Managing ambiguities across utterances in dialogue

David DeVault and **Matthew Stone**
Department of Computer Science
Rutgers University
Piscataway, NJ 08845-8019
David.DeVault@rutgers.edu, Matthew.Stone@rutgers.edu

Abstract

Task-oriented dialogue systems exploit context to interpret user utterances correctly. When the correct interpretation of a user utterance is ambiguous, a common response is to employ a special process of clarification that delays context update until important ambiguities are resolved, so that the main dialogue task can proceed with an unambiguous context. In this paper, we describe an implemented dialogue agent which instead translates ambiguities in interpretation into uncertainty about which context has resulted from an utterance. It then uses question-asking strategies, including clarification as a special case of questions about speaker meaning, to manage its uncertainty across multi-utterance subdialogues. We analyze the agent's use of these strategies in an empirical study of task-oriented dialogues between the agent and human users.

1 Introduction

Dialogue agents cannot always understand their human partners. Indeed, we ourselves do not always understand what others say to us. Nevertheless, *our* conversational abilities allow us to follow up provisional interpretations of what has been said and eventually arrive at a sufficient understanding. This paper reports work on designing dialogue agents that can do the same.

The specific problem we address in this paper is how to reason about context-dependence while working to reduce ambiguity and achieve common ground. Every utterance in conversation gets its precise meaning in part through its relationship to what has come before. This applies to the clarificatory utterances interlocutors use to acknowledge, reframe or question others' contributions just as it does to fresh contributions. The distinctive issue with such followups is that they must be formulated for a context about which speaker or addressee may be uncertain. The speaker must be able to assess that addressees will understand and respond helpfully to them no matter what the context might be.

In this paper, we present a model that frames this reasoning as ordinary collaborative language use in the presence of contextual ambiguities. We describe how dialogue agents come to be uncertain about what their interlocutors have contributed, and offer a precise characterization of how agents can formulate context-dependent utterances that help pinpoint the context and resolve ambiguity. A dialogue agent that uses such utterances can play its collaborative role in working to understand its interlocutors.

Our model is implemented in COREF, a task-oriented dialogue system that collaboratively identifies visual objects with human users. We show empirically that to interact successfully in its domain, COREF does need to work collaboratively to resolve ambiguities, and moreover that our model makes COREF to some degree successful in doing so. At the same time, we highlight qualitative aspects of COREF's behavior that depend on our new synthesis of linguistic and collaborative reasoning. For example, we show how COREF needs both linguistic reasoning and collaborative reasoning to formulate followups that offer alternative descriptions of things it judges its interlocutors might have meant.

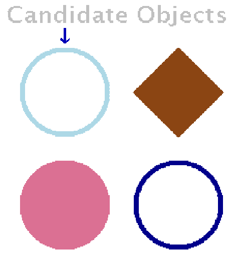
Our work is part of a larger project on reconciling linguistic reasoning and collaborative reasoning in conversation (Stone, 2004; DeVault and

Stone, 2006; Thomason et al., 2006). In particular, we build on the account of communicative intentions of Stone (2004), on the account of context update for communicative intentions of DeVault and Stone (2006), and on the model of collaboration in conversation from Thomason et al. (2006). We advance this program here by weakening many of the idealizations about mutuality that we have made explicitly or implicitly in earlier work. Thus, we are able to go significantly further towards an account of the reasoning and skills that agents use to overcome differences in achieving mutual understanding.

2 Related Work

Our work is an attempt to use the theory of collaboration to bridge two different traditions for specifying dialogue agency. The first is engineering approaches to spoken dialogue systems, where researchers have shown that systems should represent the uncertainty of their automatic speech recognition results and take that uncertainty into account in their dialogue management strategies. For example, maintaining a probability distribution over alternative recognition results can help a system to choose whether to clarify user input or proceed with a possibly incorrect interpretation (Roy et al., 2000; Horvitz and Paek, 2001; Williams and Young, 2007). It also allows statistical inference to combine evidence about user intentions from multiple utterances (Bohus and Rudnicky, 2006). Such research connects uncertainty to systems' high-level choices, but because it focuses on modeling user state rather than utterance context, it cannot connect uncertainty to principled compositional linguistic reasoning such as decision-making in natural language generation.

The other tradition is deep approaches to dialogue coherence, where researchers provide detailed models of evolving utterance context in dialogue and of the linguistic constructions that exploit this context. These models go much further in accounting for the specific utterances speakers can use in context for grounding and clarification. However, these models often create explanatory tension by running together descriptions of how utterances update the context with descriptions of how interlocutors manage uncertainty. For example, when a new utterance occurs, its content may be marked *ungrounded* to reflect the fact that its content must be acknowledged by the hearer before it can be assumed to have been understood (Traum, 1994; Poesio and Traum, 1997). However, acknowledgments in dialogue don't really always function to put specified content unambiguously onto the common ground (Clark and Schaefer, 1989). For example, Figure 1 provides a naturally occurring fragment of human–human dialogue in COREF's domain, where interlocutors treat an utterance of *okay* as ambiguous. In this interaction, S15 and S14 converse via teletype from separate rooms. S15 begins by instructing S14 to click on a certain object in S14's display. S14 does so, but S15 cannot observe the action. This leads S15 to perceive an ambiguity when S14 says *okay*: has S14 merely grounded S15's instruction, or has S14 also clicked the object? The ambiguity *matters* for this task, so S15 engages the ambiguity with a followup question.



S15:	Okay, add the light blue empty circle please.
	[S14 privately adds the object]
S14:	okay
S15:	Okay, so you've added it?
S14:	i have added it. It is in the top left position.

Figure 1: An ambiguous grounding action by subject S14 in a human-human dialogue.

Similarly, utterances that are perceived as ambiguous in important ways may be modeled as suspended until a special process of clarification resolves the relevant ambiguity (Ginzburg and Cooper, 2004; Purver, 2004). But the problem of recognizing and responding to perceived ambiguities in a collaboration is more general than the problem of clarifying utterances. For example, in the task domain of Figure 1, the question *you've added it?* serves to resolve ambiguity just like a clarification might, but it arises from the non-public nature of the "add object" action rather than from any grammatically-specified dynamics of context update (Purver, 2004).

Finally, connecting context update to the resolution of perceived ambiguities may guarantee common ground, but leaving ambiguities open can make a collaborative agent more flexible. An agent that demands a clear context but lacks the resources to clarify something may have no recourse but to take a “downdate” action—to signal to the user that their intended contribution was not understood, and discard any alternative possible contents. If the agent can proceed, however, the agent may get evidence from what happens next to resolve its uncertainty and complete the task.

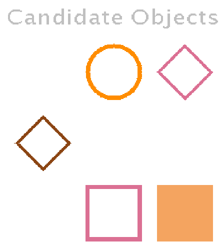
We view uncertainty management and context update as necessary but independent processes; this positions our work between the two traditions. We follow more applied work in representing uncertainty in the context probabilistically, and modeling grounding and clarification as collaborative mechanisms interlocutors can use to reduce but perhaps not eliminate this uncertainty. But we follow deeper models in using a precise dynamic semantics to characterize the evolving utterance context and its effects on utterance interpretation.

3 Technical Approach

We present our ideas through examples of referential communication. Our specific setting is based on the collaborative reference task studied in pairs of human subjects by Clark and Wilkes-Gibbs (1990). Each interlocutor perceives a collection of visual objects, as illustrated in Figures 1–2. The interlocutors perceive identical objects, but with shuffled spatial locations. One interlocutor, who we call the director, sees a target object highlighted on their display with an arrow, and is charged with conveying to their partner, who we call the matcher, which of the displayed objects is the target. The interlocutors go through the objects one by one, with the matcher attempting to identify and click on the correct target at each step.

We have implemented an agent COREF which can participate in these dialogues (DeVault and Stone, 2006). Figure 2 shows a sample interaction between COREF and a human user. We will use this interaction to illustrate how COREF frames clarification as an ambiguity management problem. Here, COREF has perceived an ambiguity in the user’s intention in uttering *it is brown*, and decides to clarify with *do you mean dark brown?*

The model that realizes COREF’s behavior here incorporates three new principles. First, the model exposes ambiguity about what the user means as uncertainty in the dialogue state that results from the user’s utterance. Here COREF assumes that



possible contexts	agent	actor
c_1	COREF:	is the target pink?
c_2	A18:	no
c_3	A18:	it is brown
c_{4l}, c_{4d}	COREF:	do you mean dark brown?
c_{5l}, c_{5d}	A18:	yes
c_6	COREF	(privately adds the brown diamond)
c_6	COREF:	done

Figure 2: COREF asks a clarification question.

the user intends to identify the color of the target object with *it is brown* and therefore finds two possible interpretations: one for the dark brown color of the empty diamond and one for the light brown color of the solid square. After the utterance, COREF is uncertain about which meaning was intended and thus which constraint the user has contributed.

Second, the model allows the specification of dialogue strategies that allow COREF to proceed with appropriate high-level dialogue moves despite having more than one alternative for what the context is. Here COREF settles on a clarification move, because we have specified a policy of clarifying ambiguities reflecting different constraints on the target object. In other kinds of uncertain contexts, COREF will proceed without clarifying.

Third, COREF plans its generation decisions so that the user will recover a specific and useful interpretation of what it says no matter what the context is. Here COREF explicitly constructs the utterance *do you mean dark brown* by carrying out an incremental derivation using a lexicalized grammar. The rich representation of the utterance context allows the system to recognize the applicability of forms that cohere with what has gone before, such as the use of the frame *do you mean* to refer to content from the previous utterance, whatever it may have been. The model predicts that this underspecification is unproblematic, but predicts that the ambiguity of *brown* must be eliminated and therefore motivates the adjunction of the modifier *dark*.

In this section, we sketch the implementation of COREF, briefly summarizing the details we carry over from previous presentations, and highlighting the differences that support the implementation of the three new principles.

3.1 Context, Tasks, Actions, and Uncertainty

We follow DeVault and Stone (2006) in understanding the utterance context at each point in a dialogue as an *objective* and *normative* product of prior interlocutor action. The context for COREF describes both the state of the ongoing referential activity and the semantic and pragmatic status of information in the dialogue. Activity is represented through a stack of ongoing tasks, drawn from an inventory including COREF's overall multi-object reference task, its single-object reference task, a yes/no question task, a reminder question task, a clarification task, and an ambiguity management task (`ManageAmbiguity`) that is automatically pushed after each utterance or action. The linguistic context, meanwhile, includes aspects of the discourse history such as specifications of recent utterances and of salient referents.

Dialogue allows interlocutors to change the context by manifesting a suitable communicative intention—in other words by taking an observable action with a specific commitment as to how the action will link up with and update the context (Stone, 2004). This is formalized by a function $\texttt{update}(c, i)$ which describes the context that results from acting with intention i in context c. However, interlocutors actually observe actions rather than intentions and so must recognize the intention from knowledge of language and of the ongoing task. Thus, while COREF tries to identify the true context at each point in time, it is sometimes uncertain about it, as when there is perceived ambiguity in its interlocutor's intentions. The basic interpretive operation in COREF is not *updating*—that is, tracking deterministic context change—but *filtering*—propagating uncertainty about the context at time t to uncertainty about the context at time $t+1$ based on an observed action.

We follow Thomason et al. (2006) in characterizing filtering in COREF's domain through *tacit actions* as well as observable actions. Tacit actions include task-relevant cognitive actions like identifying the target object or abandoning a task. A speaker is free to use tacit actions as well as observable actions to update the context. However, successful coordination requires the speaker to provide sufficient evidence in their observable actions to reconstruct any tacit actions they have committed to. Formally, for any context c and interlocutor S, we can use the next actions that could contribute to the pending tasks in c to determine a set of alternative contexts $Z(c, S)$ that could be reached by S from c just using tacit actions. We call this set of alternative contexts the *horizon*.

The horizon allows us to make an agent's filtering operation precise. Let us write $c : i$ to denote an interpretation which shows the speaker (or actor) acting in context c with a commitment to intention i. In understanding, an agent H starts from a prior probability distribution over the initial context at time t given the evidence E available so far: $P_H(c_t|E)$. H observes an action a_t (carried out by agent S), and must infer $\hat{c}_t : i_t$ to explain that action. H can assume that the new context $\hat{c}_t$ must be some element of $Z(c_t, S)$, and that i_t must match action a_t into $\hat{c}_t$ so as to contribute to the ongoing tasks. H will inevitably bring substantial background knowledge to bear, such as grammatical knowledge and interpretive preferences. However, H's evidence may still leave multiple options open. We summarize H's intention recognition as a probabilistic likelihood model $P_H(\hat{c}_t : i_t|c_t, a_t)$. (As usual, we assume the context tells you everything you need to know about the current state to interpret the action.) Filtering combines update, prior and likelihood:

$$P_H(c_{t+1}|a_t, E) \propto \sum P_H(\hat{c}_t : i_t|c_t, a_t) P_H(c_t|E)$$

where the summation ranges over all values of c_t, $\hat{c}_t$, and i_t such that $c_{t+1} = \texttt{update}(\hat{c}_t, i_t)$.

We illustrate this model through COREF's reasoning on A18's utterances *it is brown* and *yes*, the third and fifth utterances from Figure 2. For the first of these utterances, COREF starts with just one context c_3 with any probability. There are two possible interpretations i_{3l} and i_{3d} corresponding to the different colors (*light* and *dark* brown respectively) that might be picked up by *brown*; COREF's model happens to assign them equal probability. Each interpretation involves a tacit move to a context $\hat{c}_3$ which implicitly completes any discussion of the contribution of the user's previous utterance *no*. Filtering therefore results in two possible values for the next context, $c_{4l} = \texttt{update}(\hat{c}_3, i_{3l})$ and

$c_{4d} = \texttt{update}(\hat{c}_3, i_{3d})$. Each is assigned probability 0.5. Ambiguity in interpretation has been exposed as uncertainty in the context.

For the second of these utterances, *yes*, COREF starts with *two* equally probable contexts c_{5l} and c_{5d} which (as we shall see further below) are derived from taking into account the effect of COREF's tacit actions and clarification question in contexts c_{4l} and c_{4d}. Here the context-dependence of *yes* means that COREF must find an interpretation in which the user gives an appropriate affirmative answer to the salient question (in the context $\hat{c}_{5l}$ or $\hat{c}_{5d}$ following a tacit action closing discussion of COREF's meaning). That question is whether the user meant *dark brown* by *brown*. The *yes* answer is appropriate in contexts derived from c_{4d} because that is what the user meant there, but not in contexts derived from c_{4l} where the user meant something else. So across all the candidate contexts only one interpretation i_5 can be assigned nonzero probability. Accordingly filtering restores all the probability mass to $c_6 = \texttt{update}(\hat{c}_{5d}, i_5)$.

3.2 Minimizing Ambiguity

Our discussion thus far has shown how interlocutors can interpret utterances in succession as creating and resolving temporary ambiguities. Our goal, however, is to design dialogue agents that can not only deal passively with ambiguity, but can collaborate actively to resolve ambiguities with their interlocutors. This means giving agents high-level strategies that are helpful in dealing with uncertainty, and generating natural language utterances that do not exacerbate the problems of ambiguity even when used in uncertain contexts.

COREF includes a hand-built action policy that decides which contributions to the conversation would be *acceptable* for the agent to take *given its current uncertainty*. For example, COREF's policy deems it acceptable to ask for clarification any time COREF is uncertain which constraint a speaker intended to add with an utterance, as in Figure 2. Similarly, COREF's action policy deems it acceptable for the agent to ask whether a non-public action m has occurred, if some possible contexts but not others indicate that m has taken place. For example, COREF translates an ambiguous acknowledgment like that of Figure 1 into uncertainty about whether the "add object" action has tacitly occurred in the true context; COREF follows up such an *okay* by asking *did you add it?*

COREF's generation module is tasked with formulating an utterance that makes these contributions in a way its interlocutor will understand. In Thomason et al. (2006) we investigate a *strong* notion of recognizability. Each utterance must result in a *checkpoint* where speaker and hearer agree not only on a unique interpretation for the utterance but also on a unique resulting context. Enforcing this constraint supports the traditional attribution of mutual knowledge to the two interlocutors at each point in the conversation.

Here we develop a more flexible notion of *weak recognizability* that allows for uncertain contexts and makes interpretation more robust to potential differences in their perspectives. In interpreting a user utterance, COREF expects to find zero, one, or multiple interpretations in each possible context. In generation, COREF is sometimes willing to take the risk of using an action or utterance that may not be interpretable in all possible contexts. Taken together, this means new utterances can serve not only to present the speaker's intention, but also in some cases to introduce or defuse uncertainties about the true context. Checkpoints, where COREF achieves certainty about the true context, arise as side effects of this dynamic rather than as a strict requirement in the architecture. While there is no guarantee that any given speaker contribution will ever become common ground, COREF's dialogue policies are designed to try to achieve common ground when it is practical to do so.

Our formal development assumes that agents can take their own probabilistic models of interpretation as good indicators of their partners' disambiguation preferences (for example by slightly overspecifying their utterances). More precisely, we will allow each interlocutor to discard certain interpretations whose probability falls below a threshold ϵ and so are of sufficiently low probability, relative to others, that they can safely be ignored. Consider then an observable action a by S. If there were only a single possible context c, the set of recognized interpretations for a would be $R(c, a) = \{\hat{c}:i | P(\hat{c}:i|c, a) \gg \epsilon\}$. But in general, S is uncertain which of $C = \{c_1, ..., c_k\}$ is the true context, and expects that H may give any of these a high prior and take seriously the corresponding interpretations of the utterances. Indeed, S must also be prepared that S is actually making any of these contributions. In other words H and S will consider any interpretation in $R^*(C, a) = \cup_{c \in C} R(c, a)$. $R^*(C, a)$ is weakly recognizable if and only if each $c_i \in C$ is associated with at most one interpretation in $R^*(C, a)$.

The formalism explains why, in generation, COREF chooses to elaborate its utterance *do you mean brown* by adding the word *dark*. COREF's policy makes a clarification question acceptable across all of the candidate contexts after the user says *it is brown*. But *do you mean brown* is not weakly recognizable. For example, in c_{4d}, there are two interpretations, which could be paraphrased *do you mean light brown* and *do you mean dark brown*. COREF therefore chooses to coordinate more finely on the alternative interpretations of its clarification action. The utterance *do you mean dark brown* has only one interpretation in each of c_{4l} and c_{4d} and therefore represents a solution to COREF's communicative goal.

3.3 Strategically Discarding Ambiguities

To keep search tractable for real-time interaction, COREF tracks a maximum of 3 contexts. If more than 3 are possible, the 3 most probable are retained, and the others discarded. Further, after each object is completed, COREF discards all but the most probable context, to avoid retaining unilluminating historical ambiguities. In fact, according to COREF's action policy, it is acceptable to complete an object despite an ambiguous context, provided the ambiguity does not affect the agent's judgment about the target object—this is COREF's analogue of a "grounding criterion".

4 Empirical Results

We recruited 20 human subjects[1] to carry out a series of collaborative reference tasks with COREF. The study was web-based; subjects participated from the location of their choice, and learned the task by reading on-screen instructions. They were told they would work with an interactive dialogue agent rather than a human partner. Each subject worked one-by-one through a series of 29 target objects, for a total of 580 objects and 3245 utterances across all subjects. For each subject, the 29 target objects were organized into 3 groups, with the first 4 in a 2x2 matrix, the next 9 in a 3x3 matrix, and the final 16 in a 4x4 matrix. As each object was completed, the correct target was removed from its group, leaving one fewer object in the matrix containing the remaining targets. The roles of director and matcher alternated with each group of objects. Either COREF or the subject was randomly chosen to be director first.

[1]Most of the subjects were undergraduate students participating for course credit at Rutgers University.

`correct`	`no object`	`skipped`	`wrong`
75.0%	14.3%	7.4%	3.3%

Table 1: Overall distribution of object outcomes.

1 context	2 contexts	3 contexts
83.4%	6.8%	9.8%

Table 2: Number of possible contexts perceived when utterances or actions occur.

The experiment interface allows an object to be completed with one of four outcomes. At any time, the matcher can click on an object to add it to her "scene," which is another matrix containing previously added objects for the same group. An object is completed when the director presses either the `continue` or `skip` button, or when the matcher presses `skip`. An outcome is scored `correct` if the director presses `continue` after the matcher has added the correct target to her scene. It is scored `skipped` if either interlocutor presses the `skip` button.[2] It is scored `no object` or `wrong` if the director presses `continue` before the matcher adds any object, or after the matcher adds the wrong object, respectively.

Table 1 shows COREF's overall performance in the task. We would like to understand this performance in terms of COREF's uncertainty about the context. To begin, Table 2 shows the distribution in the number of alternative contexts perceived by COREF across all subjects. COREF is usually completely certain what the true context is, but is uncertain about 17% of the time.[3] To better understand how this uncertainty affects object outcomes, we investigated the agent's performance during the subdialogues associated with individual objects, which had a mean length of 5.6 utterances. Figure 3 shows the relation between the mean number of possible contexts during an object subdialogue and the outcome for that dialogue. The figure shows that high mean uncertainty has a clear negative impact on object outcomes, but a smaller degree of uncertainty is less harmful, if at all. In total, 13.1% of COREF's `correct` ob-

[2]Though note that COREF never presses `skip`.

[3]Since COREF truncates its uncertainty at 3 possible contexts, the higher frequency of 3 possible contexts relative to 2 here very likely masks a longer underlying tail.

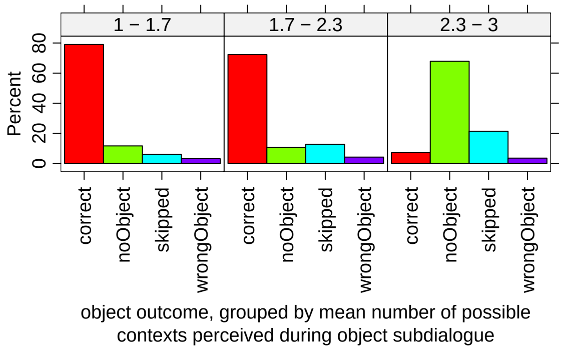


Figure 3: Object outcome vs. context uncertainty.

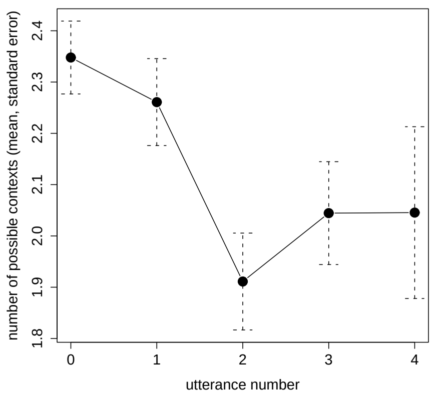


Figure 4: Effect of ambiguity management questions on COREF's uncertainty. At utterance 0, COREF faces an ambiguous context. At utterance 1, COREF has asked a question. Utterance 2 is typically an answer by the subject.

ject outcomes occur at a moment when COREF is uncertain what the true context is (9.7% two contexts, 3.4% three contexts).

While certainty about the context is not strictly necessary for a `correct` outcome, COREF nevertheless does often try to reduce its uncertainty according to its question-asking policy. Figure 4 illustrates the effectiveness of COREF's question-asking policy at reducing uncertainty. As the figure shows, when COREF asks questions in an ambiguous context, the mean reduction in the agent's uncertainty is about 0.4 contexts. Figure 2 is an example where the subject's answer eliminates a context. But the subject's answer does not always reduce uncertainty, because it may introduce a new ambiguity.[4] Figure 1 actually gives such an example in a human-human dialogue. In this dialogue, from S15's perspective, it is possible that S14 had already added the object to the scene; but it is also possible that S14 took the question as a reminder to add the object to the scene and answered in the affirmative only after correcting the error. This distinction does not matter for task success, but it does introduce a potentially lasting ambiguity into the dialogue history. When COREF's questions do not resolve an ambiguity, COREF does not force a downdate; it tries instead to proceed with the task. Figures 3 and 4 suggest that COREF's ambiguity management mechanisms are relatively successful in cases of mild or short-lived ambiguities.

5 Discussion and Future Work

We have presented a framework that allows task-oriented dialogue agents to use language collaboratively despite uncertainty about the context. We have presented empirical evidence that managing ambiguity is a key task for dialogue agents such as ours, and that it can be addressed successfully within a uniform architecture for collaboration under uncertainty. In particular, our model shows how dialogue agents can support grounding acknowledgments, clarification of ambiguous utterances, and task-oriented question asking using generic linguistic resources and goal-oriented ambiguity management strategies. For such an agent, what is distinctive about acknowledgments and clarification is simply their reference and relation to prior utterances; they play no special role in a language-specific context-update mechanism.

The proposed model is most applicable to situations in which the speaker's true intention is always among the alternative interpretations derived by the hearer. This is the case for the acknowledgments and clarifications of speaker meaning that occur frequently in COREF's domain, and that have been our focus to date. We believe our model could also be extended to clarifications of perceived ambiguities in phonology and syntax, drawing on the work of Ginzburg and Cooper (2004). Perceived phonologic or syntactic ambiguities could be translated into ambiguities in the context resulting from an utterance, entirely analogously to COREF's response to ambiguities of meaning.

However, our work does not immediately cover

[4] Other ways a question can fail to reduce uncertainty are if the user chooses not to answer the question or if the agent fails to understand the user's answer.

clarification questions that are not designed to resolve perceived ambiguities, but rather are asked in situations where *no* interpretations are found. Such examples occur; see Ginzburg and Cooper (2004) or Purver (2004) for examples. When COREF finds no interpretations for a user utterance, it notes the utterance and signals an interpretation failure (currently by saying *umm*), but it otherwise leaves its context representation as it was, and is unable to address the failure with its usual ambiguity management policy. Alternative characterizations of agents' reasoning in such cases are still required, and work such as Purver's provides a natural starting point.

Moreover, traditional classifications of grounding actions (Traum, 1999) include a variety of other cases as well. For example, we do not treat repair requests like *what?* or *what did you say?*, which can signal interpretation failure or the hearer's incredulity at the speaker's apparent (but correctly and uniquely identified) meaning. Similarly, we do not treat self-repairs by speakers. These can exclude a possible but unintended interpretation, to avoid a foreseen misunderstanding—an example in COREF's domain would be, *A: I moved it. A: I mean I moved the blue circle.* They can also correct a prior verbal mistake, as when a speaker has mistakenly used the wrong word: *A: I moved the circle. A: I mean I moved the square.* It would be interesting to explore whether richer models of domain uncertainty and dialogue context would enable us to account for these utterance types.

Ultimately, our framework suggests that agents face uncertainty from various sources, but that their experience provides quantitative evidence about what kinds of uncertainty arise and how best to resolve them. A final direction for our future research, then, is to analyze records of agents' interactions to develop decision-theoretic strategies to optimize agents' tradeoffs between asking clarification questions, resolving ambiguity to the most likely interpretation, and proceeding with an uncertain context.

Acknowledgments

We thank our anonymous reviewers for helpful comments. This work was supported by NSF HLC 0308121, IGERT 0549115 & HSD 0624191.

References

D. Bohus and A. Rudnicky. 2006. A k hypotheses + other belief updating model. In *Proceedings AAAI Workshop on Statistical and Empirical Approaches for Spoken Dialogue Systems*.

H. H. Clark and E. F. Schaefer. 1989. Contributing to discourse. *Cognitive Science*, 13:259–294.

H.H. Clark and D. Wilkes-Gibbs. 1990. Referring as a collaborative process. In P. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, pages 463–493. MIT.

D. DeVault and M. Stone. 2006. Scorekeeping in an uncertain language game. In *Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue (SemDial-10)*, pages 139–146.

J. Ginzburg and R. Cooper. 2004. Clarification, ellipsis and the nature of contextual updates in dialogue. *Linguistics and Philosophy*, 27(3):297–365.

E. Horvitz and T. Paek. 2001. Harnessing models of users' goals to mediate clarification dialog in spoken language systems. In *User Modeling Conference*, pages 3–13.

M. Poesio and D. R. Traum. 1997. Conversational actions and discourse situations. *Computational Intelligence*, 13(3):309–347.

M. Purver. 2004. *The Theory and Use of Clarification Requests in Dialogue*. Ph.D. thesis, King's College.

N. Roy, J. Pineau, and S. Thrun. 2000. Spoken dialogue management using probabilistic reasoning. In *Proc. of ACL*, pages 93–100, Hong Kong.

M. Stone. 2004. Communicative intentions and conversational processes in human-human and human-computer dialogue. In Trueswell and Tanenhaus, editors, *World-Situated Language Use*, pages 39–70. MIT.

R. H. Thomason, M. Stone, and D. DeVault. 2006. Enlightened update: A computational architecture for presupposition and other pragmatic phenomena. To appear in Byron, D., Roberts, C., and Schwenter, S., eds, Presupposition Accommodation.

D. R. Traum. 1994. *A Computational Theory of Grounding in Natural Language Conversation*. Ph.D. thesis, Rochester.

D. R. Traum. 1999. Computational models of grounding in collaborative systems. In S. E. Brennan, A. Giboin, and D. Traum, editors, *AAAI Fall Symposium on Psychological Models of Communication*, pages 124–131.

J. Williams and S. Young. 2007. Partially observable markov decision processes for spoken dialog systems. *Computer Speech and Language*, 21(2):393–422.