Disfluency: The Cost of Attending to Listener Feedback?

Hannele B.M. Nicholson

School of Philosophy, Psychology and Language Sciences,

University of Edinburgh

E-mail: hannele@ling.ed.ac.uk

ABSTRACT

Previous disfluency studies have argued that disfluencies are strategic signals to a listener in times of difficulty (Clark & Wasow, 1998). Other research suggests that disfluencies are a possible indication of cognitive overload (Bard et al., 2000). Bard et al. (2003) demonstrate that interaction with a Follower could induce more difficulties for a speaker. The current paper tests these hypotheses by analysing eye-gaze and disfluencies, instances where communication goes awry, in 96 interactive Map Task dialogues, to determine whether there is a cognitive cost associated with interaction with a listener.

1 INTRODUCTION

Within psycholinguistic literature, there are traditionally two models of how dialogue is theorised to function. The first, the 'user-modelling' design, suggests that speakers create and maintain a mental model of what their listener knows (Brennan & Clark, 1996). The speaker then specifically designs any utterances to this model of the listener while planning what to say next. This model contrasts with a second model of dialogue, the speaker cognition model, which suggests that listener-modelling is a costly process and is perhaps not necessary all the time (Pickering & Garrod, 2004; Brown & Dell, 1987, Bard & Aylett, 2001). Instead, this model suggests that cognitive resources are allocated such that the system is not over-loaded (Brown & Dell, 1987). Moreover, basic set theory logic shows that there is no need for the speaker to maintain a model of the listener since any mutual knowledge in the conversation is also knowledge that the speaker knows to be true (Barr & Keysar, 2002).

In this paper, disfluency will be used, as it has been previously, as a means for testing models of dialogue. Previous research by Clark & Wasow (1998) has described disfluency as a signal to a listener. The strategic view of disfluency is in line with the 'user-modelling' view because it suggests that the speaker intentionally hesitates in order to signal processing problems to the listener. The second view of disfluency as proposed by Bard & Aylett (2000a, 2000b) and Horton & Keysar (1996) suggests that disfluency is the result of cognitive overload. Disfluency occurs when the cognitive system is over-burdened, for example when there is a time limit or when the listener shows signs of misunderstanding and the speaker is unable to plan speech and attend to other tasks simultaneously.

A crucial difference between the models of disfluency is the prediction made with respect to whether interaction with a listener comes with a cost. The 'user-modelling' view predicts that there is no such cost and that if anything listener interaction is crucial. The cognitive burden view, on the other hand, does predict such a cost since it views cognitive resources as limited.

The MONITOR project was designed to test collaborative dialogues to determine whether there is an associated cost for interactivity. The MONITOR task is a variant of the Map Task design (Brown et al., 1983; Anderson et al., 1991). In this task, Instruction Givers (IGs) are asked to guide an Instruction Follower (IF) along the route of a cartoon map. The IG's map has a route drawn in already whereas the IF has no such route; it is the IG's job to explain the route in such a way that the IF can reproduce it on her map.

In the MONITOR version of the Map Task two variables were added as a means for testing how collaborative dialogues work. Firstly, time pressure was a variable. The IG could be forced to complete an accurate description within a one-minute time limit or the IG was allotted an unlimited amount of time. Secondly, access to listener Feedback was also a variable. In half of all trials, the IG could see a red square superimposed on the map on her computer screen as is depicted in Figure 1 below. The experimenter told the IG that this red square represented the eye-gaze of an IF who was seated in adjacent room and explained that it would move about. Unbeknownst to the IG, however, this was a 'Wizard of OZ' technique and there was no actual IF. The experimenter advanced the square to pre-determined landmarks that were scheduled either to match or to mismatch IG's expectations.

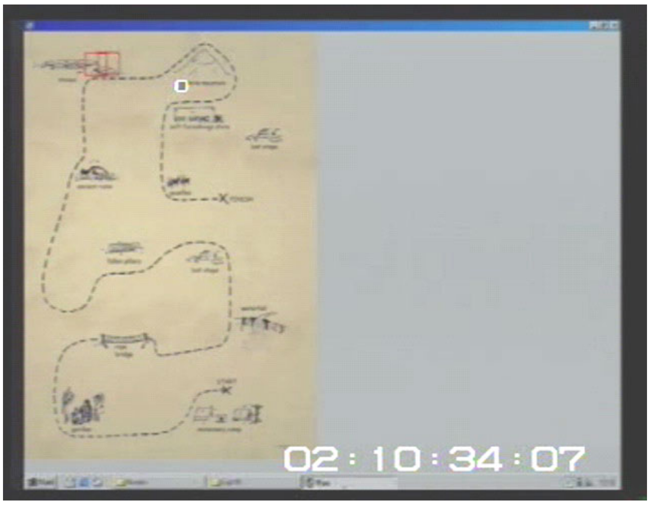

Figure 1. A snapshot of an ongoing dialogue. The black box represents the speaker's gaze while the red box represents the follower's purported location.

1.1 MONITOR HYPOTHESES

In this experiment, time-pressure and attention to feedback are critical to determining the IG's state, and the rate of disfluencies is used to indicate their effect. Three theoretical outcomes are possible. The first prediction, as shown graphically in Figure 2, is in line with the signalling or 'user-modelling' hypothesis. The signalling view predicts that only listener feedback should affect speaker disfluency rate; time-pressure should not have any effect because disfluencies are essentially discreet signals of commitment to an utterance. In Figure 2, the disfluency rate is high whenever a listener is involved (Feedback) and low when one is not (Not Feedback). Time-pressure ('timed' vs. 'untimed') does not cause the disfluency rate to differ.

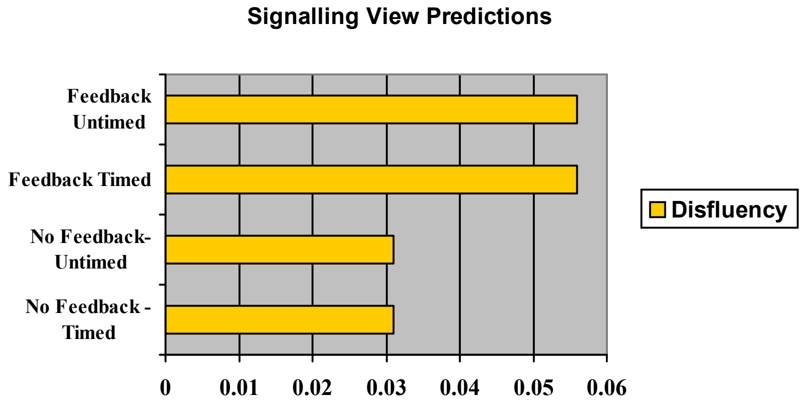

Figure 2. Graph showing the predictions of the Signalling view. Disfluency rate per fluent word is shown on the horizontal axis.

Figure 3 depicts the predictions made by the strong version of the cognitive burden view, henceforth the 'planning' view. Recall that the planning view predicts that disfluency is associated with an over-burdened system. In terms of time-pressure and feedback, the system is most taxed and the rate of disfluency highest when it has to cope with listener feedback whilst under time-pressure. Speakers under time-pressure and with listeners to attend to are predicted to have higher disfluency rates than those without because interacting with another person is particularly burdensome, therefore having feedback from the IF without time-pressure should still yield higher disfluency rates than untimed conditions where no feedback was present.

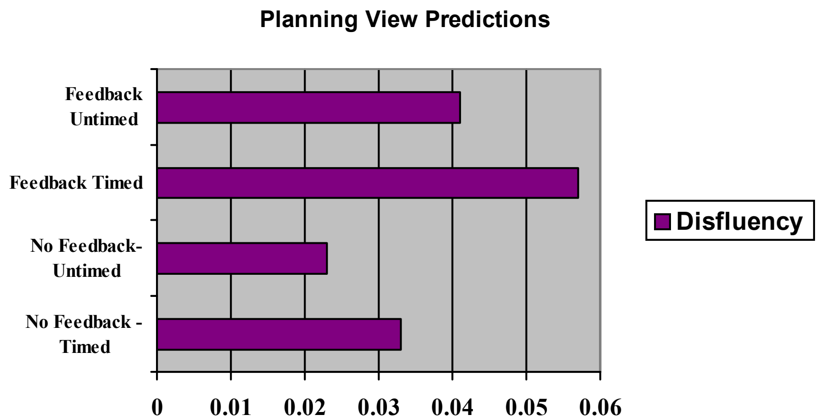

Figure 3. Graph showing the predictions of the Planning view. Disfluency rate is plotted on the horizontal axis.

Finally, the third prediction is somewhat of a hybrid. Bard & Aylett (2000) showed that longer utterances tend to have higher disfluency rates. The more words one utters, the greater the probability that one or more of those words will be disfluent. For this reason, the ‘Luxury view’ predicts that the more time one has to describe the route, the more likely one is to be disfluent. According to this view, IGs should be less disfluent under time-pressure because they speak more economically. The Luxury view still predicts that interacting with a listener is taxing and therefore disfluency rates for Feedback trials are projected to be higher than those for No-Feedback trials.

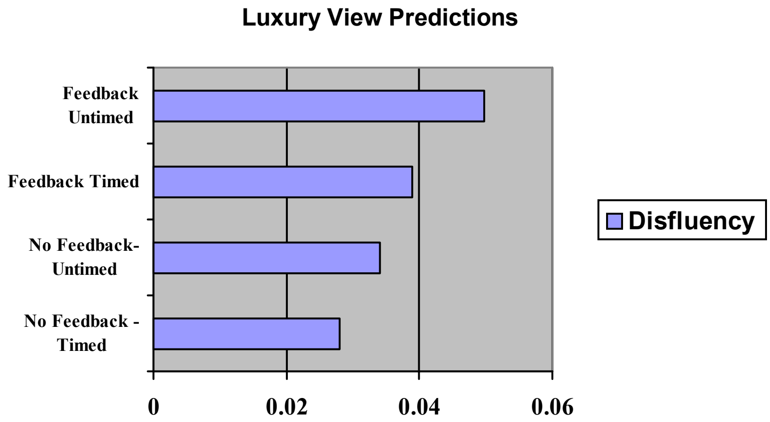

Figure 4. Graphs depicting the predictions of the Luxury View. Disfluency rate is plotted on the y-axis.

2 METHOD

2.1 Stimuli

24 native-English speakers at the University of Glasgow were recruited as Instruction Givers (IGs). All participants were seated in a room in a reclining chair and in front of a computer monitor. A cartoon map with a route was projected onto the monitor 3 feet away from the subject. An SMI (Sensory Motor Instruments) non-invasive, infra-red eye tracker recorded where IG was looking. Participants were instructed to guide the Instruction Follower (IF) along the route of the map using everyday instructions. The IG’s speech was then recorded on a Mackie 12-channel mixer and an Aiwa tape deck recorder.

Each subject completed 4 trials according to a 2 x 2 Repeated Measures ANOVA design for Feedback (2) x Time-Pressure (2). Each of the 4 maps was described by each of the 24 subjects. Each subject performed under each condition, one with visual feedback and a time limit of 1 minute (FT), one with feedback but without the time limit (FU), one with no visual feedback and a time limit (NT), and one with no feedback and no time limit (NU). Each map had 8 landmarks scheduled to be ‘correct’ (ie. on the IG’s route) and 4 landmarks scheduled to be ‘wrong’ (ie. off the IG’s route).

2.2 Speech Transcription and Coding

All 96 trials were transcribed by trained students at the University of Edinburgh using Transcriber 1.4.2 software. The coders were also trained to code the trial according to the Conversational Game-Move schema developed by Carletta et al. (1996). In this schema, every conversation can be broken down into Transactions and Moves. Transactions are stages of the talk in which IG completes a task sub-goal, e.g. succeeds in guiding the IF around a particular landmark on the map. There are several types of Transactions: normal, review, overview and irrelevant. In a normal transaction the IG simply directs the IF along the route. In a review transaction, the IG might reiterate particular instructions for the IF, whereas an overview would contain a general description of the map.

A new type of Transaction had to be introduced for the MONITOR project. Quite frequently, at least four times per map, the IF square deviated off-course and it would then become the IG's responsibility to decide whether to 'rescue' the IF. These transactions were termed 'Retrievals'.

Each Transaction can be broken down in to Games and Moves. According to Carletta et. al (1996) Moves 'are simply different kinds of initiations and responses classified according to their purposes'. A Move is basically a unit of interaction which can accumulate to form a Conservational Game. There are also several types of Moves: instruct, explain, align, query, acknowledge. The IG issues instructions in an Instruct move, gives further detail in an Explain move, asks questions about the IF's map in a Query move, reorients with the IF in an Align move, acknowledges that a previous statement was correct in an Acknowledge move. Usually Transactions comprise whole numbers of Games but since Games involve two actual participants, the monologues were coded only for Moves and Transactions.

All transcriptions and Transaction-Move codings were cross-checked by other coders and by the Research Associate in charge of the transcriptions.

2.3 Disfluency Coding

All 96 trials were coded for disfluency by the author according to Lickley's (1998) system. In problematic cases (ie. extremely disfluent cases) a cross-checker was brought in and a verdict reached[1]. In this system, a disfluency can be one of four types: Repetition, Insertion, Substitution or Deletion.

Silent pauses, filled pauses (ehm, um, uh), prolongations, inhalations, and exhalations were also coded. Only analyses of repetitions, substitutions, insertions and deletions will be presented in this paper.

[1] The author is extremely grateful to Dr. Robin Lickley for his assistance in this manner.

3 RESULTS

3.1 Words and Transactions Results

Word rate per trial increased significantly in the absence of time-pressure: trials were significantly longer when IG lacked a time-limit (319 words/trials on average) compared to when IG had a deadline(224) ($F(1,23) = 33.69$, $p < .001$). This result is in concordance with previous research by Bard & Aylett (2000a) which shows that given more time, speakers will say more.

As reported previously in Bard et al. (2003), time-pressure was a significant factor for increasing Transaction rate. The mean number of Normal transactions increased significantly in time-unlimited trials compared to the time-limited trials (9.63 Normal, 11.27 total) ($F(1,23) = 5.77$, $p = .025$ for normal; $F(1,23) = 9.95$, $p < .01$ overall). Feedback on the other hand produced no such effect for Normal Transactions.

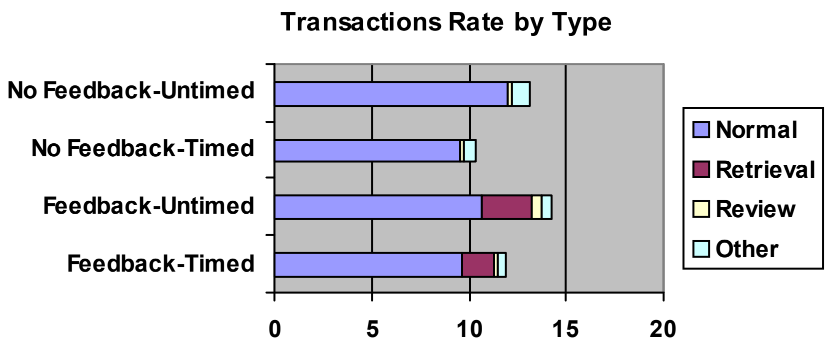

Figure 5. Graph depicting the rate per trial of transaction types for each of the four experimental conditions.

Retrieval transactions displayed a different trend. As is clearly evident from Figure 8, Retrievals occurred primarily in Feedback conditions and virtually never in No-Feedback conditions ($F_1(1,23) = 25.84$, $p < .001$). This could be due to the fact that during Feedback trials, IG could see precisely when IF diverged off-course whereas this was not possible during No-Feedback trials.

3.2 Disfluency Results

Speakers were more verbose in time-unlimited trials (319 words/untimed trial compared to 224 words/timed trial). Since it has been shown that lengthy utterances are naturally more disfluent, disfluency rate was measured in terms of the number of disfluencies per fluent word.

Overall disfluency rates were higher in Feedback conditions: speakers erred more when interacting with a listener ($F_1(1,23) = 8.86$, $p < .01$). Disfluency rate remained unaffected by time-pressure, however.

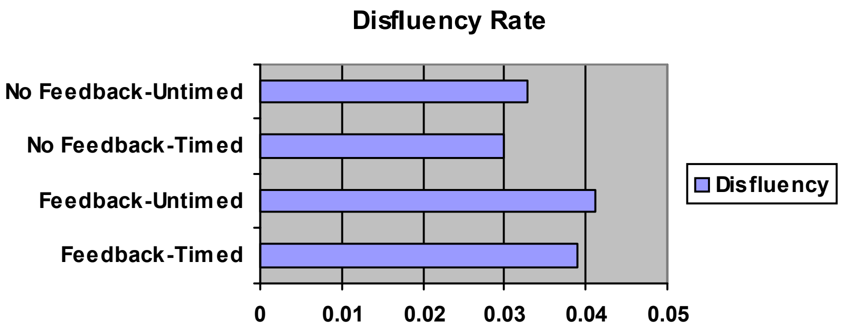

Figure 6. Graph depicting disfluency rate (number of disfluences per total words) in each of the four experimental conditions.

At first glance, these results seem in line with the predictions made by the 'signalling' view which predicts that time-pressure will not have an effect but Feedback with a listener will. A closer observation reveals a more complex picture. Clark & Wasow (1998) originally propose that signalling occurs during repetition disfluencies when the speaker re-iterates some portion of an utterance. This act of re-iteration signals a commitment to the utterance and to the listener and one might think that disfluency could be a strategic signal.

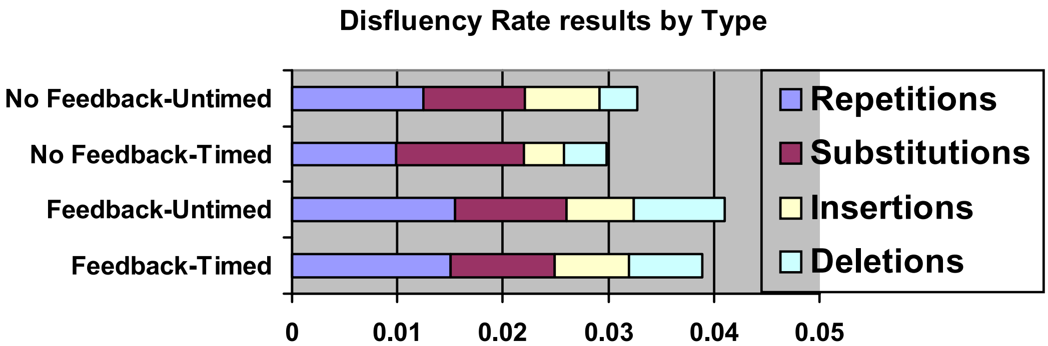

Figure 7. Graph depicting disfluency rate (disfluencies per total words) by disfluency type in each of the four experimental conditions. Disfluency rate is presented on the x-axis.

Figure 7 shows what happens when the disfluency rate presented in Figure 6 is divided into disfluency types. Notice that the boxes for repetitions, substitutions and insertions remain relatively unchanged as one moves through the different conditions. Only deletion rate (signified by light green shading) changes significantly. Indeed, deletion rate increased in the Feedback conditions (0.008) compared to the No-Feedback trials (0.004) ($F_1(1,23) = 14.61, p < .001$). No other disfluency type or sub-group of disfluency types revealed a significant trend. In terms of time-pressure, while results were largely non-significant, there was a trend for an increased deletion rate in Feedback-Untimed trials compared to Feedback-Timed trials ($F_1(1,23) = 3.59, p < .071$).

Thus far, we've seen that quantitatively time-pressure affects the amount of speech. Feedback, on the other hand, has a qualitative effect, increasing disfluency rate when present. This finding appears listener-centric or otherwise in accordance with the signalling view. On the surface, results suggest that feedback has a significant effect on the fluency of the speaker whereas time-pressure does not. Recall, however, the type analysis in light of Clark & Wasow's (1998) claims. Clark & Wasow (1998) suggest that repetitions signal commitment to a listener because some portion of the utterance is repeated when the utterance could have otherwise commenced from the interruption point. The disfluency-type analysis presented here finds that only deletion rate varies significantly for Feedback. The question one must then ask is whether a deletion signals commitment in the same fashion as a repetition. By definition, a deletion is an abandonment of the previous utterance and it could not possibly signal commitment. This observation raises serious speculation as to whether the signalling view is the best explanation for the results found here.

Nevertheless, there does seem to be a relationship between visual information and the IG's reaction. IGs retrieve more often when they possess knowledge that the IF has diverged off-course. Deletions also occur significantly more often when the IG has seen the IF's location. In order to answer this question, we must conduct an analysis of disfluency and eye-gaze.

3.3 Disfluency and Eye-gaze Results

Gaze refers to where the IG is looking on the screen, whether at the route or the location of the IF or both. In this analysis, the IF can either be on a 'Correct' or IG-anticipated location or at a 'Wrong' off-route location. The IG can either look at the IF's 'gaze' square or avoid glancing at it.

All 3 theories predicted higher disfluency rates in interactive circumstances. The theories differ as to when disfluencies actually occur. The signalling view predicts that attention to the listener is constant, and therefore disfluency would occur continuously. The planning view, however, theorizes that disfluency will occur most often after problematic interchanges or when re-alignment is essential. Contrary to predictions, Bard et al. (2003) have demonstrated that IGs attend to IFs most when the IF hovers near a 'correct' or on-route landmark. Conversely, IGs tend to avoid gazing at the IF's visual feedback when the IF has strayed from the route.

The purpose of this section is to analyse which theory makes the most accurate prediction about looking and becoming disfluent. We ask if a move where the IG looked at the IF was also a disfluent Move. The prediction is that Moves in which the IG looked at the IF on a

Wrong location will be disfluent because once the IF has deviated off-route, the IG must rescue the IF so that the route-description is successful and the IF reaches the finish point. Moves in which the both IG and IF gazed at a Correct location will be fluent because since the IF is in the predicted location, the success of the trial is not at risk and the Giver has less extra explaining or rescuing to do. Following this analysis, I will attempt to determine if IG disfluency follows any sort of a pattern with respect to eye-gaze and IF location by conducting a temporal analysis of disfluency and eye-gaze.

3.3.1 Disfluent Move and Gaze Patterns

1229 moves (ie. the total number of moves across subjects in Feedback conditions) were analysed. For each Move, a Perl script[2] determined the IF's location (either 'correct' or 'wrong'), the IG's attention to the IF (either 'looking' or 'not looking') and the fluency of the IG between the relevant Move start and end times. An ANOVA of Feedback x Gaze x Time with the disfluent move rate as dependent variable revealed that regardless of the IF's location, the IG was more disfluent when she attended to the IF ($F_1(1,23) = 13.35, p < .001$). A fine-grained analysis of this finding in terms of IF location demonstrated that seeing the IF on a 'correct' location, not a 'wrong' one as predicted, induced higher disfluency rates ($F(1,23) = 3.097, p < .02$). Time-pressure was not significant. Moreover, significant results for Wrong IF locations alone (IG looking at a lost IF vs. IG not attending to a lost IF) were not found. While the IG is more disfluent when looking at the IF, a fine-grained analysis reveals that the majority of these disfluent moves occurred when the IG saw that the IF was on-route. Since there are more scheduled correct landmarks per map and IGs prefer to gaze at Corrects, the above observation could merely be an artefact and is at best inconclusive.

3.3.2 Temporal Analysis

In this section, the results of the temporal analysis will be reported. The input to the Perl script was time spans of IF location information (508 instances where the IF was correct, 185 instances where the IF went wrong). The Perl programme then searched the IG's codings (both fluency and gaze) for an indication of the IG's behaviour 200 msec before or after the time points on the IF's gaze coding. This procedure was then replicated for both 300 and 400 msec prior to disfluency and 300 and 400 msec periods after disfluency as depicted in Figure 8.

[2] The author wishes to thank Joseph Eddy for creating and maintaining these scripts!

Temporal Slices of Disfluency and Gaze

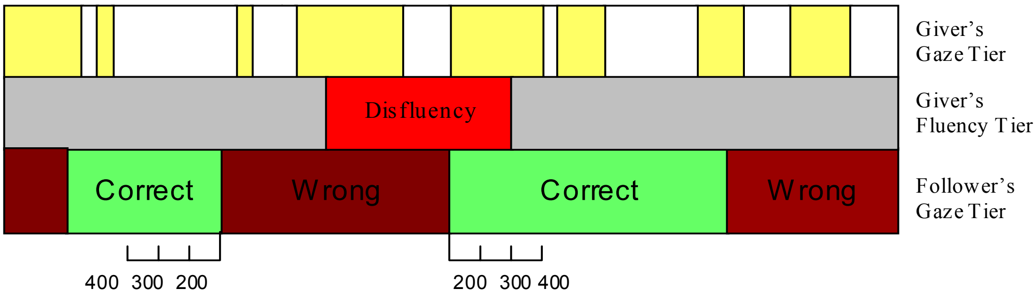

Figure 8. A pictoral diagram explaining how the Giver's Gaze Tier, the Giver's Fluency Tier and the Follower's Gaze Tier overlap. Yellow boxes on the Giver's Gaze Tier represent periods of gazing while white boxes indicate not-gazing at the Follower.

3.3.3.1. Temporal Analysis – Prequel Analysis

In this case, the dependent variable is the number of time-spans during which disfluency occurred prior to an 'IF's gaze' divided by the total number of that type of IF gaze (eg. # disfluent Corrects / Total number of Corrects). Independent variables were presence or absence of time-pressure, IG's attentiveness to the IF and the location of the IF. All results are presented in Table 1. The ANOVA was run for 24 subjects.

Independent Factors	**200 msec**	**300 msec**	**400 msec**
IG's Gaze at IF	$F(1,23)=10.197$, $p < .01$ Looking: 0.015 Not Looking: 0.059	$F(1,23)=11.311$, $p < .01$ Looking: 0.017 Not Looking: 0.061	$F(1,23)=110.749$, $p < .01$ Looking: 0.02 Not Looking: 0.062
IF's location	n.s.	n.s.	n.s.
Time-pressure	n.s.	n.s.	n.s.
IG's Gaze X IF's location	$F(1,23) =14.375$, $p =.001$	$F(1,23)=18.102$, $p< .001$	$F(1,23)=20.381$, $p< .001$

Although an ANOVA was run for windows of time varying by 100 msec (ie. 200, 300, 400 msec) the results across these windows are remarkably similar. Whatever the period of time, the IG always seemed to be gazing elsewhere (**200 msec:** $F(1,23) = 10.197$, $p < .01$; **300 msec:** $F(1,23) = 11.311$, $p < .01$; **400 msec:** $F(1,23) = 11.749$, $p < .01$) prior to the IF finding the correct location. Furthermore, for all time-spans the IG was always more disfluent before ignoring an errant IF (**200 msec:** $F(1,23) = 14.375$, $p = .001$; **300 msec:** $F(1,23) = 18.102$, $p < .001$; ***400 msec:*** $F(1,23) = 20.381$, $p < .001$).

Overall, the prequel to IF location analysis reveals an association between taxing information (ie. when the square is wrong and the IG must do something to ameliorate the situation) and disfluency. Undoubtedly, since it is unlikely that IGs are actually clairvoyant, all prequel results must be taken with a grain of salt. It is improbable that IGs could predict where the IF would be 200-400 msec before the trajectory there.

3.3.3.2. Concurrent Analysis

A slightly different pattern is obtained during the actual IF feedback periods. IGs are most disfluent when the IF veered off-route compared to when they remained on-track ($F(1,23) = 5.754, p < .05$). This supports the hypothesis that when IGs are forced to handle troublesome scenarios, they pay for feedback with disfluency.

Other results for the concurrent analysis reveal effects for time-pressure, attention and an interaction. IGs are also more prone to disfluency when they are not subject to the pressures of time ($F(1,23) = 5.276$, $p < .05$). Additionally, just looking at the IF, regardless of the IF's actual location, induced higher rates of disfluency ($F(1,23) = 6.782, p < .02$). Finally, although mean disfluency rate (0.227) is highest when IGs are gazing at a IF in the correct location, the increase in disfluency rate with looking at an IF is less when the IF is off-route (.024) than when she is correct (0.162) ($F(1,23) = 8.332$, $p < .01$).

3.3.3.3. Sequel Analysis

The sequel results echo those of the prequel findings. Speakers tended to be most disfluent after a period of not looking at the IF's feedback (**200 msec:** $F(1,23) = 12.267$, $p < .01$; **300 msec:** $F(1,23) = 13.76$, $p < .001$; **400 msec:** $F(1,23) = 9.027$, $p < .01$). An interaction between Square and Look showed that IGs are most disfluent subsequent to not attending to wrong feedback (**200 msec:** $F(1,23) = 7.244$, $p < .02$; **300 msec:** $F(1,23) = 7.575$, $p < .02$; **400 msec:** $F(1,23) = 6.69$, $p < .02$). In this situation, it could be that the IG had noticed that the IF was not in the expected location. Realising this, the IG suddenly panicked and then shifted her eye-gaze to the intended correct landmark while devising a rescue strategy. Another explanation, in concordance with the concurrent disfluency results (Section 3.3.3.2) is that the speaker did not need to attend to the location of the square after noticing it had gone astray. The IG's task then was simply to devise a means for retrieving the IF so that the dialogue could continue.

4 DISCUSSION

This paper began with three hypotheses that predicted how time-pressure and feedback would affect disfluency rate in a Map Task experiment. First, the Signalling view predicts high disfluency rates when listener feedback is involved but no effect of time-pressure. Next, the Planning view also predicts high disfluency rates when a listener is involved but highest rates when both feedback and time-pressure are present. Finally, the Luxury view offers somewhat of a hybrid view: disfluency rates are highest when the utterance is lengthy (ie. in the absence of time-pressure) and when the trial is interactive. The purpose of the eye-gaze and disfluency analysis is to determine the chronology of interaction and whether a lost IF induces disfluency.

The results from this experiment show that disfluency rate is highest in conditions in which Feedback was present. However, a breakdown by disfluency type reveals that only deletions are responsible for the higher disfluency rates with feedback. This suggests a possible link between deletion and the speaker's remedy for an errant IF.

A bipartite analysis of eye-gaze revealed that a high disfluency rate is associated with attention to the IF. First, during the majority of disfluent moves, speakers attempted to look at the IF. When the speaker looked at the IF, however, the IF was generally on-track and not off. Since there are 8 scheduled correct landmarks and 4 scheduled wrong landmarks, there is more on-track time. This reflects only the general trend that IGs prefer to look at Correct information and seem to avoid knowledge that things have gone wrong.

Second, a temporal analysis of the eye-track record both prior to and subsequent to disfluency showed again that speakers are most disfluent when gazing at the on-route IF. That said, after a deletion, speakers looked more often at the IF than they ignored her. Perhaps speakers do this as a last resort attempt to rescue the IF and the dialogue from going too far afield.

Finally, we saw that disfluency was linked to situations during which the IF was in a Wrong location. This is evidence that handling taxing information like a lost IF is more difficult than an on-route IF and that in such a situation, an IG may encounter fluency problems when trying to re-align with the IF. While this paper has not shown conclusively that there is an associative cost for attending to feedback in dialogue, it has shown at least an association between problematic interactions and disfluency.

5 CONCLUSIONS

This paper has shown that speakers pay for visual feedback with their own fluency. A speaker may be required to pay for finding and fixating on the IF. IGs were also more disfluent during periods in which the IF was lost, suggesting a link between problematic knowledge and disfluency.

While the results presented here do not meet any of the three predictions exactly, there does seem to be support for the Pickering and Garrod (2003) model of interaction and alignment. According to this model and observed in this experiment, IGs do not seem to need to align with their IF at all times. Rather, disfluency results when interlocutors are not aligned. After disfluency, comes a period of re-alignment during which the speaker may gaze at the IF and use the dialogue to re-establish a shared view of the instructions.

Planned future work will investigate if the speaker responds the same way to verbal signs of distress as he does to visual indications of a lost IF.

ACKNOWLEDGEMENTS

This research has benefited from the support of a grant awarded by the EPSRC. I would like to thank my supervisors Dr. Ellen Gurman Bard and Dr. Robin Lickley for their help and their friendly supervision. I would also like to thank the transcribers and coders involved with the MONITOR project, most especially research associates Dr. Yiya Chen and Ms. Catriona Havard.

BIBLIOGRAPHY

Anderson, A.H., Bader, M., Bard, E.G., Doherty, G.M., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H.S., and Weinert, R., 1991. The HCRC Map Task Corpus. *Language and Speech*, **34,** pp. 352-366.

Bard, E.G., Anderson, A., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. 2000a. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language,* 42, pp. 1-22.

Bard, E.G., Aylett, M., Bull, M.. 2000b. More than a Stately Dance: Dialogue as a Reaction Time Experiment. *Proc. Soc. For Text & Discourse*.

Bard, E.G. & Aylett, M. 2001. Referential Form, Word duration, and Modelling the Listener in Spoken Dialogue. Paper presented at the Proceedings of the Twenty third Annual Conference of the Cognitive Science Society.

Bard, E.G., Anderson, A.H., Flecha-Garcia, M., Kenicer, D., Mullin, J., Nicholson, H.B.M., Smallwood, L., & Chen, Y. (2003). Controlling Attention and Structure in Dialogue: The Interlocutor vs. the Clock. *Proceedings of ESCOP 2003.* Granada, Spain.

Barr, Dale J. & Boaz Keysar, 2002. Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, vol. 46, pp. 391-418.

Brennan, S. & Clark, H.. 1996. Conceptual Pacts and Lexical choice in Conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition*. Vol. 22, No. 6, pp. 1482-1493.

Brown, G., Anderson, A.H., Yule, G., Shillcock, R.. 1983. *Teaching Talk*. Cambridge: Cambridge University Press.

Brown, P. & Dell, G.S. 1987. Adapting production to comprehension – the explicit mention of instruments, *Cognitive Psychology*. 19, pp. 441-472.

Carletta, J., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G., and Anderson, A. 1996. HCRC Dialogue Structure Coding Manual. Technical Report HCRC/TR-82. http://www.citeseer.ist.psu.edu/carletta96hcrc.html

Clark, H.H. & Wasow, T..1998. Repeating words in Spontaneous Speech. *Cognitive Psychology*.37, pp. 201242.

Lickley, R.J..1998. HCRC Disfluency Coding Manual. *HCRC Technical Report 100*. http://www.ling.ed.ac.uk/~robin/maptask/disfluency-coding.html

Pickering, Martin & Simon Garrod. 2004. Why is Conversation so easy? *Trends in Cognitive Sciences,* vol. 8, Issue 1. pp. 8-11.