Hesitation disfluencies in spontaneous speech: The meaning of um.

Citation for published version:
Corley, M & Stewart, OW 2008, 'Hesitation disfluencies in spontaneous speech: The meaning of um.', *Language and Linguistics Compass*, vol. 2, no. 4, pp. 589-602. https://doi.org/10.1111/j.1749-818X.2008.00068.x

Digital Object Identifier (DOI):
10.1111/j.1749-818X.2008.00068.x

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Language and Linguistics Compass

Hesitation disfluencies in spontaneous speech: The meaning of *um*

Martin Corley and Oliver W. Stewart

Philosophy, Psychology, and Language Sciences

University of Edinburgh

Martin.Corley@ed.ac.uk

In Press, Language and Linguistics Compass

Human speech is peppered with *um*s *and uh*s, among other signs of hesitation in the planning process. But are these so-called fillers (or filled pauses) intentionally uttered by speakers, or are they side-effects of difficulties in the planning process? And how do listeners respond to them? In the present paper we review evidence concerning the production and comprehension of fillers such as *um* and *uh*, in an attempt to determine whether they can be said to be "words" with "meanings" that are understood by listeners. We conclude that, whereas listeners are highly sensitive to hesitation disfluencies in speech, there is little evidence to suggest that they are intentionally produced, or should be considered to be words in the conventional sense.

Hesitation disfluencies in spontaneous speech: The meaning of *um*

When humans communicate, their messages are conveyed by more than just the words they use. They can, for example, use gesture to indicate what a phrase such as “this one” refers to, or change their tone of voice to show that the assertion that they “love linguistics” is not to be taken at face value. As well as these (presumably deliberate) additional components of communication, the words of unprepared spoken language are likely to be accompanied by a range of unintentional errors. If, for example, a speaker exchanges the onsets of two words, by perhaps saying “darn bore” when intending to say “barn door”, it is highly unlikely that the exchange is intentional, and the occurrence of accidental speech errors like this may inform us about the nature of speech planning (e.g. Hartsuiker et al. 2005). Between these extremes of intentionality are *disfluencies,* or the false starts, repetitions, and hesitations that accompany the words that speakers plan and utter. Averaging across several studies, Fox Tree (1995) estimated that approximately 6% of words uttered are, or are affected by, some form of disfluency (see also Bortfeld et al 2001). These disfluencies may not always be accidental: It has been argued that some types of disfluency should be counted among the tools the speaker has for communicating to the listener, alongside things such as tone of voice (others, such as false starts, may be due to speakers editing their own speech). Chief among the potentially communicative disfluencies are the so-called *fillers*, such as *um* and *uh*, which (together with prolongations and pauses) mark a hesitation on the part of the speaker. In this paper, we investigate the role played by hesitation in human communication, with a particular focus on fillers and the communicative goals they may serve.

Producing Hesitation Disfluencies

Hesitation phenomena such as fillers are most likely to occur at the beginning of an utterance or phrase, presumably as a consequence of the greater demand on planning processes at these junctures (Barr 2001; Beattie 1979; Maclay & Osgood 1959). The view that cognitive load is an important predictor of disfluency is supported by the fact that disfluencies are found to occur more often before longer utterances (Oviatt 1995; Shriberg 1996), and when the topic is unfamiliar (Bortfeld et al 2001; Merlo & Mansur 2004). Cognitive load is also implicated when we look at hesitations on a word-by-word basis. Investigations of where disfluencies such as fillers occur throughout utterances have established that they are more likely to occur before content words (Maclay & Osgood 1959), such as low-frequency color names (Levelt 1983). However, Beattie and Butterworth (1979) came to a different conclusion when they investigated the distributional properties of disfluencies across a set of recordings of two-person conversations. They showed that both low-frequency content words and those rated as contextually improbable were likely to be preceded by hesitations such as fillers; when frequency was held constant, contextual probability still predicted disfluency. Rather than attributing disfluency to cognitive load, Beattie and Butterworth suggested that speakers might be aware of an element of choice in selecting words with low contextual probability, and were more likely to be disfluent for this reason. Choice was also implicated in a study by Schachter et al. (1991) in which lectures in the natural sciences, social sciences, and humanities were recorded and analyzed for numbers of fillers per minute. Disfluency differed between topics, with the natural sciences resulting in the least and the humanities in the most frequent use of fillers. However, when the lecturers were interviewed on general topics, their rates of disfluency did not differ. Schachter et al. attributed the differences in lecture disfluency rates to the fact that there were fewer linguistic options in the sciences, causing lecturers to hesitate less as they selected appropriate terms. They later corroborated their claims by measuring vocabulary size in lectures, learned articles, and topic-related journalism (Schachter et al. 1994), showing that there were indeed fewer terms used in the sciences.

So far, we have been speaking about cognitive load and choice as if they were different, but it is of course the case that a higher number of options to choose from could result in an increased cognitive load. In a detailed experimental investigation, Oomen and Postma (2001) manipulated speech rate, using a task modified from Levelt (1983; see also Martin et al. 1989). Participants were required to describe the progress of a dot which moved either quickly or slowly over a set of pictures connected by a network of paths. Although there was significant evidence of increased cognitive load at the faster speech rate (for example, participants were more likely to omit syntactically obligatory constituents), hesitation disfluencies showed an interesting pattern: Participants were more likely to repeat words, but no more likely to use fillers such as *uh*, in the fast conditions. Oomen and Postma suggest that the increase in repetitions can be explained by Blackmer and Mitton's (1991) “autonomous restart capacity”: If the cognitive processes that result in a phonetic plan fail to keep up with articulation, the articulation of the existing phonetic plan is restarted, causing words to be articulated more than once. However, the use of fillers does not appear to be subject to a default strategy such as this one. In fact, cognitive load alone does not appear to explain their production.

In a study designed to explore the issue of choice further, Schnadt and Corley (submitted) used a variant of Oomen and Postma’s (2001) network task in which the speech rate remained constant but the items pictured at the nodes of the network varied. Each picture either had one or had several potential names, and the lexical frequency of the (most likely) name was either high or low. Schnadt and Corley found that hesitation phenomena such as prolongations and fillers increased in the words just before multiple-named or low-frequency items. In a second experiment, participants completed a naming task during which they were exposed to the preferred name of each picture before doing the network task itself. This was designed to reduce choice when describing the networks. Although the general pattern of results remained the same, with more hesitation disfluency for difficult-to-name items, in this experiment the numbers of fillers didn’t increase for multiple-named pictures, presumably because participants already had a name in mind. Schnadt and Corley concluded that fillers were more likely to occur where the speaker had a choice of what to name a picture, perhaps because the speaker anticipated a longer delay in these circumstances, consistent with a proposal by Smith and Clark (1993).

The evidence above suggests that disfluencies may not necessarily all be automatic (i.e. part of an underlying mechanism) in their relation to an increased cognitive load. Some of them – notably, fillers – may be a part of the speaker’s expressive armory. Further support for this view comes from the fact that disfluency rates vary with context: Speakers tend to be more disfluent overall when addressing other humans than when addressing machines, and more disfluent still in dialog than in monologue situations (Oviatt 1995, although 77% of the variance in disfluency in these studies is accounted for by utterance length). In a detailed investigation of the circumstances under which disfluency occurs, Bortfeld et al. (2001) analyzed a recorded collection of task-oriented dialogs. Among other factors, they found that the roles participants played (of either describing images or attempting to find the images that matched the descriptions) greatly influenced the numbers of disfluencies, and particularly fillers, produced, regardless of utterance length or complexity. Additional studies have shown that the disfluencies people use do seem to indicate something about the message that is being conveyed: For example, when speakers are asked to rate their own confidence in answers they have just given to general knowledge questions, less confident answers are likely to have been marked by an utterance-initial hesitation including a filler (Smith & Clark 1993; Brennan & Williams 1995; Swerts & Krahmer 2005).

The most explicit claim that fillers serve a communicative function, effectively as words in the speaker's vocabulary, comes from Clark and Fox Tree (2002), who argue that *um* and *uh* should be considered as integral to the information the speaker is trying to convey, although they do not add to the propositional content, or *primary message*. Instead, fillers are part of a *collateral* message, in which the speaker is commenting on her performance (Clark 1994; 2002). In a detailed argument, Clark and Fox Tree (2002: 103) claim that fillers conform to the "phonology, prosody, syntax, semantics, and pragmatics of English words". They can be used to transmit a variety of interpersonal messages, such as "speakers want to keep the floor", and should therefore be considered members of the word class of interjections, like *ah* and *oh*. Moreover, Clark and Fox Tree argue that *um* and *uh* serve different functions: *Um* is used when the speaker is having greater difficulty than *uh*. This assertion is in part supported by the fact that *um*s are followed by longer pauses than *uh*s in several corpora of spontaneous speech (although the generality of these findings has been challenged by O'Connell and Kowal 2005; and it should also be noted that in two of the three corpora examined by Clark and Fox Tree, pause length was subjectively determined by the original transcribers).

Despite Clark and Fox Tree's arguments, the evidence for a communicative account of hesitation phenomena remains circumstantial. It is true that fillers tend to occur in situations where there is uncertainty or choice, whether the uncertainty is about the best word with which to name a picture or about the name of the capital of Romania. But this does not prove that they are used by speakers to signal, for example, that there will be a delay in the speech stream due to uncertainty, except in the sense that smoke signals fire. In the cases that we have considered, delays are in any case most likely highly correlated with uncertainty about what to say next. For example, in the recorded speech examined by Bortfeld et al. (2001), dialog partners describing pictures will undoubtedly have a more challenging task formulating utterances than those verifying descriptions. Similarly, finding the answer to a question may result in a substantial hesitation, which may be marked by a disfluent pause or filler. In general, it is hard to determine the reason that a speaker is disfluent, especially if the investigation is carried out after the fact from a corpus of recorded speech (although more experimental approaches have been taken by, for example, Barr 2001; Oomen & Postma 2001; Schnadt & Corley submitted). For that reason, many researchers have focused on the comprehension of disfluent language: What, of anything, do disfluencies such as fillers communicate to the listener?

Understanding Speech with Hesitation Disfluencies

A consequence of the view that fillers such as *um* and *uh* are words, as suggested by Clark & Fox Tree (2002), is that they should be treated by the listener as a part of the message, just like any other word. And just like any other word, they should contribute to the meaning of the message, in the sense that the message would not be identical if the filler were not present. This suggestion is a radical departure from earlier views of speech comprehension, which assumed that disfluencies were noise in the signal (Brennan & Schober 2001: 275). For example, computational approaches have typically focused on the identification of disfluencies so that they could be filtered from the representation of the intended utterance (e.g. Charniak & Johnson 2001).

Potential support for the suggestion that disfluencies should be classified as noise rather than words comes from Lickley and colleagues (Lickley 1995; Lickley & Bard 1996), who show that listeners tend not to be able to accurately identify the locations of fillers in sentences they have just heard, perhaps because they are produced at a lower fundamental frequency than the rest of the utterances they are part of (Shriberg & Lickley 1993). But this may not be very different from the well-known finding that listeners tend to forget the surface forms of sentences they have heard almost immediately (e.g. Jarvella 1971); there is no reason to suppose that ‘understanding’ a filler makes its surface form more memorable.

Brennan and Williams (1995) offer a straightforward demonstration of listener sensitivity to fillers. They asked listeners to rate answers to general knowledge questions which had been obtained during their production study. These were digitally manipulated such that they were sometimes preceded by either a silent pause, or by a silent pause including a filler. Based on their observation that speakers were more likely to use utterance-initial fillers in cases where they were uncertain about their answers to general knowledge questions, Brennan and Williams reasoned that this information might be conveyed to listeners. Accordingly, the listeners in their study rated the speakers as being less confident about their answers in cases where the recorded answers were preceded by silence, and less confident still when that silence contained a filler. On the face of it, this seems to be straightforward evidence that the ways in which listeners interpret utterances are affected by disfluency. But listeners were explicitly asked to rate speaker confidence, and it seems equally plausible that they were simply evaluating speaker performance, based on available evidence. Brennan and Williams' (1995) study shows that rather than being automatically filtered out, fillers such as *uh* can provide a useful source of evidence to listeners, but it does not allow us to conclude whether they are integral to the meaning of an utterance or whether they are an aspect of its performance, such as facial gestures (e.g. Swerts & Krahmer 2005) or blushing.

Rather than relying on listeners' explicit interpretations, several studies on speech comprehension have focused on the ways in which disfluencies may help listeners as a part of the ongoing comprehension process. For example, Howell and Young (1991) found that listeners rated utterances including self-repairs as more comprehensible when those repairs were preceded by pauses. In an on-line task requiring participants to move a cursor to one of three geometric shapes, Brennan and Schober (2001) observed that compared to fluent controls, between-word interruptions (*yellow-orange*), and mid-word interruptions with or without fillers (*yel-uh-orange*, *yel-orange*) led to quicker identification of the correct (repair) word (*orange*). The quickest identifications were in cases where the interruption included a filler, or a silent pause of the same length. These findings strongly suggest that pauses and fillers help the identification of upcoming words. In a similar vein, Fox Tree (2001) showed that both English and Dutch listeners were faster to identify a target word in a carrier sentence when it followed an *uh*, in comparison to a control condition without the *uh*. In a comparable set of conditions with or without an *um* there was no effect. Fox Tree suggested that different fillers have different effects because they convey information: *Uh* signals a shorter upcoming delay than *um* (as suggested by Clark & Fox Tree 2002). When confronted with a short (predictable) delay in the speech signal, listeners are able to heighten attention for upcoming speech; however, in the face of a longer delay signaled by *um*, heightening attention would not be functional, since there would be no reason to expect the next word any time soon.

Although they may seem to suggest that fillers can be communicative, the conclusion that can be drawn from these studies is rather weaker: They demonstrate that in some circumstances, fillers such as *uh* can help the listener perform a particular task, such as responding to a predetermined target word. In these experiments, the fillers are simply affecting the *process* of comprehension: They may for example be improving intelligibility, in the same way that clear pronunciation does (perhaps in the case of fillers, by separating words). Alternatively, since the pronunciation of fillers takes time, it may be the delay in the speech signal which drives listeners' responses (Brennan & Schober 2001). Fox Tree's (2001, Experiment 2) materials, for example, included substantial silent pauses before and after the fillers. In the case of *um*s, an average 1004ms silence remained even after the *um* had been excised, compared to a 380ms of silence for *uh*, suggesting that the materials with excised *um*s may have been as 'disfluent' as those which included them. In Brennan and Schober's (2001) study, extra time in the speech signal would have allowed participants more time to prepare a response, having ruled out the original (*yellow*) possibility from a limited set of potential referents. We will revisit the issue of the time taken by disfluencies below. First, we consider a number of studies which have tried to access the comprehension processes more directly, in an attempt to establish the effect that disfluencies have when the task is designed to measure the ongoing comprehension of, rather than the speed of response to, a linguistic stimulus.

To examine whether fillers directly affect comprehension requires a paradigm which is sensitive to the ongoing comprehension of the acoustic signal the listener encounters. One such paradigm was used by Barr (2001), who investigated the ways in which participants moved computer mice in response to spoken instructions. Barr first carried out a production experiment which established that, when describing an abstract shape that they hadn't described before, speakers produced 34% more fillers than when describing one which they had previously described (see also Arnold et al. 2003; Arnold & Tanenhaus in press). This is consistent with the observation that speakers find it more difficult to introduce new items into the discourse than to speak about old ones (e.g. Almor 1999). In the comprehension stage of the study, listeners were required to move a mouse cursor to one of two shapes: One which had twice been previously mentioned, and one which had been shown on-screen but not referred to in the recorded instructions. When the instructions included an *um* as opposed to length-matched background noise, participants were faster to click on the new referent; moreover, they were likely to begin moving the mouse towards the new referent during the *um* before it was described. In a similar vein, Arnold et al. (2004) used an eyetracking paradigm in which participants responded to spoken instructions to manipulate one of four objects depicted on a computer display. Of interest in each trial were two of the objects, which shared a phonological onset (e.g. camel and candle). One of the objects was given, because it had been mentioned in the previous trial, whereas the other had not been previously mentioned and was new to the discourse. Arnold et al. manipulated the fluency of the spoken instructions, such that some instructions included fillers and concomitant signs of difficulty for the speaker. Where the instructions were fluent, listeners were more likely to fixate on the given competitor 200-600ms after target onset, or as early as phonetic information about the two potential targets started to become available (200ms is generally taken as a lower bound for the planning and execution of an eye movement: e.g. Matin et al. 1993). On the other hand, in cases where the instructions were disfluent, listeners were more likely to fixate on the new item, indicating that the disfluency had been taken into account, and a mention of the given item was no longer predicted (see also Arnold et al. 2003). In a recent extension to this work, Arnold et al. (2007) compared fixations on concrete and abstract items in fluent and disfluent conditions. They found that disfluency in an utterance increased fixations on both discourse-new items, and on abstract items which were likely to be difficult to name (see also Watanabe et al. 2008). Arnold and Tanenhaus (in press) suggest that both sets of findings can be encompassed by an expectancy hypothesis. Based on evidence that listeners tend to predict what speakers are likely to say (e.g. Altmann & Kamide 1999), they suggest that the evidence of difficulty in the disfluent instructions leads listeners to predict an upcoming referent which is difficult for the speaker. When presented visually with a set of items, hesitation in an utterance prompts the listener to expect, and fixate on, the most problematic item. In these experiments, speaker difficulty would correspond either to discourse status (new items are harder for speakers to mention than old: Almor 1999) or on the ease with which a name can be found for a picture.

Taken together, Barr's (2001) and Arnold et al.'s (2004; Arnold et al. 2007) evidence presents a persuasive picture: Disfluencies which mark hesitation appear to directly affect the comprehension process. When a spoken message contains a filler, such as *um*, listeners are able to make a concrete prediction about what is likely to be mentioned next (and, importantly, this prediction isn't arbitrary: It corresponds to the item that is likely to cause the speaker to become disfluent; e.g. Barr 2001; Schnadt & Corley submitted). However, these experiments rely on the fact that the listener is visually presented with a limited set of candidate referents. This gives rise to two issues. First, it may be the case that prediction is facilitated by disfluency only when the information is freely available (participants in these experiments always know that one of as few as two items is likely to be mentioned). Second, it is difficult to determine whether participants are truly predicting anew what the speaker is likely to say, or whether they are simply abandoning the most likely option. Participants who responded to hesitations by 'predicting the difficult referent' in the Arnold and Barr studies would behave identically to those 'ruling out the easy one'.

The first of these issues has been addressed in a series of ERP studies, in which participants were required to listen to utterances, but there was no secondary task and no associated visual display of potential referents (Corley et al. 2007; Collard et al. in press). Corley et al. (2007) focused on the N400, an ERP component associated with the meaningful processing of language (Kutas & Hillyard, 1980, 1984). Where integration of a written or spoken word into the preceding context is difficult (for example because that word is not predictable from its context), it is typical to observe a negative change in voltages recorded at the scalp, particularly at the crown, relative to more easily integrated words. This difference, the N400 effect, peaks at around 400 ms after word onset. If, as Barr's (2001) and Arnold et al.'s (2004) studies suggest, disfluency in an utterance can affect what the listener is predicting, then the N400 effect should be affected when utterances contain fillers, compared to their fluent counterparts. In Corley et al.'s study, participants listened to recordings of sentences that ended with words that were either highly predictable, given their contexts (in pretesting, 84% of respondents had ended written versions of the sentences with the same words) or not at all predictable (no respondent had come up with those words, although they still made sense in context). In line with previous research, there was a reliable N400 effect relative to target onset for the predictable compared to unpredictable words. Importantly, half of the materials were disfluent: They included the filler *uh*, together with other signs of disfluency such as a lengthening of the word *the*, just before the target words. Although the N400 effect remained reliable in these cases, it was significantly reduced, suggesting that the difference in integration difficulty for unpredictable compared to predictable words had been substantially affected by disfluency. This evidence does not directly answer the second question of whether participants were predicting something difficult or had simply stopped predicting something easy (although a comparison of the ERP plots across conditions hints that the latter may be the case). However, it does show that disfluency affects the comprehension process, and most likely processes involved with predicting upcoming information, even in the absence of relevant visual material.

Note that Corley et al.'s N400 evidence is subject to the same criticism that was made earlier of studies which show improved reaction times (e.g. Brennan & Schober 2001; Fox Tree 2001): Whereas the attenuation of the N400 shows that disfluency affects the *process* of comprehension, that does not mean that the *outcome* is affected. In order to show that the differences in N400s were associated with consequences for the way in which the message was represented, Corley et al. (2007) asked participants to take part in a surprise recognition study after the listening part of the experiment. In this phase, participants were visually presented with words which either had or hadn't been in the experiment, and had to indicate which of the words they thought they had heard previously. Words which had been preceded by disfluency when they were initially encountered were more likely to be later recognized, showing that disfluency in an utterance can have direct consequences for the way in which the words of that utterance are represented.

The recognition memory effect has since been replicated by Collard et al. (in press). However, the ERP phase of Collard et al.'s experiment was designed to explore in more detail the process by which disfluency might render words more memorable. In doing this, they built on the suggestion by Fox Tree (2001) that disfluency might, in some circumstances at least, cause listeners to heighten attention. Instead of the N400 component discussed above, Collard et al. focused their study on the MMN (early frontal negativity) and the P300 complex (relative positivity maximal 300ms after stimulus onset at the center of the scalp). The MMN (mismatch negativity) is associated with the detection of change (Schröger 1997); the P300 refers to a family of components thought to index the orientation of attention and updating of memory (Polich 2004). Instead of manipulating linguistic predictability, Collard et al. (in press) used only the predictable targets from Corley et al. (2007), manipulating the acoustic characteristics of the targets such that half of them were acoustically incongruous with their contexts. As would be predicted, incongruous targets in fluent stimuli gave rise to an MMN followed by a large P300 effect: Participants noticed the incongruity, which subsequently drew their attention. In the disfluent case, an MMN was evident but the P300 effect vanished completely. Collard et al. reasoned that attention was already engaged in response to the disfluent *uh* in the stimuli, so that no further orientation of attention could be observed. Because more attention was paid to the subsequent material, it was more easily recognized during the surprise recognition tasks in this and in Corley et al.'s (2007) study. Once attention was engaged, predictive processes (as implicated by Arnold et al. 2004; Arnold & Tanenhaus in press; Barr 2001; Corley et al. 2007) might be affected.

Taken together, the findings from the studies above demonstrate clearly that listeners' comprehension can be affected by fillers like *um* and *uh*. This would be predicted if these fillers were used as interjections by speakers, as contended by Clark and Fox Tree (2002). Note however that this is not the only plausible account of the findings: It could be the case that speakers produce fillers quite unintentionally (as a byproduct of delay, for example), but at predictable junctures, and listeners are sensitive to these accidental patterns of occurrence. In other words, the evidence above is consistent with the view that fillers are words, but does not rule out other explanations. Because the evidence *for* the proposition that *um* and *uh* are words is equivocal, we turn our attention to evidence *against* this position, by considering the time that fillers take to utter.

Our earlier contention was that fillers might affect listeners not because they are interjections, like *ah* and *oh*, but because they take time (in effect, delaying the continuation of the utterance). Striking evidence that time may at least be a factor in the comprehension of disfluent utterances comes from Bailey and Ferreira (2003). Given that fillers often occur at phrase boundaries, one function they may serve is to disambiguate syntactic structures, in a similar way to optional function words such as *that* in *The horse (that) raced past the barn fell.* In a series of studies, Bailey and Ferreira (2003) investigated just such a possibility. They asked people to rate the grammaticality of spoken sentences such as *Sandra bumped into the busboy and the waiter told her to be careful*, which contains a temporary ambiguity (did Sandra bump into the waiter?). When a disfluency coincided with the start of the second clause (*the uh uh waiter told*) participants were more likely to rate the sentence as grammatical than when the disfluency occurred later in the sentence (*the waiter uh uh told*), once again suggesting that the disfluencies were influential in comprehension: Listeners were more likely to assume that a noun phrase was the start of a new clause if it was disfluent. Importantly, a replication of this experiment substituted "environmental noise" (dogs barking, telephones ringing) as plausible interruptions in place of the fillers. Once again, participants were more likely to rate the utterances as grammatical if the interruption coincided with the start of the second clause. Although some care in interpretation is warranted, given that rating for grammaticality and listening for comprehension may engage very different processes, there is little reason on this evidence to assume a special word-like role for fillers beyond the time that they add to the utterance. More mixed evidence comes from (as yet unpublished) experiments run by Lucy MacGregor in our own lab, which replicate Corley et al. (2007) with silences and disfluent repetitions (*the- the dog*) in place of the original *uh*s, and suggest that the picture is more complex: In those studies, the N400 does not appear to be greatly affected by repetitions in the utterance, whereas the results for silence are more equivocal (MacGregor, 2008). Clearly, further investigation is warranted before we are able to answer just which characteristics of a disfluency allow the listener to recognize it as disfluent; but the suggestion that time is a factor militates against the proposition that listeners are sensitive to the particular forms that hesitations take.

Conclusions

Taken together, the studies we have surveyed show that fillers occur in utterances when speakers are uncertain (Brennan & Williams 1995; Smith & Clark 1993), or when they have choices to make (Schachter et al. 1991; Schnadt & Corley submitted). They affect the comprehension process, facilitating understanding (Brennan & Schober 2001), and allowing the listener to amend their predictions about what might be said next (Arnold et al. 2004; Corley et al. 2007) or evaluate the speaker's confidence in what they are saying (Brennan & Williams 1995). Although this evidence is consistent with the view that *um* and *uh* are words, other studies suggest that they do not have such a status. Speakers may produce them automatically when there is a delay to the speech plan (Smith & Clark 1993), and it may be the case that any plausible interruption to fluent speech will affect listeners in the same way as a filler (Bailey & Ferreira 2003; Brennan & Schober 2001). One unexplored entailment of Clark and Fox Tree's (2002) claim that fillers transmit interpersonal messages is that speakers must presumably *intend* to transmit those messages. Evidence that the fillers speakers produce are designed for their audiences (e.g. are used more when they can inform listeners) would rule out a simple delay-based hypothesis of filler production, suggesting that speakers at least have some intentional control over the production of *um* and *uh,* in the way that they have control over the production of other words.

In the absence of such evidence, the picture remains murky. There is no conclusive evidence that fillers are words. What can be surmised at present is that, like a facial gesture or a tone of voice, hesitation disfluencies like *um* provide information to the listener. The information is something like "pay attention, the speaker's in trouble and the next part of the message might not be what you predicted".

Altmann, G. T. M. and Y. Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. Cognition 73.247-264.

Almor, A. 1999. Noun-phrase anaphora and focus: The informational load hypothesis. Psychological Review, 106.748-765.

Arnold, J. E., M. Fagnano, and M. K. Tanenhaus. 2003. Disfluencies signal theee, um, new information. Journal of Psycholinguistic Research 32.25-36.

Arnold, J. E., C. L. Hudson-Kam, and M. K. Tanenhaus. 2007. If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition 33(5).914-930.

Arnold, J. E., and M. K. Tanenhaus. in press. Disfluency effects in comprehension: How new information can become accessible. The processing and acquisition of reference, ed. by E. Gibson and N. Perlmutter. MIT Press.

Arnold, J. E., M. K. Tanenhaus, R. J. Altmann, and M. Fagnano. 2004. The old and thee, uh, new: Disfluency and reference resolution. Psychological Science 15.578-582.

Arnold, J. E., T. Wasow, R. Ginstrom, and T. Losongco. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. Language 76.28-55.

Bailey, K. G. D., and F. Ferreira. 2003. Disfluencies affect the parsing of garden-path sentences. Journal of Memory and Language 49.183-200.

Barr, D. J. 2001. Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. Oralité et gestualité, communication multimodale, intéraction, ed. by S. Santi, I. Guaïtella, C. Cave, and G. Konopczynski, 597-600. Paris : L'Harmattan.

Beattie, G. W. 1979. Planning units in spontaneous speech: Some evidence from hesitation in speech and speaker gaze direction in conversation. Linguistics 17.61-78.

Beattie, G. W., and B. L. Butterworth. 1979. Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. Language and Speech 22.201-211.

Blackmer, E. R., and J. L. Mitton. 1991. Theories of monitoring and the timing of repairs in spontaneous speech. Cognition 39.173-194.

Bortfeld, H., S. D. Leon, J. E. Bloom, M. F. Schober, and S. E. Brennan. 2001. Disfluency rates in spontaneous speech: Effects of age, relationship, topic, role, and gender. Language and Speech 44.123-147.

Brennan, S. E., and M. Williams. 1995. 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 34.383-398.

Brennan, S. E., and M. F. Schober. 2001. How listeners compensate for disfluencies in spontaneous speech. Journal of Memory and Language 44.274-296.

Charniak, E., and M. Johnson. 2001. Edit detection and parsing for transcribed speech. Proceedings, 2nd meeting of the North Amercian chapter of the Association for Computational Linguistics. 118-126. Pittsburgh, PA.

Clark, H. H. 1994. Managing problems in speaking. Speech Communication 15.243-250.

Clark, H. H. 2002. Speaking in time. Speech Communication 36.5-13.

Clark, H. H., and J. E. Fox Tree. 2002. Using uh and um in spontaneous speaking. Cognition 84.73-111.

Collard, P., M. Corley, L. J. MacGregor, and D. I. Donaldson. in press. Attention orienting effects of hesitations in speech: Evidence from ERPs. Language and Cognitive Processes.

Corley, M., L. J. MacGregor, and D. I. Donaldson. 2007. It's the way that you, er, say it: Hesitations in speech affect language comprehension. Cognition 105.658-668.

Fox Tree, J. E. 1995. The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. Journal of Memory and Language 34.709-738.

Fox Tree, J. E. 2001. Listeners' uses of um and uh in speech comprehension. Memory & Cognition 29.320-326.

Hartsuiker, R. J., M. Corley, and H. Martensen. 2005. The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars, Motley, and MacKay (1975). Journal of Memory and Language 52.58-70.

Howell, P., and K. Young. 1991. The use of prosody in highlighting alteration in repairs from unrestricted speech. Quarterly Journal of Experimental Psychology 43(A).733-758.

Jarvella, R. J. 1971. Syntactic Processing of Connected Speech. Journal of Verbal Learning and Verbal Behavior 10.409-416.

Kutas, M., and S. A. Hillyard. 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. Science 207.203-205.

Kutas, M., and S. A. Hillyard. 1984. Brain potentials during reading reflect word expectancy and semantic association. Nature 307.161-163.

Levelt, W. J. M. 1983. Monitoring and self-repair in speech. Cognition 14.41-104.

Lickley, R. 1995. Missing disfluencies. Proceedings, International Congress of Phonetic Sciences 4.192-195. Stockholm.

Lickley, R., and E. G. Bard. 1996. On not recognizing disfluencies in dialog. Proceedings, International Conference on Spoken Language Processing, 1876–1879. Philadelphia.

Maclay, H., and C. E. Osgood. 1959. Hesitation phenomena in spontaneous speech. Word 15.19-44

MacGregor, L. 2008. Disfluencies affect language comprehension: Evidence from event-related potentials and recognition memory. Unpublished doctoral dissertation. University of Edinburgh.

Martin, N., R. W. Weisberg, and E. M. Saffran. 1989. Variables influencing the occurrence of naming errors: Implications for models of lexical retrieval. Journal of Memory and Language 24.462-485.

Matin, E., K. C. Shao, and K. R. Boff. 1993. Saccadic overhead: Information-processing time with and without saccades. Perception & Psychophysics, 53.372-380.

Merlo, S., and L. L. Mansur. 2004. Descriptive discourse: topic familiarity and disfluencies. Journal of Communication Disorders 37.489-503.

O’Connell D. C., and S. Kowal. Uh and um revisited: Are they interjections for signaling delay? Journal of Psycholinguistic Research 34.555-576.

Oomen, C. C. E., and A. Postma. 2001. Effects of time pressure on mechanisms of speech production and self-monitoring. Journal of Psycholinguistic Research 30.163-184.

Oviatt, S. 1995. Predicting spoken disfluencies during human-computer interaction. Computer Speech and Language 9.19-35.

Polich, J. 2004. Neuropsychology of P3a and P3b: A theoretical overview. Brainwaves and Mind: Recent developments, ed. by N.C. Moore & K. Arikan, 15-29. Wheaton, IL : Kjellberg Inc.

Schachter, S., N. Christenfeld, B. Ravina, and F. Bilous. 1991. Speech disfluency and the structure of knowledge. Journal of Personality and Social Psychology 60.362-267.

Schachter, S., F. Rauscher, N. Christenfeld, and K. Tyson Crone. 1994. The vocabularies of academia. Psychological Science 5.37-41.

Schnadt, M. J., and M. Corley. submitted. Buying time in spontaneous speech: How speakers accommodate lexical difficulty.

Schröger, E. 1997. On the detection of auditory deviations: A pre-attentive activation model. Psychophysiology 34.245-257.

Shriberg, E. 1996. Disfluencies in Switchboard. Proceedings, International Conference on Spoken Language Processing, Addendum, 11-14. Philedephia.

Shriberg, E., and R. Lickley. 1993. Intonation of clause-internal filled pauses. Phonetica 50.172-179.

Smith, V. L., and H. H. Clark. 1993. On the course of answering questions. Journal of Memory and Language 32.25-38.

Swerts, M., and E. Krahmer. 2005. Audiovisual prosody and feeling of knowing. Journal of Memory and Language 53.81-94.

Watanabe, M., K. Hirose, Y. Den, and N. Minematsu. 2007. Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. Speech Communication 50.81-94.