless of whether they appeared in locations that were typical or atypical based on norming data. This pattern suggests that fillers were not leading to specific predictions about upcoming material.

These results contrast with work on online reference resolution, in which listeners do appear to make predictions on the basis of disfluency. In experiments using the visual world paradigm, listeners quickly look to referents that are new to the discourse (Arnold et al., 2004) or difficult to name (Arnold et al., 2007) when they hear a disfluency. One explanation for these differences may be that they occur at different levels of analysis. Predictions about the next referent to be mentioned may be easier to generate than predictions about a more global level of discourse structure. Effects at lower levels of processing do not always generalize to the discourse level; for instance, syntactic prominence of a noun phrase anaphor may facilitate online lexical processing but impair later recall of the discourse (Almor & Eimas, 2008).

Another, related difference is the level of constraint provided by the discourse context. In the present experiments, a filler might have allowed listeners to expect a new plot event rather than a continuation of an existing one, but it would still place relatively few constraints on the upcoming input. The universe of possible events that could have occurred next is large, especially in a fantasy story like *Alice's Adventures in Wonderland*. This stands in contrast to findings from the visual world paradigm, where predictions are relatively easy because the speaker can refer only to a few possible referents. Thus, although fillers clearly allow listeners to make predictions in certain referential contexts (Arnold et al., 2004, 2007), those effects may not generalize to cases where the referential context is much broader.

Disfluency and World Knowledge

Prior work has found that when speakers answer a question, the use of fillers correlates with lower self-rated and other-rated knowledge (Smith & Clark, 1993; Brennan & Williams, 1995). Based on this association, it was possible that listeners would have disregarded material spoken disfluently because they were not confident about its accuracy. In the two present experiments, however, there was no evidence for such an effect: fillers were remembered *better* than the fluent stories. These results are not necessarily inconsistent: Even though disfluencies make discourse more memorable, they might also undermine confidence in the knowledge of the speaker. Indeed, several participants noted during the debriefing that the story-teller "did not seem to know the stories very well." In fact, it is possible that one of the reasons attention is enhanced in response to fillers is *because* fillers often suggest that the speaker will not tell the story effectively and that comprehension may require extra effort on the part of the listener.

Some evidence does suggest that, in other cases, listeners may treat disfluent information as less reliable. In category learning tasks, learning of novel color names is facilitated when an instructor disfluently names poor exemplars of the colors and fluently names good exemplars, relative to the reverse (Barr, 2003), although that experiment did not manipulate

filler use independently of prosodic contours. One interpretation of these results is that listeners assumed the examples named disfluently were less reliable as exemplars and made less use of them. This assumption would produce the observed benefit in learning when the disfluent exemplars were indeed poor exemplars.

Conclusion

In two experiments, fillers facilitated later recall of complex discourses. This effect was observed whether or not the location of the fillers was consistent with their distribution in production, but was not observed for coughs matched to the fillers in duration. These results are most consistent with an attentional orienting account in which fillers direct attention to the speech stream but do not always result in specific predictions about the nature of upcoming material. These results also generalize past experimental findings on fillers to the level of the discourse and to later recall, demonstrating that fillers can facilitate recall even of complex discourses.

Appendix A

Stimulus stories for both experiments. * indicates the location of interruptions in Experiment 1 and of interruptions in the predictive condition of Experiment 2. † indicates the locations of interruptions in the non-predictive condition of Experiment 2.

Practice Passage

Alice is exploring Wonderland and she comes to the garden belonging to the Queen of Hearts. She sees that the Queen's gardeners are painting the roses at the entrance of the garden. One of the gardeners splashes paint on another and they start arguing about who was responsible. Alice asks them what they're doing with the paint. And the gardeners explain that they were supposed to put red roses in, but they planted white by mistake, and now they are trying to paint the roses red to correct their mistake. At this point, the King and Queen of Hearts arrive and everyone bows.

Cave Passage

In this story, Alice is in a cave somewhere in Wonderland. * She finds a golden key on a table and † it opens up a tiny door that leads to a beautiful garden. She wants into the garden but she's too tall to fit through the door. So * Alice returns to where † she found the golden key and finds a bottle that says "DRINK ME." Alice drinks the bottle and shrinks down so that she is only ten inches tall. She tries to get through the door again but she realizes that she left the key on the table and now she's too short to reach it, so she starts to cry. Then, * she finds a cake marked "EAT ME" † and she decides to eat it. The cake makes Alice start to grow so tall that she can't fit through the door any more. Alice is kind of upset so she starts to cry. Her tears form a big pool on the ground. At this point, * the White Rabbit runs by, but, as he's running, † he drops his fan and his gloves. Alice picks up the fan and gloves. But after everything that's happened today, * she's feeling not really like herself, † so she starts comparing herself to other girls that she knows to try and convince herself that she hasn't turned into one of them. * She starts to recite a poem † that she knows about a crocodile. But that ends up being all wrong, so Alice concludes that she must be another girl that she knows named Mabel and not Alice. Alice puts on one of the White Rabbit's gloves and this makes her realize that she must be shrinking again, because the glove wouldn't fit her when she was big. She realizes that fanning herself with the rabbit's fan is what's making her shrink so she drops the fan so that she doesn't completely shrink away. Now she's back to her regular size.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

House Passage

In this story, Alice is wandering around Wonderland. She sees a White Rabbit looking for a fan and a pair of his gloves. * Alice is really confused because, just a minute ago, she had been in the Great Hall † but now it's completely vanished. The White Rabbit sees Alice wandering around and mistakes her for his servant, so he calls her "Mary Ann" and orders her to find his gloves and fan. * Alice is really scared and confused so she runs away. Alice eventually comes to the White Rabbit's house † and she decides to help him by looking for his missing fan and gloves in his house. So she looks through the house and she finds the gloves and fan on a table upstairs. * She also finds a bottle on that same table † with the fan and the gloves and she decides to drink it. After drinking the bottle, Alice starts growing and growing and she becomes so tall that she can't fit in the house so she has to lay down in order to stay inside the house and she gets stuck inside. Alice starts to try and talk to herself to calm herself down. Meanwhile, * the White Rabbit returns home but he can't get through the front door because Alice is in the way. † So he tries to climb through a window but he can't get in that way either. So * the White Rabbit has one of his servants try and climb down the chimney † but Alice just kicks him out. Then * the White Rabbit † starts throwing pebbles through the window at Alice. Alice finds a cake and she eats it and it makes her tiny again. Finally, she's able to get out of the house and she runs off into the woods. Now she's safe again.

Duchess Passage

In this story, Alice is wandering through the woods in Wonderland. She starts looking at the Duchess's house. * While she's standing there, a footman who looks like a fish † comes out of the woods and knocks on the door of the house. The door is answered by another footman who looks like a frog. The first footman delivers a message saying that the Duchess is invited to play croquet with the queen. After he delivers the message, * the two footmen bow to each other, † which causes their curly-haired wigs to get tangled together. Alice starts laughing at this and has to run into the forest to hide so that they don't hear her. When Alice comes out of the forest, she knocks on the door to the Duchess's house. But the frog-looking footman tells her she won't be able to get inside because he is still outside and can't let her in. * But Alice just opens the door † and goes inside. Inside the house, the Duchess is nursing a baby and her cook is cooking up soup with too much pepper, which is causing everyone to sneeze. Alice looks around and * she sees the Cheshire Cat † sitting on the hearth and grinning. Meanwhile, * the cook keeps hurling plates † and other items at the Duchess and the baby, but the Duchess doesn't even seem to mind. * Instead, the Duchess just keeps singing to the baby † while tossing it up and down rather violently. Finally, the Duchess leaves to go play croquet and she throws the baby to Alice and tells her to take care of it. Alice is worried about the baby, so she takes it outside, but once she's outside, she realizes that the baby has turned into a pig. She releases the pig into the woods and that's the end of the story.

Appendix B

Plot points scored in recall from each passage.

Cave Passage

1. Alice finds a door that she can't fit through.
2. The door leads into a garden.
3. Alice drinks a bottle labeled "DRINK ME."

4. Alice shrinks
5. Alice is too short to reach the key on the table.
6. Alice cries.
7. Alice eats a cake marked “EAT ME.”
8. Alice grows until she is too tall to fit through the door.
9. Alice’s tears form a pool.
10. The White Rabbit runs by and drops his fan and gloves.
11. Alice compares herself to other girls she knows to make sure she’s not turned into 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 fanning herself is making her shrink.

House Passage

1. The White Rabbit is looking for a fan and a pair of gloves.
2. The great hall Alice had been in has vanished.
3. The rabbit mistakes Alice for a servant and orders her to find his missing items.
4. Alice runs away.
5. Alice goes into the White Rabbit’s house.
6. Alice finds the missing items.
7. Alice drinks a bottle.
8. Alice grows until 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 inside.
11. One of the White Rabbit’s servants tries to climb down the chimney.
12. The White Rabbit throws pebbles at Alice.
13. Alice eats a cake that makes her smaller.
14. Alice runs into the woods.

Duchess Passage

1. Alice is at the Duchess’s house.
2. One footman knocks at the door.
3. Another footman answers the door.
4. The first footman delivers an invitation to the Duchess to play croquet with the Queen.
5. The two footmen bow and their wigs 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.
8. Alice goes inside.
9. The Duchess is nursing a baby and the cook is cooking soup with too much pepper.
10. Alice sees the Cheshire Cat sitting on the hearth.
11. The Cook hurls plates and other items at the Duchess and the baby.
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.

Appendix C

lme4 (Bates, Maechler, & Dai, 2010) model formulae for the final statistical models from both experiments. 1 denotes the intercept, : denotes an interaction between two factors, and (**x** | b)denotes random effects of **x** by b.

Experiment 1

CorrectRecall ~ 1 + FillerVsFluent + CoughVsFluent + Manipulated + Manipulated:FillerVsFluent + Manipulated:CoughVsFluent + (1 + FillerVsFluent + CoughVsFluent|Participant) + (1|PlotPoint), family=binomial

Experiment 2

CorrectRecall ~ 1 + PresenceOfFillers + FillersArePredictive + Manipulated + Manipulated:PresenceOfFillers + Manipulated:FillersArePredictive + (1 + FillersArePredictive | Participant) + (1|PlotPoint), family=binomial

Acknowledgments

We thank Sarah Brown-Schmidt and members of the Communication and Language Lab for their comments and suggestions, Jessica George for recording stimulus materials, and Lisa Brannan, Shefali Khanna, Sujin Park, and Amie Roten for data collection.

This work was supported by National Science Foundation Graduate Research Fellowship 2007053221 to Scott H. Fraundorf and a NIH grant from the National Institute on Deafness and Other Communication Disorders R01DC008774.

References

Almor A, Eimas PD. Focus and noun phrase anaphors in spoken language comprehension. Language and Cognitive Processes. 2008; 23:201–225.

Arnold JE, Hudson Kam CL, Tanenhaus MK. If you say *thee uh*- you're describing something hard: the on-line attribution of disfluency during reference comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2007; 33:914–930.10.1037/0278-7393.33.5.914

Arnold JE, Tanenhaus MK, Altmann RJ, Fagnano M. The old and thee, uh, new: Disfluency and reference resolution. Psychological Science. 2004; 15:578–582.10.1111/j.0956-7976.2004.00723.x [PubMed: 15327627]

Bailey KGD, Ferreira F. Disfluencies affect the parsing of garden-path sentences. Journal of Memory and Language. 2003; 49:183–200.10.1016/S0749-596X(03)00027-5

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Bard EG, Lickley RJ. On not remembering disfluencies. Proceedings of the 5th European Conference on Speech Communication and Technology. 1997:2855–2858.

Barr D. Paralinguistic correlates of conceptual structure. Psychonomic Bulletin & Review. 2003; 10:462–467. [PubMed: 12921425]

Barr D, Seyfeddinipurr M. The role of fillers in listener attributions for speaker disfluency. Language and Cognitive Processes. 2009; 25:441–455.10.1080/01690960903047122

Bates, D.; Maechler, M.; Dai, B. lme4: Linear mixed-effects models using s4 classes [Computer software manual]. 2010. Retrieved from http://lme4.r-forge.r-project.org (R package version 0.999375-37)

Bell A, Jurafsky D, Fosler-Lussier E, Girand C, Gregory M, Gildea D. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. The Journal of the Acoustical Society of America. 2003; 113:1001–1024.10.1121/1.534836 [PubMed: 12597194]

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

Brainard DH. The Psychophysics Toolbox. Spatial Vision. 1997; 10:433–436.10.1163/156856897X00357 [PubMed: 9176952]

Brennan SE, Schober MF. How listeners compensate for disfluencies in spontaneous speech. Journal of Memory and Language. 2001; 44:274–296.10.1006/jmla.2000.2753

Brennan SE, Williams M. The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. Journal of Memory and Language. 1995; 34:383–398.10.1006/jmla.1995.1017

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

Clark, HH.; Clark, EV. Psychology and language: An introduction to psycholinguistics. New York: Harcourt Brace Jovanovich; 1977.

Clark HH, Fox Tree JE. Using *uh* and *um* in spontaneous speaking. Cognition. 2002; 84:73–111. [PubMed: 12062148]

Clark HH, Wasow T. Repeating words in spontaneous speech. Cognitive Psychology. 1998; 37:201–242.10.1006/cogp.1998.0693 [PubMed: 9892548]

Collard P, Corley M, MacGregor LJ, Donaldson DI. Attention orienting effects of hesitations in speech: Evidence from ERPs. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2008; 34:696–702.10.1037/0278-7393.34.3.696

Corley M, Hartsuiker RJ. Hesitation in speech can… um… help a listener understand. Proceedings of the Twenty-Fifth Meeting of the Cognitive Science Society. 2003

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 Linguistic Compass. 2008; 2:589–602.10.1111/j.1749-818X.2008.00068.x

Duez D. Perception of silent pauses in continuous speech. Language and Speech. 1985; 28:377–389. [PubMed: 3842875]

Fan J, McCandliss BD, Sommer T, Raz A, Posner MI. Testing the efficacy and independence of attentional networks. Journal of Cognitive Neuroscience. 2002; 14:340–347.10.1162/089892902317361886 [PubMed: 11970796]

Fox Tree JE. The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. Journal of Memory and Language. 1995; 34:709–738.10.1006/jmla.1995.1032

Fox Tree JE. Listeners' uses of um and uh in speech comprehension. Memory & Cognition. 2001; 29:320–326.

Fox Tree JE, Clark HH. Pronouncing "the" as "thee" to signal problems in speaking. Cognition. 1997; 62:151–167. [PubMed: 9141905]

Fraundorf, SH.; Watson, DG. Dimensions of variation in disfluency production in discourse. In: Ginzburg, J.; Healey, P.; Sato, Y., editors. Proceedings of LONDIAL 2008, the 12th Workshop on the Semantics and Pragmatics of Dialogue. London: King's College London; 2008. p. 131-138.

Freeman E, Heathcote A, Chalmers K, Hockley W. Item effects in recognition memory for words. Journal of Memory and Language. 2010; 62:1–18.10.1016/j.jml.2009.09.004

Graesser AC, Millis KK, Zwaan RA. Discourse comprehension. Annual Review of Psychology. 1997; 48:163–189.

Kintsch W, van Dijk TA. Toward a model of text comprehension and production. Psychological Review. 1978; 85:363–394.10.1037/0033-295X.85.5.363

Labov, W.; Ash, S.; Boberg, C. The atlas of North American English: phonetics, phonology, and sound change. New York: Walter de Gruyter; 2006.

Landis JR, Koch G. The measurement of observer agreement for categorical data. Biometrics. 1977; 33:159–174.10.2307/2529310 [PubMed: 843571]

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

Macdonald MC, Christiansen MH. Reassessing working memory: Comments on Just and Carpenter (1992) and Waters and Caplan (1996). Psychological Review. 2002; 109:35–54.10.1037//033-295X.109.1.35 [PubMed: 11863041]

MacGregor LJ, Corley M, Donaldson DI. Not all disfluencies are equal: The effects of disfluent repetitions on language comprehension. Brain & Language. 2009; 111:36–46.10.1016/j.bandl.2009.07.003 [PubMed: 19700188]

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

Novick JM, Thompson-Schill SL, Trueswell JC. Putting lexical constraints in context into the visual-world paradigm. Cognition. 2008; 107:850–903.10.1016/j.cognition.2007.12.011 [PubMed: 18279848]

Pelli DG. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision. 1997; 10:437–442.10.1163/156856897X00357 [PubMed: 9176953]

R Development Core Team. Vienna, Austria: 2008. R: A language and environment for statistical computing [Computer software manual]. Retrieved from http://www.r-project.org

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.

Smith VL, Clark HH. On the course of answering questions. Journal of Memory and Language. 1993; 32:25–38.10.1006/jmla.1993.1002

Swerts M. Filled pauses as markers of discourse structure. Journal of Pragmatics. 1998; 30:485–496.10.1016/S0378-2166(98)0014-9

Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JC. Integration of visual and linguistic information in spoken language comprehension. Science. 1995; 268:1632–1634.10.1126/science.7777863 [PubMed: 7777863]

Watanabe M, Hirose K, Den Y, Minematsu N. Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. Speech Communication. 2008; 50:81–94.10.1016/j.specom.2007.06.002

NIH-PA Author Manuscript

NIH-PA Author Manuscript

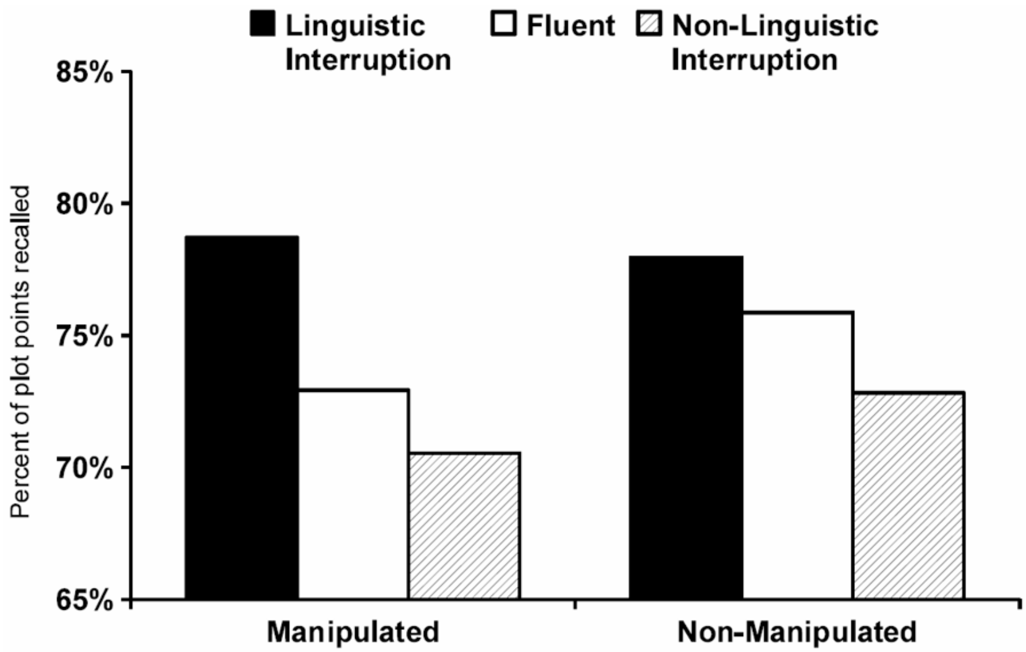

Figure 1.
Recall of manipulated and non-manipulated plot points in Experiment 1 as a function of interruption condition.

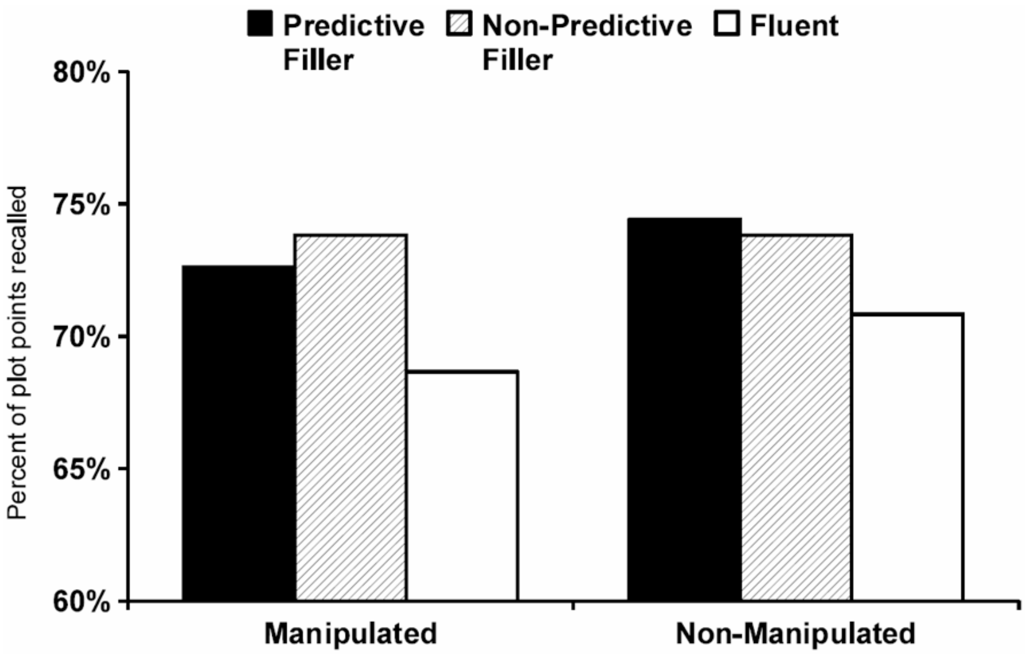

Figure 2.
Recall of manipulated and non-manipulated plot points in Experiment 2 as a function of filler condition.

Table 1

Fixed Effect Estimates for Multi-Level Logit Model of Recall of Plot Points in Experiment 1 (N = 3021, log-likelihood: -1328).

Fixed effect	Coefficient	*SE*	Wald *z*	*p*
Intercept	1.69	0.25	6.73	<.001
Story contains fillers (vs. fluent)	0.45	0.23	1.96	<.05
Story contains coughs (vs. fluent)	-0.45	0.19	-2.34	<.05
Plot point is manipulated	-0.06	0.42	-0.14	.89
Manipulated plot point x filler	0.25	0.30	0.84	.40
Manipulated plot point x cough	-0.09	0.29	-0.32	.75

Table 2

Summary of Random Participant and Item Effects and Correlations in Model of Recall of Plot Points in Experiment 1.

Random effect	s^2	Correlation with random intercept	Correlation with effect of fillers
Subject			
Intercept	1.38		
Story contains fillers	2.15	-0.31	
Story contains coughs	1.07	0.08	-0.78
Plot point			
Intercept	1.70		

Table 3

Fixed Effect Estimates (Top) for Multi-Level Logit Model of Recall of Plot Points in Experiment 2 (N = 1764, log-likelihood: -798).

Fixed effect	Coefficient	*SE*	Wald *z*	*p*
Intercept	1.60	0.31	5.11	<.001
Story contains fillers (vs. fluent)	0.32	0.14	2.32	<.05
Fillers are in predictive location	0.14	0.27	-0.51	.61
Plot point is manipulated	-0.01	0.51	-0.01	.99
Manipulated plot point x filler	-0.05	0.37	-0.15	.88
Manipulated plot point x predictive filler	0.24	0.33	-0.71	.48

Table 4

Summary of Random Subject and Item Effects and Correlations in Model of Recall of Plot Points in Experiment 2.

Random effect	s^2	Correlation with random intercept
Subject		
Intercept	1.42	
Fillers in predictive location	1.76	-.54
Plot point		
Intercept	2.39	