Building Representations of Common Ground with Intelligent Agents: Papers from the 2011 AAAI Fall Symposium (FS-11-02)

Modeling Expert Effects and Common Ground Using Questions Under Discussion

Alex Djalali, David Clausen, Sven Lauer, Karl Schultz†, and **Christopher Potts**
Stanford University †University of Massachusetts Amherst
Stanford CA 94305 Amherst MA 10003
[djalali, clausend, sven.lauer, cgpotts]@stanford.edu, *kschultz@cs.umass.edu*

Abstract

We present a graph-theoretic model of discourse based on the Questions Under Discussion (QUD) framework. Questions and assertions are treated as edges connecting discourse states in a rooted graph, modeling the introduction and resolution of various QUDs as paths through this graph. The amount of common ground presupposed by interlocutors at any given point in a discourse corresponds to graphical depth. We introduce a new task-oriented dialogue corpus and show that experts, presuming a richer common ground, initiate discourse at a deeper level than novices. The QUD-graph model thus enables us to quantify the experthood of a speaker relative to a fixed domain and to characterize the ways in which rich common ground facilitates more efficient communication.

1 Introduction

The presuppositions of a speaker's utterance are often a rich source of insights into her understanding of the context — what she assumes to be mutual knowledge of the discourse participants, what she regards as the current goals, what kinds of information she regards as relevant to those goals, and so forth. These inferences are an important part of robust utterance understanding and thus vital for research in linguistic pragmatics and artificial intelligence (Thomason 1990; Stone, Thomason, and DeVault 2007).

In this paper, we present a method for deriving (some of) these inferences based on a speaker's utterances and a partial understanding of the context. We begin from the notion that discourse is collaborative inquiry into the state of the world (Stalnaker 1974; 1998; 2002). This inquiry is guided by a partially ordered set of abstract questions under discussion (QUDs), which determine what is relevant, help to define the domain of discourse, and influence the common ground (Groenendijk and Stokhof 1984; Groenendijk 1999; Ginzburg 1996; Roberts 1996; Cooper and Larsson 2001; Roberts 2004; Büring 1999; Büring 2003; Stone 2002; van Rooy 2003).

In very well organized (idealized) inquiry, the participants begin from the most general questions and then move systematically to the most specific ones. For example, if A is trying to find out where B would like to vacation, A should first resolve whether B would like to go to Europe before inquiring about specific European countries. If, and only if, A and B arrive jointly at a positive resolution of the Europe question should they press on to inquire about Germany, Hungary, and so forth.

Of course, most discourse does not march from general to specific in this transparent way. Speakers often begin with very specific questions; A might in fact begin the vacation discussion by asking which of Germany or Hungary they should visit. In such cases, the speaker presupposes that the more general issues have already been settled, or else acts as though the addressee will accommodate resolutions of those more general issues that make sense given the specific question posed. Where it works, this can be an efficient strategy of inquiry, allowing the participants to resolve many issues at once. Where it fails, it leads to confusion and a rapid retreat back to more general issues.

We present a formal model of these accommodation processes. Questions and assertions are treated as edges in a rooted graph, connecting potential discourse states. The edges thus model the introduction and resolution of various QUDs. The information that a speaker presupposes with an utterance U can be characterized via the position of U in the QUD graph, and the amount of information presupposed with U is given by its depth in the QUD-graph.

To evaluate this proposal, we introduce a new task-oriented dialogue corpus that represents not only the participants' utterances, but also their goals and actions, thereby allowing us to precisely model the context of utterance at any stage of play. In particular, if a player poses or addresses a task-oriented question Q, then we can identify where Q sits in the QUD-graph determined by the nature of the task, which provides an estimate of the amount and kind of information contributed by Q. We show that expert players tend to initiate discourse at greater depth in the QUD graph than do novices. This behavior is explicable in terms of the assumption that experts are more likely to presume rich common ground than novices, which allows them to bypass general questions in favor of more specific ones.

2 The QUD model

We now develop a graph-theoretic model of discourse, which we refer to as the *QUD graph model*. This model is based on those of Roberts (1996), Ginzburg (1996) Groenendijk (1999), and Büring (2003), though we have made various simplifications and modifications. The intuitive idea is that any discourse can be viewed as a sequence of questions and their answers, all of which address sub-issues of the current topic of conversation, until all such issues are exhaustively resolved. At any point in the discourse, there is a current QUD, which interlocutors attempt to resolve, either directly or by tackling a series of sub-questions, whose answers, taken together, fully resolve that QUD. The overarching topic of conversation can be viewed as a (typically quite general) question, e.g., *What is the state of the world?*, and all moves in a discourse are in the service of resolving its sub-questions.

To make this precise, we assume a possible-worlds framework. Our most basic discourse notion is the *common ground*, which includes vast quantities of basic world knowledge as well as information about the public commitments of the discourse participants, the salience of various objects, a record of previous utterance events, and so forth.

Definition 1 (Common ground). Let W be a set of possible worlds and D a set of discourse participants. The *common ground* for D is the subset of W in which all and only the public commitments of all the individuals in D are true.

A question is a partition on W. Intuitively, each cell of the partition is a complete resolution of the issue (Groenendijk and Stokhof 1984). For example, if p is the proposition that it is raining, then $R = \{p, W-p\}$ is the question of whether it is raining. The question of what the weather is like corresponds to the set of propositions $C = \{p, q, r, \ldots\}$, where p is the proposition that it raining (and nothing else), q is the proposition that it is snowing (and nothing else), r is the proposition that it is hailing (and nothing else), and so on for all the mutually exclusive weather propositions.

We say that Q_1 is a *sub-question* of Q_2 relative to common ground CG iff every complete answer to Q_1 in CG excludes (at least) some possible answers to Q_2 (Groenendijk and Stokhof 1984; 1997; van Rooy 2003):

Definition 2 (Contextual sub-question). Q_1 is a *sub-question* of Q_2 relative to CG, written $Q_1 \sqsubset_{CG} Q_2$, iff $\forall p \in Q_1, \exists q \in Q_2$ such that $CG \cap (p \cap q) = \emptyset$.

Even without the contextual relativization, the question $R = \{p, W-p\}$ (is it raining?) is a subquestion of $C = \{p, q, r, \ldots\}$ (what is the weather like?): $p \in R$ excludes all of $q, r, \ldots \in C$, and $W-p \in R$ excludes $p \in C$. We see the effects of contextual relativization when we look at more complex interactions. For example, the question of whether it is summer, $S = \{s, W-s\}$, will not generally be a sub-question of C, but if the common ground entails that rain implies winter and hail implies summer, then we have $S \sqsubset_{CG} C$, because s contextually excludes rain and $W-s$ contextually excludes hail.

The common ground and the QUD are essential to the flow of discourse at all points, so we define *discourse states* in terms of them:

Definition 3 (Discourse state). Let W be a set of possible worlds and D a set of discourse participants. A discourse state is a pair $\langle CG, QUD\rangle$ where CG is a common ground defined in terms of D and W (def. 1) and QUD is a partition on W.

Equipped with these ancillary notions, we define a *QUD graph* as a graph whose vertices are discourse states and whose edges represent assertions and questions:

Definition 4 (QUD Graph). A *QUD graph* is a directed graph $G = \langle V, A, Q\rangle$, where V a set of discourse states (vertices); A is a set of edges corresponding to answers; and Q is a set of edges corresponding to questions. Then for any two states m and n,

1. mQn iff $CG_m = CG_n$ and $QUD_m \sqsubset_{CG_m} QUD_n$
2. mAn iff $CG_m \supset CG_n$ and QUD_m is the closest unresolved QUD in the path from m to the root node.

Question-edges relate two states (mQn) iff the first contains a sub-question of the second. Answer-edges relate two states (mAn) iff the QUD of the first is (at least partially) answered by the common ground of the second and the QUD of the second is constrained to be the closest unresolved question in the graph.

As we discussed above, speakers will typically not obey the strict structure of such a QUD graph in terms of their actual utterances. However, their utterances will indicate where in the graph they currently take themselves to be in the discourse, and they will expect the other discourse participants to reason in terms of the graph, by quietly resolving super-questions in a way that delivers sensible results for the sub-questions they pose, and by properly relating assertions to the corresponding QUDs. Thus, the length of the shortest path between two explicit moves can serve as a measure of how far the interlocutors have progressed in a given discourse with respect to resolving a QUD, and, indirectly, how much knowledge they take to be in the CG.

3 Experiment

We now present our evaluation of the QUD-graph model. We first introduce the Cards corpus and describe the additional annotations that were used to construct the *Cards graph*, a QUD-graph informed by the structure of the task. We then show that experts leave more information implicit than do novices, which is expected in our model, since experts have richer common ground to rely on.

3.1 The Cards corpus

The Cards corpus is built around a Web-based, two-person collaborative search task, partly inspired by similar efforts (Thompson et al. 1993; Allen et al. 1996; Stoia et al. 2008). The game-world consists of a maze-like environment in which a deck of cards has been randomly distributed. The players are placed in random initial positions and explore using keyboard input. A chat window allows them to exchange information and make decisions together. Each player can see his own location, but the location of the cards and the other player are limited by distance and line-of-sight. Players can pick up and drop cards, but they can hold at most three cards at a time. In addition, while most of the walls are visible, some appear to a player only when within that player's line-of-sight.

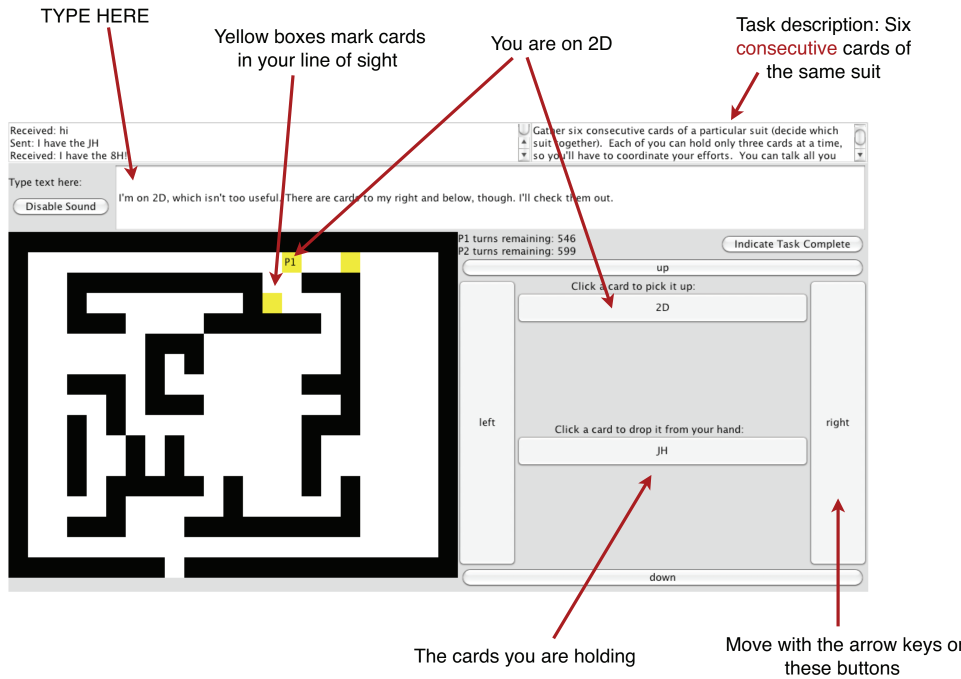

Figure 1: A annotated version of the Cards game-board.

When players sign on, they are presented with the following underspecified task description:

> Gather six consecutive cards of a particular suit (decide which suit together), or determine that this is impossible. Each of you can hold only three cards at a time, so you'll have to coordinate your efforts. You can talk all you want, but you can make only a limited number of moves.

The players must decide together, via the chat interface, which sequence to choose, which they usually do based on initial random exploration. The following is a typical exchange right after this initial phase:

P1: I have 9 clubs and K clubs
P1: want to look for clubs?
P2: ok

In this transcript, the players then find various clubs, checking with each other frequently, until they gain an implicit understanding of which specific sequences to try for (either 8C-KC or 9C-AC):

P1: so you are holding Jc and Kc now?
P2: i now have 10d JC and KC
P2: yes
P1: drop 10d and look for either 8c or Ace of clubs

This snippet also highlights the value of limiting the players to holding three cards at a time: they are compelled to share information about the locations of cards, and their solutions are necessarily collaborative.

In some versions of the game, there are sub-regions of the environment that are walled off. This can make the game unsolvable: required cards might be unreachable by one or both of the players. This led to different kinds of collaboration, in which players shared extensive information about the nature of the game-board in order to jointly arrive at the verdict of impossibility. In the following snippet, the players have already divided up the task of searching out the remaining possible solutions:

P1: i see nothing in my colors. do you see any in yours?
P2: don't think so – double check hearts for me, ok?
P1: ok
P1: i don't see it
P1: so unsolvable
P2: unsolveable, agree?

The corpus consists of 439 transcripts. Each transcript records not only the chat history, but also the initial state of the environment and all the players' actions (with timing information) throughout the game, which permits us to replay the games with perfect fidelity. In all, the corpus contains 12,280 utterances (mean length: 5.28 words), totaling 64,900 words, with a vocabulary size around 3,000. Most actions are not utterances, though: there are 175,503 movements, 8,330 card pick-ups, and 6,105 card drops. The median game-length is 392 actions, though this is extremely variable (s.d.: 263 actions).

The transcripts were collected in two phases: 62 at the University of Pennsylvania in 2008, and the rest via Mechanical Turk during the summer of 2010. Fig. 1 provides the annotated game-board that was presented to the Turkers before they began playing, to acquaint them with the interface. Each player was paid \$1.00 per game, with \$0.50

bonuses paid to especially thoughtful players. Feedback received from the players suggested that they generally enjoyed playing.

In our experiments, we use the number of transcripts each player contributed as an objective measure of expertise. We have this data only for 324 of the transcripts, all from the Mechanical Turk section of the corpus, so we limit attenton to them. These were created by 102 distinct players. While 33 played just once, 36 played 10 or more times; the median number of games played was 4 (mean: 6.7; s.d.: 5.98). Fig. 2 depicts this distribution as a boxplot.

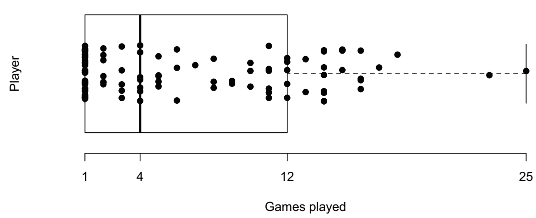

Figure 2: Number of games played by each player. The data points have been jittered randomly along the y-axis to make their clustering evident.

3.2 The Cards graph

We constructed the Cards graph around seven sets of questions. Each set can be thought of as an individual subgraph whose root node is a high-level QUD. The restricted nature of the game permits exhaustive identification of these sets. For example, the questions *What is the configuration of the game board?* and *What is the expertise of my fellow player?* are members of a subgraph rooted at *What is the state of this particular game?*. Utterances that do not pertain to the game (e.g., *What's your name?*) are excluded.

For the purposes of our experiment, the relative order of the members of each individual subgraph is not consequential. However, the subgraphs themselves are linearly ordered as in def. 4 (assuming a rich set of background assumptions about the nature of play), with the root node of the entire graph being the QUD *How do we finish the game successfully?* Each subgraph is assigned a *depth*: the number of other subgraphs intervening between the root node and that graph. Fig. 3 summarizes the subgraphs used with their respective depths. We assume that each question in the graph comes with a common ground that encodes for (i) the rules of the game each player is exposed to prior to playing the game, and (ii) answers to all the questions of a lower depth.

There are two ways in which a subgraph can be deeper than another. (i) The questions in the first subgraph need to be completely resolved before the questions in the second can be tackled. This is true, for example, for the questions of depth 1 and those of depth 2 in fig. 3: it is impossible to sensibly discuss what the goals of the game are unless one understands what the game world is like. Thus, a speaker whose initial utterance relates to an issue in depth 2 will presuppose that all issues in depth 1 are answered. (ii) The questions in the second subgraph can be concerned with resolving a sub-issue of the first. This is the relation that holds between the subgraph in depth 5 and the one in depth 6 – determining how to gather the decided sequence involves determining how to gather the specific cards in the sequence.

- **Depth 1**
  - *How do I interact with the game world?*
  - *What are the meanings of the various technical terms?*
- **Depth 2**
  - *What is the goal of the game generally?*
- **Depth 3**
  - *What is the configuration of the game board?*
  - *What is the expertise of my fellow player?*
- **Depth 4**
  - *What is the goal of this game specifically?*
- **Depth 5**
  - *How do we achieve this goal generally?*
- **Depth 6**
  - *What cards do we need to achieve this goal specifically?*
- **Depth 7**
  - *Have we obtained a particular winning sequence?*

Figure 3: The clusters used to approximate the initial depth of players in the QUD graph.

For each player in each transcript, we identified the player's first initiating utterance, i.e., a question that raises a new (sub-)issue or an assertion that resolves a previously unmentioned (sub-)issue. We then identified the subgraph of the Cards graph that the utterance belonged to and coded its depth. We did this by considering the first six utterances of each player, thus allowing us to take into account various contextual factors that might affect the interpretation of those players' utterances. For example, the utterance "6D" (six of diamonds) could either be placed at depth 4 if it was an attempt by a player to negotiate the six-card sequence the two players should collect; or it could be placed at depth 6 if it was used to indicate that a player had picked up the six of diamonds after the sequence had been settled upon.

Responses to initiating moves of the other player were not considered initiating moves. Thus if Player 1 started the conversation with *Where [on the game board] are you?* and Player 2 responded with *On the lower left*, Player 1's initial depth was coded as 3, but Player 2's initial depth was determined by a later utterance.

The following transcript excerpt shows each utterances with the depth of its associated subgraph. The first utterances by P1 and P2 were identified as their initiating utterances; the rest are included to help convey what the coding is like:

P2:	what suit do we want?	(depth 4)
P1:	I hit a KD.	(depth 3)

	I think we should see what we get, and keep the most promising suit	(depth 3)
P2:	i have a JD	(depth 3)
P1:	That works	
P2:	so we are looking for Ds?	(depth 4)
P1:	I vote Ds.	(depth 4)
P2:	okay i have 10D , 9D and JD	(depth 6)
P1:	7D	(depth 6)
P1:	okay do you think my cards work?	(depth 5)
P1:	So we're looking for 8D, and 6D or QD	(depth 6)
P2:	You should be good (I'm slow at this...)	(depth 6)

3.3 Results

We expect experts to be more likely than novices to assume that their partners will accommodate rich contextual knowledge. In the context of the QUD model, this predicts that experts will initiate discourse at a deeper level in the Cards graph than novices. Furthermore, these effects should be amplified if it is mutual knowledge between the two players that they are both experts.

To test this at the individual level, we first fit an ordinary least-squares regression, using the number of games played up to and including the present one (GamesPlayed) to predict initial utterance depth (Depth). The coefficient for GamesPlayed in the fitted model is 0.02 with a standard error of 0.01 ($p = 0.04$). Thus, the association is significant but weak. This is arguably because the game is not complicated; one successful task completion might be enough to show the presuppositional behavior we seek to characterize. Thus, we classified a player–transcript pair (P, T) as Novice if T was the first transcript that P contributed, and Expert otherwise. A linear regression using this Novice/Expert distinction (Expert) to predict Depth revealed a significant positive correlation; the coefficient for Expert, summarizing the predicted difference between the two groups, is 0.54 with a standard error of 0.11 ($p < 0.0001$).

At the transcript level, we grouped the discourses into three types: *Novice games* (novice–novice pairings), *Mixed games* (novice–expert pairings) and *Expert games* (expert–expert pairings). Novice games had a mean word-count of 328.85, mixed games had a mean word-count of 190.03, and Expert games had a mean word-count of just 113.01. These successively shorter values are expected on a model where rich common ground licenses speakers to presuppose more and say less overtly. In addition, we coded the Novice/Mixed/Expert distinction as a variable Pairings (1 for Novice, 2 for Mixed, and 3 for Expert) and fit a linear regression using this variable to predict the mean initial utterance depth of the two players. The coefficient for Pairings was 0.35 with a standard error of 0.07 ($p < 0.001$); expertness once again correlates with increased utterance depth.

4 Discussion

The annotated Cards graph revealed that experts start with general QUDs far less often than their novice counterparts, supporting the idea that experts presuppose more. It is likely this is because experts are more familiar with the game world and had opportunities to develop optimal strategies.

Here is a typical start to a Mixed game interaction:

P2:	Hi?
P1:	hey
P2:	Hey, I'm in the bottom right corner what suit are we looking for?
P1:	uh idk
P2:	okay..let's then browse for a while
P1:	im so confused
P1:	on how to play this
P2:	I've got ah kh and 10 h do you wont to go for H?
P1:	sure
P2:	ok then :)
P2:	we have to find 6 consecutive cards of a particular suit
P1:	ok so we need 9h 10h jh qh kh ah

The novice (P1) starts at depth 1, the expert (P2) at depth 4 (*What suit are we looking for?*). Moreover, we see that P2 (belatedly) briefly accommodates to P1's initial level by explaining the basic goal of the game, i.e., *we have to find 6 consecutive cards* .... P1, in turn, signals that (s)he has resolved the lower-level questions, by initiating discussion of a question at depth 5/6. This suggests that a discrepancy in the amount of presupposed knowledge leads to re-negotiation of the QUD.

The following complete transcript is an extreme example of efficient communication between experts:

P2:	hi–what side r u on?
P1:	outside
P2:	inside
P2:	i have 3H, 5H and 6H
P2:	can u find 4H, and other 2 cards that surrond the set on either side?
P1:	ok
P1:	hey
P1:	i mean i got it
P1:	4,7,8 H
P2:	nice

As can be seen, not only do both players start at a deep level in the Cards graph (depth 3), but they also exchange a minimum of information, relying on each other to fill in the content of their underspecified utterances. This shows how players, after they have recognized each other as experts, take advantage of the rich common ground by drastically reducing the length and number of their utterances.

Interestingly, while expert players optimized their linguistic behavior by using fewer words and moving swiftly to deep levels in the QUD graph, they did not reliably complete the games in fewer moves than novices. We found no reliable correlations between overall transcript length and expert levels. In many cases, experts moved too quickly to a very specific strategy, one that was insensitive to the overall layout of the game world, and this meant more searching around for the required cards than would have been necessary if they had been flexible in their approach.

5 Conclusion

This paper provided a metric for quantifying the amount of common ground presupposed by interlocutors. In order to achieve this, we developed the QUD graph model, in which the common information presupposed by the interlocutors corresponds to graphical depth.

The Cards corpus was collected in a controlled environment where the interlocutors' information state and utterances were recorded at every point in the discourse, which facilitated the construction of a detailed QUD graph. The graph revealed that experts, able to rely heavily on the common ground, require their interlocutors to accommodate more information than novices do. This is expected on a model of pragmatics in which accommodation is a strategic decision in communication (Thomason 1990; Stone, Thomason, and DeVault 2007).

We believe that the QUD-graph approach will generalize to other tasks and domains, since it is founded in very general pragmatic theories (Groenendijk and Stokhof 1984; Groenendijk 1999; Ginzburg 1996; Roberts 1996; Büring 2003; Stone 2002; van Rooy 2003). It should be said, though, that few corpora have the rich metadata of the Cards corpus, so it is far from straightforward to apply the ideas elsewhere. What is needed are automatic methods for inferring the QUD graph based on more superficial aspects of the discourse, an area that is ripe for further systematic inquiry.

Acknowledgements

Special thanks to Victoria Schwanda, Chriz Czyzewicz, and the SUBTLE team for help collecting the data. This research was supported in part by ONR grant No. N00014-10-1-0109 and ARO grant No. W911NF-07-1-0216.

References

Allen, J. F.; Miller, B. W.; Ringger, E. K.; and Sikorski, T. 1996. A robust system for natural spoken dialogue. In *Proceedings of the 1996 Annual Meeting of the Association for Computational Linguistics*, 62–70. ACL.

Büring, D. 1999. Topic. In Bosch, P., and Sandt, R. v., eds., *Focus — Linguistic, Cognitive, and Computational Perspectives*. Cambridge: Cambridge University Press. 142–165.

Büring, D. 2003. On D-trees, beans, and B-accents. *Linguistics and Philosophy* 26(5):511–545.

Cooper, R., and Larsson, S. 2001. Accommodation and reaccommodation in dialogue. In *Gothenburg Papers in Computational Linguistics*. Department of Linguistics, Göteborg University.

Ginzburg, J. 1996. Interrogatives: Questions, facts and dialogue. In Lappin, S., ed., *The Handbook of Contemporary Semantic Theory*, Blackwell Textbooks in Linguistics. Blackwell. 385–422.

Groenendijk, J., and Stokhof, M. 1984. *Studies on the Semantics of Questions and the Pragmatics of Answers*. Ph.D. Dissertation, Universiteit van Amsterdam, Amsterdam, The Netherlands.

Groenendijk, J., and Stokhof, M. 1997. Questions. In van Bethem, J., and ter Meulen, A., eds., *Handbook of Logic and Language*. Amsterdam: Elsevier Science. chapter 19: Questions, 1055–1124.

Groenendijk, J. 1999. The logic of interrogation. In Matthews, T., and Strolovitch, D., eds., *Proceedings of SALT IX*, 109–126. Ithaca, NY: Cornell University.

Roberts, C. 1996. Information structure in discourse: Towards an integrated account of formal pragmatics. In *Paper in Semantics, OSU Working Papers in Linguistics*, volume 49. Columbus, OH: Department of Linguistics, the Ohio State University.

Roberts, C. 2004. Context in dynamic interpretation. In Horn, L. R., and Ward, G., eds., *Handbook of Pragmatics*. Oxford: Blackwell. 197–220.

van Rooy, R. 2003. Questioning to resolve decision problems. *Linguistics and Philosophy* 26(6):727–763.

Stalnaker, R. 1974. Pragmatic presupposition. In Munitz, M., and Unger, P., eds., *Semantics and Philosphy*. New York University Press. 197–213.

Stalnaker, R. 1998. On the representation of context. *Journal of Logic, Language, and Information* 7(1):3–19.

Stalnaker, R. 2002. Common ground. *Linguistics and Philosophy* 25(5-6):701–721.

Stoia, L.; Shockley, D. M.; Byron, D. K.; and Fosler-Lussier, E. 2008. Scare: A situated corpus with annotated referring expressions. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*.

Stone, M.; Thomason, R.; and DeVault, D. 2007. Enlightened update: A computational architecture for presupposition and other pragmatic phenomena. To appear in Donna K. Byron; Craige Roberts; and Scott Schwenter, *Presupposition Accommodation*.

Stone, M. 2002. Communicative intentions and conversational processes in human–human and human–computer dialogue. In Trueswell, J., and Tanenhaus, M., eds., *World Situated Language Use: Psycholinguistic, Linguistic, and Computational Perspectives on Bridging the Product and Action Traditions*. Cambridge, MA: MIT Press.

Thomason, R. H. 1990. Accomodation, meaning, and implicature: Interdisciplinary foundations for pragmatics. In Cohen, P. R.; Morgan, J.; and Pollack, M. E., eds., *Intentions in Communication*. Cambridge, MA: MIT Press. chapter 16, 325–363.

Thompson, H. S.; Anderson, A.; Bard, E. G.; Doherty-Sneddon, G.; Newlands, A.; and Sotillo, C. 1993. The HCRC map task corpus: Natural dialogue for speech recognition. In *HLT '93: Proceedings of the workshop on Human Language Technology*, 25–30. Morristown, NJ: Association for Computational Linguistics.