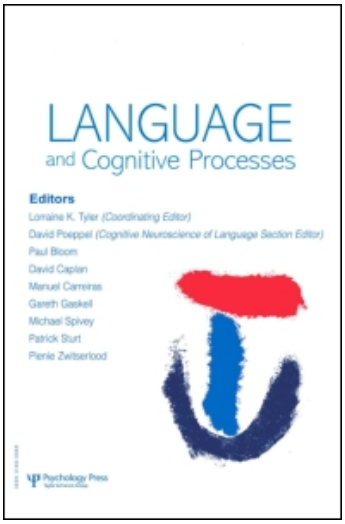

Language and Cognitive Processes
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683153

The role of fillers in listener attributions for speaker disfluency

Dale J. Barr [a];Mandana Seyfeddinipur [bc]
[a] University of California, Riverside, CA, USA [b] Stanford University, Palo Alto, CA, USA [c] Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

First published on: 08 July 2009

To cite this Article Barr, Dale J. andSeyfeddinipur, Mandana(2010) 'The role of fillers in listener attributions for speaker disfluency', Language and Cognitive Processes, 25: 4, 441 — 455, First published on: 08 July 2009 (iFirst)
To link to this Article: DOI: 10.1080/01690960903047122
URL: http://dx.doi.org/10.1080/01690960903047122

LANGUAGE AND COGNITIVE PROCESSES
2010, 25 (4), 441–455

Psychology Press
Taylor & Francis Group

The role of fillers in listener attributions for speaker disfluency

Dale J. Barr
University of California, Riverside, CA, USA

Mandana Seyfeddinipur
Stanford University, Palo Alto, CA, USA, and Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

When listeners hear a speaker become disfluent, they expect the speaker to refer to something new. What is the mechanism underlying this expectation? In a mouse-tracking experiment, listeners sought to identify images that a speaker was describing. Listeners more strongly expected new referents when they heard a speaker say *um* than when they heard a matched utterance where the *um* was replaced by noise. This expectation was speaker-specific: it depended on what was new and old for the current speaker, not just on what was new or old for the listener. This finding suggests that listeners treat fillers as collateral signals.

Keywords: Common ground; Dialogue; Disfluency; Fillers; Perspective taking.

In spontaneous discourse, speakers often fail to maintain fluent speech. Speakers hesitate, repeat words (the the red one), prolong vowels (theeee other one) or produce so-called 'fillers' *uh* or *um* (it's *um* blue). How do these disfluencies impact comprehension? Traditionally, disfluencies were seen as potentially harmful to the comprehension process (e.g., Martin & Strange, 1968). However, recent psycholinguistic research has shown that disfluencies

Correspondence should be addressed to Dale J. Barr, Department of Psychology, University of California, Riverside, CA 92521, USA. E-mail: dale.barr@ucr.edu

Thanks to Lama Alsibai, Shreya Banerjee, Monica Cervantes, Evangelina Garcia, Steve Holman, Christine Lee, Golnaz Pajoumand, Danielle Pearson, Vanna Vuong, and Marita Zamora for assistance with data collection.

http://www.psypress.com/lcp DOI: 10.1080/01690960903047122

can actually benefit comprehension. Hearing a disfluency can help listeners avoid integrating potentially erroneous material into an ongoing parse (Brennan & Schober, 2001; Fox Tree, 2001) or can relax contextually driven expectations about upcoming words (Corley, MacGregor, & Donaldson, 2007). Disfluencies can also lead listeners to expect a dispreferred syntactic structure (Bailey & Ferreira, 2003) or syntactically complex descriptions (Watanabe, Hirose, Den, & Minematsu, 2008). In addition, disfluencies influence listeners' discourse-level expectations. Following a disfluency, listeners may expect the speaker to shift topic (Swerts, 1998; Swerts & Geluykens, 1994). Listeners also expect that a disfluent speaker may be about to refer to something that is difficult to describe, whether this difficulty stems from the referent's novelty (Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Barr, 2001), its lack of a conventional name (Arnold, Hudson Kam, & Tanenhaus, 2007), or its atypicality (Barr, 2003). Finally, disfluencies impact listeners' attributions of a speaker's level of certainty (Brennan & Williams, 1995), production difficulty, honesty, and comfort with the topic (Fox Tree, 2002).

Despite the substantial evidence that disfluencies benefit comprehension, questions regarding the nature of this benefit remain. First, disfluent speech contains diverse elements: silent pauses, repeated words, prolonged vowels, and the fillers *uh* or *um*. Which of these elements is responsible for the observed comprehension benefits? Fillers might play a special role, given the theoretical view that they are collateral signals used to manage the conversation (Clark, 1996; Clark & Fox Tree, 2002). Speakers pause longer after *um* than after *uh*, suggesting that they use *um* to signal major delays and *uh* to signal minor delays (Clark & Fox Tree, 2002; Smith & Clark, 1993). Finding that listeners treat fillers as signals of delay would be supporting evidence for the collateral signal account (although this need not imply that speakers actually produce fillers with the intent to signal delay).

Currently, the evidence is equivocal as to whether fillers by themselves influence language comprehension. To find such independent effects, it is necessary to control for other characteristics of the disfluencies that contain fillers. Some studies have compared the processing of disfluent utterances containing a filler to the processing of fluent utterances (e.g., Arnold et al., 2004, 2007; Corley et al., 2007). But in these studies, the disfluent utterances differed from the fluent utterances not only by having fillers, but also by having a pause following the filler and possibly even other differences in prosody. As such, these studies do not isolate the effects of fillers from the effects of other characteristics. Studies that have controlled for the influence of other characteristics have not yielded uniform findings. Some studies that have controlled for the length of surrounding pauses using splicing techniques suggest that fillers do not impact comprehension above and

beyond the extra processing time they buy for the listener (Bailey & Ferreira, 2003; Brennan & Schober, 2001). Yet other studies do suggest independent effects. Fox Tree (2001) found that *uh* aided listeners in recognising upcoming words, whereas *um* did not. Also, Barr (2001) found that listeners more strongly expected the speaker to refer to something new after *um* than after noise (e.g., a cough or a sniffle). However, this effect could be interference from the noise rather than facilitation due to *um*. Finally, Fox Tree (2002) also found effects of *um* using stimuli matched for duration, but only looked at overhearers' offline ratings and not at their online comprehension processes.

If independent effects of fillers can be demonstrated, then this would raise the question of how they influence comprehension processes. The claim in the literature to date has been that disfluencies constrain the moment-by-moment processing of referring expressions. In apparent support of this hypothesis, eyetracking studies have found effects of disfluency over a time window corresponding to the processing of the referring expression (Arnold et al., 2004, 2007). For example, Arnold et al. (2004) found effects of the disfluency 200–600 ms after the onset of the word 'camel' in the disfluent utterance 'click on thee uh... camel'. However, such analyses can be misleading about the time course of effects, because they confound effects that emerge during the time window with 'anticipation' effects that may have emerged earlier and that persist over the time window (Barr, 2008a, 2008b). It is possible, then, the expectation for the new referent emerged entirely during the 'click on thee uh...' part of the utterance, making listeners more likely to look at new referents, but that it did not constrain how the word 'camel' was processed. In sum, because these studies do not control for such anticipation effects, they cannot rule out an alternative interpretation for their data: namely, that disfluency creates an expectation for something new (causing listeners to look more at a new referent overall) without this expectation constraining the moment-by-moment processing of the referring expression itself. Distinguishing these possibilities requires an analysis that can separate these different processing events.

Time-course issues aside, fillers can cause listeners to expect new referents based on two different underlying mechanisms. One possibility is that listeners who hear an *um* derive expectations by taking the speaker's perspective. This *perspective taking* account is consistent with the assumption that fillers have a core meaning of delay, and that listeners work out the reasons for the delay from contextual evidence and cooperative assumptions (Clark & Fox Tree, 2002). This implies that listeners' expectations should be speaker-specific; in other words, listeners should be guided by their beliefs about what would cause trouble for this particular speaker in this particular context.

However, an alternative possibility is that listeners rely on co-occurrences between features of disfluencies and certain events in speech, making perspective taking unnecessary. Under this *distributional learning* account, listeners do not make attributions to speakers who delay. Instead, their expectations are driven by a more passive memory process that generates predictions based on statistical regularities. For instance, listeners might associate the presence of a filler with new information, given that speakers are more likely to produce a filler when they refer to something new (Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Barr, 2001). Such knowledge about co-occurrences could enable listeners to benefit from fillers without requiring perspective taking, because it is often the case that what is new for the speaker is also new for the listener (because they share a conversational history).

Against this account, there is evidence that listeners who hear disfluencies containing fillers, silent pauses, and prolongations generate expectations that are specific to the person speaking. In a study by Arnold et al. (2007), listeners who heard a disfluent utterance from a normal speaker (*click on thee uh*...) expected the speaker to refer to something difficult to describe (an unusual shape) rather than an everyday object (e.g., an apple). But when listeners were told that the speaker had difficulty recognising everyday objects (object agnosia), they also expected that the speaker would have difficulty naming them, and so they no longer anticipated reference to the difficult object. Although this demonstrates that listeners make attributions that are speaker-specific, it is not yet clear whether these attributions are driven by fillers themselves. Furthermore, Arnold et al. (2007) did not find evidence that listeners could change their expectations on a trial-by-trial basis, so it is unclear how flexibly listeners can make these attributions.

In sum, we have identified three outstanding questions in the disfluency literature with respect to fillers: (1) Do fillers influence comprehension above and beyond other characteristics of disfluencies? (2) Do fillers influence the processing of referring expressions, or only create an anticipation effect? (3) Are effects of fillers driven by perspective taking or by distributional learning? We report the results of an experiment that investigated these questions by examining listeners' expectations for new referents following disfluency (Arnold et al., 2004; Barr, 2001). We focused on disfluencies that take place at the beginning of a speaking turn. This allowed us to cleanly isolate the effects of fillers on comprehension from other characteristics of disfluent speech (e.g., speaking rate, prosody, etc.).

We conducted a referential communication experiment in which participants interpreted utterances from two different speakers while movements of a computer mouse were tracked. Mouse tracking provides sensitive information about the time course of cognitive processing

(Brennan, 1990, 2005; Magnuson, 2005; Spivey, Grosjean, & Knoblich, 2005). In each trial of the experiment, the listener heard a speaker describe one of two abstract images presented on a computer screen, and used the mouse to click on the corresponding image. We manipulated whether the descriptions contained a filler (*um*) or noise (a cough or a sniffle). If there are independent effects of fillers, then listeners should expect new information more when they hear an *um* than when they hear noise. To localise the effect of the filler, we analysed two subsequent time windows. The first began at the onset of the *um* (or noise) and ended at the onset of the referring expression. The second began where the first ended and extended into the early moments of the referring expression. We also manipulated whether the listener's and speaker's perspectives matched or mismatched in terms of what images were old and what images were new. If listeners take the speaker's perspective into account, following an *um* they should expect something that is new for the speaker, and not something that is only new for themselves. However, if listeners rely on distributional learning, they should expect something new for themselves independently of what is old and new for the speaker.

METHOD

Subjects

Ninety-two undergraduates (58 females) from the University of California, Riverside participated for course credit. All identified themselves as native speakers of English.

Task

In each trial in the experiment, listeners saw two abstract images and heard pre-recorded speech describing one of the images. The speech came from one of two speakers, a male or a female. The listeners' task was to click on the image the speaker was describing (the target).

The trials were organised into blocks. Each block had four trials followed by a single test trial (see Figure 1). The first four trials set up what images were old or new for the listener and for a given speaker. The images used in each block were drawn from a set of three, which we refer to as A, B, and C. For example, the listener might hear the male speaker describe image A in the first trial, image B in the second, image B again in the third, and then image A again in the fourth. So, after the fourth trial, image A and B had been described two times each and were therefore OLD for both the male speaker and the listener. In contrast, image C had only appeared as the

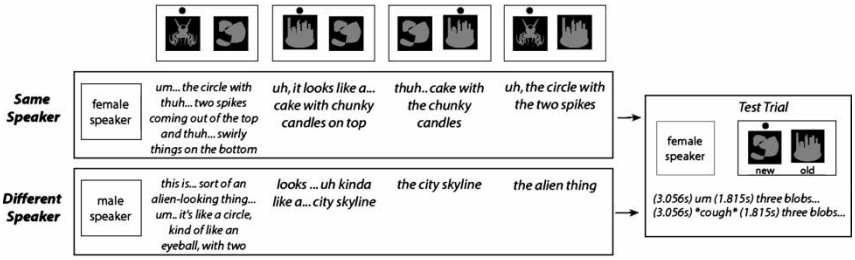

Figure 1. Sample block from the experiment, including four filler trials followed by a test trial. The black dot denotes the target image for a given trial (the dot and the words 'new' and 'old' are for expository purposes – they were not visible to the subject).

context image (in the first three trials) but had never been described. This made it NEW for both the male speaker and the listener.

In the test trial that immediately followed, we introduced our two main manipulations. In this trial, images B and C appeared, and listeners heard a disfluent description of image C that either contained an *um* or noise (a cough or sniffle). The description was given either by the same speaker who gave the four filler trials, or by a different speaker. This speaker manipulation determined whether the perspectives of the speaker and listener matched or mismatched. If the speaker at test was the same speaker who gave the descriptions for the first four trials, the perspectives matched: for both the listener and the speaker, B was old and C was new. If it was a different speaker, then the perspectives mismatched: for the listener, B was old and C was new, but for the speaker, both B and C were new, since that speaker had not yet described any images for that block.

Procedure

We told listeners that the speech was from previous participants who had completed the task on separate sessions with different participants playing the role of listener. At the beginning of each trial, the mouse cursor appeared at the exact centre of the screen, halfway between the inner edges of the two images. Playback of the audio file began simultaneously with the appearance of the images. We encouraged listeners not to wait to the end of a description but to click on the image as soon as they thought they knew which one it was.

The four trials preceding the test trial presented image pairs A-C, B-C, B-C, A-B (with the first of each listed pair being the target). The order of these four non-test trials varied across blocks, and the position of the target on the screen (right or left side) also varied from trial to trial and from block to block. The test trial for each block presented image pair C and B. Image

C, which was new for both the speaker and listener, was always the target. The position of the target varied from block to block.

There were 24 blocks of five trials each in the experiment, for a total of 120 trials. Twelve of the blocks were 'experimental' and ended with a test trial; the remaining 12 were filler blocks. These blocks were structured identically to the experimental blocks, with the exception that the image mentioned on the fifth trial was image B. This made it impossible to predict which image would be mentioned on the fifth trial (B or C).

Because our critical measurements were taken during the test trial, we wanted to make this trial as comparable as possible across conditions. Thus, we kept the speaker for test trial constant: it was always the female speaker. We created the speaker manipulation by varying the speaker for the first four trials in the block (see Figure 1). For half of the experimental blocks, the speaker for the first four trials was the male speaker, thus creating the Different Speaker condition. For the other half, it was the female speaker, creating the Same Speaker condition. The filler blocks were constructed identically. The fifth trial was always given by the female speaker, and the preceding four trials were given either by the male speaker (six blocks) or by the female speaker (six blocks).

At the beginning of each block, and just before the fifth trial of the same block, listeners were informed of the identity of the upcoming speaker. The name of the female or the male speaker appeared centred on the screen for 2500 ms.

Materials

The displays for each block presented pairs of images drawn from a set of three images that were seen only in that block. The two images in each display were centred vertically and positioned horizontally at the left and right sides of a 15-inch computer monitor with the resolution set to 1024×768 pixels. The dimensions of each image were 400×400 pixels. The inner edge of each image was 100 pixels away from the vertical midline of the screen (where the mouse cursor initially appeared).

The speakers recorded the stimuli without a script. Their recordings were made independently, so they would lack knowledge of one another's descriptions. Given that the test trials involved descriptions by the female speaker, the characteristics of the male speaker's descriptions are not described further.

The female speaker's description of image C for each block was digitally altered for use in the test trial. We altered the hesitation at the beginning of the utterance to fit certain parameters observed in a pilot production study. Each sound file in the *um* condition began with a 3056 ms interval of background noise, followed by an *um* that was spliced in from another

utterance of the speaker (unless the speaker naturally produced an *um* for that description, in which case the silences around the *um* were altered to match the desired parameters). Additional background noise was spliced in after the *um* such that the onset of the referring expression would begin at 4871 ms. For the baseline condition, the *um* was replaced with incidental vocal noise (e.g., a sniffle or a cough), and the trailing silence was altered so that description onset would take place at 4871 ms.

We created four lists so that each experimental block would appear an equal number of times in each of the four conditions (Same Speaker-*um*, Same Speaker-noise, Different Speaker-*um*, Different Speaker-noise). Each listener was presented with the stimuli from one of these four lists. The blocks were presented in a random order.

Analysis

Our analyses only considered the horizontal coordinates of the cursor, because the listener had to move the cursor horizontally but not vertically to select a picture. The cursor position was recorded in screen pixels, but we converted the measure into a more meaningful distance score: the proportion of the total distance travelled by the cursor, with the total distance defined as the distance the midline to the border of either picture. Positive distances reflect positioning in the direction of the target (which was always new for both the speaker and the listener). Negative values reflect positioning in the direction of the image that was old for the listener. A value of zero means that the cursor was centred at the midline. Any time the cursor moved beyond the inner edge of a picture, the distance was coded as 1.0 (for new image) or -1.0 (for the image that was old for the listener).

RESULTS AND DISCUSSION

For all analyses, trials in which the incorrect referent was selected were discarded (29 out of 1104, or 2.6%). We analysed the data using linear mixed-effects regression. We estimated models using the *lmer* and *pvals.fnc* functions from packages *lme4* (Bates, 2007) and *languageR* (Baayen, 2007) of the R software environment (R Development Core Team, 2007). Each mixed-effect model included subjects and items as crossed random effects, and *p*-values were obtained using Markov Chain Monte Carlo (MCMC) sampling (Baayen, 2008; Baayen, Davidson, & Bates, 2008). The two independent variables, Speaker (Same, Different) and Disfluency Condition (*um*, noise) were dummy coded and then mean-centred so that parameter estimates would correspond to main effects and interactions in an ANOVA model. Since both factors were administered within subjects and within

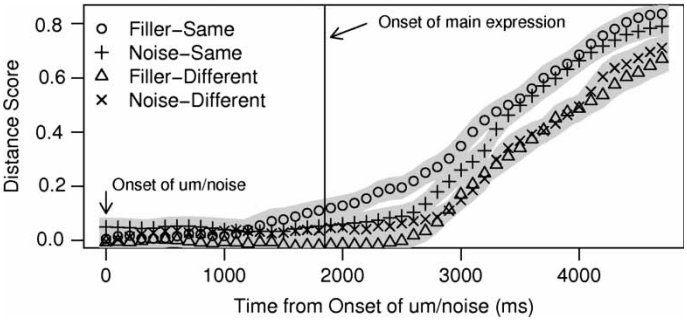

Figure 2. Time-course data for mouse cursor position. Shaded areas represent ±1 standard error of the difference between the *um* and noise conditions for a given level of Speaker.

items, with multiple observations per subject and item for each treatment combination, the model included not only by-subject and by-item random intercepts, but also by-subject and by-item random slopes for the main effects as well as for the interaction.[1]

When interpreting the results, it is important to recall that the target image was always new for both the speaker and the listener and the context image was always old for the listener. Whether the context image was new or old for the speaker depended on whether the speaker was the same or different. The distributional learning account predicts that listeners will identify the target faster following *um* than following noise, regardless of what is new or old for the speaker. In contrast, the collateral signal account predicts that listeners should identify the target faster following *um* than noise when the target is new and the context is old for the speaker (Same Speaker condition), but not when both images are new for the speaker (Different Speaker condition).

The time-course of mouse movement, averaged over subjects and items, is presented in Figure 2. The first analysis (Table 1) examined the distance that the mouse travelled toward the target during the filled interval (from the onset of the *um* or noise up to the onset of the referring expression). This analysis supported the collateral signal account: the filler *um* produced a speaker-specific expectation for new information prior to the first word of the referring expression (i.e., there was a significant interaction; $t = 2.37$, $p = .011$). When listeners heard the same speaker say *um*, they moved the mouse about 10% more toward the target than following noise ($t = 2.83$, $p = .014$).

[1] At the time of writing, *lme4* version 0.999375-28 did not provide MCMC *p*-values for models including correlations among the random effects, so these correlations were not included in any of the models. To check on our results, we reproduced each analysis with models including these correlation parameters, and used the regular *p*-values (from Wald *z* tests). All effects that were significant in the original analyses were also significant using the regular *p*-values.

TABLE 1
Mean distance travelled by the mouse cursor

	Noise		*um*	
	M	*SD*	*M*	*SD*
Filled interval (0–1871 ms)				
Same speaker	.01	.38	.11	.51
Different speaker	.04	.32	.01	.38
Referring expression (1871–4471 ms)				
Same speaker	.68	.56	.67	.56
Different speaker	.59	.61	.57	.72

In contrast, when they listened to a new speaker, for whom both images were new, listeners did not significantly move closer to the target after hearing *um* than after hearing noise ($t = 1.05$, $p = .319$); in fact the numerical trend was 3% in the opposite direction. No other effects were significant.

Given that the filler produced an anticipation effect, did it also facilitate processing of the referring expression? If so, then during the referring expression itself, listeners who hear the same speaker say *um* should move toward the new target faster than those who hear noise. This effect, in turn, should be greater than that for listeners who listen to a different speaker. For each trial, we conducted an analysis of the change in distance starting where the previous analysis left off (i.e., at the end of the silent interval following *um* or noise), and ending 2600 ms into the referring expression. We chose a 2600 ms window because this corresponds to the length of the shortest referring expression. This way, none of the effects in this window could be due to any post-processing taking place after the expression had ended. Also note that this analysis also controls for the anticipation effects that were already present at the beginning of the window.

The analysis indicated that when listeners processed the referring expression, they were only sensitive to the identity of the speaker, but not to the presence of an *um* versus noise (Table 1). Listeners moved faster toward the new referent when listening to the same rather than a different speaker ($t = 1.90$, $p = .053$). This main effect of speaker is not surprising, as it only indicates that listeners pay a processing cost when they listen to a speaker different from the one who gave descriptions on the previous four trials. Critically, during the processing of the referring expression there was no evidence for any speaker-specific effect of the filler (interaction term, $t = 0.152$, $p = .879$), nor of any overall effect of *um* versus noise ($t = 0.517$, $p = .631$). In sum, the filler produced an anticipation effect, but there was

no evidence that it facilitated the processing of the referring expression itself.

GENERAL DISCUSSION

The findings reported in this experiment contribute to the understanding of how people process disfluent speech by providing greater detail about the source of the effects and their time-course, as well as offering insight into the underlying mechanism. First, the filler *um* can cause listeners to expect new information independently of other characteristics of the disfluency. Second, the mechanism underlying this effect is a perspective-taking process that is deployed upon hearing a filler. These findings support the predictions of the collateral signal account for language comprehension. Finally, we found that fillers induced an anticipation effect, but did not appear to influence the moment-by-moment processing of the referring expression itself.

Replicating Barr (2001), listeners expected new referents when they heard a pause including an *um*, but not when they heard a pause that was matched for duration and that included noise (e.g., a cough) instead of the *um*. But although the effect in Barr (2001) could be explained as a consequence of disruption due to the noise, rather than facilitation due to the filler, the current effect cannot be explained in this way. Such an explanation account would predict that the noise would disrupt comprehension equally across the two speaker conditions. Instead, there was clear evidence for a difference between *um* and noise only when the speaker was the same. The time-course data suggested that this speaker-specific effect of *um* emerged one second after the onset of the filler, even before the first word of the referring expression.

Not only did we find an independent effect of *um*, but we also showed that perspective taking is the mechanism underlying this effect. Listeners who heard an *um* followed by a long pause directed attention toward referents that were new for the person speaking. When both referents were new for the speaker, listeners did not expect the speaker to make reference to information that was new for themselves. This finding cannot be explained by a distributional learning account, which assumes an association between *um* and information that is new for the listener.

Perhaps one could save the distributional learning account by assuming that listeners initially interpreted the *um* from their own perspective, and then corrected this interpretation. This explanation would be in line with the Perspective Adjustment model of Keysar and colleagues (Barr & Keysar, 2006; Epley, Morewedge, & Keysar, 2004; Keysar, Barr, Balin, & Brauner, 2000). In the Different Speaker condition, listeners who heard *um* may have initially considered the image that was new for themselves, and only later

realised that both images were new for the speaker. However, this explanation seems implausible. A central assumption of Perspective Adjustment is that the adjustment process is insufficient, resulting in an interpretation that is biased toward the listener's knowledge. So, listeners should show some expectation for the target following an *um* even when they listened to a different speaker. Contrary to this prediction, there was no suggestion whatsoever of an expectation for the target when the speaker was different; in fact, the trend was in the opposite direction.

It is interesting that the speaker-specific effect of the filler was wholly localised to the interval prior to the onset of the referring expression. Although hearing an *um* led listeners to expect reference to something new for the speaker, surprisingly, this expectation did not appear to modulate how they processed the referring expression itself. In that respect, this 'anticipation without integration' is consistent with recent findings that listeners can use information about a speaker's perspective to anticipate what the speaker will refer to, but are unable to integrate this information into lexical processing (Barr, 2008b).

Viewed in this way, the results from our mouse-tracking study would appear to be inconsistent with claims made from eye-tracking studies that disfluencies modulate moment-by-moment referential processing (Arnold et al., 2004, 2007). However, as mentioned above, the analyses in these studies confound effects of disfluency on anticipation with effects on referential processing itself. It is possible, then, that the effects in these studies are largely anticipatory in nature. Further research is needed to disentangle these possibilities.

Timing issues aside, the finding that effects of *um* are speaker-specific further supports the idea that listeners interpret disfluencies flexibly (Arnold et al., 2007). However, our findings indicate an even greater flexibility than has been previously shown. Arnold et al. (2007) found that listeners were less likely to expect something difficult to describe when they believed the speaker had a cognitive impairment (object agnosia). However, they did not find that listeners were less likely to expect something difficult to describe when the attribution varied from trial to trial; specifically, when listeners could attribute the disfluency to distraction provoked by an external source (loud noise, beep) rather than to difficulty arising internally from the encoding process itself. In contrast, listeners in the current study were able to make the appropriate attribution to the speaker even when the possible attributions varied from one trial to the next.

It is possible to explain this discrepancy in two distinct ways. First, in Arnold et al. (2007), listeners received evidence that the speaker was distracted within one second of the disfluency. They suggest that this may not have given listeners enough time to attribute the disfluency to the distraction. In contrast, in our study, listeners knew who would be speaking

prior to the beginning of the trial. Therefore, they had plenty of time to access given-new status from the speaker's perspective. Indeed, listeners probably routinely track given-new status because it is relevant for many linguistic processes. So in that sense, the attribution may have been easier for listeners to make.

The collateral signal account offers a second explanation for the discrepancy with Arnold et al. (2007): namely, listeners should not expect speakers to use fillers to account for disruptions that speakers themselves are not responsible for, particularly when the reasons for the disruption are available to the listener. For example, imagine a speaker who decides to stop speaking because of loud environmental noise (e.g., a jackhammer or a passing plane). Although the speaker will delay speaking until the noise has ceased, it is not self-evident that she would mark her anticipation using a filler. She does not need to account for the delay because the reasons for the delay would be obvious to the listener. Although listeners in the Arnold et al. (2007) experiment indicated on a post-experiment questionnaire that they believed the noise to have been distracting, the presence of fillers in the speech may still have led them to believe that it was an unobserved difficulty, rather than the observed distraction, that produced the disfluency.

In closing, our findings support the collateral signal account of Clark and Fox Tree (2002) for language comprehension. Whether or not speakers actually produce fillers with the intention to signal delay, it is clear that listeners interpret fillers as delay signals, and infer plausible reasons for the delay by taking the speaker's perspective. One interesting possibility suggested by these results is that fillers are used not only to manage the conversation, but also have a kind of metacognitive function, drawing the listener's attention to the mind of the speaker. It is an open question whether such a metacognitive function of fillers can explain the various benefits that disfluencies bring to language comprehension.

Manuscript received August 2008
Revised manuscript recevied May 2009
First published online July 2009

REFERENCES

Arnold, J. E., Hudson Kam, C., & Tanenhaus, M. (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*, 914–930.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new. *Psychological Science*, *15*, 579–582.

Baayen, R. H. (2007). languageR: Data sets and functions with 'Analyzing Linguistic Data: A practical introduction to statistics'. [Computer software manual].

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Bailey, K. G., & Ferreira, F. (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. In C. Cavé, I. Guäitella, & S. Santi (Eds.), *Oralité and gestualité: Interactions et comportements multimodaux dans la communication* (pp. 597–600). Paris: L'Harmattan.

Barr, D. J. (2003). Paralinguistic correlates of conceptual structure. *Psychonomic Bulletin and Review*, *10*, 462–467.

Barr, D. J. (2008a). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457–474.

Barr, D. J. (2008b). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, *109*, 18–40.

Barr, D. J., & Keysar, B. (2006). Perspective taking and the coordination of meaning in language use. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 901–938). Amsterdam: Elsevier.

Bates, D. (2007). lme4: Linear mixed-effects models using S4 classes [Computer software manual]. (R package version 0.99875-9).

Brennan, S. E. (1990). *Seeking and providing evidence for mutual understanding*. Unpublished doctoral dissertation, Stanford University.

Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the Language-as-Product and Language-as-Action traditions*. Cambridge, MA: MIT Press.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, *44*, 274–296.

Brennan, S. E., & Williams, M. (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.

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, *84*, 73–111.

Corley, M., MacGregor, L., & Donaldson, D. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, *105*, 658–668.

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, *40*, 760–768.

Fox Tree, J. E. (2001). Listeners' uses of *um* and *uh* in speech comprehension. *Memory and Cognition*, *29*, 320–326.

Fox Tree, J. E. (2002). Interpretations of pauses and ums at turn exchanges. *Discourse Processes*, *34*, 37–55.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*, 32–38.

Magnuson, J. S. (2005). Moving hand reveals dynamics of thought. *Proceedings of the National Academy of Sciences USA*, *102*, 9995–9996.

Martin, J. G., & Strange, W. (1968). The perception of hesitation in spontaneous speech. *Perception and Psychophysics*, *3*, 427–438.

R Development Core Team. (2007). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from http://www.R-project.org (ISBN 3-900051-07-0).

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, *32*, 25–38.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences USA*, *102*, 10393–10398.

Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, *30*, 485–496.

Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech*, *37*, 21–43.

Watanabe, M., Hirose, K., Den, Y., & Minematsu, N. (2008). Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication*, *50*, 81–94.