

## Towards a psycholinguistics of dialogue: defining reaction time and error rate in a dialogue corpus

Ellen Gurman BARD  
 HCRC & School of Philosophy,  
 Psychology, & Language  
 Sciences, U. of Edinburgh  
 George Square  
 Edinburgh, UK EH8 9LL  
 ellen@ling.ed.ac.uk

Matthew P. AYLETT  
 HCRC, U. of Edinburgh &  
 Rhetorical Systems, Ltd.  
 4 Crichton's Close  
 Edinburgh, UK EH8 8DT  
 matthewa@rhetorical.com

Robin J. LICKLEY  
 HCRC, U. of Edinburgh &  
 Dept of Speech Science,  
 Queen Margaret U. College  
 Clerwood Terrace  
 Edinburgh, UK EH12 8TS  
 rlickley@qmuc.ac.uk

### Abstract

This study uses the multi-level coding of a designed corpus of unscripted task-oriented dialogues to demonstrate that time to respond (Inter-Move Interval, IMI) and rate of disfluency behave like psycholinguistic measures, reaction time and error rate, in reflecting the speakers' cognitive burdens. Multiple-regression analyses show that IMI is sensitive to social distance between interlocutors, to the difficulty of the task which the dialogue serves, and to comprehension of the prior utterance and production of the current one. Rate of simple overt disfluency, in contrast, shows social and task effects, with most of the uniquely explained variance associated with planning and producing the current utterance. The results suggest that coded corpora may be useful in developing models of human interlocutors.

### Introduction

Traditionally, the study of dialogue and the study of psycholinguistic processes have been separate fields. Recently, however, a growing body of research is creating a psycholinguistics of communicative speech. Pickering and Garrod (*in press*) have claimed that interactive language use gives us much better clues to psycholinguistic processes than we can find by examining decontextualized 'laboratory speech' or text. Even in laboratory studies, there is a trend towards covert observation of natural linguistic behaviours and away from discrete non-linguistic responses. At the same time, the design of computer-assisted dialogues is often supported by ad hoc studies using exactly the sorts of off-line measures which psycholinguists are at pains

to avoid. The large samples of natural dialogue behaviour which are now available in the form of speech corpora yield rich observational data but are often thought to be irrelevant to psycholinguistics, because they do not follow balanced experimental designs. Without some experimental controls, it is difficult to tell which natural dependent variables might reflect the cognitive demands of natural language use. Using a carefully designed corpus, however, this paper shows that two classic psychological measures, reaction time and error rate, can be defined in terms which are likely to be available for a coded corpus of spoken dialogues. The results help to open the way to a naturalistic, on-line, psycholinguistics of dialogue, and allow strategic design of dialogue applications.

### 1 Motivation

There is good reason to believe that both reaction time and error rate measures will function in dialogue. In laboratory paradigms, reaction time, measured from the presentation of a stimulus to the onset of a verbal response, is used to study perception, comprehension, and problem solving (see Bock (1996) for a summary). Though the dialogue tradition is more interested in what might induce a speaker to initiate a contribution to a verbal *pas de deux*, psycholinguistic studies of monologues (Beattie, 1983; Butterworth, 1980; Wheeldon and Lahiri, 1997) reveal a natural reaction time measure: pause lengths between utterances or between larger discourse segments are related to the complexity of the unit being planned. If time taken to begin speaking in dialogue has these same characteristics, then turn-taking time is a kind of reaction time.

Error rate, usually defined as incorrect answers in forced choice tasks, seems to have a

natural counterpart in spontaneous speech: overt disfluencies, those which include genuine on-line editing of something already said<sup>1</sup>. Fluency is known to be affected by planning and production processes. Disfluencies tend to occur early in utterances, when planning of later stages is incomplete (Clark and Wasow, 1998). Disfluencies are more common in longer utterances (Clark and Wasow, 1998), in more complex constituents (Clark and Wasow, 1998; Oviatt, 1995), and when response choices are complex (Oviatt, 1995). We propose that disfluencies should be viewed as edited errors, because they indicate the speaker's failure to design an utterance portion conforming to his or her current conversational goals before beginning to speak. Since self-monitoring and self-correction appear to be possible before articulation begins (Blackmer and Mitton, 1991; Levelt, Roelofs, and Meyer, 1999; Levelt, 1983; Maclay and Osgood, 1959; Postma and Kolk, 1992), overt disfluencies comprise failures of those earlier processes of quality control.

Although these results suggest dependent variables for psycholinguistic studies of dialogue, they have not been checked for effects of many of the plausible independent variables, the possible sources of difficulty for the speaker. Disfluency, for example, has been associated with the process of planning and producing an utterance, but it might also be affected by the other cognitive demands on an interlocutor. In addition to producing their own contributions, interlocutors must also interpret one another's contributions, keep some record of the dialogue so far, and provide appropriate replies promptly enough to make it plain that they wish to take the floor. In task-oriented dialogue, at least, they must use the interaction to achieve an extra-linguistic goal. As with any other task, participants may improve over time or tire as they continue. Reaction time and error rate could be affected by any of these processes.

Most simply, the difficulty of conducting a dialogue should be increased by the complexity and the novelty of the *task* which it pursues. If so, reaction times and error rates should as usual be worse for difficult tasks and early in the course of mastering a task.

Second, there should be effects of the *prior utterance*, because preparing a reply depends on processing the interlocutor's contribution on

some level. How hard it is to do this may be critical because the time between conversational turns is so short (with a mode around 150 msec for the materials discussed below) that production processes are likely to start before listening stops entirely. Comprehension difficulty should give rise to slower responses, as it does in laboratory studies. Furthermore, if self-monitoring in speech production actually uses the same comprehension processes which we use for comprehending what is said to us, (Levelt, Roelofs, and Meyer, 1999), then difficulty in comprehending an interlocutor's utterance could interfere with early self-monitoring, increasing rates of overt disfluency. Disfluent input could be particularly damaging. Difficulty in perceiving disfluent utterances (Bard and Lickley, 1998) could delay production of replies. Cross-speaker syntactic priming (Branigan, Pickering, and Cleland, 2000) might even encourage disfluent replies to disfluent utterances.

Third, *production effects*, as outlined above, could independently yield slower responses and more disfluencies where the speaker's planning burden rises. Longer, more complex utterances should yield slower, more disfluent production, as should the need to plan discourse segments.

Finally, *interpersonal factors*, like the number of sensory channels carrying the communication or the familiarity of the interlocutors could affect delicate processes of listener-modelling, planning, and feedback (Branigan, Lickley, and McKelvie, 1999; Doherty-Sneddon, Anderson, O'Malley, Langton, et al., 1997), with better knowledge facilitating responding.

To test for these separate sources of concurrent difficulty, we exploit existing multi-level coding of a dialogue corpus and use multiple regression analysis to determine whether each set of processes makes a statistically independent contribution to reaction time and error rate.

## 2 Method

### 2.1 Materials

#### 2.1.1 Corpus.

Materials came from the HCRC Map Task Corpus (Anderson et al., 1991), 128 unscripted dialogues in which 32 pairs of Glasgow University undergraduates communicated routes defined by labeled cartoon landmarks on schematic maps of imaginary locations. Instruction Giver's (hereafter 'IG') and Follower's

<sup>1</sup> We exclude mere hesitations, filled pauses, or rapid self-corrections designed for effect or humour.

(IF) maps for any dialogue matched only in alternate landmarks. Participants knew that their maps might differ but not where or how. Interlocutors could not see each other's maps. Familiarity of participants (within subjects) and ability to see the interlocutor's face (between subjects) were counterbalanced. Each participant served as IG for the same route to two different IFs and as IF for two different routes, yielding 8 dialogues for every group of 4 co-participants. Channel-per-speaker digital stereo recordings were orthographically transcribed and word-segmented. Like the coding systems described below, the segmentations and design variables form part of an XML corpus database.

### 2.1.2 Unit of analysis.

The word-segmented corpus is framed as a series of Conversational Game Moves (Carletta., Isard, Isard, Kowtko, et al., 1997), turns or parts of turns whose purpose in moving the dialogue forward can be determined by their form and context. Moves are stages of Conversational Games, which are themselves usually stages in completing Transactions, sections of the task (here route communication) which the dialogue serves (Isard and Carletta, 1995). We examined only those Moves which were likely to be a replies to the previous speaker's Move: all were produced by a different speaker from the speaker of the prior Move; none began too early to respond to that prior Move (each included Move began less than 1 sec before the end of the prior speaker's Move and more than 350 msec after the beginning of that Move) and none merely resumed an earlier incomplete Move by the same speaker (all began at least 300 msec after the end of a previous Move by the same speaker and none were coded as continuations of the previous Move) (Bard, Aylett, and Bull, 2000; Bull, 1998; Bull and Aylett, 1998). Bull and Aylett (1998) showed that these criteria were conservative relative to human judgments about the relationship between successive Moves.

## 2.2 Dependent Variables.

### 2.2.1 Reaction: Inter-move interval (IMI).

IMI is the time in milliseconds, positive or negative, between the offset of one speaker's Move and the onset of the next starting Move if the latter is produced by the other speaker and is a 'real reply' (as defined above) to the former (Bard et al, 2000; Bull, 1998). IMI ranges from – 999 msec to over 13 sec (mean = 506 msec, *s.d.*=

795 msec; mode between 100 and 200 msec)..

### 2.2.2 Error rate: Disfluency rate

Disfluency rate is the number of simplex disfluencies per Move. Disfluency annotation (Levelt, 1983) was performed on the whole corpus using Xwaves/Entropic xlabel to display the speech waveform and spectrograms where necessary. Disfluencies involved an overt editing of uttered words or part words. Each word or part-word in a disfluent Move was assigned to one of the five possible parts of a disfluency. The optional *original utterance* was composed of pre-error words which should stand after correction. The *reparandum*, the words or partial words which constituted an error, was followed by an *interruption* with or without an optional *filler*, then in most cases by a *repair* and *continuation*.

The current paper reports on only simple disfluencies of 4 types. In *repetitions* the speaker repeats a string verbatim, with no additions or deletions: e.g. [*we're going*] *we're going left of the camera shop*. In *insertions* the speaker repeats a string and inserts a word or words within the repeated string: e.g. [*go left*] *go just left of the camera shop*. In *substitutions* a word or string is replaced by another with no major syntactic alteration: e.g. *go [left] right of the camera shop*. In *deletions*, the speaker interrupts an utterance and either restarts without repeating or directly substituting, or simply surrenders the floor to the other speaker: e.g. [*you're away f-*] *right see the wee bit that's jutting out?* Combinations of these types, complex forms of disfluency, and mere pauses, filled or otherwise, are ignored. There are 0-4 disfluencies per Move, mean = .13, *s.d.* = .40.

## 2.3 Independent (Predictor) Variables

### 2.3.1 Interpersonal

Social distance between IG and IF is reflected in *Eye-contact* condition, (the presence or absence of a flimsy barrier blocking the line of sight between interlocutors), *Familiarity* (whether the pair have just met or are friends), and *Sex of Players* (IG-sex and IF-sex). If speakers can be more efficient at framing their messages when they know more about their listeners, then familiar pairs with eye-contact should have fast, fluent responding. We include *Sex of Players* both because men are more likely to be disfluent than women (Bortfeld et al., 2001; Branigan, Lickley, and McKelvie, 1999; Shriberg, 1994)

and because there is a premium on fluency in initial meetings between different sex pairs (McLaughlin and Cody, 1982).

### 2.3.2 Task.

*Deviation score* (mismatch in cm<sup>2</sup> between the route printed on IG's map and the version drawn by IF) measured the dyad's success in communicating the printed route. *Drawing* (whether the IMI included an attempt to draw part of the route) indicated the presence of an ongoing physical task. For the disfluency equations, *IMI* was added so that disfluency rate could be assessed for effects of fast responding (see Carletta, Caley, and Isard, 1995)

### 2.3.3 Order

Two levels are marked, *Conversation* (among the 8 dialogues produced by each set of 4 the speakers) and *Ordinal Position of the Move* in the dialogue. Both could yield practice or fatigue effects.

### 2.3.3 Prior Move Comprehension.

For any Move IMI, the immediately prior Move was coded for number of referring expressions (classed as *New* introductions of landmarks *Shared* by IF and IG maps, later *Given* mentions of *Shared* items, *New Unshared* or *Given Unshared*), for each of the 4 categories of *Disfluencies*, and for *Length in Words* excluding any reparamandum. Prior moves with higher scores on any of these variables could be more difficult to process.

### 2.3.4 Current Move Planning and Production.

Planning burden is measured by Speaker's Role (because IGs usually determine the structure of the dialogue), and by the *Size of the Unit Begun* by the Move (Transaction, Game, and Game Internal Move). The production burdens are measured in terms of *Referring Expressions* and *Length in Words* (under the system used for prior Moves).

## 2.4 Procedure

For IMI, we report results for regression equations for the 64 dialogues whose video record made it possible to eliminate IMIs during which IF drew parts of the route (remaining Moves,  $N = 5543$ ). Predictor variables were first checked for collinearity. No high cross-correlations were found. To determine likely interactions, all predictor variables were entered into a single equation for the 'no-drawing' materials and removed individually, so that the

effect on the contribution of the other variables could be examined. Where the effect was significant and the result interpretable, individual interactions were coded and all the new variables were tested by backward-stepping regression ( $F$  to remove = 4.0). The resulting model was run for the other 64 dialogues ( $N = 7444$ ) and for the corpus as a whole. We report results only for variables whose sign and significance was maintained in all three equations.

For disfluency rate, we made it possible to examine the effects of the drawing activity by using all the response Moves in the 64 videotaped dialogues except for those with complex disfluencies ( $N = 6882$ ). We report results borne out for all the data ( $N = 14389$ ). To allow comparison with IMI, we used the IMI regression equation as a starting point. No meaningful interactions were found. Contributions of the sets of predictors described above were assessed by noting fall in explained variance (Multiple- $R^2$ ) as each set was omitted from the full model.

## 3. Results

The regression equations for IMI (Multiple  $R^2 = .09$ ;  $p < .0001$ ) and disfluency rate (Multiple  $R^2 = .14$ ,  $p < .0001$ ) show significant effects of psychological variables, but different patterns of effects. Results are stated as  $\beta$ -values, standardized regression coefficients, with  $p < .01$ , unless otherwise noted. They represent the contribution of a predictor variable with the effects of other predictors statistically removed. Because they are standardized,  $\beta$ s may be directly compared. Tables show raw means.

### 3.1 IMI

In line with decades of commentary on turn-taking, IMI is significantly affected by *interpersonal* factors: as Table 1 shows, the shortest IMIs are in dialogues between different sex pairs who have just met (Familiarity, .08; Familiarity by sex-difference, -.05).

Table 1. Interpersonal effects on IMI: interlocutor familiarity and sex match

Sex match	Familiarity	
	Familiar	Unfamiliar
Same	384	435
Different	440	279

Table 2. Order effect on IMI: position of Move in dialogue

Quartile	Ordinal Position	Mean IMI

1	1-50	447
2	51-100	414
3	101-150	378
	151-200	357
4	201-250	274
	251-300	346
	301-350	334
	351-400	251
	401-450	312
	451-586	364

In line with our hypothesis that IMI is a reaction time measure, all the cognitive variable types have effects. There is a typical *order* effect: IMI falls later in dialogues (-.04) (Table 2). There is also a speed x communication-accuracy trade-off: though IMIs are longest in dialogues where the IF draws the least accurate routes (.10), they are longer in dialogues with the most accurate routes than in those of intermediate accuracy (Table 3).

Table 3. Task effect on IMI: deviation score

Deviation score quartile	1	2	3	4
Mean IMI	411	289	370	463

There is also an effect of *comprehension*. IMIs are shorter after references to landmarks already mentioned in the dialogue (for those on only one map, -.10, for those on both, -.04). They are longer after Moves introducing New landmarks only if the New landmark is not on the responder's map and no reference is made to previously mentioned shared items (-.09) in the introducing Move. In other words, this is a long reaction time during which the speaker conducts an extensive unsuccessful search for a landmark which s/he lacks (Table 4a and 4b).

Finally there are effects of *production* and planning in dialogue much like those known in monologue. IMIs are longer before longer utterances (.06) (Table 5) and those which begin larger units of dialogue: Transaction-initial moves follow longer intervals than Game-initial (.03,  $p < .05$ ), which in turn follow longer

intervals than Game-internal Moves (.14) (Table 6).

Table 4. Effects of comprehension on Mean IMI: (a) referring expressions in prior Move by status of first referring expression; (b) interaction of references to New Unshared and Given Unshared map landmarks

(a)	Occurrence of referent on players' maps	
	Shared	Unshared
Status in dialogue		
Given	227	217
New	277	514

(b)	References to Unshared Given	
	No	Yes
References to Unshared New		
No	378	350
Yes	557	398

Table 5. Production effects on IMI: current Move length in words (excluding reparanda)

Length (words)	% cases	Mean IMI
1	70	319
2	5	338
3	4	347
5-6	4	410
7-8	4	393
9-12	4	478
13-52	5	505

Table 6. Production effect on IMI: size of unit initiated by current Move

Unit initiated	Mean IMI
Game-internal Move	304
Game	535
Transaction	617

### 3.2 Disfluency.

Disfluency also responds to putative sources of cognitive difficulty, but the effects differ considerably from the pattern we found for IMI. Among *interpersonal* effects, the sex effect appears in higher rates of disfluency for male IGs (.04). Since the same men take the IF role in other dialogues, this effect amounts to a sex x role interaction, with males more disfluent only when running dialogues. In contrast to IMI, disfluency displays a simple Familiarity effect which does not depend on the sexes of the dyad: speech to unfamiliar interlocutors is more disfluent overall than speech to familiar (.174 vs .140 disfluencies/Move;  $-.03, p < .05$ ).

The *task* effects now are restricted to a local order effect: though IMI falls across a dialogue, disfluency rates rise (.03,  $p < .05$ ). (Table 7). The increase does not appear to be associated with any other measures of task difficulty.

Table 7. Order effect on disfluency rate: position of Move in dialogue

Quartile	Ordinal Position	Disfluencies / Move
1	1-55	.147
2	56-110	.135
3	111-185	.169
4	> 185	.167

In contrast to IMI, also, disfluency rates are insensitive to *comprehension* variables.

Table 8. Production effect on IMI: size of unit initiated by current Move

Unit initiated	Disfluencies / Move
Game-internal Move	.116
Game	.219
Transaction	.294

Disfluencies do show the predicted robust effects of *production and planning*. IGs, who usually take responsibility for directing the dialogue, are more disfluent than IFs (.181 v .087:  $\beta = -.07$ ) Just as there were longer IMIs at the onsets of larger units of dialogue, there are more disfluencies (.02,  $p < .05$ ) (Table 8). Disfluencies are affected by the complexity of the current Move: longer Moves attract more disfluencies (.28) (Table 9), as do Moves including more referring expressions (.06 for

references to Given Shared landmarks, .03,  $p < .05$ , for New Unshared and .03 for Given Unshared) (Table 10).

Table 9. Production effects on disfluency rate: current Move length in words (excluding reparaanda)

Length (words)	Disfluencies / Move
<1	.063
2-5	.099
6-9	.229
10-13	.378
14-17	.495
>17	.709

Table 10. Production effects on disfluency rate: number of referring expressions in current Move (excluding reparaanda)

Referring expressions	Disfluencies / Move
0	.09
1	.22
2-5	.47

## 4. Discussion

These results show that objectively codable characteristics of a dialogue corpus reliably reflect psychological processes. For both IMI and disfluency, multiple regression results confirm the findings of studies exploring single cause-effect relationships. The current work goes beyond earlier studies in testing a range of potential sources of psycholinguistic difficulty against each dependent variable. As Figure 1 shows, the proportions of explained variance attributable to each group of predictors pattern quite differently in the two cases. IMI looks like a general measure of cognitive difficulty with independent contributions made by order, task difficulty, comprehension, production, and interpersonal factors. Disfluency, on the other hand, shows largely the effects of planning for production, with the greater part of the uniquely explained variance attributable to characteristics of the disfluent current Move itself.

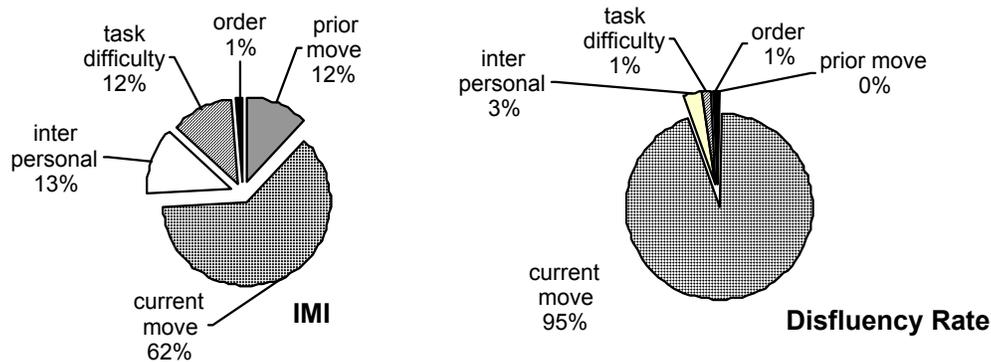


Figure 1. Percentages of uniquely explained variance in IMI and disfluency rate associated with sets of predictor variables

These outcomes have theoretical, methodological and technological implications. Models of speech production, for example, will have to consider why disfluency rate is insensitive to attested difficulties of comprehension: features of a prior utterance which ought to affect comprehension, and do adversely affect IMI, do not increase the rate of disfluency. Nor does it seem that the IMI absorbs the effects of comprehension difficulty, leaving processors clear for language production. If it did, then shorter IMIs, some of which might mark premature onset of the current Move, should make for more disfluency. But IMI is not a significant predictor of variance in disfluency rate and its removal from the equation does not enhance the apparent effects of prior-Move difficulty. Overt self-monitoring appears to be unrelated to factors that are supposed to be related to covert self-monitoring. Either the two kinds of monitoring are unrelated, with only covert monitoring processes interacting with comprehension, or the claim that comprehension processes participate in speech production is overstated.

Instead, the results confirm Clark and Wasow's finding for the effects of sentence length and add pragmatic, discourse and role effects: even when Move length in words is statistically controlled, disfluencies reflect the number of referring expressions being generated, the size of the discourse unit being initiated, or the degree of responsibility which the speaker bears for structuring the dialogue. It is tempting to conclude that all of these facets of speech production belong in a complete psychological

model of the process.

It is also of interest that IMI is affected by prior-Move discourse position, length, and referential character, but not by prior move disfluency. Whatever difficulty listeners may have in recognizing words in reparanda (Bard and Lickley, 1998) or whatever advantage they may accrue in interpreting signalled disfluency (Brennan and Schober, 2001), none of this seems to affect the time taken to produce a reply. An explanation awaits further study of the comprehension of real speech in real communicative contexts.

The current techniques should help make such studies more viable. Our multiple regression results confirm several studies examining single cause-effect relationships. If basic results are evident whether or not other sources of variance are partialled out, then large corpora with simpler coding systems might still have something to offer psycholinguists in search of naturalistic on-line data. At the same time, designers of human-computer dialogues can use coded corpora to look for direct effects on delay to response or fluency of response with reasonable confidence that their observations can profit from our wider understanding of how interlocutors work.

### Acknowledgements

The authors acknowledge the support of the Economic and Social Research Council, U.K., via HCRC and of the Engineering and Physical Sciences Research Council, U.K. (Speech and Language Technologies grant GR/L50280).

## References

- Anderson A., Bader M., Bard E. G., Boyle E., et al., (1991) *The H.C.R.C. Map Task Corpus*. Lang. & Speech, 34, pp. 351-366.
- Beattie G. (1983) *Talk*. Open University Press, Milton Keynes.
- Bard E. G., Aylett M., and Bull M. (2000) *More than a Stately Dance: Dialogue as a Reaction Time Experiment*. Proc. Soc. for Text & Discourse.
- Bard E. G. and Lickley R. J. (1998) *Graceful failure in the recognition of running speech*. Proc Cognitive Science Soc., pp. 108-113.
- Blackmer E. and Mitton J. (1991) *Theories of monitoring and the timing of repairs in spontaneous speech*. Cognition, 39, pp. 173-194.
- Bock K. (1996) *Language production: Methods and methodologies*, Psychon Bull & Rev., 3, pp. 395-421.
- Bortfeld H., Leon S., Bloom J., Schober M., and Brennan S. (2001) *Disfluency rates in conversation: effects of age, relationship, topic, role and gender*. Lang. & Speech, 44, pp. 123-147.
- Branigan H., Pickering M., and Cleland, A. (2000) *Syntactic coordination in dialogue*. Cognition, 75, pp. 813-825.
- Branigan H., Lickley R., and McKelvie D. (1999) *Non-linguistic influences on rates of disfluency in spontaneous speech*. Proc. 14th ICPHS.
- Brennan S., and Schober M. (2001) *How listeners compensate for disfluencies in spontaneous speech*. J. Mem. & Lang., 44, pp. 274-296.
- Bull M. (1998) *Inter-turn intervals in spontaneous speech*. PhD dissertation, U. of Edinburgh, U.K.
- Bull M. and Aylett M. (1998) *An analysis of the timing of turn-taking in a corpus of goal-oriented dialogue*. Proc. ICSLP'98.
- Butterworth B. (1980) "Language production: Vol 1., Speech and Talk". Academic Press.
- Carletta J., Caley R., and Isard S. (1995) *Simulating time-constrained language production*. Lang. & Cognitive Proc., 10, pp. 357-361.
- Carletta J., Isard A., Isard S., Kowtko J., et al. (1997) *The reliability of a dialogue structure coding scheme*. Computational Linguistics, 23, pp. 13-31.
- Clark H. H., and Wasow T. (1998) *Repeating words in spontaneous speech*. Cognitive Psych., 37, pp. 201-242.
- Doherty-Sneddon G., Anderson A. H., O'Malley C., Langton S., et al., (1997) *Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance*. J Exp. Psych.: Applied, 3, pp. 105-125.
- Isard A., and Carletta J. (1995) *Transaction and Action Coding in the Map Task Corpus*. Research paper HCRC/RP-65. HCRC, U. of Edinburgh, U.K.
- Levelt W.J.M., Roelofs A., and Meyer, A. S. (1999) *A theory of lexical access in speech production*. Brain & Behav.Sci., 22, pp. 1-45.
- Levelt W.J.M. (1983) *Monitoring and self-repair in speech*. Cognition, 14, pp. 14-104.
- Lickley R. J. (1994) *Detecting disfluency in spontaneous speech*. PhD Thesis, U. of Edinburgh, U.K.
- Lickley R.J. (1998) *HCRC Disfluency Coding Manual*. Technical Report HCRC/TR-100, HCRC, U. of Edinburgh, U.K.
- Lickley R. (2001) *Dialogue moves and disfluency rates*. Proc. of Disfluency in Spontaneous Speech '01, Edinburgh, U.K.
- McLaughlin M., and Cody M. (1982) *Awkward silences: Behavioral antecedents and consequences of the conversational lapse*. Human Communication Res, 8, pp. 299-316.
- Maclay H. and Osgood S. (1959) *Hesitation phenomena in spontaneous English speech*. Word, 15, pp. 19-44.
- Oviatt S. (1995) *Predicting disfluencies during human-computer interaction*. Computer Speech & Lang., 9, pp. 19-35.
- Pickering M. and Garrod S. (in press) *Toward a mechanistic psychology of dialogue*. Brain & Behav. Sci.
- Postma A. and Kolk H. (1992) *The effects of noise masking and required accuracy on speech errors, disfluencies, and self-repairs*. J. Speech & Hearing Res., 35, pp. 537-544.
- Shriberg E. (1994) *Preliminaries to a theory of speech disfluencies*. PhD Thesis, U. of California at Berkeley.
- Wheeldon L., and Lahiri A. (1997) *Prosodic units in speech production*. J. Mem. & Lang, 37, pp. 356-81.