A: An Experimental Investigation into...
B: ...Split Utterances

Christine Howes, Patrick G.T. Healey and Gregory J. Mills
Queen Mary University of London
Interaction, Media and Communication Research Group, London, E1 4NS
{chrizba, ph, gj}@dcs.qmul.ac.uk

Abstract

A distinguishing feature of dialogue is that more that one person can contribute to the production of an utterance. However, until recently these 'split' utterances have received relatively little attention in models of dialogue processing or of dialogue structure. Here we report an experiment that tests the effects of artificially introduced speaker switches on groups of people engaged in a task-oriented dialogue. The results show that splits have reliable effects on response time and on the number of edits involved in formulating subsequent turns. In particular we show that if the second half of an utterance is 'misattributed' people take longer to respond to it. We also show that responses to utterances that are split across speakers involve fewer deletes. We argue that these effects provide evidence that: a) speaker switches affect processing where they interfere with expectations about who will speak next and b) that the pragmatic effect of a split is to suggest to other participants the formation of a coalition or sub-'party'.

1 Introduction

Split utterances, defined simply as utterances which are split between speakers[1], are known to occur in dialogue, as evidenced by Conversational Analysis (CA) studies, based on the analysis of naturally occuring dialogues. In addition to numerous analyses of split utterances in generic English dialogues, there are cross linguistic studies, and observations of conversations with aphasics. In Finnish, split utterances within a single clause conform to the strict syntactic constraints of the language (which has a rich inflectional morphology), despite the change in speaker (Helasvuo, 2004). Similarly, in Japanese, a verb-final language, speakers also engage in "co-participant completions" (Hayashi, 1999; Lerner and Takagi, 1999). There is also evidence of split utterances in conversations with aphasics (Oelschlaeger and Damico, 1998), demonstrating that the phenomenon is pervasive in dialogue. However, with the possible exception of Szczepek (2000) who analysed some 200 splits from 40 hours of recorded English conversation, these studies tend to be unconcerned with frequencies of occurrence; that split utterances occur at all renders them worthy of study.

Split utterances are a clear and canonical example of coordination in dialogue. In order for one person to continue an utterance which has been begun by another person requires the hearer to have coordinated with the initial speaker up to the point at which they take over the role of producer[2].

Analysis of split utterances, when they can or cannot occur and what effects they have on the coordination of agents in dialogue, is therefore an area of interest not only for conversational analysts wishing to characterise sytematic interactions in dialogue, but also linguists trying to formulate grammars of dialogue, and psychologists interested in alignment mechanisms in dialogue.

[1] What we call split utterances have been variously referred to as *collaborative turn sequences* (Lerner, 1996; Lerner, 2004), *collaborative completions* (Clark, 1996) *co-constructions* (Helasvuo, 2004), *co-participant completions* (Hayashi, 1999; Lerner and Takagi, 1999) *collaborative productions* (Szczepek, 2000) and *anticipatory completions* (Fox and others, 2007) amongst others.

[2] Note that this says nothing about whether such a continuation is the same as the initial speakers intended continuation.

Proceedings of SIGDIAL 2009: the 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue, pages 79–86,
Queen Mary University of London, September 2009.

In this regard, studies of split utterances, in both spontaneous dialogues and experimentally, as below, provide a complementary way of studying structural alignment to the traditional experimental set up exemplified by Branigan and colleagues (Branigan et al., 2000; Branigan et al., 2003; Branigan et al., 2006). Indeed, Poesio and Rieser (In preparation) claim that "[c]ollaborative completions . . . are among the strongest evidence yet for the argument that dialogue requires *coordination* even at the sub-sentential level" (italics original).

Broadly speaking, there have been two types, or levels, of explanations of split utterances offered; pragmatic accounts and processing accounts. Pragmatic accounts are favoured by Conversational Analysts, with various aspects of split utterances analysed. However, in line with CA assumptions, these analyses are almost exclusively concerned with the conditions under which split utterances can occur. Lerner (1991), for example, identifies a number of 'compound' turn-constructional units, such as the IF-THEN construction (whereby the second participant is in some sense licensed to provide the THEN part of the structure). However, Lerner's insistence on identifying the circumstances in which split utterances usually occur misses the important generalisation that, syntactically, they can be anywhere in a string (his *opportunistic completions*). His claim that an *anticipatory completion* is ordinarily "designed as a syntactic continuation of the utterance part it follows at the point of onset", seems to hold for all split utterances.

The occurrence of split utterances also has implications for the organisation of turn-taking, as outlined in Sacks et al. (1974). According to Schegloff (1995), turn-taking operates, not on individual conversational participants, but on 'parties'. For example, if a couple are talking to a third person, they may organise their turns as if they are one 'party', rather than two separate individuals. Lerner (1991) suggests that split utterances can clarify the formation of such parties; "collaboratively produced sentences reveal a relationship between syntax and social organisation. It provides evidence of how syntax can be mobilised to organise participants into "groups"."

The processing approach towards split utterances is exemplified by the interactive alignment model of Pickering and Garrod (2004). They claim that;

> . . . it should be more-or-less as easy to complete someone else's sentence as one's own, and this does appear to be the case.
> (Pickering and Garrod, 2004, p186)

According to this model, speaker and listener ought to be interchangeable at any point, and this is also the stance taken by the grammatical framework of Dynamic Syntax (Cann et al., 2005). In Dynamic Syntax (DS), parsing and production are taken to use exactly the same mechanisms, leading to a prediction that split utterances ought to be strikingly natural (Purver et al., 2006). Additionally, for a third person to process an utterance that appears to come from two separate speakers ought not be more difficult than processing the same utterance from a single speaker, regardless of where in a string the changeover occurs.

According to Poesio and Rieser (In preparation), "the study of sentence completions can shed light on a number of central issues. . . this type of data may be used to compare competing claims about coordination – i.e. whether it is best explained with an intentional model like Clark's. . . or with a model based on simpler alignment models like Pickering and Garrod's." As they see intentions as crucial to dialogue management, they conclude that a model which accounts for intentions (such as their PTT account) better captures their task specific split utterance data (See Poncin and Rieser (2006) for details of the German data they are modelling).

If this is the case, it ought to be more difficult to process an utterance that appears to be split between speakers, as opposed to one that comes from one source, because the intentions of the two different agents have to be considered in arriving at an interpretation, and they may appear to have formed a 'party' with respect to the subject of the utterance. Additionally it ought to be more disruptive to the conversation if the utterance is attributed to someone other than the person who genuinely contributed it, because the hearer would falsely attribute intentions to the wrong interlocutor. This ought to be especially clear in cases where the 'conversational momentum' appears to be with the 'wrong' interlocutor. Contrarily, if a processing model such as the interactive alignment model is correct, then no such differences should

be observed[3].

To test these predictions, an experiment was set up to alter genuine single-turn utterances into split utterances at an arbitrary point in the string. Different types of intervention were introduced, in a 2 x 2 factorial design, in order to separate out the effects of an utterance appearing to come from two different participants from effects caused by an apparent change of floor.

2 Method

The effects of seeing an utterance split between speakers or not were tested using the Dialogue Experimentation Toolkit (DiET) chat tool, as described in Healey et al. (2003), which enables dialogues to be experimentally manipulated.

The DiET chat tool allows interventions to be introduced into a dialogue in real time, thus causing a minimum of disruption to the natural 'flow' of the conversation. In this case, a number of genuine turns in a three way conversation were artificially split into two sections, with both parts either appearing to originate from the genuine source, or one or both parts being falsely attributed to another participant.

2.1 Materials

2.1.1 The Balloon Task

The *balloon task* is an ethical dilemma requiring agreement on which of three passengers should be thrown out of a hot air balloon that will crash, killing all the passengers, if one is not sacrificed. The choice is between a scientist, who believes he is on the brink of discovering a cure for cancer, a 7 months pregnant woman, and her husband, the pilot. This task was chosen on the basis that it should stimulate discussion, leading to dialogues of a sufficient length to enable an adequate number of interventions.

2.1.2 The DiET Chat Tool

The DiET chat tool itself is a custom built java application consisting of two main components, which will be outlined in turn; the user interface, and the server console.

[3]This is, of course, an oversimplification, and note that in contrast to pragmatic accounts, no claims are made regarding higher level discourse effects of the split utterance, as the focus is on the mechanisms which allow split utterances to occur. Additional mechanisms could of course be posited in processing models to account for any such differences.

2.1.3 User interface

The user interface is designed to look and feel like instant messaging applications e.g. Microsoft Messenger. It consists of a display split into two windows, with a status bar, indicating whether any other participant(s) are actively typing, between them (see figure 1). The ongoing dialogue, consisting of both the nickname of the contributor and their transmitted text, is shown in the upper window. In the lower window, participants type and revise their contributions, before sending them to their co-participants. All key presses are time-stamped and stored by the server.

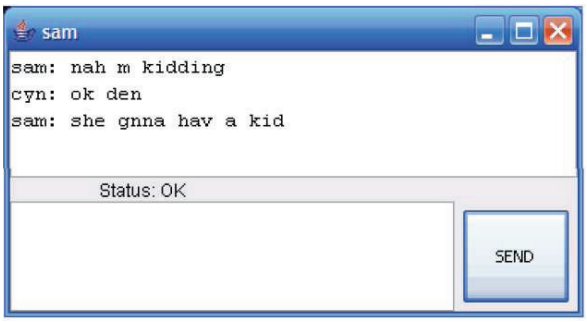

Figure 1: The user interface chat window (as viewed by participant 'sam')

2.1.4 Server Console

All text entered is passed to the server, from where it is relayed to the other participants, not relayed directly between participants. Prior to being relayed, some turns are altered by the server to create fake split utterances.

This is carried out automatically such that a genuine single-person turn is split around a space character near the centre of the string. The part of the turn before the space is relayed first, followed by a short delay during which no other turns may be sent. This is followed by the part of the turn after the space, as if they were in fact two quite separate, consecutive turns. In every case, the server produces two variants of the split utterance, relaying different information to both recipients. Each time an intervention is triggered, one of the two recipients receives both parts from the *actual* source of the utterance (henceforth referred to as an *AA*-split). The other recipient receives one of three, more substantial, manipulations; the first half could appear to be from the actual origin with the second part of the split appearing to originate from the other recipient (an *AB*-split), or

the inverse could be the case (a *BA*-split), or both parts could be wrongly attributed to the other participant (a *BB*-split). This design was in order to separate the effects of a change in conversational momentum (floor change) from the effects of splitting per se, hence the inclusion of the BB condition where who apparently has the floor is altered without the utterance being attributable to different participants. This contrast is shown in table 1.

Table 1: Comparison of split types

A types:		
Should we start now		
B sees (AA intervention):		
A: Should we A: start now		
C sees (one of):		
AB intervention:	BA intervention:	BB intervention:
A: Should we B: start now	B: Should we A: start now	B: Should we B: start now

The intervention is triggered every 10 turns, and restricted such that the participant who receives the non AA-split is rotated (to ensure that each participant only sees any of the more substantially manipulated interventions every 30 turns). Which of the three non AA-splits they see (AB, BA or BB) is, however, generated randomly.

2.2 Subjects

41 male and 19 female native English speaking undergraduate students were recruited for the experiment, in groups of three to ensure that they were familiar with each other. All had previous experience of internet chat software such as Microsoft Messenger and each was paid £7.00 for their participation.

2.3 Procedure

Each of the triad of subjects was sat in front of a desktop computer in separate rooms, so that they were unable to see or hear each other. Subjects were asked to follow the on screen instructions, and input their e-mail address and their username (the nickname that would identify their contributions in the chat window). When they had entered these, a blank chat window appeared, and they were given a sheet of paper with the task description on. Participants were instructed to read this carefully, and begin discussing the task with their colleagues via the chat window once they had done so. They were told that the experiment was investigating the differences in communication when conducted using a text only interface as opposed to face-to-face. Additionally, subjects were informed that the experiment would last approximately 20-30 minutes, and that all turns would be recorded anonymously for later analysis. Once all three participants had been logged on, the experimenter went to sit at the server machine, a fourth desktop PC out of sight of all three subjects, and made no further contact with them until at least 20 minutes of dialogue had been carried out.

3 Results

A post experimental questionnaire and debriefing showed that participants felt the conversations went as smoothly as face-to-face dialogue. With the exception of one subject, who had taken part in a previous chat tool experiment and was therefore aware that interventions may occur, none of the participants reported awareness of any interventions.

As production and receipt of turns sometimes occurs in overlap in text chat, it is not possible to say definitively when one turn is made in direct response to another[4]. We therefore chose two separate measures; *next turn* – the first turn, by the first recipient to start and complete a response, after receipt of the intervention, and *global* – all the turns produced by both recipients between the most recent intervention and the next intervention, averaged to produce one data point per recipient per intervention. This means that in the next turn condition, only one datapoint is analysed for each intervention, despite two different people seeing an intervention (and both usually producing a response). This was to try and isolate the initial response to an intervention; for the other person who saw a split but did not respond first, it is not clear if they are responding to the split utterance, or to

[4] In online chat, participants can compose their next turns simultaneously, and turns under construction when another is received can be subsequently revised, prior to transmission. This means that a genuine response to a split utterance might have a negative start time. However, the inclusion of cases where the whole turn was constructed after receiving the split (an arbitrary cut-off point, which would catch some turns that were responses to earlier turns in the dialogue, and miss some which were begun before the intervention was received and subsequently revised) should impose the same level of noise in all cases.

the person who *already* responded to the split utterance. In the global condition, in contrast, there are two datapoints for each intervention (one for each of the participants who saw a split utterance).

Of the 253 interventions to which at least one recipient responded, 89 were AA/AB splits, 99 were AA/BA splits and 65 AA/BB splits. Table 2 shows the n values in each case.

Both next turn and global measures were analysed according to two factors in a 2 x 2 factorial design; *split* – whether both parts of the utterance had appeared to come from the same person, or from different sources ([AA and BB] vs [AB and BA]), and *floor change* – who appeared to have produced the second part of the split, the genuine source, or the other participant ([AA and BA] vs [AB and BB]).

Measures selected for analysis were *typing time of turn* (The time, in milliseconds, between the first key press in a turn and sending the turn to the other participants by hitting the return key) and *length of turn in characters* as measures of production; *deletes per character* (The number of keyed deletes plus one (to prevent null values) divided by the total number of characters) as a measure of revisions; and *typing time per character* as a measure of speed. Data in tables are displayed in the original scale of measurement. However, as inspection of the data showed that they were not normally distributed, logarithmic transformations (using log_e) were applied to the data prior to all formal analyses.

2 x 2 ANOVAs show a main effect of floor change on the typing time of turn (see table 2). This holds for next turns ($F_{(3,249)} = 7.13$, $p < 0.05$) and globally ($F_{(3,486)} = 3.78$, $p < 0.05$), with participants taking longer over their turns in the AB and BB conditions. There was no main effect of split, and no effect of interaction. This effect is greater locally than globally, with participants who respond first after seeing a floor change condition taking more than 40% longer over their turns than those who saw a non-floor change condition. Globally the difference is in the order of 10%.

There was a main effect of split on the number of *deletes per character* , which also held both in the next turn condition ($F_{(3,249)} = 6.26, p < 0.05$) and globally ($F_{(3,486)} = 9.23$, $p < 0.05$), with subjects seeing a split condition (AB or BA) using *fewer* deletes per character than those seeing a non-split condition (see table 3). There was no main effect of floor change or interaction effect. This effect is also stronger in the next turn condition, with those not seeing a cross-person split using over 50% more deletes. In the global condition, this difference is still 40%, though the overall proportion of deletes is approximately 25% lower, from 0.334 per character in the next turn condition to 0.244 globally.

Table 2: Typing time of turn by type of intervention

Condition		Mean (s.d.)		N (poss N)	
Next Turn	AA	9475.54	(12258.5)	136	(253)
	AB	14560.70	(18863.9)	37	(89)
	BA	6968.24	(6437.0)	51	(99)
	BB	14812.59	(20367.8)	29	(65)
Global	AA	11122.27	(14413.5)	246	(253)
	AB	12500.98	(10944.6)	89	(89)
	BA	9800.77	(8810.3)	92	(99)
	BB	11561.67	(10138.4)	63	(65)

Table 3: Deletes per character by type of intervention

Condition		Mean (s.d.)	
Next Turn	AA	0.435	(1.63)
	AB	0.152	(0.30)
	BA	0.202	(0.25)
	BB	0.324	(0.61)
Global	AA	0.288	(0.83)
	AB	0.192	(0.28)
	BA	0.145	(0.18)
	BB	0.287	(0.37)

Additional analyses showed an effect of floor change on *length of turn in characters* (table 4) in the next turn condition ($F_{(3,249)} = 5.57$, $p < 0.05$) such that turns are longer in the AB and BB conditions (note that though this might be thought to be confounded by the typing time of turn, as you would expect longer turns to take longer to type, there are no significant effects when ANOVAs are performed on *typing time per character*). There is no main effect of split, or interaction effect. In the global condition, however, there is a main effect of split ($F_{(3,486)} = 4.08, p < 0.05$) such that turns are longer after seeing an utterance that appears to be split between two different people (AB and BA conditions). There is no main effect of floor change, and no effect of interaction.

As the experiment was looking for generic effects of splitting on coordination, the location of the splits was random. A post-hoc analysis was therefore carried out to ascertain whether the standalone coherence (as judged by the authors) of the two separate parts of the utterance was a possible confounding factor. Examples of coherence judgements are shown in table 5.

Table 4: Length of turn in characters by type of intervention

Condition		Mean (s.d.)	
Next Turn	AA	23.95	(22.0)
	AB	37.76	(34.9)
	BA	23.92	(18.4)
	BB	26.52	(21.5)
Global	AA	26.41	(20.4)
	AB	32.12	(23.9)
	BA	28.27	(18.4)
	BB	25.78	(13.6)

Table 5: Examples of standalone coherence judgement examples

Part of Split		Coherent	
First	Second	1st	2nd
what the hell	is that	Y	N
the woman is pregnant	she should stay	Y	Y
these people said	you did something	N	Y
I think this is also	the wish of the doctor	N	N

2 x 2 ANOVAs showed that in the next turn condition, there are no main effects of first or second part coherence, but there was an interaction effect of first part coherence by second part coherence on deletes ($F_{(3,249)} = 4.05$, $p < 0.05$), such that if *both* parts are independently coherent, or if *neither* part is independently coherent, there are fewer deletes used in the turn immediately following the intervention (see table 6). There are no significant global effects.

Table 6: Deletes per character by first and second part standalone coherence (next turn condition)

Coherence		Mean (s.d.)	
1st	2nd		
Y	Y	0.198	(0.38)
	N	0.651	(2.26)
N	Y	0.304	(0.66)
	N	0.206	(0.30)

Running a 2 x 2 x 2 x 2 ANOVA with these additional factors does not alter the main effects observed for floor change or split, as detailed above. There are no additional interaction effects on any of the measures.

4 Discussion

As this is the first experimental study into split utterances using the DiET chat tool, what follows is necessarily exploratory. This discussion presents our current hypotheses as to how best to interpret the data, as summarised in table 7, below.

Table 7: Summary of significant effects

Effect of	Condition	on and direction
Floor Change	Next Turn and Global	Typing Time $(AB \wedge BB) > (AA \wedge BA)$
Floor Change	Next Turn	Number of Chars $(AB \wedge BB) > (AA \wedge BA)$
Split	Next Turn and Global	Deletes $(AA \wedge BB) > (AB \wedge BA)$
Split	Global	Number of Chars $(AB \wedge BA) > (AA \wedge BB)$

Taking longer over the production of a turn (independently of typing speed) indicates a lack of confidence in the conversation (misattributing the second part of the utterance thus reducing confidence), and is also indicative of local organisation of turn-taking. If a participant who has seen a floor change intervention (Participant C) responds first, then they may be taking longer over their turns because there is less pressure on them to take a turn. This is because of the C's expectations. They will falsely believe that the fake source (Participant B) has just completed a turn, and will therefore not expect them to take the floor, and the genuine source (Participant A) will not be taking the floor because they have just completed a turn (though C does not know this). It is probable that in the turn immediately following a floor change intervention both these factors are at play, whereas globally it is the weaker effect of generic confidence loss that is observed. This compounding of effects in the next turn condition would also help explain the divergent effects on the length of turn in characters in next turn and global conditions.

Regardless of the precise reasons for it, this effect of floor change on typing time clearly demonstrates that changing the apparent speaker is disruptive, perhaps because it alters the forward mo-

mentum of the conversation.

More interestingly, independently of a change of floor, seeing an utterance that appears to be split between speakers also has an impact on the conversation, seen in the amount of revision undertaken in formulating a response (deletes). One reason why participants might worry less about precisely formulating their turns following a cross-person split is that the production of a cross-person split could have the effect on the recipient of suggesting that the two other participants have formed a 'party' (Schegloff, 1995) with respect to the decision of who to throw out of the balloon. This might be understood as signalling the formation of a strong coalition between the other two participants, therefore making the recipient behave as though they are resigned to the decision of this coalition. This is not the same as the effect on the typing time of turn, whereby participants are less rushed when seeing a change of floor. Deletes, on the other hand, demonstrate how carefully participants are constructing their turns. Excerpt 1, taken from the transcripts shows an example where this appears to be the case.

Excerpt 1 AB-Split showing apparent coalition between 'Bhups' and 'Dan' ('fake' part of split shown in bold)

Bhups:	and he can tell his formula
Dan:	**to tom and susie**

If we take split utterances as an indicator of coordination then it is likely that if we believe our two conversational partners to be in coordination, we will worry less about precisely formulating our own contributions. This also backs up the idea that people are not interchangeable.

The interaction of first and second part coherence also underlines the effect of split on revisions as outlined above. In the case were both parts of the split could potentially stand as independent utterances, they are treated as such and the number of deletes per character is in line with the global average (i.e. they are treated as normal dialogue). In the other non ambiguous case, where neither part could be interpreted as an utterance on its own, there are also fewer deletes, in line with the result that there are fewer deletes in strong split cases. Interestingly, the most disruptive case is that where the first part could have been a standalone utterance, but the second part could not. This could be seen as analogous to a garden path effect, and provides some indication that that the building up of interpretations is incremental, and not concerned with who supplies the input.

These results do not, of course, prejudice the claim that, at a purely mechanistic level, people could anticipate the structures needed to complete a turn, as the interactive alignment model suggests, because they are not concerned with the actual production of a split utterance, rather on the effect it has on the conversation. They do indicate that in terms of the effects of seeing split utterances, the pragmatic approach offers a more feasible level of analysis. For example, if we wish to treat a jointly produced split utterance as signalling especially strong alignment, then we need to account for more than simply syntax.

There is an issue with the design of the experiment which means that the floor change effects might be caused by a confounding variable; in essence, because one of the recipients always received an AA-split, in the cases which have been labelled as cases of floor change, the two recipients will have been left with the impression that a different person made the final contribution. This means that there may well be a an effect of confounded listener expectation (though see Schober and Brennan (2003) for discussion), although it should be noted that this does not have any bearing on the observed differences after an utterance split between speakers. It is also possible that split utterances might be particularly marked in a chat environment, though preliminary results of a corpus study show that, perhaps surprisingly, split utterances also occur naturally and as frequently in text-based chat (Eshghi, in prep) as they do in face-to-face dialogue (Purver et al., 2009). Because of these issues, and the already noted potential problems of linearity in text-based chat, a follow-up study using a character-by-character chat tool interface is underway. This more directly enforces turn-taking, as it does not allow participants to formulate their turn before communicating it; each character is transmitted as and when it is entered.

5 Conclusions

The experiment reported here offers clues towards an understanding of split utterances as an example of dialogue phenomena, and provides evidence

that speaker switches affect processing where they interfere with expectations about who will speak next and that the pragmatic effect of a split is to suggest to other participants the formation of a coalition or sub-'party'. It also clearly demonstrates that this type of experiment provides a fruitful line of future research in the ongoing attempt to adequately characterise dialogue, though further developments are needed.

References

H. Branigan, M. Pickering, and A. Cleland. 2000. Syntactic co-ordination in dialogue. *Cognition*, 75(2):13–25.

H. Branigan, M. Pickering, J. Pearson, J. McLean, and C. Nass. 2003. Syntactic alignment between computers and people: The role of belief about mental states. In *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society*.

H. Branigan, M. Pickering, J. McLean, and A. Stewart. 2006. The role of local and global syntactic structure in language production: Evidence from syntactic priming. *Language and cognitive processes*, 21(7-8):974–1010.

R. Cann, R. Kempson, and L. Marten. 2005. *The Dynamics of Language*. Elsevier, Oxford.

H. Clark. 1996. *Using Language*. Cambridge University Press.

A. Eshghi. in prep. *Uncommon ground: the distribution of dialogue contexts*. Ph.D. thesis, Department of Computer Science, Queen Mary University of London.

A. Fox et al. 2007. Principles shaping grammatical practices: an exploration. *Discourse Studies*, 9(3):299.

M. Hayashi. 1999. Where Grammar and Interaction Meet: A Study of Co-Participant Completion in Japanese Conversation. *Human Studies*, 22(2):475–499.

P. G. T. Healey, M. Purver, J. King, J. Ginzburg, and G. J. Mills. 2003. Experimenting with clarification in dialogue. In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*.

M. Helasvuo. 2004. Shared syntax: the grammar of co-constructions. *Journal of Pragmatics*, 36(8):1315–1336.

G. Lerner and T. Takagi. 1999. On the place of linguistic resources in the organization of talk-in-interaction: A co-investigation of English and Japanese grammatical practices. *Journal of Pragmatics*, 31(1):49–75.

G. Lerner. 1991. On the syntax of sentences-in-progress. *Language in Society*, pages 441–458.

G. Lerner. 1996. On the semi-permeable character of grammatical units in conversation: Conditional entry into the turn space of another speaker. In E. Ochs, E. A. Schegloff, and S. A. Thompson, editors, *Interaction and grammar*, pages 238–276. Cambridge University Press.

G. Lerner. 2004. Collaborative turn sequences. In *Conversation analysis: Studies from the first generation*, pages 225–256. John Benjamins.

M. Oelschlaeger and J. Damico. 1998. Joint productions as a conversational strategy in aphasia. *Clinical linguistics & phonetics*, 12(6):459–480.

M. Pickering and S. Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–226.

M. Poesio and H. Rieser. In preparation. Completions, coordination, and alignment in dialogue. to appear.

K. Poncin and H. Rieser. 2006. Multi-speaker utterances and co-ordination in task-oriented dialogue. *Journal of Pragmatics*, 38(5):718–744.

M. Purver, R. Cann, and R. Kempson. 2006. Grammars as parsers: Meeting the dialogue challenge. *Research on Language and Computation*, 4(2-3):289–326.

M. Purver, C. Howes, P. G. Healey, and E. Gregoromichelaki. 2009. Split utterances in dialogue: a corpus study. In *SigDial 2009 workshop proceedings*.

H. Sacks, E. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, pages 696–735.

E. Schegloff. 1995. Parties and talking together: Two ways in which numbers are significant for talk-in-interaction. *Situated order: Studies in the social organization of talk and embodied activities*, pages 31–42.

M. Schober and S. Brennan. 2003. Processes of interactive spoken discourse: The role of the partner. *Handbook of discourse processes*, pages 123–64.

B. Szczepek. 2000. Formal Aspects of Collaborative Productions in English Conversation. *Interaction and Linguistic Structures (InLiSt), `http://www.uni-potsdam.de/u/inlist/issues/17/index.htm`*.