



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Disfluency in dialogue

Citation for published version:

Finlayson, I.R. & Corley, M. 2012, 'Disfluency in dialogue: an intentional signal from the speaker?'
Psychonomic Bulletin & Review, vol 19, no. 5, pp. 921-8., 10.3758/s13423-012-0279-x

Digital Object Identifier (DOI):

[10.3758/s13423-012-0279-x](https://doi.org/10.3758/s13423-012-0279-x)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Author final version (often known as postprint)

Published In:

Psychonomic Bulletin & Review

Publisher Rights Statement:

© Finlayson, I. R., & Corley, M. (2012). Disfluency in dialogue: an intentional signal from the speaker?.
Psychonomic Bulletin & Review, 19(5), 921-8. 10.3758/s13423-012-0279-x

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Running head: DISFLUENCY IN DIALOGUE

Disfluency in dialogue: An intentional signal from the speaker?

Ian R. Finlayson

CASL, Queen Margaret University

Martin Corley

Psychology, PPLS, University of Edinburgh

In press, *Psychonomic Bulletin & Review*

Final version available from www.springerlink.com

DOI: 10.3758/s13423-012-0279-x

Abstract

Disfluency is a characteristic feature of spontaneous human speech, commonly seen as a consequence of problems with production. However the question remains open as to *why* speakers are disfluent: Is it a mechanical by-product of planning difficulty, or do speakers use disfluency in dialogue to manage listeners' expectations? To address this question, we present two experiments investigating the production of disfluency in monologue and dialogue situations. Dialogue affected the linguistic choices made by participants, who aligned on referring expressions by choosing less frequent names for ambiguous images where those names had previously been mentioned. However, participants were no more disfluent in dialogue than in monologue situations, and the distribution of types of disfluency used remained constant. Our evidence rules out at least a straightforward interpretation of the view that disfluencies are an intentional signal in dialogue.

Disfluency in dialogue: An intentional signal from the speaker?

Around six per hundred spoken words are affected by *disfluencies*, including fillers such as *uh* and *um*, prolongations of both open and closed class words, repairs, and whole or part-word repetitions (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Fox Tree, 1995). Such disfluencies tend to occur when the topic of the speech is unfamiliar (Bortfeld et al., 2001; Merlo & Mansur, 2004) or is associated with a larger vocabulary (Schachter, Christenfeld, Ravina, & Bilous, 1991). They are often found at the beginnings of longer phrases (Oviatt, 1995; Shriberg, 1996), and before words with low contextual probability (Beattie & Butterworth, 1979).

These findings suggest that disfluencies reflect the difficulty that the speaker is having in retrieving the appropriate words to say. Open to question, however, is the issue of *why* difficulties in speech planning result in disfluency, rather than in some other accommodation. One possibility is that a disfluency is a mechanical by-product of the difficulty itself (e.g., Blackmer & Mitton, 1991). Alternatively, disfluencies may be used to communicate to the listener that the speaker is in difficulty (Clark & Fox Tree, 2002). Given that speech occurs most often in the form of dialogue, the resolution of this question is important in exploring the ways in which interlocutors communicate with each other. In the present paper, we address the issue with two experiments which compare the situational effects of dialogue vs. monologue on the production of disfluencies and of words.

According to Clark and Fox Tree (2002), speakers utter particular disfluencies in order to inform the listener, for example about the length of an anticipated interruption to speech (Clark & Fox Tree, 2002; Fox Tree & Clark, 1997). In line with this view, investigations based on corpora of transcribed speech show that “thee” is followed by silence more often than “thuh” (Fox Tree & Clark, 1997); and longer silences follow *um*

than *uh* (Clark & Fox Tree, 2002), consistent with earlier speech comprehension findings which suggest that *uh* and *um* have different effects on listeners (Fox Tree, 2001).

Although this view has been challenged O’Connell and Kowal (2005), evidence from recorded speech which is consistent with Clark and Fox Tree’s (2002) findings has been reported elsewhere (e.g., Barr, 2001; Fox Tree, 2001).

Further support for view that disfluencies are used communicatively appears to come from a study of patterns of disfluency in the speech of adults with autistic spectrum disorders (Lake, Humphreys, & Cardy, 2011). Lake et al. suggested that speakers with autism would be less likely to produce disfluencies that were specifically listener-oriented. Accordingly, participants with autism produced fewer fillers than did matched controls, but appeared to trade these off against disfluent repetitions and silent pauses. It should be noted that the reported findings are not directly compatible with Clark and Wasow’s (1998) suggestion that fillers and repetitions serve functionally similar communicative purposes; nor do they match evidence showing that listeners are similarly affected by *uh* and silent pauses, (Corley, MacGregor, & Donaldson, 2007; MacGregor, Corley, & Donaldson, 2010), but not by repetitions (MacGregor, Corley, & Donaldson, 2009). But the study supports a general suggestion that different disfluencies may be produced for different reasons.

However, the facts that fillers tend to precede silence, or that different people produce different patterns of disfluency, do not lead to the conclusion that disfluencies are intentionally chosen to serve as signals to the listener, any more than the smoke which accompanies fire (or not) is ‘chosen’. Moreover, although disfluencies affect listeners, both immediately and in the longer term (e.g., Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Corley et al., 2007; Fox Tree, 2001; Swerts & Krahmer, 2005), one cannot conclude from this that speakers use them to communicate, any more than the fact that a hand is withdrawn from the flame proves that the fire uses pain to affect behaviour. Although

evidence is consistent with the view that disfluencies are uttered with communicative intent, it remains possible that they are simply a consequence of delays to the speech plan, co-occurring with them automatically in ways which listeners can stochastically exploit.

In contrast to disfluencies, there is little room for doubt that the words that constitute an utterance (and convey its primary message) are chosen by the speaker. According to Pickering and Garrod’s (2004) *interactive alignment* model, alignment at all levels of dialogue (from the choice of individual words to that of syntactic structure) is at the root of successful communication. Because alignment is fundamental, “the production of a word or utterance in dialogue is only distantly related to the production of a word or utterance in isolation” (Pickering & Garrod, 2004, p. 183). Speakers in dialogue are highly likely to refer to things using the same words that their interlocutors have just used.

Whereas word choice in dialogue is well understood, there has to date been no direct experimental investigation of the role that disfluency plays in dialogue. In this paper, we present a study designed to investigate whether disfluencies are used communicatively or whether they are an automatic consequence of difficulty in the formulation of speakers’ utterances, by comparing the production of disfluencies across monologue and dialogue situations. By manipulating the ease with which pictures can be named (see Schnadt & Corley, 2006; Hartsuiker & Notebaert, 2010) in a card-sorting task, we ensure that there will be difficulties in lexical selection: Of interest is whether these difficulties automatically result in disfluency, or whether disfluencies are only found in dialogues, where they would be informative to the listener.

The monologue/dialogue manipulation is similar to that used by Bavelas, Gerwing, Sutton, and Prevost (2008) in their investigation of the production of non-verbal gestures. In that study, face-to-face dialogues were compared with telephone dialogues and monologue production. While gestures were produced in all three settings, they occurred in greater frequency in the two dialogue conditions than in the monologue condition. If we

assume that disfluencies serve a communicative purpose, then, as for gestures, we may reasonably expect fewer disfluencies to be produced in monologue.

To show that participants in the present study are affected by the monologue/dialogue manipulation, a subset of the pictures used have more than one name. By manipulating the name which one (confederate) party in the dialogue has just used for each of these pictures, we should be able to show that the participants align in dialogue, by tending to choose the same names. This manipulation serves as a demonstration that, in common with other confederate-dialogue tasks (e.g., Cleland & Pickering, 2003), the participants are sensitive to their interlocutors, and their word choices are governed by the principles of alignment. If word-choice is affected by the presence of an interlocutor but the production of disfluency is not, it will be harder to argue that disfluency is produced with communicative intent.

Experiment 1

Participants were asked to perform two tasks. In one, they were provided with grids containing pictures of objects and were instructed to name them in sequence (monologue condition). In the other, they used similar grids to play a picture matching task with a confederate of the experimenter (dialogue condition). In each condition half of the images the participant named were *disfluency images*, used to establish how disfluent the speaker was, and half were *alignment images*, used to measure alignment. Disfluency images were selected based on the difficulty with which they could be named. Other things being equal, images which were difficult to name were expected to elicit more disfluencies than those which were easy. Alignment images each corresponded to pairs of names which were used either frequently (preferred) or infrequently (dispreferred) in pretests. We predicted that participants would be unlikely to use the dispreferred names except in cases where they had previously been used by the confederate.

Participants

Twenty native British English-speaking undergraduate students from the University of Edinburgh volunteered to take part in the study.

Materials

Images were chosen from the International Picture Naming Project (IPNP: Szekely et al., 2004), which provides information about the naming of 520 black-and-white line drawings of common objects, some of which are freely downloadable. Where images could not be obtained directly from the IPNP, suitable images were selected from a commercial clip art package.

Thirty-two disfluency images were classified as either difficult or easy (16 of each), based on the findings of Schnadt and Corley (2006). Difficult images had low codability (they corresponded to several possible names), with H values (Snodgrass & Vanderwart, 1980) of above 0.85 (mean = 1.60, SD = 0.39) in the IPNP; CELEX frequencies (Baayen, Piepenbrock, & van Rijn, 1993) of the dominant names were kept below 25 counts per million (cpm: mean = 4.00, SD = 4.75). Easy images had high codability and high frequency, with H below 0.15 (mean = 0.06, SD = 0.07) and CELEX frequencies of the dominant names above 75 cpm (mean = 255, SD = 167). Example images are given in Figure 1.

Ten raters were asked to name each of an additional 40 images, and to rate alternative image names for appropriateness. The alternative names were infrequently-used names for each image taken from the Beckman Spoken Picture Naming Norms (Griffin & Huitema, 1999). Eight images were discarded because the most common name was used by less than 80% of the participants, or the selected alternative name had a mean appropriateness rating of less than 2.5 out of 5. The remaining 32 images constituted the alignment images, each associated with a commonly-used (*preferred*)

name, and an alternative (*dispreferred*) name (see Figure 1).

Finally, 32 filler images were selected. These were not subject to any constraint other than that they would be easily recognised as depicting objects named by the confederate.

Four 4×4 grids were created and the images were randomly assigned to each, with the constraint that each grid included 8 disfluency images and 8 alignment images. An additional four grids containing printed names in lieu of images (and therefore serving as scripts) were created for use by the confederate. Eight of the names corresponded to alignment items on the relevant picture grid (5 were dispreferred names, to increase the opportunity for alignment). In lieu of the disfluency items, each of the confederate's grids included the names of 8 filler items. For the matching tasks, participants and the confederate were each given 4 blank 4×4 grids on which to arrange cards depicting the images named by their interlocutors. All grids were numbered 1–16, starting in the top left corner.

Procedure

In order to prevent participants from realising that their performance in monologue and dialogue would later be compared, a cover story was created that they would be performing two separate experiments for two different experimenters (only one of whom was able to be present). To reinforce this they were given two different instruction sheets and signed two different consent forms. When performing in the monologue condition, participants were told that the researcher needed recordings of phonemes obtained from arbitrary natural speech for use in a further project. These instructions were designed to minimise the communicative aspect of the task. In the dialogue condition, participants were told that they were involved in a study investigating the ways in which people work together to perform a task. The order of conditions was counterbalanced across

participants, and upon completion of both they were informed as to the true nature of the experiment.

Each of the four grids was used equally often in the monologue and dialogue conditions. In the monologue condition, participants were shown each of two grids in turn, and asked to name the pictures in sequence. In order to imitate spontaneous speech, it was suggested that participants name each image in a sentence, although no guidance was given about the structure of the sentence. If participants asked, they were simply instructed to ensure that they stated the number of the square and its contents.

In the dialogue condition, the confederate acted as if they were a second naïve participant. Experimental participants were introduced to the confederate and both were seated at a table with a partition separating them. This prevented the participant and confederate from seeing each other or the other's grids but did not restrict them from hearing each other. Both were given grids and instructed that they should take turns to name in sequence each item and its position in their grids, and were provided with an example of what they might say: "In box one I have a dog". Upon hearing their partner naming an image in their grid, each had to place the matching individual image on to the appropriate square of a blank grid. The confederate always went first, reading from the appropriate 'script' grid. This ensured the participant heard the preferred or dispreferred name for a given item before it was his or her turn to name the relevant picture. However the confederate never produced a 'name' immediately before the participant named the same image, ensuring that the participant could not simply echo what the confederate said at any stage of the experiment. Once all of the images in a grid had been named, the procedure was repeated with a second grid.

Each participant's speech was recorded throughout the experiment using an iRiver H120 digital recorder.

Transcription and Coding

Transcription and coding of was carried out by the first author. Due to experimenter error, recordings of a single grid were missing for each of two participants. Thus the analysis was based on recordings of 78 grid descriptions.

Each grid description was first divided into 16 utterances describing the location of each picture, which tended to consist of two parts: A description of the numeric location, followed by an image description. Example transcriptions of fluent and disfluent utterances locating pictures are given in (1).

(1a) On five there is a leaf

(1b) In the fifth box there is a: [pause] um [pause] tape recording device

The 1248 resulting utterances were then coded as follows. For the 624 *alignment images*, we recorded whether each image was given the preferred or dispreferred name (23 utterances used other names and were discounted from further analysis). Where participants used more than one name, the first name used was recorded.

Coding for the 624 *disfluency images* was restricted to the image description part of each relevant utterance, which included the image name and preceding function, but not content, words (e.g., “there is a . . . device” in 1b). A data-driven approach was taken to generating categories of disfluency. Disfluencies in the first ten sets of transcriptions were used to generate categories. Each utterance was scored as fluent (no discernible disfluency) or as disfluent, and numbers of disfluencies in each category (prolongation, *uh*, *um*, hesitation, or repetition) were additionally noted.

Results

We conducted two independent analyses: The first, focusing on the alignment images, established whether the names that participants chose for these images were

affected by the names a confederate used. The second, using the disfluency images, investigated whether the disfluencies participants produced were influenced by the presence of a confederate.

Because our dependent variables were binomial (whether or not the dispreferred name had been used; whether or not there was a disfluency) we modelled outcome likelihood, using logit mixed effects models (Breslow & Clayton, 1993; DebRoy & Bates, 2004). All analyses were carried out in R (R Development Core Team, 2011) using the lme4 package (Bates, Maechler, & Bolker, 2011). All predictors were sum coded, with values of $-.5$ and $.5$ chosen as levels (confederate absent/present, preferred/dispreferred name scripted, easy/difficult, respectively), allowing odds ratios (OR) to be readily calculated without additional manipulation of model coefficients. For each analysis we constructed a full model (with maximal random effect structure) and report the coefficients for each fixed effect, together with the likelihood that each coefficient equals zero, derived from Wald's Z .

Influence of Confederate on Naming

Table 1 shows the proportions of trials in which participants chose dispreferred names for the alignment images. In conditions where a confederate was present, 63% of the alignment images would have been previously referred to using a dispreferred name (as, for each grid, the confederate's script included a dispreferred name for 5 out of 8 alignment images). In cases where there was no confederate, these images are still referred to as being in a 'dispreferred' condition; since the experiment was fully counterbalanced, we can compare cases where the dispreferred name was previously mentioned (in dialogue) to cases where there was no confederate present to mention it.

Participants were found to be over six and a half times more likely to use a dispreferred name when the confederate was present ($p < .001$), $\beta = 1.884$, $SE = 0.448$

($e^{1.884} = 6.578$), and over 17 times as likely when a dispreferred name was scripted ($p < .01$), $\beta = 2.866$, $SE = 0.974$. These two factors were found to interact ($p < .001$), $\beta = 3.876$, $SE = 1.137$, showing that participants were sensitive to the name previously used by their partner when it was their turn to name the image.

Influence of Confederate on Disfluency

Because of different views on the communicative function of silent pauses, we analysed the disfluencies produced first with and then without including the ‘silence’ category. The proportions of trials in which participants used a disfluency in naming disfluency images are shown in table 2. Including silences, images classified as difficult were 3 times as likely to be associated with disfluency as easy images ($p < .01$), $\beta = 1.125$, $SE = 0.409$ ($e^{1.125} = 3.080$). However no effect was found for the presence of a confederate ($p < 1$), $\beta = 0.239$, $SE = 0.420$, suggesting that participants were no more (or less) likely to be disfluent when a partner was present. There was no evidence of any interaction between these factors ($p < 1$), $\beta = 0.414$, $SE = 0.540$. Disfluencies other than silences were over two times as likely to be produced when naming difficult images ($p = .02$), $\beta = 0.858$, $SE = 0.353$. Without silences, no other effect reached significance ($ps > .89$).

To test whether the distributions of participants’ disfluencies were affected by the presence of a confederate, we tabulated the total numbers of disfluencies in five categories observed across the experiment. Table 3 shows the totals observed in the presence and absence of a confederate. As can be clearly seen, the distribution of disfluencies was not affected by the presence of a confederate, a fact confirmed by Fisher’s exact test ($p = .95$).

Discussion

Experiment 1 showed that, while word choice differed between monologue and dialogue, the use of disfluency did not. However, given that participants named items disfluently on less than 10% of occasions, it is possible that the lack of disfluency effect

reflects a scarcity of observations. To address this issue, and to ensure that the null effect obtained in Experiment 1 could be replicated, we ran an additional experiment, which was identical to Experiment 1 except that the images used were digitally manipulated to make them harder to recognise, and therefore more likely to result in disfluent descriptions.

Experiment 2

For Experiment 2, the 96 images used for Experiment 1 were blurred using a Gaussian algorithm ($\sigma = 6$ pixels). In all other respects, the experimental was identical to Experiment 1; participants were 24 native British-English speaking undergraduate students from the University of Edinburgh, who participated in return for course credit.

Speech was recorded using a ZOOM H4n digital recorder.

Transcription and Coding

Two raters each transcribed and coded the recordings for 15 participants. Raters were instructed to count the occurrences of each of the five categories of disfluency identified in Experiment 1. For the six participants which were rated by both raters, there was 86.4% agreement on disfluencies. For each of these six participants, one rater's coding was selected at random for analysis.

Results

Influence of Confederate on Naming

Table 1 shows the proportions of trials in which participants chose dispreferred names for the alignment images. Participants were over 6 times more likely to use dispreferred names when a confederate was present ($p < .001$), $\beta = 1.840$, $SE = 0.553$ ($e^{1.840} = 6.297$). When a dispreferred name had been scripted participants were over twenty times more likely to use it themselves ($p < .01$), $\beta = 3.019$, $SE = 1.072$; as in Experiment 1, the two factors interacted ($p < .001$), $\beta = 6.699$, $SE = 1.372$.

Influence of Confederate on Disfluency

Table 2 shows the proportions of disfluent trials. Including silences, difficult images were almost seven times as likely to be associated with disfluency as easy images ($p < .001$), $\beta = 1.917$, $SE = 0.459$ ($e^{1.917} = 6.800$). No effect was found for the presence of a confederate ($p < 1$), $\beta = 0.216$, $SE = 0.247$ and these two factors did not interact ($p < 1$), $\beta = -0.044$, $SE = 0.528$. Disfluencies other than silences were almost five and a half times as likely to be associated with disfluency ($p < .001$), $\beta = 1.700$, $SE = 0.414$. Without silences, no other effect reached significance ($ps > .43$).

Counts for each category in the presence and absence of a confederate are shown in Table 3. A Fisher's exact test showed that the presence of a confederate did not influence the distribution of disfluencies ($p = .47$).

A final analysis combined the data from both experiments. A regression model was constructed which included a fixed effect for experiment, which was allowed to interact with all other fixed effects, and an experiment-by-items random slope. Speakers were no more likely to be disfluent in the presence of a confederate ($p < 1$), $\beta = 0.166$, $SE = 0.171$. A main effect of experiment showed that using blurred images made participants over one and a half times more likely to be disfluent ($p < .05$), $\beta = 0.466$, $SE = 0.207$. A marginal interaction between experiment and difficulty suggested that the effect of blurring on disfluency was larger for difficult images ($p = .09$), $\beta = 0.693$, $SE = 0.410$. No other interactions with experiment were significant (all $ps < 1$). An analysis excluding silences confirmed this pattern of results, although the effect of experiment became marginal ($p = .06$), $\beta = 0.450$, $SE = 0.242$.

General Discussion

The present study was designed to investigate whether or not disfluencies are used by speakers to signal difficulty to their interlocutors. We manipulated whether a task was

performed communicatively (in a dialogue condition) or non-communicatively (in a monologue condition) and investigated the effects of this manipulation on the production of disfluency. As a precondition to being able to interpret our findings, we had to show that in the dialogue condition, speakers were in fact producing language which took their listeners into account. Results from the alignment items show unequivocally that this was true. In line with previous similar work (Cleland & Pickering, 2003; Clark & Wilkes-Gibbs, 1986), when the experimental confederate referred to a picture using a dispreferred name, participants were many times more likely to choose that name to refer to the same picture than they were in cases where the more common, preferred, name had previously been used.

Having established that participants' language choices were affected by the presence of an interlocutor, the question remains of what factors caused them to be disfluent. Participants were much more likely to refer disfluently to images when those images corresponded to several names (cf. Hartsuiker & Notebaert, 2010) and the most commonly-used name was low-frequency. These effects were exacerbated when the images were blurred (cf. Schnadt & Corley, 2006). This suggests that disfluencies reflect cognitive difficulty, either in selecting a particular name (cf. Vitkovich & Tyrrell, 1995) or in retrieving a low-frequency name (cf. Caramazza, Costa, Miozzo, & Bi, 2001; Jescheniak & Levelt, 1994). However, there was no evidence at all that the presence of an interlocutor in the dialogue condition affected the likelihood of being disfluent. Moreover, this finding is not the consequence of conflating different types of disfluency. If particular disfluencies are viewed as communicative signals of upcoming difficulty (cf. Fox Tree, 2001), we might expect participants to use them more with a listener present: But there was no suggestion that non-silent disfluencies were used more often, nor that the distributions of disfluency types used differed, between monologue and dialogue conditions.

There are three potential interpretations of these findings. First, participants might

not have had awareness of the confederate in any significantly communicative sense, and might instead have viewed each condition as comprising a monologue. According to this view, lexical alignment with the confederate would be attributed to straightforward priming, and disfluency levels across conditions would remain constant because the conditions were communicatively equivalent.

We would not wish to contest that priming has a role to play, given Pickering and Garrod's (2004) view that priming mechanisms are fundamental to alignment in dialogue. However, evidence suggests that, at least at the lexical level, the names chosen for images are influenced by beliefs about one's interlocutor (Branigan, Pickering, Pearson, MacLean, & Brown, 2011), as part of a general tendency for speakers to take into account what they believe their listeners to know (Isaacs & Clark, 1987), and we see no reason to believe that our participants were not sensitive to these factors. Moreover, if both conditions were perceived as monologues, proponents of the 'disfluency as signal' view would need to account for the fact that disfluencies were uttered throughout both experiments (and 12% of these where either *um* or *uh*).

A second interpretation of the present findings relies on the observation that dialogue is the most common form of speech, while monologue is a special case (Garrod & Pickering, 2004; Pickering & Garrod, 2004). It is possible that participants continued to use disfluency as a signal in the monologue conditions, either out of habit, or perhaps even because they lacked a special set of communicative strategies which were more suitable for monologue. Anecdotally, there do appear to be occasions where disfluency rates are adapted for monologue—in public speaking, for example—but if one accepts the view that the use of disfluencies as signals is a habit that is hard to break, then testing Clark and Fox Tree's (2002) suggestion that disfluencies are used as communicative signals is likely to be difficult. One possibility may be to explore the developmental evidence: Whereas there is evidence that children as young as two can infer that an adult is likely to refer to

a novel object after a filler (Kidd, White, & Aslin, 2011), the distinction reported by Clark and Fox Tree (2002) between pauses following *um* and *uh* does not appear in the speech of 3–4-year-olds (Hudson Kam & Edwards, 2008).

For Clark and Fox Tree (2002), the use of particular disfluencies is clearly seen as intentional. But determining whether speakers are doing something intentionally is difficult, particularly when they are not consciously aware of doing it. Thus the claim that speakers use disfluencies to communicate remains uncontested, not because it is right (or wrong), but because it is difficult to verify. In the absence of a direct solution to this problem the present paper provides the first example of experimental disfluency research focusing specifically on the case of dialogue. We replicated previous findings on lexical alignment, and showed that the production of disfluencies was affected by the ease with which words in the intended message could be selected. However, we found no evidence to suggest that the disfluencies a speaker produces are influenced by the presence of a listener. Whereas this finding does not rule out the possibility that disfluencies are created intentionally, it does not provide evidence to support this claim. The third, and simplest, account of the existing evidence is therefore that disfluencies do not serve a communicative purpose, other than in the sense that listeners are able to exploit their occurrence in predictable circumstances. Instead, they are by-products of difficulty in speech, whether there is someone present to whom the difficulty can be communicated or not.

References

- Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The Old and Thee, Uh, New: Disfluency and Reference Resolution. *Psychological Science*, *15*, 578–582.
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX Lexical Database*. <http://celex.mpi.nl/>.
- Barr, D. J. (2001). Oralité et gestualité: Interactions et comportements multimodaux dans la communication. In S. Santi, I. Guatella, C. Cave, & G. Konopczynski (Eds.), (pp. 597–600). L'Harmattan.
- Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes [Computer software manual]. Available from <http://cran.r-project.org/package=lme4> (R package version 0.999375-39)
- Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, *58*(2), 495–520.
- Beattie, G. W., & Butterworth, B. (1979). Contextual Probability and Word Frequency as Determinants of Pauses in Spontaneous Speech. *Language and Speech*, *22*, 201–211.
- Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, *39*(3), 173–194.
- Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, *44*(2), 123.
- Branigan, H. P., Pickering, M. J., Pearson, J., MacLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, *121*, 41–57.
- Breslow, N. E., & Clayton, D. G. (1993). Approximate Inference in Generalized Linear

- Mixed Models. *Journal of the American Statistical Society*, 88, 9–25.
- Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The Specific-Word Frequency Effect: Implications for the Representation of Homophones in Speech Production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 1430–1450.
- Clark, H. H., & Fox Tree, J. E. (2002). Using Uh and Um in Spontaneous Speaking. *Cognition*, 84, 73–111.
- Clark, H. H., & Wasow, T. (1998). Repeating Words in Spontaneous Speech. *Cognitive Psychology*, 37(3), 201–242.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a Collaborative Process. *Cognition*, 22, 1–39.
- Cleland, A. A., & Pickering, M. J. (2003). The Use of Lexical and Syntactic Information in Language Production: Evidence from the Priming of Noun-Phrase Structure. *Journal of Memory and Language*, 49, 214–230.
- Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's The Way That You, Er, Say It: Hesitations in Speech Affect Language Comprehension. *Cognition*, 105, 658–668.
- DebRoy, S., & Bates, D. M. (2004). Linear Mixed Models and Penalized Least Squares. *Journal of Multivariate Analysis*, 91, 1–17.
- 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 and Cognition*, 29(2), 320.
- Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "The" as "Thee" to Signal Problems in Speaking. *Cognition*, 62, 151–167.
- Garrod, S., & Pickering, M. J. (2004). Why is Conversation so Easy? *Trends in Cognitive Sciences*, 8, 8–11.

- Griffin, Z. M., & Huitema, J. (1999). *Beckman Spoken Picture Naming Norms*.
<http://langprod.cogsci.uiuc.edu/~norms/>.
- Hartsuiker, R. J., & Notebaert, L. (2010). Lexical Access Problems Lead to Disfluencies in Speech. *Experimental Psychology*, 57(3), 169–177.
- Hudson Kam, C. L., & Edwards, N. A. (2008). The use of uh and um by 3-and 4-year-old native English-speaking children: Not quite right but not completely wrong. *First Language*, 28(3), 313.
- Isaacs, E. A., & Clark, H. H. (1987). References in Conversation Between Experts and Novices. *Journal of Experimental Psychology. General*, 116, 26–37.
- Jeschaniak, J. D., & Levelt, W. J. M. (1994). Word Frequency Effects in Speech Production: Retrieval of Syntactic Information and of Phonological Form. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, 824–843.
- Kidd, C., White, K. S., & Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental Science*, 14(4), 925–934.
- Lake, J. K., Humphreys, K. R., & Cardy, S. (2011). Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders. *Psychonomic Bulletin & Review*, 1–6.
- MacGregor, L. J., Corley, M., & Donaldson, D. I. (2009). Not all disfluencies are equal: The effects of disfluent repetitions on language comprehension. *Brain and Language*, 111(1), 36–45.
- MacGregor, L. J., Corley, M., & Donaldson, D. I. (2010). Listening to the sound of silence: Investigating the consequences of disfluent silent pauses in speech for listeners. *Neuropsychologia*, 48, 3982–3992.
- Merlo, S., & Mansur, L. L. (2004). Descriptive Discourse: Topic Familiarity and Disfluencies. *Journal of Communication Disorders*, 37, 489–503.
- O'Connell, D. C., & Kowal, S. (2005). Uh and Um Revisited: Are They Interjections for

- Signaling Delay? *Journal of Psycholinguistic Research*, 34, 555–576.
- Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language*, 9, 19–35.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169–190.
- R Development Core Team. (2011). 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)
- Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech Disfluency and the Structure of Knowledge. *Journal of Personality and Social Psychology*, 60, 362–367.
- Schnadt, M. J., & Corley, M. (2006). The Influence of Lexical, Conceptual and Planning Based Factors on Disfluency Production. In *Proceedings of the twenty-eighth meeting of the Cognitive Science Society*. Vancouver, Canada.
- Shriberg, E. E. (1996). Disfluencies in Switchboard. In *Proceedings of the International Conference on Spoken Language Processing, Addendum* (pp. 11–14). Philadelphia, PA.
- Snodgrass, J. G., & Vanderwart, M. (1980). A Standardized Set of 260 Pictures: Norms for Name Agreement, Image Agreement, Familiarity, and Visual Complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 174–215.
- Swerts, M., & Krahmer, E. (2005). Audiovisual Prosody and Feeling of Knowing. *Journal of Memory and Language*, 53, 81–94.
- Szekely, A., Jacobsen, T., D’Amico, S., Devescovi, A., Andonova, E., Herron, D., et al. (2004). A New On-Line Resource for Psycholinguistic Studies. *Journal of Memory and Language*, 51, 247–250.
- Vitkovich, M., & Tyrrell, L. (1995). Sources of Disagreement in Object Naming. *Quarterly Journal of Experimental Psychology*, 48A, 822–848.

Author Note

The authors wish to thank Ewan Keith and Genevieve Warriner-Gallyer for help with Experiment 2.

Table 1

Proportions of trials in which participants used the dispreferred name to refer to alignment images for Experiments 1 and 2. Standard errors are given in parentheses. Where the confederate was present, a preferred or dispreferred name was scripted and would previously have been heard by the participant; where the confederate was absent, the scripted name was nominal only, in that no name was actually heard before the participant named each item.

		Confederate Absent	Confederate Present
Exp. 1	Preferred Name Scripted	0.10 (0.03)	0.03 (0.02)
	Dispreferred Name Scripted	0.11 (0.02)	0.47 (0.04)
Exp. 2	Preferred Name Scripted	0.18 (0.04)	0.04 (0.02)
	Dispreferred Name Scripted	0.12 (0.02)	0.59 (0.03)

Table 2

Proportions of trials in which participants referred disfluently to disfluency images for Experiments 1 and 2. Standard errors are given in parentheses.

		Confederate Absent	Confederate Present
Exp. 1	Easy Images	0.09 (0.02)	0.09 (0.02)
	Hard Images	0.18 (0.03)	0.23 (0.03)
Exp. 2	Easy Images	0.10 (0.02)	0.11 (0.02)
	Hard Images	0.31 (0.03)	0.35 (0.03)

Table 3

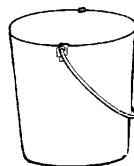
Total numbers of disfluencies observed in each of five categories across the experiment.

		Prolongation	Uh	Um	Repetition	Silence
Exp. 1	Confederate Absent	41	7	8	4	54
	Confederate Present	35	7	10	5	51
Exp. 2	Confederate Absent	79	3	21	12	141
	Confederate Present	66	8	18	8	112

Figure Captions

Figure 1. Examples of easy-to-name (car) and a hard-to-name (llama) images, and two alignment images (for each image the preferred name is given, followed by the dispreferred name used by the confederate in bold), as used in (a) Experiment 1 and (b) Experiment 2.

(a)



(b)



Car

Llama

Bucket/**Pail**

Magician/**Conjurer**