Participation, Precedence and Co-ordination in Dialogue

Patrick G.T. Healey (ph@dcs.qmul.ac.uk),
Greg J. Mills (gj@dcs.qmul.ac.uk)

Interaction, Media and Communication Research Group
Department of Computer Science
Queen Mary, University of London,
London E1 4NS.

Abstract

A key problem for models of dialogue is to explain how semantic co-ordination is achieved. The collaborative model highlights pair-specfic co-ordination processes. The interactive alignment model emphasises priming processes. We report a 'Maze-task' experiment that investigates the effects of: a) apparent origin of an utterance (primary vs. peripheral participants) and b) prior exposure (priming) on the development of inter-speaker semantic co-ordination. The results provide evidence of both local and global semantic co-ordination phenomena that are not captured by pair-specific or priming processes. We argue that mechanisms that are sensitive to semantic differences between different forms of co-ordination are required and sketch a repair-driven approach.

Introduction

An obvious point about conversation is that it involves more than one person. As Goffman (1979) pointed out, this is more than the recognition that 'speaker' and 'hearer' roles differ. Different combinations of, say, direct and indirect addressees, over-hearers and bystanders can each have different effects on the way people formulate their utterances. Differences in participant status also lead to different levels of understanding and differences in the kinds of clarificiation question that can be asked (Ginzburg & Fernandez, 2005).

An important part of the experimental evidence for these differences comes from work by Clark and co-workers comparing the ability of primary (speaker and hearer) and peripheral (side-participant and overhearer) participants to understand a conversation (Wilkes-Gibbs and Clark, 1992; Schober and Clark, 1989). For example, when primary participants make repeated references to the same item, the referring expressions they use become progressively abbreviated. Peripheral participants who have full access to the entire interaction but do not actively participate are less accurate at interpreting these referring expressions than the primary participants. Participants with different degrees of involvement in an interaction thus develop different levels of communicative co-ordination with one another.

What mechanisms could account for this? In Clark's (1996) grounding model, differences in participant status are analysed as differences in the opportunities and obligations people have to engage in the grounding cycle. Thus, peripheral participants are unable to co-ordinate effectively with primary participants because they are unable to signal their acceptance (or otherwise) of the contributions of the primary participants. As a result, primary participants treat peripheral participants as more or less equivalent to naïve partners who will need to work through a new cycle of presentations and acceptances in order to establish mutual understanding (Wilkes-Gibbs and Clark, 1992).

An alternative account is provided by the interactive alignment model (Pickering and Garrod, 2004). This aims to account for the development of co-ordination through a priming mechanism that couples comprehension and production. Participants use the same interpretive scheme – including a situational model, syntax, semantics and lexicon – used for the last input utterance to formulate their next output utterance. All things being equal, the same mechanism operates both within and between speakers. Coupling comprehension and production in this way entails that multiple participants should tend to converge with one-another. They will all favour the interpretative scheme which they have had the most collective exposure to. (Garrod and Anderson, 1987; Garrod and Doherty, 1994).

Prima facie, the interactive alignment account has difficulties explaining the differences between primary and peripheral participants because, by definition, the priming mechanism operates independently of differences in participant status. However, Carletta, Garrod and Fraser-Krauss (1998) argue that the same mechanism can apply to multi-party interactions. The key point is that peripheral participants have fewer opportunities to address problems they have comprehending a primary participant's contribution. As a result they will be less strongly co-ordinated because they are, in effect, exposed to fewer instances of each input.

In summary, both the grounding model and the interactive aligment model attribute differences in semantic coordination achieved in multi-party exchanges to differences in participant's opportunities for interaction. Both predict that primary participants will co-ordinate more strongly and more quickly than peripheral participants. However, they differ in their predictions about how people respond to these differences. The grounding model predicts that speakers actively track the different levels of co-ordination that develop with different participants. The interactive alignment model predicts that speakers respond instead to the cumulative exposure to particular inputs independently of their origin in the conversation.

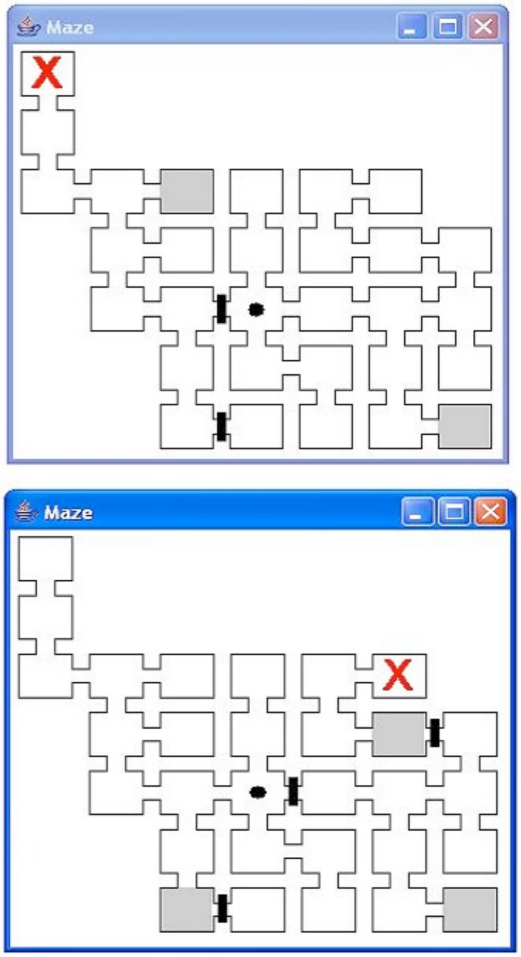

Figure 1: Example pair of Maze Configurations. The solid black circle shows the player's current position, the cross represents the goal point that the player must reach, solid bars the gates and grey squares the switch points.

To test these predictions a "Maze Game" experiment was set up using a text based chat tool. This approach enables us to investigate whether people's responses to what is ostensibly the *same* question vary according to its apparent origin (peripheral or primary participant).

Methods

A modified version of the "Maze Game", devised by Garrod and Anderson (1987) was developed. This task creates a recurrent need for pairs of participants to produce location descriptions. These descriptions can be reliably classified into four broad categories (see below). Alignment of the category of description used by participants can then be used as an index of semantic coordination (Garrod and Anderson, 1987; Garrod and Doherty, 1994).

To support turn-level experimental manipulations of the dialogues the chat-tool technique described by Healey, Purver, King, Ginzburg, and Mills (2003) was used. This allows probe turns –in this case questions about location– to be introduced into an exchange without overt disruption of the dialogue. First we describe the implementations of the maze game and chat tool used in the experiment and then go on to explain how the probe location questions are produced.

Materials

The Maze Game Application This custom-built Java application displays a simple maze consisting of a configuration of nodes that are connected by paths to form grid-like mazes (see Figure 1). The mazes are based on a 7x7 grid but with different configurations of nodes in each instance. Participants move location markers (a small black circle) from one node to another via the paths. Each move is recorded and relayed to the server where it is time-stamped and stored.

The game requires both participants to move their location markers from a starting location to a goal that is marked with a cross. Although the topology of the maze is the same for both participants, each subject has a different starting location, goal and marker, none of which are visible to the other. Some paths are blocked by gates (solid black lines) which can be opened by switches (grey coloured nodes). The locations of switches and gates are different for each subject. If a subject's location marker moves onto a switch on the *other's* screen, all the other subject's gates open. They close when the subject moves off the switch. This forces participants to use the chat-tool to collaborate. In order for subject 1 to open his/her gates, he/she has to guide subject 2 onto a node that corresponds to a switch that is only visible on subject 1's screen.

Successful completion of each maze (when both participants' position markers are on their respective goals) therefore requires participants to exchange descriptions of gate, switch, goal and position marker locations. Once both participants have reached their respective goals the next maze, with a new configuration, is automatically started.

The Chat Tool This is a custom-built Java application similar to desktop messaging systems. The display is split into an upper window, a status bar and a lower window (see Figure 2). 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, displays the activity status of the other participant.

The Chat Server The server generates artificial probe questions which appear, to participants, to originate either from each other (primary probe) or from an overhearer (peripheral probe). Each turn is preceeded by the name of apparent source (either the other participant's chosen nickname or the experimenters name) followed by a colon. Six interventions were used, each designed to elicit a spatial description and to be plausible in a range of dialogue contexts:

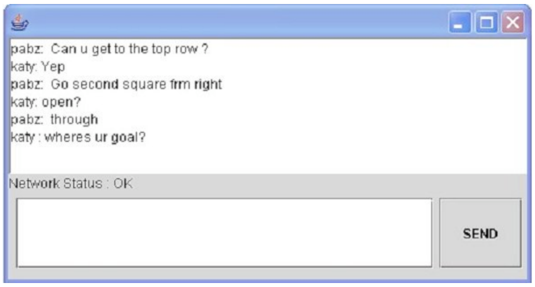

Figure 2: Chat Tool Client Window

1. Where's your goal?
2. Where's the gate nearest to your goal?
3. Where's your nearest gate?
4. Where's the switch nearest to your goal?
5. Where's your nearest switch?
6. Where did you start from?

Probes are sent simultaneously to both participants. This helps to minimise disruption to the dialogue by keeping both participants engaged in interaction and ensuring both experience similar patterns of questioning during the task. Probes were dynamically modified to mimic the speed of typing and spelling / 'txt' conventions used by each subject.

The responses to the probes are captured by the server. The probe and the subjects response are displayed only in their own chat-window, the other participant does not see them. In order to co-ordinate the resumption of the interaction after a probe, the server monitors whether one participant starts typing before their partner has finished responding to the probe. If this occurs an error message is displayed and further text-entry is prevented until either the partner responds or a pre-defined time-out expires. Subsequent turns are relayed as normal. To ensure error messages did not cue the interventions a small number of random error messages were also introduced at other points.

Subjects:

31 pairs of participants were recruited, 38 male and 24 female, from undergraduate students. They were recruited in pairs to ensure that they were familiar with each other. Only participants who had some previous experience of using internet chat software such as ICQ or Microsoft Messenger were selected for the experiment. Each subject was paid £10.00 for participating in the experiment.

Procedure:

Pairs of participants were seated in separate rooms in front of a desktop PC. On each PC a window containing the maze (same configuration but different features see Figure 1) and a chat-tool window (Figure 2) are displayed. Participants were asked to select a nickname to be used in identifying chat turns and then wait for further instructions. Except when giving the initial verbal instructions the experimenter was seated at a third PC with screens to prevent any visual contact.

Participants were told that the experiment was investigating the effects of a novel chat-tool on how people interact with each other. They were informed that their interaction would be recorded anonymously for subsequent analysis. Participants were advised that they could request the log to be deleted and were free to leave at any time but would still receive payment in full. They were given a written description of the maze game and told that the experiment involved solving twelve mazes. They were told that the experimenter could see the maze configuration but not their positions in it or the layout of features. They were further told that the experimenter would occasionally ask questions about features of the maze. No information was given about the artifical probe questions generated by the server. At the end of the experiment the full nature of the experimental interventions was explained.

To ensure participants had some experience of responding to queries apparently originating from the experimenter (peripheral participant) an intitial probe asked participants if they could read the text. On receipt of an acknowledgement, a turn instructing participants to start the experiment was sent. Twelve mazes were presented in random order to each pair. Probe questions were introduced only in the first four and the last four mazes. This created two conditions: early and late which indexed different levels of exposure to the dialogue. Half of the probes had the other (primary) participant as the apparent source and half the experimenter (peripheral) as the apparent source. 32 randomly selected probe questions were used with a maximum of two interventions per subject per maze. In addition, no intervention from same apparent origin was repeated in the first four or last four games. Overall, this resulted in a factorial design with two within-subjects factors: Source (primary vs. peripheral) crossed with Exposure (early vs. late)

Results

On debriefing, no participant reported detecting the artificial probe turns. Times from the log files provided two measures of response to the interventions. Firstly, turn completion time, the time from the onset of typing of a response to a probe turn to its completion was calculated. This was analysed in a 2×2 analysis of variance with Exposure (early vs. late) and Source (Primary vs. peripheral) as within-subjects factors. There was a main effect for Exposure ($F_{(1,412)} = 8.61$, p = 0.04), no main effects of Source ($F_{(1,412)} = 0.09$, p = 0.77) and no interaction ($F_{(1,412)} = 0.25$, p = 0.62). Overall, participants became faster at producing their responses over time, taking an average of 23 seconds in the early trials and 18 seconds in the late trials.

The second measure of task performance used was latency of response to the probe turns: the time between the onset of an inter-vention and the initial onset of typing the response (re-gardless of whether there was subsequent deletion) was extracted from the logs. A 2×2 analysis of variance with Exposure and Source as within-subjects factors showed no effect of Exposure ($F_{(1,412)} = 0.08$, p = 0.77) a main effect of Source ($F_{(1,412)} = 10.56$, p = 0.001) and no interaction. Initiation of responses to peripheral probes took twice as long (12 secs) as responses to primary probes (5.9 secs) and this difference was consistent throughout.

Description Types:

For comparison with previous work a total of 9755 turns, including both descriptions generated in normal dialogue and 684 responses to probe turns were classified according to the criteria developed by Garrod and Anderson (1987). This categorises location descriptions into four basic classes corresponding to different underlying mental models of the maze:

Figural: a heterogeneous category of relatively concrete descriptions that draw on some specific element of the overall configuration or distribution of particular features to identify a target location.

A : You see the sticking out bit at the top ?
B : Yep
A : I'm on the bottom right one and the switch is right above it

Path: involves identifying a route to be traversed through the maze to the target location. Path descriptions are sensitive to the specific layout of boxes and connections in the maze.

A : Where are you now ?
B : See where your switch is ?
A : Yep
B : Go up 1 , 2 right , 1 down

Line: classifies the maze into a set of line elements corresponding to rows, columns or diagonals. The target line is described first, followed by the target box as a position along it.

A : I'm in the bottom box in the second column from the right.
A : I'm in the third row, fifth to the left

Matrix: introduces a Cartesian coordinate system with locations identified via the specification of two vectors either as rows and columns or in terms of numbers or letters for each axis.

A : My switches are at 4,6 5,4 and I'm on 3,4.
A : My goal is b2
A: I need to get to 2nd row 5th column

Transcription Results

Figure 3 illustrates the change in the baseline pattern of use of description types in spontaneous dialogue (i.e., excluding responses to experimental interventions) from Early to Late trials.[1]

The distribution of description types used in responses to interventions is illustrated in Figure 4. This suggests a different pattern of responses to Primary and Peripheral participants. Multinomial regression shows a significant main effect of both Exposure ($Chi^2(3) = 14.5$, p = 0.00) and Source ($Chi^2(3) = 15.6$, p=0.00) on choice of description types.

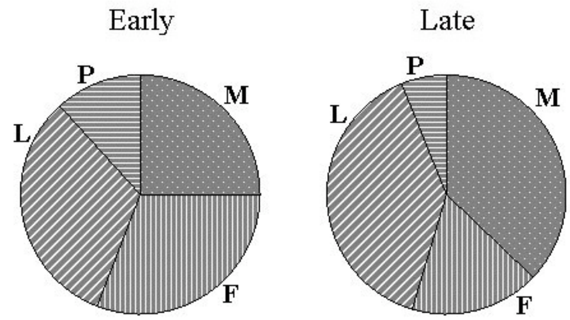

Figure 3: Distribution of Description Types in Spontaneous Dialogue (Early vs. Late). F = Figural, P = Path, L = Line and M = Matrix

To provide focused tests of the hypotheses three additional comparisons were carried out using multinomial regression. Firstly, the prediction that Peripheral participants level of co-ordination should change over time was tested by comparing the distribution of description types produced for Peripheral participants only in early vs. late trials. This showed no reliable difference ($Chi^2(3) = 0.82$, p=0.66). Secondly, a test of whether the description types produced for Primary and Peripheral participants differ in the Early trails ($Chi^2(3) = 15.5$, p=0.00). Finally, a test for whether Peripheral and Primary differ in the Late trials ($Chi^2(3) = 3.98$, p = 0.26). Overall, the profile of description types produced for Primary participants evolves over trials whereas those produced for Peripheral does not change. This pattern is illustrated in Figure 4.

Discussion

The distributions of description types observed here replicate the basic patterns observed in the original, oral, Maze game studies (Garrod and Anderson, 1987; Garrod and Doherty, 1994). There is a general migration from the relatively concrete descriptions (Figural and Path) that depend on the specific details of each maze, towards more abstract description types (Line and Matrix) that invoke schemata that generalise across instances. This is true both of the descriptions produced in the spontaneous dialogue (Figure 3) and for those produced in response to the experimental interventions with the primary participant as their apparent source (Figure 4).

The study reported here is the first experimental analysis of the effects of participant status on co-ordination in the Maze game. As noted in the introduction, Carletta et. al. (1998) propose that that the same basic mechanisms of alignment, i.e. input-output co-ordination, should operate in dyadic and in multi-party interaction –modulo opportunities to interact. Participants should use the lexical, semantic and syntactic forms which are

[1] Pie charts are used because this is compositional data.

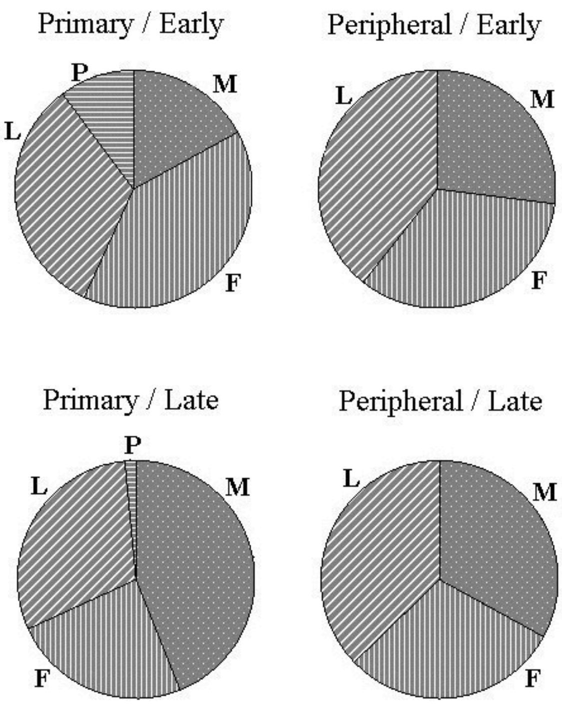

Figure 4: Distribution of Description Types in Response to Interventions

most strongly primed by the preceding dialogue (Garrod and Doherty, 1994; Pickering and Garrod, 2004).

The results reported here show, however, that maze game participants reliably distinguish between primary and peripheral participants. The same questions receive different responses depending on whether the apparent source of the question is a primary or peripheral participant. Specifically, they take twice as long to initiate responses to probes that appear to originate from a peripheral participant and, in the early exchanges, they use different description types. As a result the local choice of description type in the responses to probes is not explained only by reference to priming from the preceding dialogue.

The grounding model appears to be better equipped to deal with these observations since it assumes that levels (or forms) of co-ordination will be explicitly indexed to different participants (Brennan and Clark, 1996). However, all things being equal, levels of co-ordination should be lower where participants have not explicitly engaged in a cycle of grounding (Clark, 1989; Wilkes-Gibbs and Clark, 1992). In the experiment reported here the peripheral participants provide no positive evidence of acceptance and on this basis should be more weakly co-ordinated. However, the pattern of description types used indicates that peripheral participants are actually treated as more highly co-ordinated in the early games than primary participants.

One possible explanantion for this is a 'lab coat' effect. Participants might assume that the peripheral participant –the experimenter– is a maze game expert who can understand more complex forms of description (Clark, 1996a). If participants have a conception of what constitutes 'expert' language they could use it with the experimenter from the start of the task. However, it them becomes difficult to explain why they do not do the same thing with their task partner since, by hypothesis, they can both already understand and produce the more 'expert' descriptions and, unlike the experimenter, are actually able to ground them with their partner.

The key problem, as both Clark (1996b) and Pickering and Garrod (2004) emphasise, is that the selection of description types cannot be modeled as autonomous choices: interaction plays an essential role in co-ordination. In the present results this is reflected in the fact that co-ordination builds up between the primary participants, who do interact, but remains unchanged with the peripheral participant. We return to this point below.

In addition to the problems with local patterns of co-ordination, the Maze game data also display more global patterns of co-ordination. As noted above, it is consistently found that participants migrate from the more concrete instance-specific description types to more abstract ones that capture generalisations about the underlying grid structure of the Maze (Garrod and Anderson, 1987; Garrod and Doherty, 1994; Healey, 1997).

These more global trends also pose a problem for models of dialogue co-ordination. They are not accounted for by priming since, as Figures 3 and 4 illustrate, they run counter to local precedence. Over time people do not converge on the description type that they are most frequently exposed to. These trends also have a different character to the patterns of contraction or abbreviation of referring expressions that are the primary concern of the grounding model (Wilkes-Gibbs and Clark, 1992). Specifically, the Line and Matrix schemes do not emerge as abbreviated versions of Figural and Path descriptions, rather they involve changes of semantic model (Garrod and Anderson, 1987).[2]

Repair Driven Co-ordination?

The underlying problem, we propose, is that the primary co-ordination mechanisms articulated in existing models are, in effect, 'semantically neutral'. They focus on co-ordination processes that are not sensitive to the different kinds of semantic co-ordination implied by changes in description type. As a result, they operate in the same way regardless of whether participants are using Figural, Path, Line or Matrix schemes. We propose that the use of different description schemes in fact implies different levels of co-ordination.

[2] We note, though, that the different Maze location description types are distinguishable in terms of the density of grounding cycles typically used in each case. As the examples in the Methods section illustrate, Figural and Path descriptions are normally grounded explicitly and incrementally whereas Line and Matrix schemes often involve no direct feedback. This observation reinforces the assumption that the Line and Matrix descriptions reflect higher levels of co-ordination.

The abstract schemes (Line and Matrix) invoke a relatively systematic, compositional model of the maze domain - one which abstracts away from each instance of the maze to an array of possible locations. The figurative schemes (Figural and Path) depend much more on the concrete details of the layout in each case. For example, whereas Figural and Path descriptions only refer to 'missing' boxes as "gaps" or "holes", in abstract descriptions they are integrated into the counts of rows or boxes used to specify a location. Boxes and gaps are thus given the same abstract ontological status. As a result, co-ordination of abstract descriptions is more difficult since, by hypothesis, it involves co-ordinating on a semantic model that is not directly manifest in any particular instance of the Maze.

Our proposal is that co-ordination is built on the opportunities interaction creates for identifying and addressing *differences* in interpretation. Initially participants try out whatever description type occurs to them. Once they detect a problem with interpretation they systematically exploit less abstract schemes as one means of repairing the problem. This allows them to take advantage of the local context –in particular the maze in front of them– to resolve uncertainties about what location was intended. By repeatedly resolving ambiguities in this way participant's semantic models can progressively converge. This is, in effect, a semantic variant of Clark and Marshall's concept of vertical repair. However, our evidence suggests that the basic strategy for dealing with communication problems is to switch to less, not more, specific forms of description.

On this account the initial difference between the primary and peripheral participants is due to the fact that the primary participants are providing feedback to each other about problems they are encountering. In response they switch to more basic description types and over time build up their co-ordination. Convergent evidence for this claim comes from the finding that when participants encounter problems with co-ordination they switch away from abstract description types back to more concrete ones (Healey, 1997).

Conclusion

We have argued that neither interactive alignment nor grounding provide an adequate account of the co-ordination phenomena observed in the Maze game. There are both local and global patterns of co-ordination that are not explained by the primary co-ordination mechanisms they provide. We share their assumption that opportunities for interaction are the key difference between primary and peripheral participation. However we have argued that people's responses to problems with interpretation, and their sensitivity to different forms of semantic co-ordination, are key dialogue processes.

Acknowledgments

We thank EPSRC for supporting this work through Mills' studentship and Matt Purver, Jonathan Ginzburg and the anonymous referees for helpful comments on earlier versions.

References

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 1482–1493.

Carletta, J., Garrod, S., & Fraser-Krauss, H. (1998). Communication and placement of authority in workplace groups – the consequences for innovation. *Small Group Research*, *29*(5), 531–559.

Clark, H. H. (1996a). Communities, commonalities, and communication. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 324–355). Cambridge: Cambridge University Press.

Clark, H. H. (1996b). *Using language*. Cambridge: Cambridge University Press.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. Joshi, B. L. Webber, & I. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, *13*, 259–294.

Garrod, S. C., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*, 181–218.

Garrod, S. C., & Doherty, G. (1994). Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition*, *53*, 181–215.

Ginzburg, J., & Fernandez, R. (2005). Action at a distance: the difference between dialogue and multilogue. In C. Gardent & B. Gaiffe (Eds.), *Proceedings of dialor '05: Ninth workshop on the semantics and pragmatics of dialogue.* LORIA, Nancy.

Goffman, I. (1979). Footing. *Semiotica*, *25*, 1–29.

Healey, P. (1997). Expertise or expert-*ese*?: The emergence of task-oriented sub-languages. In M. Shafto & P. Langley (Eds.), *Proceedings of the 19th annual conference of the cognitive science society* (pp. 301–306).

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

Pickering, M., & Garrod, S. (2004). The interactive alignment model. *Behavioral and Brain Sciences*, *27*(2), 169–189.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*, 211–232.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, *31*, 183–194.