Proceedings of DiSS'03, Disfluency in Spontaneous Speech Workshop, 5–8 September 2003, Göteborg University, Sweden.
Robert Eklund (ed.), Gothenburg Papers in Theoretical Linguistics 90, ISSN 0349–1021, pp. 17–20.

The intentionality of disfluency: Findings from feedback and timing

Hannele Nicholson[1], Ellen Gurman Bard[1], Robin Lickley[2],
Anne H. Anderson[3], Jim Mullin[3], David Kenicer[3] & Lucy Smallwood[3]

[1] University of Edinburgh, Edinburgh, Scotland
[2] Queen Margaret University College, Edinburgh, Scotland
[3] University of Glasgow, Glasgow, Scotland

Abstract

This paper addresses the causes of disfluency. Disfluency has been described as a strategic device for intentionally signalling to an interlocutor that the speaker is committed to an utterance under construction [14, 21]. It is also described as an automatic effect of cognitive burdens, particularly of managing speech production during other tasks [6]. To assess these claims, we used a version of the map task [1, 11] and tested 24 normal adult subjects in a baseline untimed monologue condition against conditions adding either feedback in the form of an indication of a supposed listener's gaze, or time-pressure, or both. Both feedback and time-pressure affected the nature of the speaker's performance overall. Disfluency rate increased when feedback was available, as the strategic view predicts, but only deletion disfluencies showed a significant effect of this manipulation. Both the nature of the deletion disfluencies in the current task and of the information which the speaker would need to acquire in order to use them appropriately suggest ways of refining the strategic view of disfluency.

1. Introduction

Disfluency is known to be more common in dialogue than in monologue [19]. Explanations for this fact fall into two categories. One ties disfluency to active strategies for cultivating common ground, the accumulating knowledge that interlocutors are mutually conscious of sharing [9, 13, 21], while the other sees disfluency as an accidental result of cognitive burdens [6], which necessarily increase when a speaker must process a listener's utterances while composing his or her own.

In the strategic view, disfluency is one of a number of intentional strategies which speakers employ to maintain mutuality. Clark & Wasow [14] argue that repetition disfluencies are strategically deployed to signal ongoing difficulty in producing an utterance to which the speaker is nonetheless committed. Evidence of prosodic cues that signal strategic intention has been obtained for repetitive repair [21].

In the alternate view, conversation is a cognitively taxing process and competition is high for production resources [3, 4, 9, 15, 16]. A speaker must design the sub-goals of any task which a dialogue helps the interlocutors to pursue, plan the sections of the dialogue which correspond to these goals, and attend to the contributions of the interlocutor, while micro-planning his/her own utterances [4, 5]. Disfluencies may occur when this burden becomes so great that errors in planning or production are not detected and edited covertly before articulation begins. Increases in disfluency accompanying increased complexity of any of the cognitive functions underlying dialogue are taken to support this view. Long utterances, which tend to be more complex than short, certainly tend to be disfluent more often [14]. Bard and her colleagues have shown that even with utterance length taken into account, production burdens correlate with disfluency: formulating multi-reference utterances and initiating new sections of the dialogue both tend to encourage disfluency. In contrast, no characteristics of the prior interlocutor utterance have any independent effect on disfluency rate. This account of disfluency joins other models of dialogue phenomena in ascribing to the speaker's own current needs many of the behaviours which are often thought to be adaptations to a developing model of the listener's knowledge [See 2, 3, 4, 5, 8, 20].

This paper presents the first group of results from a series of experiments designed to discover whether speakers are more concerned with attending to their listeners' knowledge or completing their own production tasks. The experiments use a variant of the map task [1, 11]. In the original task, players have before them versions of a cartoon map representing a novel imaginary location. The Instruction Giver communicates to the Instruction Follower a route pre-printed on the Giver's map. The current series uses only Instruction Givers and manipulates both time-pressure and feedback from a presumptive Follower.

The time-pressure variable contrasts instructions composed in the Giver's own time with a time-limited condition. If disfluencies are a basic signaling device and important to the conduct of a dialogue, then this manipulation will not affect them. If disfluencies are failures of planning, time-pressure should increase their rate of occurrence. If, on the other hand, disfluencies are a luxury, a rhetorical device available to speakers but not required for the process of maintaining mutual knowledge, then they may be more common when interlocutors have the time to indulge in them, that is, in the untimed condition.

The feedback variable contrasts monologue map tasks, supposedly transmitted to a listener in another room, with tasks for which there is minimal feedback in the form of a square projected on the map to represent the direction of the Follower's gaze. If modeling the listener's knowledge is critical to the process of dialogue, then this is the most important kind of feedback, for it tells one interlocutor what the other knows about the map and how s/he interprets the instructions. If speakers treat these tasks as interactive, and if disfluency is an intentionally helpful signal, then disfluency should be more common in this condition than in pure monologue. For example, repetition disfluency should be induced by the availability of the listener [14].

The interactions of these two manipulations are of particular interest. A pure strategic model demands a main effect of feedback but would sit well with enhanced rates of disfluency in the feedback condition with time pressure, where most difficulties would arise. A pure cognitive difficulty model predicts enhanced rates of disfluency under time pressure, but particularly again where feedback and time-pressure both add to the speaker's cognitive burdens. Associated with the cognitive difficulty model are a set of results which could support a hybrid view: that listener-centric behaviour in dialogue is a luxury [15, 16] which will be abandoned when the speaker has more pressing tasks to pursue. This model predicts that disfluencies will appear at a higher rate where feedback makes the task interactive and where ample time permits the consideration of the listener's needs.

2. Method

2.1. Task

Disfluencies are obtained from the MONITOR corpus currently under collection [7]. This corpus employs a variant of the map task [1, 11]. In this version of the MONITOR task, subjects are seated before a computer screen displaying a map of a fictional location which includes a route from a marked start-point to buried treasure. Labelled landmarks and map designs are adapted from the HCRC Map Task Corpus [1]. Subjects are requested to help a distant listener reproduce the route. Subjects' instructions were recorded onto the video record by a close-talking microphone and their gaze direction was recorded by a screen-mounted eye-tracker. At the beginning of each trial, the tracker was calibrated.

2.2. Experimental Design

The experiment crossed feedback (2) and time-pressure (2). In the no feedback conditions, subjects saw only the map. In the feedback condition, a small moving square was superimposed on the map and subjects were told that this represented the current direction of their Instruction Follower's gaze. Unbeknownst to the subjects, there was no actual Follower. The feedback gaze-square followed a pre-programmed sequence. It remained on the landmarks determining the route until the first two or three had been successfully negotiated. Subsequently, feedback gaze wandered off-course at least once every other landmark The pattern of incorrect gaze-responses corresponded roughly to the distribution of landmarks which did not match across Giver and Follower maps in [1]. In four cases in each map, the feedback square did not go to the intended landmark, but instead moved to a second, but distant, copy of that landmark or to a space on the map which would have hosted a landmark on the Follower's version of the corresponding HCRC map. In each case, once the subject had introduced the next route-critical landmark, an experimenter in another room advanced the feedback gaze square to its next scheduled target. The square moved about its target landmark in a realistic fashion, with sorties of random radius and angle.

Crossed with feedback was the time-pressure variable. In half of the trials, speakers were permitted only one minute to complete the task; otherwise time was unlimited.

Subjects with normal uncorrected vision were recruited from the Glasgow University community. All were paid for their time. All encountered all 4 conditions. Four different basic maps were used, counter-balanced across conditions over the whole design. Subjects were eliminated if any single map trial failed to meet criteria for feedback or capture quality. The feedback criterion demanded that the experimenter advance the feedback square between the introduction of the pertinent landmark and the onset of the following instruction in all cases where where the feedback was scheduled to be errant and in 70% where the square's movement was scheduled to be correct. The capture criterion demanded that at least 80% of the eye-tracking data was intact. Fifty-four subjects were run before 24 remained with valid sessions in all conditions and with a balanced design in total.

3. Results

3.1. Dialogue Structure

Each monologue was transcribed verbatim and then coded for transaction [12]. A transaction is a block of speech in task-oriented dialogue which accomplishes a task sub-goal. Accordingly, in this task Normal transactions are periods of standard instruction giving. Review transactions recount the route negotiated thus far. Overviews describe the route or map in general. Irrelevant transactions are all off-task remarks.

A fifth type of transaction, Retrievals, was identified in the present monologues and can be used to show that the feedback conditions were in fact interactive. In a Retrieval the speaker neither gives new instructions nor reviews the route but instead moves the presumed IF to a previously named landmark where s/he should be but apparently is not. Figure 1, which divides Transactions by type in each of the four conditions, shows that Retrievals occurred in the two feedback conditions (13% of all Transactions in Feedback-Timed; 18% in Feedback-Untimed) but very rarely otherwise (0.8% of all No Feedback Timed Transactions and 0.3% of No Feedback Untimed: by-subjects 2 × 2 repeated measures ANOVA main effect for Feedback, $F_1(1,23) = 25.84$, $p < .001$). The imbalance suggests that Retrievals are unlikely to be mere clarifications, independent of the IF's behaviour. Since each speaker encountered 4 off-route gaze locations per dialogue, the average number of Retrieval transactions per dialogue, 1.58 for Feedback Timed; 2.58 for Feedback Untimed, shows fairly good uptake of the feedback square's 'mistakes'. The effect of Time-pressure approached significance ($F_1(1,23) = 4.12$, $p = .054$). but only because of an increase in Retrievals in Feedback conditions (interaction: $F_1(1,23) = 5.40$, $p = .029$).

As Figure 1 also shows, Retrievals do not follow the general trends for volume of transactions. Both Normal transactions and total number of transactions are more numerous in the Untimed conditions (11.40 Normal transactions, 13.83 in total per trial) than in the Timed (9.63 Normal, 11.27 total) ($F_1(1,23) = 5.77$, $p = .025$ for normal; $F_1(1,23) = 9.95$, $p < .01$, overall), with no effect of feedback. Other transaction types were unaffected by the experimental variables.

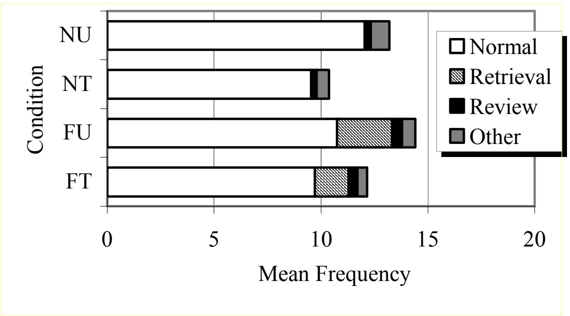

Figure 1: Mean numbers of transactions per trial by type and experimental condition (N = No Feedback; F = Feedback; T = Timed; U = Untimed).

3.2. Words

Word counts included whole and part-words. Again results show less speech with time-pressure (224 words/trial on average) than without (319): ($F_1(1,23) = 33.69$, $p < .001$). There was a non-significant tendency for speakers to resist the effect of time-pressure more with feedback (FT: 238 words/trial; FU: 316) than without (NT: 209; NU: 320): ($F_1(1,23) = 3.31$ $p = .082$).

3.3. Disfluencies

Disfluencies were first labeled according to the system devised by Lickley [18]: as repetitions, insertions, substitutions or deletions. The disfluency coder used Entropic/Xwaves software to listen, view and label disfluent regions of speech. Spectrograms were analyzed whenever necessary. Each word within a disfluent utterance was labeled as belonging to the onset, reparundum, repair, or continuation [17].

Because disfluencies are more common in longer utterances [3, 14, 21], raw disfluency counts may reflect only opportunities for disfluency. To provide a measure of disfluency rate, we divided the number of disfluencies in a monologue by its total number of fluent words, that is by the total number of words less the words in reparanda.

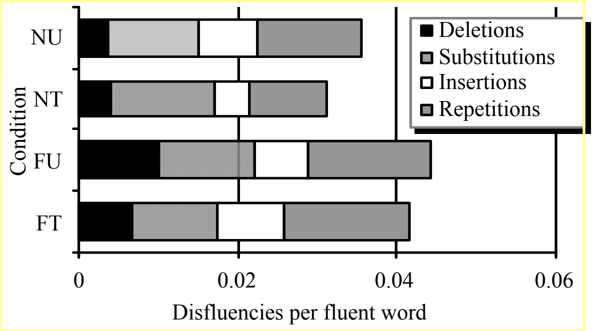

Figure 2: Rates of disfluency by type and experimental condition

The data in Figure 2 display a pattern which would be predicted from an strategic model of disfluency: Speakers were more disfluent in conditions with feedback (0.044) than in conditions without feedback (0.034), ($F_1(1,23) = 8.66$, $p = .007$), but were unaffected by time pressure ($F_1(1,23) = 1.87$, $p = .185$) or by any interaction ($F_1(1,23) < 1$). Because transaction-initial utterances are prone to disfluency, the effects were recalculated with number of transactions in the trial as a covariate. Again, only feedback affected disfluency ($F_1(1,22) = 11.33$, $p < .003$).

3.4. Disfluency Type

Figure 2 also displays the breakdown of disfluencies by type across experimental conditions. Only the rate of deletions showed any significant effect of feedback: an increase in the feedback conditions (.008) over no feedback (.004): ($F_1(1,23) = 14.61$, $p = .001$; $F_1(1,22) = 14.24$, $p = .001$ with transactions as covariate). There was no overall effect of time pressure on deletion ($F_1(1,23) = 2.44$ $p > .10$), though there was a non-significant tendency ($F_1(1,23) = 3.59$, $p = .071$; $F_1(1,22) = 3.62$, $p = .070$ with transactions as covariate) towards the 'disfluency as luxury' pattern: deletions tended to be more common in Feedback Untimed (0.010) than in Feedback Timed (0.007) trials, with no corresponding effect of time pressure in the No Feedback conditions (0.004 in both cases). No other type of disfluency and no combination of other types showed significant effects, though the rate of all non-deletion disfluencies was numerically higher (0.035) with feedback than without (0.030) ($F_1(1,23) = 3.21$, $p = .086$).

4. Discussion and Conclusions

The literature provided us with two major proposals for the causes of disfluency. One suggests that interlocutors intentionally employ disfluencies to warn each other of local difficulty. An interactive situation should encourage more disfluency, and if the signal function is critical, it should be maintained or even increase as the speaker's difficulties are augmented with increasing time pressure. An alternative view suggests that disfluency is an accident of heightened cognitive burden. If so, time pressure should promote disfluency particularly when feedback complicates the speaker's task. A third prediction stresses the fragility of listener-centric behaviour. If disfluency is listener-centric and all such behaviour is at best an option available to speakers when time or attention permit, disfluencies should be more frequent when speakers are not under time pressure but are interacting with listeners.

The experiment reported above successfully manipulated the interactive quality of the speaker's task and the pressure to complete it efficiently. Feedback in the form of a visual representation of a presumptive listener's gaze changed speakers' strategic treatment of the route communication task. A novel type of transaction, provides circumstantial evidence that subjects took seriously the task of tracking and redirecting their listener's gaze when it appeared to have strayed off-course. Retrievals were almost exclusive to the Feedback trials. Time pressure affected how much subjects said, with fewer transactions and fewer words under the one-minute limit.

With the manipulations effective in altering speakers' behaviour, we can return to the predictions for disfluency rate. At first glance, disfluency seems to operate as an important strategic tool, with higher rates in the conditions with feedback and no effect of time-pressure. Yet, when disfluencies are subdivided by type, only deletion disfluencies were significantly more common in feedback trials. This fact is not just a result of sparse data in certain disfluency sub-types. Taken together, all the other kinds of disfluency still failed to respond robustly to feedback. Deletions alone support the strategic view.

Subject 10. Feedback Untimed	
Start	*Utterance*
70.4340	ehm go around and do a big circle ehm like just do a big loop down, **not**
71.4250	oh sorry there was
72.1388	<breath
72.2730	two stone creeks
72.4504	breath>
75.1890	ehm so yeah you're in the right place
Subject 19. Feedback Timed	
Start	*Utterance*
55.6070	and then you take a right across the farmed land
56.4686	< breath
56.7157	breath>
57.8160	**doing a s-**
58.8550	no you go right right at the farmed land

Figure 3: Deletion examples. Deletion disfluency in boldface.

It cannot yet be said that they support it conclusively. First, there was a nearly significant interaction of the type which would be predicted if disfluency were a luxury: disfluency rates were highest in the untimed feedback trials rather than in the timed, where there ought to have been more problems to report. Though we are unable to conclude definitively that deletions result from some optional rhetorical strategy, their content invites further investigation.

The examples in Figure 3 are typical. Subject 10 appears to be abandoning an utterance because he encountered difficulties in reading the map, and resumed with more accurate instructions. His deletion marks 'Giver failure'. Subject 19, on the other hand, interrupts the flow of speech and begins anew because the feedback gaze square did not move in the correct direction. This is an instance of 'Follower failure': the 'Follower's' action appears to have induced the subject to abandon an instruction which the Follower was in no position to obey.

Though deletions are indicators of interaction, it would be difficult to see them as signalling commitment to an utterance, as is thought to be the case for repetitions [14]. Instead, by abandoning an utterance, the speaker is expressing either the inadequacy of his/her own description or inappropriacy of the Follower's response. Whether the two functions are equally likely in both timing conditions we do not yet know.

It is plain, however, that both of these actions would require visual attention beyond what is needed for tracking the route to the next landmark and describing it. Our preliminary analyses of the eye-tracking data captured during these trials indicate that subjects' gaze primarily at the landmarks which are critical to the route [7]. The operations which appear to underlie deletions would produce two different patterns of off-route speaker gaze: scanning the map in the case of Giver failures and monitoring the feedback square's location in the case of Follower failures. If digressions are more common with feedback than without, and if they predominantly track the feedback square, then we may have a visual substrate for Follower failure deletions. If digressions are more common in untimed trials than in timed, then time to acquire the knowledge which underlies any deletion may be the real luxury afforded by our paradigm. Exactly how such a luxury is used – for better scanning of the map or tracking of the interlocutor, we do not yet know. At present, we are examining Giver gaze data to determine which patterns accompany disfluency.

5. Acknowledgements

The authors would like to thank Maria Luisa Flecha-Garcia and Yiya Chen who devised and administered the transcription and dialogue coding systems. This work was supported by EPSRC Research Grant GR/R59038/01 to Ellen Gurman Bard and and GR/R59021/01 to Anne H. Anderson.

6. References

[1] Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Gwyneth Doherty, Simon Garrod, Steve Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Cathy Sotillo, Henry S. Thompson, and Regina Weinert, 1991. The HCRC Map Task Corpus. *Language and Speech***,** vol. 34*,* pp. 352–366.

[2] Anderson, Anne H., Ellen Gurman Bard, Cathy Sotillo, Alison Newlands & Gwyneth Doherty-Sneddon, 1997. Limited visual control of the intelligibility of speech in face-to-face dialogue. *Perception and Psychophysics*, vol. 59(4), pp. 580–592.

[3] Bard, Ellen Gurman, Anne H. Anderson, Cathy Sotillo, Matthew Aylett, Gwyneth Doherty-Sneddon & Alison Newlands. 2000. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, vol. 42, pp. 1–22.

[4] Bard, Ellen Gurman, Matthew Aylett & Matthew Bull. 2000. More than a stately sance: Dialogue as a Reaction Time experiment. *Proceedings of the Society for Text and Discourse*.

[5] Bard, Ellen Gurman & Matthew Aylett, 2001. Referential Form, Word duration, and Modelling the Listener in Spoken Dialogue. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society.*

[6] Bard, Ellen Gurman, Matthew Aylett & Robin Lickley,2002. Towards a Psycholinguistics of dialogue: Defining Reaction time and Error Rate in a Dialogue Corpus. *EDILOG 2002. Proceedings of the 6th workshop on the semantics and pragmatics of dialogue.* Edinburgh: The University of Edinburgh.

[7] Bard, Ellen Gurman, Anne H. Anderson, Marisa Flecha-Garcia, David Kenicer, Jim Mullin, Hannele B.M. Nicholson, Lucy Smallwood & Yiya Chen, 2003. Controlling Structure and Attention in Dialogue: The Interlocutor vs. the Clock. *Proceedings of ESCOP, 2003*, Granada, Spain.

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

[9] Brennan, Susan. & Herbert H. Clark, 1996. Conceptual Pacts and Lexical choice in Conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol. 22(6), pp. 1482–1493.

[10] Brown, P. & Gary S. Dell, 1987. Adapting production to comprehension – the explicit mention of instruments, *Cognitive Psychology*, vol 19, pp. 441–472.

[11] Brown, Gillian, Anne H. Anderson, George Yule, Richard Shillcock, 1983. *Teaching Talk*. Cambridge: Cambridge University Press.

[12] Carletta, Jean, Amy Isard, Steve Isard, Jacqueline Kowtko, Gwyneth Doherty-Sneddon, and Anne H. Anderson, 1997. The reliability of dialogue structure coding scheme. *Computational Linguistics*, vol. 23, pp. 13–31.

[13] Clark, Herbert H. and Catherine Marshall, 1981. Definite reference and mutual knowledge. In Aravind K. Joshi, Bonnie L. Webber, and Ivan A. Sag (eds.), *Elements of discourse understanding*. Cambridge: Cambridge University. Press.

[14] Clark, Herbert H. & Thomas Wasow, 1998. Repeating words in Spontaneous Speech. *Cognitive Psychology*, vol. 37, pp. 201–242.

[15] Horton, W. & Boaz Keysar, 1996. When do speakers take into account common ground? *Cognition,* vol. 59, pp. 91–117.

[16] Keysar, Boaz, 1997. Unconfounding common ground. *Discourse Processes*, vol. 24, pp. 253–270

[17] Levelt, Willem J.M., 1989. Monitoring and self-repair in speech, *Cognition*, vol. 14, pp. 14–104.

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

[19] Oviatt, Sharon, 1995. Predicting disfluencies during human-computer interaction. *Computer Speech and Language*, vol. 9, pp. 19–35.

[20] Pickering, Martin & Simon Garrod, in press, Towards a mechanistic theory of dialogue: The interactive alignment model. *Behavioral & Brain Sciences*.

[21] Plauché, Madelaine & Elizabeth Shriberg, 1999. Data-Driven Subclassification of Disfluent Repetitions Based on Prosodic Features. *Proceedings of the International Congress of Phonetic Sciences,* vol. 2, pp. 1513–1516, San Francisco.