What is Conversation? Distinguishing Dialogue Contexts

Arash Eshghi (arash@dcs.qmul.ac.uk)
Queen Mary University of London
Interaction, Media and Communication Research Group
School of Electronic Engineering and Computer Science
Mile End Road, London E1 4NS

Patrick G.T. Healey (ph@dcs.qmul.ac.uk)
Queen Mary University of London
Interaction, Media and Communication Research Group
School of Electronic Engineering and Computer Science
Mile End Road, London E1 4NS

Abstract

Conversations are a basic unit of analysis in studies of human interaction. These units are conventionally distinguished by reference to the set of ratified participants who take part, often by appeal to their physical proximity/orientation. We show that within such conversational units there are distinct dialogue contexts which are more fine-grained than this. We present experimental evidence that demonstrates how these dialogue contexts are distinguished not in terms of topic but in terms of the participants who are actively contributing. More specifically, we show that changes in the set of primary participants alters how easily participants can access a prior turn in the conversation using a context dependent elliptical expression. So even within small groups of fully ratified participants, distinct dialogue contexts, indexed to specific sets of participants, can emerge. This leads to the idea that in multi-party dialogue, context/adjacency has a tree like structure, so that under certain conditions, presented here, utterances on either side of a sub-dialogue (a stretch of talk during which the set of primary participants remains unchanged) can be regarded as 'pragmatically next' to each other. This has implications for our understanding of how conversations are structured and for how models of dialogue track what is salient for whom at any given point in conversation.

Keywords: Conversation, Dialogue, Multi-party Interaction, Ellipsis, Adjacency Structure

Motivation

Theories of human interaction use the notion of 'conversation' as a basic building block. It groups together the sets of people, utterances, gestures or other communicative signals which are considered together in an analysis. In the case of two-person, dyadic, interactions this choice is relatively unproblematic. However, in gatherings of three or more people the boundaries between conversations become both more important for understanding the structure of the unfolding interaction and more difficult to track.

Conversation as a unit has typically been defined in terms of physical criteria that are intended to capture who the ratified participants are (i.e. those who are officially part of the conversation). For example, Goffman (1981), one of the first researchers to recognise the problems created by multi-party interactions, highlighted the way physical orientation, movement and the use of 'ritual brackets' such as greetings and farewells can mark the boundaries of interaction. Kendon (1990) described more subtle ways in which people use body position and orientation to maintain interactional units or 'F-formations'. An F-formation is also a ratificational device. Think for example of how a group standing together in a circle may or may not move aside to 'let another person in'. Such gatherings are instances of what Goffman (1963) and Goodwin (1981) called focused encounters, in which participants work to maintain a joint focus of attention.

Even within a single focused encounter different participants may stand in different relations to each contribution (Goffman, 1981). For example one person may be the direct addressee of a joke that is designed to mock another person standing next to them (the non-addressed recipient or 'side-participant'). It is known that these differences in participant status can lead to systematically different patterns of common ground being established between the different parties to a conversation (Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992).

These observations have important ramifications for formal and computational theories of discourse and dialogue that aim to characterise syntactic and semantic dependencies across multiple turns, i.e. how a contribution relates to the prior context of the dialogue (e.g. (Ginzburg, 1996; Kamp & Reyle, 1993)). For these theories it is critical to be able to specify what constitutes the shared context for a conversation.

To clarify what is at issue, consider the representation of context in Ginzburg's KoS framework (Ginzburg, 1996). This has 3 components: FACTS (the set of propositions mutually accepted/believed so far by the participants), QUD (Questions Under Discussion) and LatestMove. QUD is a partially ordered set of questions that are salient or currently under discussion. This is meant to capture the highly constrained set of 'next moves' that a participant can felicitously choose from. In particular, utterances need to be 'specific' to a salient question in QUD. So we may say that context is shared if the same set of constraints/obligations are imposed on what each participant can say/do next, i.e. the participants in question have the same *discursive potential* at that point. In multi-party dialogues, it is clear that differences in participant status, result in different sets of obligations for the participants (see (Ginzburg & Fernandez, 2005) for more detail), and accordingly, also affects the acceptability - in terms of relative

positioning - of *elliptical turns* and the possibilities of their resolution from context.

Nevertheless, the characterisations of conversation as a unit in terms of the set of ratified participants, alluded to above, are contradicted in discontinuous strands of interaction when conversational contexts are 'left open' - questions left unresolved - and taken up later on; in the sense that what you say later on, counts as 'pragmatically next' or a second pair part, to what was said before the conversation was 'broken' e.g. in multi-focus gatherings where one person may be a participant in more than one parallel conversation.

In this paper we present experimental evidence that even within a single conversation more than one dialogue context can emerge as a result of fluctuations in different participants' levels of participation. These contexts, we propose, are more fine-grained than an 'interactional unit' defined solely in terms of who the ratified participants are and are also independent of topic.

Ellipsis as window on shared-context

Ellipsis provides a useful window on the accessibility of different contributions at different points in a dialogue (Eshghi & Healey, 2007). An elliptical expression involves the omission of one or more words or phrases from an utterance which are, by implication, presumed to be recoverable from the local context. In dialogue this context is often another turn/utterance referred to as the *Antecedent* of the ellipsis. Interpretability of elliptical utterances - as manifest in the relevant response to them - demonstrates that the context required to recover the missing bits (the antecedent) is sufficiently *salient* or pragmatically 'close' or 'in focus', for the responder.

In dyadic dialogue[1], the antecedent of an ellipsis is almost always found in the immediately preceding turn (Purver, 2004), e.g. "A: Where did you go yesterday evening? B: Cinema". In multi-party dialogues however, the situation is more complex. Consider dialogue 1 from the AMI corpus below:

C: What does cutest spelling mean? (1)

B: oh, she spelled cutest um with an I‿, (2)

C: oh, okay. (3)

B: so that that's just something I pointed out. (4)

D: oh yeah. (5)

A: **Cutest?** [*Clarification Ellipsis, Antecedent Distance of 5*](6)

D: E‿S‿T‿ (7)

A: Thank you.[laugh] (8)

Dialogue 1: From the AMI corpus (Carletta, 2006)

In the above example, A's elliptical request for clarification (turn 6) is, in effect, just as local to C's utterance as B's (at

turn 2) despite differences in the amount of time and number of turns that have elapsed. The adjacency structure of multi-party dialogues is thus generally more complex than dyadic dialogues. In multi-party dialogues, after the initiation of any move, multiple non-speakers could then react to it elliptically without necessarily causing any trouble in understanding; hence adjacency structure in multi-party dialogues is not linear, but looks more like a tree: Given any utterance, there can be **more than one** subsequent turn, constructed as 'pragmatically next' to it.

We can also find examples in which multiple conversational strands, with overlapping groups of participants, are interleaved. Consider Dialogue 2 below which comes from a corpus of on-line chat community interactions in an environment called Walford (described by (Healey, White, Eshghi, Reeves, & Light, 2007)). The residents of Walford interact using a single text chat screen on which all the messages they send and receive are displayed. In Walford people can conduct multiple concurrent conversations with different, potentially overlapping, groups of individuals and they routinely take advantage of this fact. These conversations all take place in the same chat window with no visual cues to help distinguish the different conversational strands or groups of participants. Dialogue 2 shows a sequence of turns seen by ' Naomi'. Note that we have used indentation to indicate the separate conversational tracks but in the actual chat window there is no indentation:

SONIA: Hopws you? (1)

NAOMI: tired (2)

NAOMI: you? (3)

SONIA: Just got in from work (4)

NAOMI: ahh I see, kinda sick and a bit pissed (5)[*Antecedent of CE*]

NAOMI: heya hun (6)

NAOMI: Logged in from home? (7)

FRANK: ok, I?m off (8)

GARRY: yep, just about to start my studying (9)

FRANK: Yeah Yeah Yeah (10)

SONIA: **Drunk pissed?** (11)[*Clarification Ellipsis(CE), distance of 6*]

NAOMI: Angry pissed. (12)

Dialogue 2: From the Walford Corpus

Naomi is conducting a dyadic exchange with Sonia, a second dyadic exchange with Garry - which Sonia does not see - in parallel with a third, broadcast 'group' interaction with Frank that is seen by everyone. Despite this mixture of turns and participants it is clear that the clarification ellipsis at turn 11 has its antecedent in turn 5. It is as if the conversation/shared-context between Naomi and Sonia is 'suspended' for the duration of turns 6-10, the intervening sub-dialogue/shared-context. Sonia's subsequent CE is then successfully understood by Naomi as referring back to turn 11.

---

[1]It may be worth noting that "dialogue" is derived from 'dia-' meaning 'through' and not 'di-' meaning 'two' and so covers interactions between multiple participants

Accessibility of the Antecedent

What these examples make clear is that the accessibility of an antecedent is not a simple function of time elapsed, turns elapsed, nor recency in memory. They suggest rather, that people track different conversational contexts that may be maintained in parallel over surprisingly long periods especially in discontinuous strands of interaction where there is an 'open state of talk' (Goffman, 1981). An initial generalisation suggested by these examples is that distinct dialogue contexts may be distinguished in terms of the set of primary participants, i.e. those actively making contributions (jointly or individually). This leads to the prediction that an elliptical turn that is separated from its antecedent by a different dialogue context, i.e. when there's been a change in primary participants, should be more easily resolved than one which is separated by the same number of turns from the same dialogue context.

In this paper we report an experimental test of this prediction by looking at the relative ease with which participants are able to resolve elliptical clarification questions within or across dialogue contexts.

Method

Task and Materials

The experiment uses a variation on the 'Tangram' (Chinese abstract figures) tasks used in various experimental studies of common ground (e.g. (Wilkes-Gibbs & Clark, 1992)). In these studies, a *Director* has a sequence of pictures of these figures in a fixed order in front of him. The director describes these one by one to a *Matcher*. The matcher has the same figures, but unordered, which he needs to put into the same order as the director. They cannot see one another's cards, but they can talk as much as they like about them.

3-Way Tangram Task We use an on-line, 3-person variant of the same task. The task involves 2 Directors and 1 Matcher (D1, D2 and M). The D's share a set of 12 Tangram images in a single target order but half are assigned to one D and half to the other. At the start of the task all the figures are hidden from the D's but they can see which items in the sequence they will deal with and which ones the other will deal with (see Figure 1). The M, by contrast, can see all the figures but, as with the original task, does not know the target order they will have to put them in (see Figure 2).

The directors proceed through the sequence of figures one at a time by clicking, as appropriate, in the window. For each figure the 'active' director provides a description so that the matcher can identify it and drag and drop it into the appropriate slot (see Figure 2). The 'passive' director does not see the image and therefore has no direct involvement in describing it. They are, however, fully ratified side-participants in the exchange. In this manner, for each round of Tangrams, we gather a sequence of interactions between D1 and M, and some between D2 and M with the sequence of participants controlled experimentally.
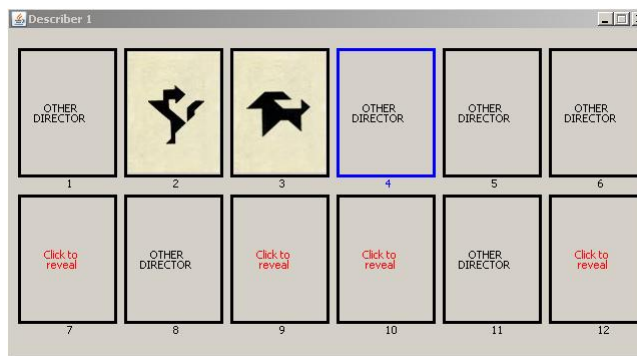

Figure 1: Director 1's view

Figure 2: Matcher's view

In order to enable the insertion of the experimental probe clarification questions all interaction between the D's and M is via the DiET Chat Tool, described below.

The DiET Chat Tool This is a custom built Java application which allows 2 or more participants to communicate via a shared chat window similar to proprietary instant messaging systems (http://www.dcs.qmul.ac.uk/research/imc/diet/index.php). The display is split into an upper window, a status bar and a lower window (see Figure 3). The upper window displays the ongoing conversation, and the lower window is used for typing. All keys pressed are recorded and relayed to the server where they are time-stamped and stored. The status bar, a prominent single line of text that is controlled by the server indicates whether other participants are currently typing. Everything that the participants say to each other, passes through the DiET server. The server stores all the relevant details about the utterances sent, including timing and structural information. This allows us to parse each turn for target expressions and provides control over what each participant sees on their local chat window.

In this experiment we parse utterances on the fly, in order to identify possible fragments from the D's turns that can serve as target sources/antecedents for long-distance RF's (Reprise Fragment, a special class of Clarification Ellipsis. E.g. A: Did Mary show up last night? B: Mary? [RF]). The criteria for selecting a potential target in this experiment were that it should be a noun phrase, verb phrase or prepositional phrase,
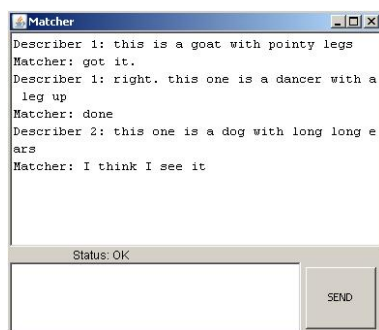
Figure 3: Chat Display

probe RF's occurred, on average, at equivalent distances from their targets in terms of time, topics and turns elapsed but at different distances in terms of dialogue context boundaries (within context in the former condition and across contexts in the latter). Exactly 4 RF's (2 in each condition) were sent to the D's per round of 12 Tangrams resulting in a *within-subjects* manipulation.

Participants

60 participants were recruited from among undergraduate and postgraduate students from various disciplines at Queen Mary, University of London. 33 were male, 27 were female. The experiments took approximately 60 minutes to complete and each was paid 7 pounds for taking part.

Procedure

Participants were randomly assigned to the D and M roles. They were given written and oral instructions about the task together and had an opportunity to ask questions about the procedure. They were then seated in separate rooms, each at a desk with a PC displaying the appropriate task window (Figure 1 for the D's and Figure 2 for the M) and the chat client window (Figure 3). They did as many rounds of tangrams that they could in 1 hour and were then debriefed about the experimental manipulations together.

Results

The principal question for our analysis is whether the dialogue context manipulation affects the responses Directors give to the spoof RF probe questions from the matcher. In particular, our prediction is that the distant antecedents should be *more accessible/salient* - and hence the RF, *easier* to interpret - when they are from a different dialogue context - because they are pragmatically 'nearer' for the Director who produced them - than when they are from the same dialogue context, i.e. were produced by the director who is describing the current item. In contrast to this, if accessibility is solely a matter of recency, number of topics (items) or total amount of intervening turns then there should be no difference between these conditions.

Because both Directors saw each RF probe question both had the opportunity to respond regardless of who actually produced the target antecedent. Responses were therefore classified according to who responded: Owner (of the target expression), Not Owner (i.e. the other director) and None where nobody responded. As Figure 4 shows, the distribution of responses was different in the two conditions. The probe question was equally likely to receive no response in the two conditions but where the target crosses the context boundaries ('different dialogue context' condition) the owner was less likely to respond than the current speaker (this is construed as *Momentum of Dyads* in the discussion below). A multinomial regression analysis confirmed the reliability of this pattern ($Chi_2^2 = 46.41$, p =0.00)

Where a response was given, it was also classified as to whether it displayed a *Local* (i.e. incorrect) or *Distant* (cor-
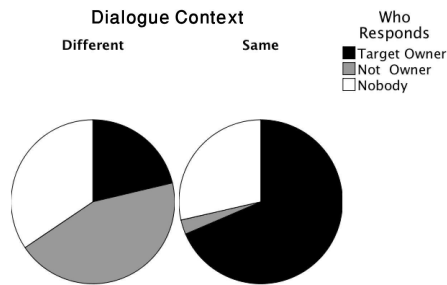
that is not subsequently repeated prior to its use in a spoof RF (in order to factor out the element of ambiguity with respect to the current figure). Given a suitable target, the server can then generate a spoof clarification question that appears to originate from the Matcher, but is not relayed to the Matcher's screen. Both D's see the RF and either may respond. The response is trapped by the server (the matcher does not see the response but the D's do) and the dialogue resumes (see Dialogue 3 for an example).

D1: the big fat lady. [Antecedent of RF]
M: ok got it.
D2: ok my one now.
D2: camel with the triangular back.
M: ye got it
M: sorry err fat lady? [Fake server probe (RF).]
D1: yeh that was number 2 [Trapped Response]

Dialogue 3: Example from the Experimental Corpus

This whole process can be carried out without disrupting the dialogue and without being detected by any of the participants. For a more detailed description of this technique see (Healey, Purver, King, Ginzburg, & Mills, 2003).

The probe RF interventions are triggered when the Matcher places a figure in the appropriate slot, and just before everybody moves on to discuss the next figure. The target antecedent of the RF is always from the *previous* item (relative to the one just placed by the M), i.e. the fragment queries something said previously by one of the D's about the penultimate figure described so far.

Experimental Design

As noted, in the experiment reported here the probe spoof RF is always generated on an antecedent (target) expression that has been used to describe the *previous* Tangram figure in the sequence. The experimental contrast is between a 'same' dialogue context condition where the current D also described the last-but-one item (i.e. there has been no change in the primary participants over the two items) and a 'different' dialogue context condition where the other D described the last-but-one item (i.e. there has been a switch in primary participants over the two items). This created a situation in which

Figure 4: Distribution of responses across conditions

Figure 5: Distribution of the Directors' interpretation of probe Reprise Fragments across conditions

rect) interpretation, or no interpretation at all - if the response itself was a request for clarification. This was possible, because the figures being described and the order in which they were described by the directors, were known at the time of analysis. In addition, the whole dialogue between Directors and Matcher prior to the sending of the RFs, was recorded. So the response often revealed clearly, whether the Director had taken the RF as clarifying something about the current figure (Local/incorrect), or the previous one (Distant/correct). In particular, the responses were coded as Distant, if:

- part of the original distant antecedent from which the RF was taken, was repeated by the Director.

- there was reference in the response to the previous item e.g. "D1: yea that was number 2".

Otherwise, the responses were coded as displaying a Local interpretation.

As Figure 5 shows, where the responses indicated a distant or a local interpretation, the pattern of interpretations was systematically different in the two conditions. If the owner responds then a distant interpretation is only favoured where the antecedent was in a different dialogue context. If the owner responds and the antecedent comes from the same dialogue context, i.e. was produced by the person who also produced the most recent description they strongly favour a local (incorrect) interpretation. This pattern is confirmed by a multinomial regression analysis which shows a main effect of both Dialogue Context condition ($Chi^2_1 = 18.26$, p =0.00) and of who responds ($Chi^2_1 = 21.25$, p =0.00) on the likelihood of a distant (correct) interpretation.

Discussion

Tendency to interpret a fragment against local context

The most obvious pattern in the results of this experiment is that there is a strong preference for local antecedents for ellipses. Even where the owner of the distant antecedent responds to the probe RF they frequently interpret it as being about the current item and not the (distant) target. This is despite the fact that the fragment used in the probe question has not been repeated since. This suggests that, all things being
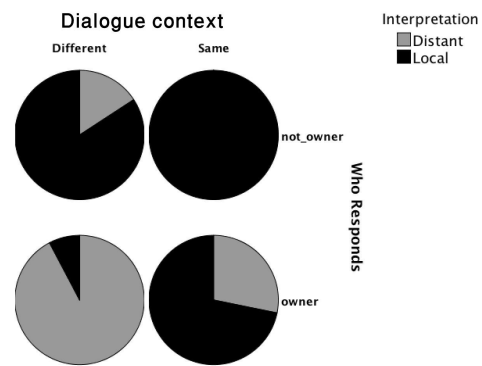
equal, there is a strong presumption that elliptical expressions have local antecedents.

Momentum of dyads

In the 'same dialogue context' condition the current speaker is also the owner of the distant antecedent (i.e. the 2 consecutive figures were described by the same director). In the 'different dialogue context' condition, the current speaker is not the owner of the distant antecedent. As figure 4 indicates, there is a strong preference for the non-owner (coincident with the current speaker) to respond in the 'different dialogue context' condition. In both conditions then, the current speaker is more likely to respond. Therefore, a possible generalisation here is that by default, in the absence of any address cues, utterances are taken to be addressed at the last speaker; dyads therefore have momentum. This is consistent with the results found in (Parker, 1988). The significant finding in that study is that a random utterance in a small focused group discussion is most likely (by a large margin) to be within a dyadic turn-taking pattern (A-B-A). Even when this pattern is broken, a new one is quickly established and sustained for an extended period. This means that even interactions with a small set of fully ratified participants, tend to lapse into stretches of dyadic talk during which the rest of the participants take on secondary roles. We argue below, that it is exactly such imbalance in the participants' level of participation, which results in the emergence of more than one distinct dialogue context; hence the need to index shared-context to specific sets of participants.

Emergence of distinct dialogue contexts

Importantly for our purposes, the results show that the structure of context, understood here in terms of changes in primary participants, has a significant effect on the accessibility of the antecedent of an ellipsis, prior to the change of primary participants. As predicted from the corpus examples, the long distance antecedents are more easily accessed where they are from a *different* dialogue context. It's crucial here to see that it is not just any change in the primary participants

across contexts that makes the distant antecedent more accessible. For the change in the primary participants to have such an effect, it is essential that:

1. the participant producing the elliptical turn, say *A*, and the owner/producer of the antecedent, say *B*, are *primary participants* in the antecedent containing, distant dialogue context i.e. they have grounded the antecedent and

2. either *A* or *B* is NOT a primary participant in the intervening dialogue context.

Support for 1 above, comes from the observation that the preference for the incorrect local interpretation is strongest when the non-owner responds. This seems to be because the non-owner is not a primary participant in the distant dialogue context containing the antecedent, and therefore has not grounded the antecedent to a level that would allow him to interpret the ellipsis as distant. So we may say that the context that *A* shares with *B* (in virtue of the fact that they are both primary participants in it) prior to the intervening dialogue context, is 'suspended' as a result of either *A*'s or *B*'s lack of active participation in the intervening context, so that subsequent to it, the distant context/antecedent is still sufficiently salient. And this is despite the fact that *A* or *B* may be a ratified side-participant to the intervening exchange.

All this supports the idea that multi-party dialogues can lead to conversational contexts that have a tree like structure. Utterances that are placed on either side of a sub-dialogue, corresponding to a single dialogue context in our sense, can be understood as pragmatically 'next' to each other. The experimental approach allows us to factor out from this effect the possible influence of number of turns, number of topics (in terms of Tangrams described) and the amount of time elapsed between the ellipsis and its antecedent. This provides a very useful notion of context locality/saliency as indexed to specific participants - contingent upon their level of participation - which accounts for the adjacency structure of turns, not only in multi-focus, discontinuous situations (such as Dialogue 2 above) but also focused gatherings with sets of fully ratified participants.

Conclusion

The experimental results presented here show that even within small groups of fully ratified conversational participants distinct conversational contexts can emerge as a result of changes in the primary participants, and hence they are, as indexed by ellipses, anaphora and other context sensitive expressions, more fine-grained than interactional units such as an F-formation (Kendon, 1990). Moreover, we have argued for the need to index these contexts to specific sets of participants and provided conditions, in terms of the participants' statuses during a stretch of talk/sub-dialogue (corresponding to a single dialogue context), under which utterances on either side of that sub-dialogue may be regarded as 'pragmatically next' to each other.

If correct, attempts to scale-up computational/formal models of dyadic dialogue, to multi-party, need to take these findings into account. They change our assumptions about how to track and update content as the conversation moves forward and what is salient for whom at any given point. This should also lead to a systematic loosening of constraints on when and from whom ellipsis is possible.

References

Carletta, J. (2006). Announcing the ami meeting corpus. *The ELRA Newsletter*.

Eshghi, A., & Healey, P. (2007). Collective states of understanding. In *Proceedings of the 8th sigdial workshop on discourse and dialogue* (pp. 2–9). Association for Computational Linguistics.

Ginzburg, J. (1996). Dynamics and the semantics of dialogue. *Logic and Computation*.

Ginzburg, J., & Fernandez, R. (2005). Scaling up from dialogue to multilogue: some principles and benchmarks. In *Proceedings of the 43rd annual meeting of the association for compuational linguistics*.

Goffman, E. (1963). *Behaviour in public places: Notes on the social organisation of gatherings*. The Free Press.

Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York, Academic Press.

Healey, P., Purver, M., King, J., Ginzburg, J., & Mills, G. (2003). Experimenting with clarification in dialogue. In *Proceedings of the 25th annual conference of the cognitive science society* (pp. 539–544.). Mahwah, N.J.: LEA.

Healey, P., White, G., Eshghi, A., Reeves, A., & Light, A. (2007). Communication spaces. *Computer Supported Cooperative Work*.

Kamp, H., & Reyle, U. (1993). *From discourse to logic*. Dordrecht: Kluwer.

Kendon, A. (1990). *Conducting interaction: Patterns of behaviour in focused encounters*. Cambridge: Cambridge University Press.

Parker, K. C. H. (1988). Speaking turns in small group interaction: A context-sensitive event sequence model. *Journal of Personality and Social Psychology*.

Purver, M. (2004). The theory and use of clarification requests in dialogue. *PhD Thesis At King's College University of London*.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*.

Wilkes-Gibbs, D., & Clark, H. (1992). Coordinating beliefs in conversation. *Memory and Language*.