Computational Models of Grounding in Collaborative Systems

David R. Traum
UMIACS
University of Maryland
A. V. Williams Building
College Park, MD. USA
traum@cs.umd.edu

Abstract

In this paper, I present in detail two models of grounding, Clark and Schaefer's contribution model, and the grounding acts model from my previous work. Description of each model is accompanied with discussion of it's limitations, particularly with respect to use within a collaborative system. Also presented is an approach taking factors such as media costs into account.

Overview

The concept of *common ground*, or similar notions, such as *mutual belief* or *a shared conception* have been very important in Cognitive Science theories of collaboration/cooperation.[1] Indeed, the two currently dominant accounts of intentional collaboration in AI, *Joint Intentions/Teamwork* (Cohen & Levesque 1991b) and *SharedPlans* (Grosz & Sidner 1990; Grosz & Kraus 1993) both include *mutual beliefs* as key components in their definitions. However, few have taken helpful stands on *how* mutual belief is established in dialogue. Most agree that acknowledgment plays some role, but there are proofs that no amount of acknowledgments can insure perfect mutual belief in a noisy environment (Halpern & Moses 1990). Most AI dialogue researchers, including those above[2], have settled for the opposite extreme — that in virtue of copresence and sometimes other assumptions, information will automatically become part of common ground.

The use of acknowledgments has also been seen in HCI systems, as well as human conversation, in a variety of ways. However there has been little systematicity in most systems as to when an acknowledgment should appear in a dialogue, leading to system behaviors which can be very frustrating to users (e.g., on a PC, being forced to click "okay" to an incomprehensible message before being allowed to continue).

In this paper I will briefly discuss two models of grounding, the process by which collaborators can establish (assumed) common ground. As well as a description of the model itself, a discussion of some of the deficiencies will be presented, particularly with an eye toward using the model within a collaborative system to enable performance of grounding behavior.

Clark and Schaefer's Contribution Model

Clark and Schaefer's *Contribution* model of grounding (Clark & Schaefer 1989) is one of the first and clearest of how this achievement of mutual understanding is itself a collaborative process, brought about by active participation by multiple participants.This model is very appealing and, at a coarse level, very influential and important for viewing dialogue as a collaborative process. While the main ideas are important considerations both for designers of collaborative systems and for those systems, themselves to attend to, the specific details are not particularly well suited to direct implementation.

Clark and Schaefer's model involves the augmentation of common ground as the production of *contributions*, composed of two parts. First, the contributor specifies the content of his contribution and the partners try to register that content. Second, the contributor and partners try to reach the *grounding criterion*, which Clark and Schaefer state as follows, "The contributor and the partners mutually believe that the partners have understood what the contributor meant to a criterion sufficient for the current purpose" (Clark & Schaefer 1989, p. 262). Clark and Schaefer divide the contribution into two phases as follows (for two participants, **A** and **B**) (Clark & Schaefer 1989, p. 265):

Presentation Phase: **A** presents utterance **u** for **B** to consider. He does so on the assumption that, if **B** gives evidence **e** or stronger, he can believe that **B** understands what **A** means by **u**.

Acceptance Phase: **B** accepts utterance **u** by giving evidence **e**′ that he believes he understands what **A** means by **u**. He does so on the assumption that, once **A** registers evidence **e**′, he will also believe that **B** understands.

[1] Several authors have taken stands on distinguishing the concepts of *cooperation* and *collaboration*, e.g. (Roschelle & Teasley 1995), however others use them interchangeable y, or with opposite definitions *c.f.,* (Allwood 1976). I will not distinguish them here but will use the term collaboration without trying to engage in this ongoing debate.

[2] (Cohen & Levesque 1991a) talked about the desire for mutual belief as a motivation to perform acknowledgments, though they did not provide a model of the role the acknowledgment was supposed to play in actually achieving this mutual belief.

Clark and Schaefer claim that once both phases have been completed, it will be common ground between **A** and **B** that **B** understands what **A** meant. Each element of the contribution may take multiple conversational turns and may include whole embedded contributions. Rather than a straightforward acceptance, **B** can instead pursue a repair of **A**'s presentation, or ignore it altogether. **B**'s next turn, whether it be an acceptance, or some other kind of utterance, is itself the presentation phase of another contribution. Thus **A** must accept **B**'s acceptance, and so on.

There are different types of evidence which can be given to show understanding. The main types considered by Clark and Schaefer are shown in Table 1, in order from strongest to weakest.

The strength of evidence needed for grounding depends on several factors, including the complexity of the presentation, how important recognition is, and how close the interpretation has to be. They try to avoid infinite recursion in accepting acceptances by invoking the following **Strength of Evidence Principle:** The participants expect that, if evidence $\mathbf{e_0}$ is needed for accepting presentation $\mathbf{u_0}$, and $\mathbf{e_1}$ for accepting presentation of $\mathbf{e_0}$, then $\mathbf{e_1}$ will be weaker than $\mathbf{e_0}$.

Deficiencies of the Contribution Model

Although the contribution model is perhaps the first explicit model of how grounding takes place and why acknowledgments occur, it still is lacking in a number of particulars, especially when used to design a collaborating system.

How much acceptance is enough? Since Clark and Schaefer assert that each signal (including acceptances) must itself be a presentation which needs acceptance, it is not clear that contributions are *ever* really complete. For example, in the simplest case, *contributions by turns*, Speaker A's first turn is a presentation part of a first contribution. Speaker B's following turn is an acceptance part of that contribution, but also is the presentation part of a next contribution. What is unclear is whether this second utterance must be accepted in order to fulfill its acceptance function. If so, as Clark and Schaefer seem to imply, then not even the next utterance by A will completely ground the first contribution: this acceptance of the acceptance will itself need to be accepted, and so on, ad infinitum. If it is possible, as they suggest, that some acceptances need not be accepted themselves, then this opens the possibility that the acceptance part of an utterance need not be itself accepted (though any actual intended next contribution which is part of the same utterance would still have to be accepted). Some utterances which are merely acceptances might not be presentations as well. Once one accepts this basic premise, the question arises as to which kind of utterances do not need to be accepted. This question will be addressed further, below.

A firm supporter of Clark and Schaefer's model might retort that all acceptances *do* require acceptance, but the strength of evidence principle takes care of this by allowing for diminishing degrees of evidence. A problem with this is that eventually it will have to bottom out to zero, and how does one distinguish "little or no evidence needed" from "evidence not needed"?

Problems with Graded Evidence of Understanding Clark and Schaefer's graded types of evidence, summarized in Table 1 has been adapted by some in the community. However it has several problems. First, there's not great evidence for the ordering itself. E.g., "demonstration" is actually usually greater evidence of understanding than "display", since it shows some level of understanding as well as just perception (Allwood, Nivre, & Ahlsen 1992). A bigger problem, however, is with "initiation of next relevant contribution." The issue is when this can actually be seen as acceptance behavior, and when as mere obliviousness – e.g., if the next contribution is just as relevant as it would be if the putative contribution had never been made. This is often the case with acknowledgments, especially in the form of backchannels. Positive feedback in the form of backchannels is generally not itself acknowledged explicitly, so how does one tell whether continuing on is acceptance of the backchannel or not? Likewise, different styles of evidence of understanding may be based on other factors than a signal of strength of understanding. For example, (Walker 1993) shows how some *informationally redundant utterances* can be used to overcome resource bounds such as memory limitations in task-oriented dialogue, and lead to better task performance.

A more fruitful model would consider *conditional relevance* and *conditional non-relevance* as respective forms of positive and negative feedback. Unconditional relevance would be a much weaker sign of acceptance — at best, lack of negative feedback or statement of a problem. Likewise, unconditional irrelevance would not particularly be a signal of non-perception/understanding. The idea is this: consider a sequence of utterances, α, β, γ performed by A,B, and A, respectively. The question is whether γ, which does not refer specifically to β counts as acceptance (and to what degree). If γ is relevant after β but would not be relevant after α (without β), then we can say it is conditionally relevant to β, and is very strong evidence of having understood β. A good example is a short answer to a question, which would be incoherent if the question had not been asked and understood. If, on the other hand, γ is just as relevant after α as after β (e.g., when β is a backchannel), then it is weak evidence, at best, coming only from the lack of explicit objection to β. Likewise, if γ is not relevant as a follow-up to β, it can only be taken as a strong signal of not having perceived or understood β if it *is* relevant as a follow-on to α (or some misunderstanding of β). A true interruption can not be taken as particularly strong evidence one way or the other.

Off-line Nature of Phases It is often hard to tell whether a particular utterance is part of the presentation phase or the acceptance phase. Self-initiated self-repair (Schegloff, Jefferson, & Sacks 1977) and other-agent completions are considered part of the presentation phase, but other-repair and other-initiated self-repair are part of the acceptance phase. Either one can have embedded contributions, in the form of

1	**Display**	B displays verbatim all or part of A's presentation.
2	**Demonstration**	B demonstrates all or part of what he has understood A to mean.
3	**Acknowledgment**	B nods or says "uh huh", "yeah", or the like.
4	**Initiation of relevant next contribution**	B starts in on the next contribution that would be relevant at a level as high as the current one.
5	**Continued attention**	B shows that he is continuing to attend and therefore remains satisfied with A's presentation.

Table 1: [Clark and Schaefer, 1989, p. 267]: Types of Evidence of Understanding

insertion sequences or clarification sub-dialogues, so, in the case of an other-initiated self-repair, it's hard to tell whether it is part of the presentation phase or the acceptance phase. We often need to look at large segments of the conversation, both before and afterwards before deciding how a particular utterance fits in. The model is thus of little use to an agent in the midst of a conversation deciding what to do next based on what has happened before.

From a processing point of view, the main deficiency of Clark and Schaefer's contribution model of grounding is that there is no easy way to tell the "state" of the current contribution while engaged in a conversation. Although we might represent a contribution as a transition network such as that in Figure 1, with a grounded contribution being one in the final state, F, this is not sufficient to monitor on-line conversation.

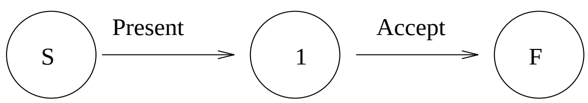

Figure 1: Transition Network for Contributions

We know that to start a contribution, a presentation must be performed primarily by one agent, whom we will call the *Initiator* (I) and then an acceptance must be performed primarily by the other agent, whom we will call the *Responder* (R), but what is less obvious is how to tell when a presentation has been performed. Another way of looking at this question is: given an utterance by the initiator, how does it function as part of the current contribution? Does it start, continue, or complete a presentation? Unfortunately, there is no way to recognize whether a presentation is complete, just by looking at an utterance itself. Consider the following sequences in examples (1) and (2).

(1)
1 I: Move the boxcar to Corning
2 I: and load it with oranges
3 R: ok

(2)
1 I: Move the boxcar to Corning
2 R: ok
3 I: and load it with oranges
4 R: ok

Since, according to Clark and Schaefer, the sequence of the first two utterances in (1) is a single contribution, that means that the presentation phase must encompass both of the utterances by the initiator. However, in (2), there are two contributions with two separate presentations by I,[3] and thus the first utterance by the initiator is a complete presentation. Since these sequences are identical up to the second utterance, there is, in general, no way to tell whether a presentation is complete until another action starts. This becomes a more serious matter because a contribution must be composed of actions by both participants, and thus there must be some way for the individual participants to determine an appropriate next action, given the current state.

While Clark and Schaefer's model is not ideally suited to the design of collaborative systems, I believe those that are better suited are those that are directly inspired by it (e.g., the work presented in the next section) or very similar in spirit (e.g., (Allwood, Nivre, & Ahlsen 1992)).

The Grounding Acts Model

In (Traum & Allen 1992; Traum & Hinkelman 1992; Traum 1994), I presented an on-line reformulation of the contribution model, which was directly used as part of a collaborative dialogue agent (Allen *et al.* 1995). This model collapses the different types of acceptance, but extends the building blocks of the units of common ground to those that could be realized with a single utterance, thus allowing an agent to track progress with each communication and without requiring lookahead.

Rather than the two phases of presentation and acceptance (each of which could contain recursively defined contributions), the basic building blocks are a set of *Grounding Acts*, each of which is identified with a particular utterance unit, and performs a specific function toward the achievement of common ground. The set of grounding acts are presented in Table 2. In this model, the units of grounded content are called *Discourse Units* (or DUs, for short), rather than *Contributions*.[4] Individual grounding acts could add to or

[3] Clark & Schaefer would also analyze the acknowledgments by R as the presentation phases of separate contributions.

[4] This was done for several reasons. First, it is not so clear that the two concepts are really modeling the same thing, given that Contributions are potentially recursive and always serially constructed, and can cover large spans of a dialogue, while DUs are purely local and modeled by finite state transitions. Also, the term "contribution" is somewhat unfortunate, since many others (e.g., (Grice 1975; Allwood 1995) use it for what Clark calls a *presentation*. Of course the term *Discourse Unit* is similarly

change content of the unit, which included the information that would be added to the common ground if the unit was seen as completed.

Label	Description
initiate	Begin new DU, content separate from previous uncompleted DUs
continue	same agent adds related content to open DU
acknowledge	Demonstrate or claim understanding of previous material by other agent
repair	Correct (potential) misunderstanding of DU content
Request Repair	Signal lack of understanding
Request Ack	Signal for other to acknowledge
cancel	Stop work on DU, leaving it ungrounded and ungroundable

Table 2: Grounding Acts

Acts are also distinguished as to whether they are performed by the initiator of a unit (I) or the responder (R)[5] As well as the three states in the transition network in Figure 1, representing starting grounding (S), acceptance/acknowledgment needed for grounding (1), and grounded material (F), the model in (Traum & Allen 1992), contains four other states, representing the need for a repair by initiator and acknowledgment by responder (2), need for acknowledgment by initiator (3), need for repair by responder and acknowledgment by initiator (4), and a "dead state" (D), in which the unit was deemed to be ungroundable (though the same content could, of course be reintroduced and grounded in a subsequent unit). Table 3 shows the transition diagram, given each state and possible grounding acts (superscripts represent the performing agent, with respect to the initiator of this unit).

A couple of other remarks on this model are in order. The grounding act, *acknowledge* is meant to cover the whole spectrum of Clark and Schaefer's *evidence of understanding*, rather than just the middle category. The term *acceptance* was not used, since that can cause confusion with accepting the main informational content, rather than just indicating that it was understood. Also, although grounding acts are associated with utterances, there is not necessarily a one-to-one relationship – as with Clark and Schaefer's Schaefer's model, a single utterance could correspond to multiple grounding acts related to different discourse units. A distinction, though, is that it is also possible for an utterance to be only a part of one unit.

The grounding acts model has clear (though not necessarily correct) positions with respect to the difficulties with

Next Act	In State						
	S	1	2	3	4	F	D
initiate^I	1						
continue^I		1			4		
continue^R			2	3			
repair^I		1	1	1	4	1	
repair^R		3	2	3	3	3	
ReqRepair^I			4	4	4	4	
ReqRepair^R		2	2	2	2	2	
ack^I				F	1	F	
ack^R		F	F			F	
ReqAck^I		1				1	
ReqAck^R				3		3	
cancel^I		D	D	D	D	D	
cancel^R			1	1		D	

Table 3: DU State Transition Diagram

the contributions model, mentioned in the previous section. By using utterance-level grounding acts, and with reference to the FSA in Table 2, it is clear, within the model, how the current state of grounding would be represented, without needing subsequent utterances to clarify the situation. E.g., the examples in (1) and (2) are analyzed as in (3) and (4), respectively. Thus, the parts that are the same are coded in the same way, while still maintaining the distinction in numbers of units and size of grounding unit from the contribution model.

(3)

utt:	**Grounding Act**	**DU1**
1:	$\mathrm{init}^{\mathrm{I}}(1)$	1
2:	$\mathrm{cont}^{\mathrm{I}}(1)$	1
3:	$\mathrm{ack}^{\mathrm{R}}(1)$	F

(4)

utt:	**Grounding Act**	**DU1**	**DU2**
1:	$\mathrm{init}^{\mathrm{I}}(1)$	1	
2:	$\mathrm{ack}^{\mathrm{R}}(1)$	F	
3:	$\mathrm{init}^{\mathrm{I}}(2)$	F	1
4:	$\mathrm{ack}^{\mathrm{R}}(2)$	F	F

The problem of graded evidence of understanding has been mostly side-stepped. There is only one degree of understanding feedback (termed *acknowledgment*), and only two degrees of grounding: *grounded* and *ungrounded*. The grounding criterion is reduced to a decision about which content needs to be grounded, and which can be assumed to be grounded without explicit acknowledgment. Grounding acts, themselves, are not seen as part of the content, and do not need to initiate recursive DUs to ground recognition of their occurrence. In addition it is possible to add some content to the newly grounded unit as part of the content of the acknowledgment, though it is also possible to put this content as part of an initiation of a new unit. There are actually several different instantiations of the model possible, depending on what constraints are put on the allowance of adding new material to a (now) grounded DU with the acknowledgment. In (Traum & Hinkelman 1992), no additional material was added, all new material being the content of an *initiate* act of a new DU. In (Traum 1994),

vague and subject to multiple usage. For this reason we used the term *Common Ground Unit* (CGU) in (Nakatani & Traum 1999).

[5]No attempt has been made to try to model grounding among groups of larger than two persons. The contribution model explicitly talks about partner*s*, though it was also applied primarily to pairs. (Novick, Walton, & Ward 1996) adapt the contribution model to larger groups.

simple kinds of “acceptance”, e.g., that could be achieved with a short answer like “yes”, or “ok” were included as content of the acknowledgment (and did not start a new DU which needed further grounding). A current hypothesis is that no *backward function* (Discourse Resource Initiative 1997) including answers and expressions of (dis)agreement requires acknowledgment (however, the same utterance will often also have a forward function, which *will need to be acknowledged*.

The grounding acts model also allows (but does not explain a need for) multiple acknowledgments of material already deemed to be grounded (e.g., the transitions from F to F, given an ack, in Table 3). Likewise, until a fixed number of subsequent DUs have been opened, it is also possible to repair a unit previously thought to be grounded, so the F state is a current “best guess” about what was grounded, rather than a final, irrevocable determination.

Deficiencies of the Grounding Acts Model

While the grounding acts model addressed some of the specific deficiencies of Clark and Schaefer’s contribution model, it is by no means a complete and accurate model of the grounding process. Several simplifications were made, and several open problems remain.

Degrees of Groundedness The binary “grounded or ungrounded” distinction in the grounding acts model is clearly an oversimplification. While there are difficulties with assigning strength of evidence directly to presentation type, it does seem that there are both different amounts of understanding and differences in effects of particular signals and the resulting (assumed) understanding. However, without a model of what the relevant degrees are, or how they are achieved, it may be better to err on the side of simplicity, hence the decisions made here.

Granularity of Utterance Units Although the grounding acts model does not require any particular sized utterance unit, the determination of the quantity, sequence, and identity of these acts is highly dependent on how a communication is carved up into utterance units. Most trivially, a communication broken into larger units will have fewer *continues* than the same communication broken into smaller-sized units. Likewise, if a disfluency or speech-repair occurs inside a unit, one would generally not mark a distinct grounding act, however, if the whole unit is the repair, then it would be marked as a *repair* grounding act. This can be a difficulty for arriving at reliable coding, using different sized units, but is not really a problem once a unit size has been settled on, since the definitions of the acts will be clear in either case. Most work has assumed units roughly the size of intonation phrases (Pierrehumbert & Hirschberg 1990), with long pauses being a secondary factor (as seems to be indicated by the analysis in (Traum & Heeman 1997)). However, both larger units (roughly the size of complete speech acts – (Zollo & Core 1999)), and smaller units (*micro conversational events* (Poesio & Traum 1997), roughly the size of words or the the meaning units of (Hobbs 1985) – Chapter 7 of (Traum 1994)) have been used.

Discreteness of Grounding Acts Although the grounding acts in Table 2 are well-defined in terms of context of occurrence (currently accessible DUs and states) and effects (on DU state and material in DU), it can often be difficult to recognize or decide which act has been performed, when the evidence is less clear about what the speaker is trying to do. As an example, a repetition or reformulation of what has just been said could be the performance of any of three grounding acts: acknowledgment, repair, or request repair. That is, it could be a signal of having understood correctly (and demonstrating that fact), having heard something that needs changing (and presenting the needed change), or asking if that is what was said. Each of these acts would result in a different DU state, with different implications for what else must be performed for a grounded unit: for an acknowledgment, the DU is (seen as) grounded and nothing more must be done; for a repair, an acknowledgment by the initiator is required; for a request repair, the initiator must repair (in this case, perhaps simply confirming the hypothesis), and then responder acknowledge.

For many of these sorts of utterances, often called *verifications* (Smith & Hipp 1994), it may actually be an in-between sort of act, where it is an acknowledgment if it happens to be correct, but a request for repair if it happens to be incorrect. Moreover, the same sequence can often result, regardless of the initial attribution (and attribution of subsequent acts). It may be better to allow some sort of vague act that encompasses these multiple possibilities, in which disambiguation/precisification is necessary only if one would follow a sequence that is prohibited by one of the possibilities (e.g., not responding is okay if the previous act was an acknowledgment, but not if it was a request for repair). A difficulty is that this would require a slightly different computational model to accommodate this sort of vague act (e.g., perhaps a probabilistic FA, or at least one with more states to capture the different states for vague acts).

A related issue is cases in which a followup utterance seems to indicate understanding of some *part* of the preceding content, while also indicating a lack of understanding of other parts. Many requests for repair are actually of nature. E.g, in (5), the request repair in 57.1 explicitly requests repair of the destination, but gives evidence of understanding that a route is involved. This is not just a symptom of the choice of utterance unit size, since there is no necessary sequential order in which information can be signaled as understood or not understood. A solution would be to dynamically *split* DUs into multiple parts, which enter different states. Thus, a new act would be needed to split a DU (as well as the subacts of acknowledge and request repair for each of the new DUs).

(5)

56.2	S: and then which route do you wanna take
57.1	M: took to Avon

Insensitivity to Media and Content Criteria The grounding acts model is not particularly well-suited to explaining behavior in modalities other than spoken language. While the grounding acts themselves are not defined in a media-dependent way, the particular finite state transition network and interpretation of a DU as grounded vs. ungrounded does not line up particularly well with observed behavior in other media. As (Clark & Brennan 1991) point out, grounding behavior varies widely on the media used. For example, in the trains-93 corpus, concerning spoken problem-solving dialogues, over 93% of substantive turns evoked responsive grounding behavior by the other participant (Traum & Heeman 1997). On the other hand, in a mystery solving task in a virtual world, using MOO communication (Curtis 1993) and a shared whiteboard, only 41% of communications evoked responsive grounding behavior (Dillenbourg *et al.* 1997). Some of this may be accounted for by the grounding criterion, in terms of what is important to be grounded, but more important are the different costs and affordances of the media involved. E.g., for a persistent medium like the shared whiteboard, something much closer to the mutual co-presence assumptions of (Clark & Marshall 1981) seems to be operative, rather than a model like Clark and Schaefer's contributions or the grounding acts model.

Towards a more complete psychological model for collaborative systems

Clark and Schaefer's contribution model provides a powerful technique for evaluating what content in dyadic conversation can be considered grounded. However, as discussed above, it is not particularly useful as a computational online model to assess the state mid-contribution and help a computational system decide what to do next.[6] The grounding acts model was more applicable in this regard, and directly usable as part of the dialogue modeling and management components of a dialogue system (Traum 1994; Allen *et al.* 1995), but did not have much to say about how to realize particular grounding acts, nor the subtleties occurring from different media and types of content.

In recent work, we have been developing a model that can serve both as descriptive of the kinds of generalized grounding acts that agents perform in complex tasks, using multiple media, and as a resource for an agent to operate in such a setting (Dillenbourg, Traum, & Schneider 1996; Traum & Dillenbourg 1996; 1998). We represent these generalized grounding acts in the form: $\alpha \rightarrow \mu$, meaning that act α is performed in order to contribute to the grounding of content μ.

Several factors are important in the determination of the utility of performing a particular grounding act. The first consideration does not particularly involve the individual act, but rather the *grounding criterion*. We term this **GC**(μ). If this criterion is low, there is no need for the information to be grounded, and thus a low utility for an agent to perform some action to ground it. The importance for grounding a particular piece of information also depends on the cost (with respect to the task) of non-grounding – how will task performance degrade if the particular μ is not grounded. An important factor in this is the persistence of the validity of the content – it is a waste of time to ground highly transient information that changes before it's current value can be of use.

The next consideration is how much performance of α will increase the groundedness of μ. We represent this as a differential between the groundedness given performance of α: $\mathbf{G}_\alpha(\mu)$, and the prior groundedness: $\mathbf{G}(\mu)$.[7] If something is already well grounded (with respect to the grounding criterion, $GC(\mu)$, we are not likely to need any additional action to ground it further. Likewise, if performance of α will not increase the groundedness of μ, then it will not be very useful (as a grounding act). Another important factor is the perplexity of the information – the probability that some information believed to be grounded could be misunderstood or disagreed upon.

Finally, we consider the cost of performing α: $\mathbf{C}(\alpha)$. This includes not only the effort required to perform the action, but also the affiliated costs of understanding α as conveying μ, including potential further actions (repairs) which may be required. Also important are *opportunity costs* relating to not performing some other action at this point, which may become more expensive later, due to implicatures and/or a lack of coherence with current context. If these costs are high, there is not as much utility in performing the action, while if they are low, the action may be performed, even when grounding is not particularly crucial. $C(\alpha)$ depends not only on the features of the medium itself, as described by (Clark & Brennan 1991) but also on the matching between the actor and the medium, e.g., how familiar the actor is with the medium. For example sometimes collaborators use an apparently more expensive medium simply because they know how to use it and are reluctant to learn something new (e.g., a new computer interface), which requires an extra learning cost.

Our initial attempt at a predictive and normative account of grounding behavior is given in (6), where the left side represents the utility of performing a particular action α which communicates μ.

(6) $U(\alpha \rightarrow \mu) \propto \frac{GC(\mu) * (G_\alpha(\mu) - G(\mu))}{C(\alpha)}$

(6) shows only the isolated utility of performing α to ground μ. Other considerations must also be taken into account when selecting an action, such as the utility of other actions α', that might help ground μ, as well as what other uses α and other actions might have, independently from their effect on the grounding of μ. For the former, the grounding criterion and prior groundedness of μ will still

[6]See (Cahn & Brennan 1999) for a computational reformulation.

[7]We are simplifying slightly here. A temporal argument is omitted here, which will be important for calculating prior groundedness as well as the effect of the action. Also, we only care about grounding up to the grounding criterion, so these terms should really be min($GC(\mu), G_\alpha(\mu)$), and min($GC(\mu), G(\mu)$), respectively.

be useful, while for the latter some more global arbitration strategy must be used to set priorities.

Questions for Discussion

I want to conclude with three questions for discussion.

1. **What does the grounding criterion really mean?** The grounding criterion is an important concept, both for escaping from the dilemma that no amount of acknowledgment can guarantee total mutual knowledge, and to help explain the differences in observed grounding behavior. There are actually two different but important criteria related to a formulation such as (6). One is related to a scale of of grounding – how much grounding is enough (for the present purposes). A second is related to a scale of importance – how important is it that grounding reach this level. These two can, of course, be combined into a function of utility given a particular degree of grounding and current purposes. Related to the first point is the whole question of degrees of groundedness. As mentioned above, the types of evidence in Table 1 does not make a particularly good scale. There are a number of scales around, though none seem completely satisfactory. Another is based on accessibility,perception,understanding, and agreement (Dillenbourg, Traum, & Schneider 1996; Allwood, Nivre, & Ahlsen 1992). Another could be the degree of truncation heuristics from (Clark & Marshall 1981). others could allow for explicitness of information, persistence of media and similar notions.
2. **How can we provide empirical validation/disconfirmation for different theories of grounding?** Grounding *should* be an empirical question, although it is very difficult to test concepts such as *mutual belief*. There have been experiments related to performance of grounding behavior and other behavior related to assumptions of grounding (e.g., (Brennan 1990)), however most of these are also subject to other possible interpretations, e.g., that only one-sided understanding was necessary rather than mutuality. In this sense, computer systems may be good test-beds for particular theories, since their representations as well as behavior can be evaluated.
3. **Should grounding models be explicit in computational HCI systems?** While it is clear that effective collaborative systems must employ the use of grounding-related feedback, what is less clear is whether there must be an explicit model of grounding that is referred to in the system's performance and interpretation of communications, or whether a system could be designed to behave properly without such an explicit model.

References

Allen, J. F.; Schubert, L. K.; Ferguson, G.; Heeman, P.; Hwang, C. H.; Kato, T.; Light, M.; Martin, N.; Miller, B.; Poesio, M.; and Traum, D. R. 1995. The TRAINS project: a case study in building a conversational planning agent. *Journal of Experimental and Theoretical Artificial Intelligence* 7:7–48.

Allwood, J.; Nivre, J.; and Ahlsen, E. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics* 9.

Allwood, J. 1976. *Linguistic Communication as Action and Cooperation*. Ph.D. Dissertation, Göteborg University, Department of Linguistics.

Allwood, J. 1995. An activity based approach to pragmatics. Technical Report (GPTL) 75, Gothenburg Papers in Theoretical Linguistics, University of Göteborg.

Brennan, S. 1990. *Seeking and Providing Evidence for Mutual Understanding*. Ph.D. Dissertation, Stanford University.

Cahn, J. E., and Brennan, S. E. 1999. A psychological model of grounding and repair in dialog. In *Working Notes AAAI Fall Symposium on Psycholiogical Models of Communication in Collaborative Systems*.

Clark, H. H., and Brennan, S. E. 1991. Grounding in communication. In Resnick, L. B.; Levine, J.; and Teasley, S. D., eds., *Perspectives on Socially Shared Cognition*. APA.

Clark, H. H., and Marshall, C. R. 1981. Definite reference and mutual knowledge. In Joshi, A. K.; Webber, B. L.; and Sag, I. A., eds., *Elements of Discourse Understanding*. Cambridge University Press. Also appears as Chapter 1 in (Clark 1992).

Clark, H. H., and Schaefer, E. F. 1989. Contributing to discourse. *Cognitive Science* 13:259–294. Also appears as Chapter 5 in (Clark 1992).

Clark, H. H. 1992. *Arenas of Language Use*. University of Chicago Press.

Cohen, P. R., and Levesque, H. J. 1991a. Confirmations and joint action. In *Proceedings IJCAI-91*, 951–957.

Cohen, P. R., and Levesque, H. J. 1991b. Teamwork. *Nous* 35.

Curtis, P. 1993. LambdaMOO programmer's manual. Xerox Parc.

Dillenbourg, P.; Jermann, P.; Schneider, D.; Traum, D.; and Buiu, C. 1997. The design of moo agents: Implications from a study on multi-modal collaborative problem solving. In *Proceedings of the 8th World Conference on Artificial Intelligence in Education (AI-ED 97)*, 15–22.

Dillenbourg, P.; Traum, D.; and Schneider, D. 1996. Grounding in multi-modal task-oriented collaboration. In *Proceedings of the European Conference on AI in Education*.

Discourse Resource Initiative. 1997. Standards for dialogue coding in natural language processing. Report no. 167, Dagstuhl-Seminar.

Grice, J. P. 1975. Logic and conversation. In Cole, P., and Morgan, J. L., eds., *Syntax and Semantics*, volume 3: Speech Acts. Academic Press. 41–58.

Grosz, B. J., and Kraus, S. 1993. Collaborative plans for group activities. In *Proceedings IJCAI-93*, 367–373.

Grosz, B. J., and Sidner, C. L. 1990. Plans for discourse. In Cohen, P. R.; Morgan, J.; and Pollack, M. E., eds., *Intentions in Communication*. MIT Press.

Halpern, J. Y., and Moses, Y. 1990. Knowledge and common knowledge in a distributed environment. *Journal of the ACM* 37(3):549–587.

Hobbs, J. 1985. Ontological promiscuity. In *Proceedings ACL-85*, 61–69.

Nakatani, C. H., and Traum, D. R. 1999. Coding discourse structure in dialogue (version 1.0). Technical Report UMIACS-TR-99-03, University of Maryland.

Novick, D. G.; Walton, L.; and Ward, K. 1996. Contribution graphs in multiparty conversations. In *Proceedings of the International Symposium on Spoken Dialogue (ISSD-96)*, 53–56.

Pierrehumbert, J., and Hirschberg, J. 1990. The meaning of intonational contours in the interpretation of discourse. In Cohen, P. R.; Morgan, J.; and Pollack, M. E., eds., *Intentions in Communication*. MIT Press.

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

Roschelle, J., and Teasley, S. 1995. The construction of shared knowledge in collaborative problem solving. In O'Malley, C., ed., *Computer-Supported Collaborative Learning.* Springer-Verlag.

Schegloff, E. A.; Jefferson, G.; and Sacks, H. 1977. The preference for self correction in the organization of repair in conversation. *Language* 53:361–382.

Smith, R. W., and Hipp, D. R. 1994. *Spoken Natural Language Dialog Systems: A Practical Approach*. Oxford University Press.

Traum, D. R., and Allen, J. F. 1992. A speech acts approach to grounding in conversation. In *Proceedings 2nd International Conference on Spoken Language Processing (ICSLP-92)*, 137–40.

Traum, D. R., and Dillenbourg, P. 1996. Miscommunication in multi-modal collaboration. In *Proceedings of the* AAAI-96 Workshop on Detecting, Repairing, and Preventing Human-Machine Miscommunication.

Traum, D. R., and Dillenbourg, P. 1998. Towards a normative model of grounding in collaboration. In *Proceedings of the* ESSLLI98 workshop on Mutual Knowledge, Common Ground and Public Information.

Traum, D. R., and Heeman, P. 1997. Utterance units in spoken dialogue. In Maier, E.; Mast, M.; and Luperfoy, S., eds., *Dialogue Processing in Spoken Language Systems — ECAI-96 Workshop*, Lecture Notes in Artificial Intelligence. Springer-Verlag, Heidelberg. 125–140.

Traum, D. R., and Hinkelman, E. A. 1992. Conversation acts in task-oriented spoken dialogue. *Computational Intelligence* 8(3):575–599. Special Issue on Non-literal language.

Traum, D. R. 1994. *A Computational Theory of Grounding in Natural Language Conversation*. Ph.D. Dissertation, Department of Computer Science, University of Rochester. Also available as TR 545, Department of Computer Science, University of Rochester.

Walker, M. A. 1993. *Informational Redundancy and Resource Bounds in Dialogue*. Ph.D. Dissertation, University of Pennsylvania.

Zollo, T., and Core, M. 1999. Automatically extracting grounding tags from bf tags. In *Proceedings of the ACL-99 workshop Towards Standards and Tools for Discourse Tagging*, 109–114. Association for Computational Linguistics.