Proceedings of DiSS-LPSS Joint Workshop 2010,
25-26 September 2010, Tokyo, Japan, pp. 89-92.

"Um…I don't see any": The Function of Filled Pauses and Repairs

Hannele Nicholson [1], *Kathleen Eberhard*[1], *Matthias Scheutz*[2]

[1] Department of Psychology, University of Notre Dame, U.S.A.
[2] Department of Computer Science and Informatics, University of Indiana-Bloomington, U.S.A.
hnichol1@nd.edu, keberhar@nd.edu, mscheutz@indiana.edu

Abstract

We investigate disfluency distribution rates within different moves from an interactive task-oriented experiment to further explore the suggestion by Bortfeld et al. [1] and Nicholson [2] that different types of disfluencies may fulfill varying functions. We focus on disfluency types within moves, or speech turns, where a speaker initiates something compared to a response to such a move. We find that filled pauses (FPs) such as *um* or *uh* fulfilled an interpersonal role for participants while repairs occurred out of difficulty.

Index Terms: disfluency, dialogue, dialogue moves, language production

1. Introduction

Dialogue is a dynamic process wherein several complex behaviors interact simultaneously. Speakers will often ensure that their listeners are paying attention to them before they speak and to be polite, listeners will often indicate that they are paying attention by gazing at the speaker and providing backchannel responses [3]. During a task-oriented dialogue, however, both speakers and listeners spend more time attending to the task-related objects and less time gazing at the speaker. Previous work on disfluency has shown that disfluencies may occur because of task difficulty [2, 4, 5]. A speaker may also become disfluent if they notice that their interlocutor isn't paying attention [3] or because information that the listener has given necessitates a reformulation of an utterance in progress [2].

Previous corpus studies have suggested that different types of disfluencies may fulfill different functions [1,2]. For example, Bortfeld et al. [1] found that speakers in an executive role used more FPs and restarts than those participants who were following directions. Noting a different distribution of FPs to restarts and repeats, Bortfeld [1] suggests that FPs may not only occur for reasons of cognitive load but instead may be related to interpersonal factors in communication. FPs may be used to gain time, for example, when a participant has just been asked a question [6]. Repairs, then, may occur because of an overburdened system.

Lickley [6] reported results from the Map Task corpus showing which types dialogue moves were the most susceptible to being disfluent. In the Map Task corpus, an Instruction Giver communicated directions so that an Instruction Follower could recreate the route of a cartoon map on a blank map. A dialogue move is roughly equivalent to the individual turns a speaker takes in a conversation [7]. For example, a speaker might ask a question or give an explanation in which she elaborates on a previous instruction in an **Initiation** category move or a speaker might provide an affirmative reply to something her partner said in a **Response** category move.

In a task-oriented dialogue like the map task, Lickley reports that Instruction Givers were more disfluent per word than Instruction Followers. Lickley controlled for length of a move by using disfluency rate per word and by conducting a detailed analysis of a subset of moves, i.e. moves that were only 4-6 words in length. Instruct moves had high repair rates because of the planning involved in giving route descriptions and the selection of novel referents. Response moves (Reply-W, Reply-Y and Clarify) had high FP and repetition rates which may suggest that speakers used these disfluency types to buy time.

Due to the size of the Map Task corpus, a detailed analysis which broke down disfluency by type was not conducted. Although Lickley [6] did report differences between repairs and filled pauses, he did not report whether there was any difference between types of repairs. We aim to further Lickley's findings by investigating repairs found within the Indiana Cooperative Remote Search Task (CReST) Corpus [8] to determine whether one particular move type is more prone to a particular type of repair. We will conduct this task by looking only at Initiation versus Response move categories. We predict a) those in an executive role should be more disfluent than those following directions, b) Initiation moves will have higher repair rates as the speaker encounters production difficulties, c) Response moves that answer a question will contain more FPs in order to buy time for the speaker.

2. Corpus and Methods

Results come from the Indiana CReST corpus [8]. In this corpus, dyads of American English speaking adults collaboratively performed tasks involving colored boxes placed throughout several office rooms. One member of the dyad was the "Director" while the other was the "Searcher". The Director sat in front of a computer screen with a map of the office environment.

As shown in Figure 1, the map detailed the location of variously colored boxes. The Director helped the Searcher locate boxes in the environment and to complete different tasks with each type of box. A specific design feature of the project was to enable Searchers with the chance to act autonomously. The Searcher was responsible for reporting back the location of the green numbered boxes so that the Director could mark their location on the map.

Six dyads participated in the experiment. All six dyads were either undergraduate or graduate students at the University of Notre Dame. None of the participants were familiar with the office environment prior to the experiment. By chance, all dyads were same-sex; there were 3 female dyads and 3 male dyads. Also by chance three dyads (2 male and 1 female) were friends prior to the experiment while the remaining three dyads (2 female and 1 male) were unacquainted prior to the experiment. Each participant was paid $5.00 or $10.00 according to their performance. Pairs who

completed the tasks for 12 or more of the 24 boxes were each paid $10; otherwise, they were paid $5 each.

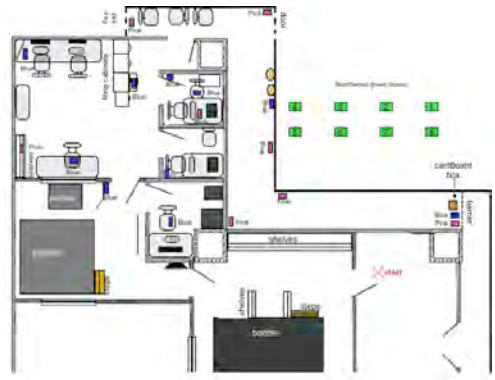

Figure 1*: Map of Search Environment. Color labels were not present for Directors but appear only for explanation.*

2.1. Annotation of Dialogue Moves

Carletta et al. [7] describe a dialogue coding system developed for classifying task-oriented dialogues. According to the scheme, moves are sub-units of the overall dialogue that further the pair's progression towards a goal. Moves fall into three major categories: ***Initiation, Response*** or ***Ready***. Each category consists of several individual move types as described by [7]. For example, initiation moves can consist of Instructions, Explains, or Queries. Response moves may be replies to questions, Acknowledgements that a previous move was heard and understood, or a clarification of an answer. **Ready** moves indicate that the speaker is prepared to begin a new turn (e.g. ***okay****, take a left).*

We adopted this scheme to code the turns taken in the Indiana CReST corpus. Every turn uttered by a Director or Searcher was classified according to this scheme by two independent coders. A third coder resolved any disagreements. Finally, a move that was not completed was classified as an **Incomplete** move. The analysis presented here will only focus on **Initiation** or **Response** moves and will not include **Incomplete or Ready** move types.

2.2. Annotation of Disfluencies

Disfluencies were annotated according to the HCRC Map Task Disfluency Coding Manual [10]. Table 1 shows examples of each type of disfluency. A disfluency was considered a ***repetition*** if one or more words was repeated verbatim. In a ***substitution***, the speaker repeats and utterance but replaces one word for another.

Table 1. Disfluency Type Examples

	Transcript
Repetition	Okay, **go…go** straight ahead
Substitution	Is there a yellow bo-…I mean yellow block?
Insertion	So **in front…right in front** of you should be an open door?
Deletion	**Well you're-**…Go through that room

In an ***insertion***, a speaker inserts one or more words either between or before words that are repeated verbatim. Finally, the speaker abandons an utterance in a ***deletion***. Filled pauses (*uh, um*), silent pauses and prolongations were also annotated.

Additionally, since FPs are known to be more common utterance initially than utterance medially [11], we coded FPs according to their position within an utterance. A FP was deemed utterance-initial if it was the first word in a move or if it was preceded only by a single discourse marker (*and, so, now)*. All other FPs were considered utterance-medially. Repairs were considered move-initial if the first word of the reparandum began at the beginning of the move or after one discourse marker; all other repairs were considered move-medial. Table 2 illustrates common examples from our corpus.

Table 2. Disfluency Position Examples

TYPE	
Initial	D3: um..yeah, if you want D5: and **um…the- the** green box was-..it wasn't in the far corner of the room
Medial	D4: and then there's **uh** the second one which um has a green box in it D1: it's right next to **the: second- the door..um the first door** I would go through

3. Results

3.1. Move Distribution

In total, there were 1,186 moves in the corpus (excluding 47 Incomplete and 176 Ready moves). The distribution of moves with respect to Directors and Searchers is shown in Figure 2. As expected, Directors were prone to make **Initiation** moves (73%) than Searchers (31%). On the other hand, Searchers made more **Response** moves (71%) than Directors (29%).

3.2. Disfluency rate Distribution & Frequency

The CReST corpus consisted of 438 Total disfluencies (251 Repairs, 187 Filled pauses). The Repair total by type consisted of 58 Repetitions, 79 Substitutions, 49 Insertions, and 65 Deletions. Following [6], we take as the dependent variable disfluency rates per 100 words. Word counts excluded filled pauses but included words appearing in the reparandum. A Univariate ANOVA with total disfluency rate as the dependent variable and Speaker role as an independent variable (Director vs. Searcher) revealed no significant differences between Directors and Searchers ($F(1,10) < 1$). FPs (2.25) were more frequent per word than repetitions (.746), substitutions (1.05), insertions (.635) or deletions (.864) (Disfluency Type: $F(4,50) = 5.65$, $p = .001$; Pairwise Comparisons, $p < .01$). So, overall, FPs were more common than any other type of disfluency.

Figure 4 depicts the distribution of disfluency types Initiation, Response or Incomplete Move initially or medially. Incomplete moves were only included because over half of all deletions (54%) occurred during an abandoned move. FPs were more equally distributed between initial (27%) and within (25%) positions in Initiation moves. Repetitions (46%), Substitutions (57%) and Insertions (58%) were more common within Initiation moves.

We conducted a Univariate ANOVA of Disfluency Type (FP vs. Repair), Move Category (Initiation vs. Response) and Role (Director vs. Searcher) with disfluency rate per word as

the dependent variable. There was no significant difference between Directors and Searchers. Within-subject ANOVAs revealed that Directors produced more repairs per word (2.6) than FPs (1.4) ($F(1,30) = 5.79$, $p < .05$). Searchers used more FPs per word (3.8) when responding to the Director than they did when initiating a move (1.9) ($F(2,10) = 6.24$, $p < .02$). For Searchers, response moves contained more FPs per word (3.8) than repairs per word (1.5). For Directors, FPs were equally common in response moves (1.54) than in initiation moves (1.50) ($F < 1$). These results lend support to the prediction that response moves will contain more FPs.

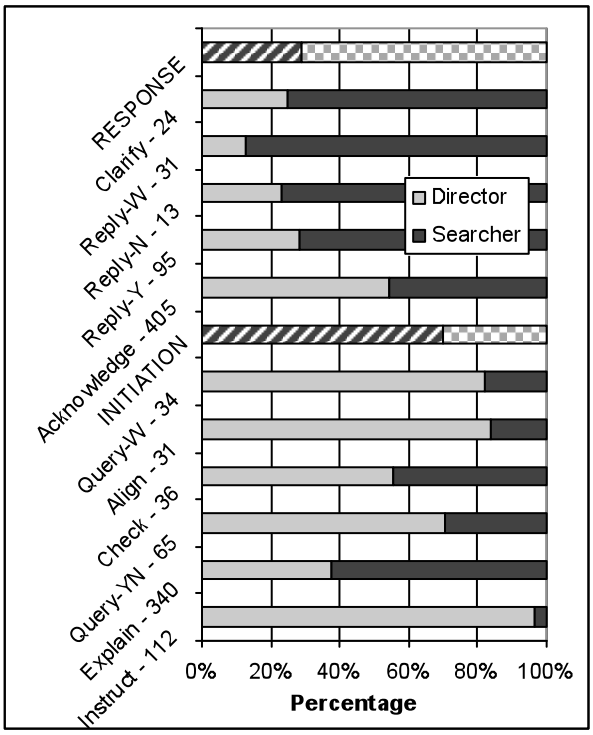

Figure 2: *Percentage of move types by role (Director vs. Searcher)*

3.3. Disfluency Position in Moves

Disfluencies are known to be more common utterance-initially [11]. For this reason, we conducted a position-based analysis where disfluency rates were compared with regards to their occurrence at the beginnings of moves versus within moves. We conducted a Univariate analysis where disfluency rate per 100 words within Initiation or Reply moves was the dependent variable. Independent variables were Disfluency Type (5) x Move Category (2: Initiation vs. Reply) x Position (2: Beginning vs. Within) x Speaker Role (2: Director vs. Searcher).

3.3.1. Filled Pause Rate

FPs were more frequent at the beginning of Initiation moves (.939) than they were within Response moves (.015) ($F(4,200) = 3.01$, $p < .02$; Bonferroni, $p < .005$; $\alpha = .005$). Also, FP rate was higher within Initiation moves (.834) than within Response moves (.015). We predicted that FP rate would be highest before Response moves and while this is certainly true numerically (7.49) compared to within Response moves (0.15), before Initiation moves (.939), or within Initiation moves (.834), the difference is not significant by Bonferroni standards. Furthermore, there were no significant differences between FP rates in any location compared to Repairs in any location.

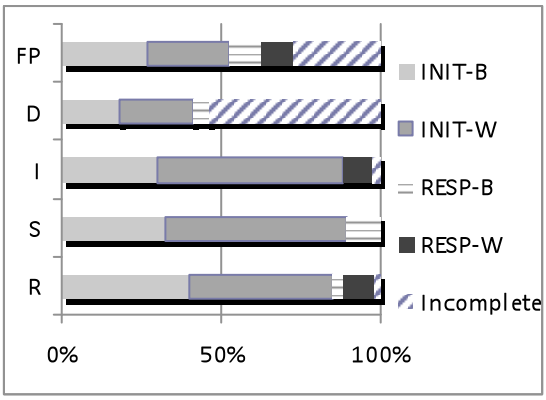

Figure 3. *Distribution of Disfluency Types move-initially within Initiation and Response Moves. B = move-initially, W = move-medially, R = Repetition, S = Substitution, I = Insertion, D = Deletion.*

To better understand how FPs were being used, a qualitative analysis of the Searchers' headcam video showed that FPs were used 50% of the time as place-holders prior to Initiation moves so the Searcher could physically walk to another location in the environment or search for a box.

S: um [Searcher walking through rooms]
alright, I'm turning right into uh . the next small room?

The other major function of Searchers' move-initial FPs seemed to be to cushion the blow of a response that might be dissatisfactory for the Director:

D: There should be a blue box there?
S: um . I don't see any.

In this case, the Searcher does not physically move anywhere or move her head to check something. She simply stares straight at the object suggesting that the FP was used to 'cushion the blow' rather than solely to buy time for utterance planning. This lends further support to the notion that FPs seem to fill an interpersonal function in some cases rather than a purely cognitive one.

3.3.2. Repair Rate

Bortfeld et al. [1] found evidence to suggest that FPs occurred for interpersonal reasons while repairs occurred for cognitive reasons. Generally, we find support for this claim although we find some evidence to suggest that occasionally repetitions patterned like FPs at the beginnings of moves. This suggests that both repetitions and FPs could be used as a mechanism for buying time for the speaker.

Substitutions were more common within Initiation moves (.709) than they were within Response moves (.000). Insertion rate was higher within Initiation moves (.391) than before Response moves (.000). Finally, deletion rate was higher within Initiation moves (.192) than within Response moves (.000) (Bonferroni t-tests, $p < .003$, $\alpha = .003$). From these results we find some support for our prediction that repair rates

would be higher within Initiation moves than within response moves.

Repetition rate at the beginning of Initiation moves (.380) was significantly higher than substitution rate within Response moves (.000), insertion rate within response moves (.000), or deletion rate within response moves (.000) (Disfluency Type x Move Category x Position: $F(4,200) = 3.01$, $p < .02$). Thus, it appears that repetitions patterned slightly differently from the other types of repairs as repetitions tended to occur more frequently at the beginning of Initiation moves while other repair types were more common within moves.

4. Discussion

We predicted that those with an executive role to fill would be more disfluent based on previous findings. This prediction was tested by comparing disfluency rates in Initiation moves (i.e. moves that filled an executive role) to Response moves (i.e. moves that replied to initiations). We found no significant between-subject differences between Searchers and Directors. However, within-subjects tests revealed that Directors produced more repairs than FPs. Searchers on the other hand showed a proclivity for making more FPs than repairs in response moves. These findings taken together with the fact that the Searcher had concrete knowledge of the search environment while Directors could only conjecture suggests that FPs were used to buy time. In a few cases, Searchers did not need to move anywhere to find a box but nevertheless used an *'um'* prior to a negative response. We suggest that in these circumstances the FP is intended to 'cushion the blow' of the negative response.

Overall, FPs were the most frequent type of disfluency. FP rate was higher both at the beginnings and middles of Initiation moves than it was within Response moves. Insertions and Substitution repairs on the other hand tended to be more frequent per word within Initiation moves. There is some suggestion that repetitions patterned like FPs as they too were more equally distributed between initial and medial positions in Initiation moves. We conclude from this, in line with [1] and [6] that FPs seem to fulfill an interpersonal function while repairs occur because of production difficulty.

5. Conclusions

Our results yielded a surprising result. Disfluency rate was near equal between speaker roles. We explain this by the fact that both participants filled an executive role. Nevertheless, Directors were prone to make repairs while Searchers produced more FPs. In terms of location, FP rates, on the other hand, were more equally distributed within move types. Both facts support Bortfeld et al.'s [1] suggestion that perhaps they occurred for interpersonal reasons instead of purely cognitive ones. A qualitative analysis revealed that Searchers tended to use FPs for two interpersonal reasons: either as a place-holder while they walked through the environment or to cushion the blow of a negative response. The current study was limited by a small subject pool but we plan to conduct future studies with a larger subject pool to confirm these results.

Acknowledgements

We would like to thank Susan Gundersen for running participants and her assistance with coding the data. We also thank Won Jae Shin for assistance in running the experiment.

6. References

[1] Bortfeld, H., Leon, S., Bloom, J., Schober, M., Brennan, S. "Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role and Gender". *Language and Speech*, 44(2). 2001

[2] Nicholson, H., Bard, E.G., Lickley, R. J. *Proc. of DiSS'05: Disfluency in Spontaneous Speech, ISCA Tutorial and Research Workshop*, Aix-en-Provence. 2005.

[3] Clark, H. H. *Using Language*. Cambridge University Press: Cambridge. 1996.

[4] Fox Tree, J. & Clark, H.H, "Pronouncing "the" as "thee" to signal problems in speaking". *Cognition*, 62: 151-167. 1997.

[5] Bard, E.G., Lickley, R., Aylett, M. Is Disfluency just difficulty? . *Proc. of DiSS'01: Disfluency in Spontaneous Speech, ISCA Tutorial and Research Workshop*, Edinburgh. 2001.

[6] Lickley, R. J. "Dialogue moves and disfluency rates" *Proc. of DiSS'01: Disfluency in Spontaneous Speech*: *ISCA Tutorial and Research Workshop*, Edinburgh. 2001.

[7] Carletta, J. C., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G., Anderson, A., "The reliability of a dialogue coding structure coding scheme", *Computational Linguistics*, 23:13-31, 1997.

[8] Eberhard, K., Nicholson, H., Kübler, S., Gundersen, S., Scheutz, M. "The Indiana Cooperative Remote Search Task" (CReST) Corpus". *Proc. of LREC 2010: Language Resources and Evaluation Conference.* Malta. 2010.

[9] Boersma, P. Weenink, D. Praat: a system for doing Phonetics by computer. (Version 5.1.34). Retrieved January 2009. from http://www.praat.org

[10] Lickley, R.J. "HCRC Disfluency Coding Manual" *HCRC/TR-100*, HCRC, University of Edinburgh, 1998.

[11] Oviatt, S. "Predicting spoken disfluencies during human-computer interaction". *Computer Speech and Language*, 9: 19-35, 1995.