Towards a Normative Model of Grounding in Collaboration

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

Pierre Dillenbourg
Pierre Dillenbourg@tecfa.unige.ch
TECFA, FPSE
Université de Genève
9, Route De Drize, BAT D
CH-1227 Carouge, Switzerland

We investigate the process by which individuals add information is to the common ground. This process, called *grounding* [Clark and Schaefer, 1989] involves performing *communicative actions* of various sorts. The actions include those which present content to be added, as well as feedback [Allwood *et al.*, 1992], displaying the reaction to that content. Feedback can be either positive — an acknowledgment of understanding, or negative — repairing or displaying a lack of comprehension. We will call any action (whether presentation or feedback) that contributes to grounding, a *grounding action*.

There have been several theories in the cognitive science literature about what minimal set of grounding actions and conditions is necessary for content to be considered grounded. [Clark and Marshall, 1981] suggested that copresence of some sort, along with other assumptions, are one means to justify the assumption of common ground. [Clark and Schaefer, 1989], on the other hand, contended that each presentation must be accepted, with an appropriate level of positive feedback. [Traum and Allen, 1992], in presenting a computational account, took a somewhat intermediate approach, claiming that acknowledgments *themselves* did not require additional feedback to have their grounding effect. These theories can provide justification for an assumption about whether or not some content is part of the common ground, given some sequence of communicative acts. However, there has been comparatively less work on explaining exactly what kind, if any, grounding act should be made by a communicator at a particular time.

Corpus studies have shown that the level and type of feedback varies widely with the medium of communication, among other things. E.g., [Traum and Heeman, 1997] report that in a corpus of task-oriented spoken dialogue, over 95% of speaker turns begin with some form of feedback on the previous turn[1]. On the other hand, we have seen in multi-modal computer-mediated collaboration experiments that typed "chat" messages received a direct acknowledgment only 41% of the time, while information displayed on a shared whiteboard received significantly fewer acknowledgments [Dillenbourg *et al.*, 1997a]. Media factors can be used to explain which content can be considered grounded, with reference to some of the assumptions like those of [Clark and Marshall, 1981], which are present in the media. For example, speech has no temporal persistence, and so is either perceived immediately or is gone, while messages on a shared whiteboard are present until explicitly removed. However, there is still a problem: communicators do not always (fully) ground all relevant content — how is a communicator, or the designer of an artificial agent that communicates, to decide whether to perform any grounding actions at a given time?

Several factors are important in the determination of the utility of performing a particular grounding act. We represent a grounding act as $\alpha \rightarrow \mu$, meaning that act α is performed in order

[1] In the case where the previous turn consisted of more than a simple acknowledgment such as "ok".

to contribute to the grounding of content μ. One consideration does not particularly involve the individual act, but rather how important it is that the content μ, becomes grounded, the *grounding criterion* of [Clark and Wilkes-Gibbs, 1986]. Clark and Wilkes-Gibbs point out that this criterion varies with the overall purposes and local state of the collaborators. We term this $\mathbf{GC}(\mu)$. 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 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: $\mathbf{G}(\mu)$.[2] 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. 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 and 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 (1), where the left side represents the utility that a particular action α which communicates μ will be performed.

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

(1) 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 be useful, while for the latter some more global arbitration strategy must be used to set priorities. In some domains, time, focus, and local context are also important factors – sometimes one must perform some acts before the balance of costs and benefits have been changed, while other acts have more flexibility.

Several authors (e.g., [Allwood *et al.*, 1992, Clark, 1994, Dillenbourg *et al.*, 1996] have noted that common ground, itself, is not a monolithic phenomenon, but that there are several different aspects, all of which must be coordinated. The levels picked out by these authors can be summarized as:

1. **access** (generalized from **contact**): do the collaborators have access to each others communicative actions.

[2] 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.

2. **perception:** do the collaborators perceive the communicative actions that are performed
3. **understanding** do the collaborators understand what is meant by the communicative actions
4. **agreement** do the collaborators reach agreement about the communicated facts or plans.

Most work specifically on grounding has focused on levels (2) and (3). (1) is seen as a pre-requisite to grounding, but not much considered by prior work, beyond serving as a motivation for *summons* actions and certain very weak kinds of feedback. (4) is, in many respects, the whole point of communication. There has been a fair amount of work on the negotiation process by which collaborators reach agreement about the facts of the world, or plans for action, as well as some computational accounts of what kinds of language acts can be used to achieve this agreement (e.g., [Baker, 1994, Sidner, 1994]).

There are at least two ways in which these levels of common ground can be applied to the formulation in (1). First, One could see these levels as degrees of common ground as applied to the meaning of a communicative act, $\alpha \rightarrow \mu$, e.g.:

0. α inaccessible
1. α accessible but not perceived
2. α perceived but not understood as μ
3. μ understood but not agreed upon
4. μ agreed

In this case, one would treat the grounding criterion (GC) and groundedness (G) as relative to these states. One would also want to add additional levels for situations in which μ is inferable from other information available or provided, even if not directly communicated.

Another, and perhaps better, way to view these levels, however is to notice that the formula itself can be applied to each of these levels. E.g., for each of them, there can be different grounding criteria, prior and posterior groundedness, and costs. While [Clark and Wilkes-Gibbs, 1986] applied their grounding criterion specifically to understanding (and perception), we think the same notion applies just as well for accessibility and agreement: for information specifically related to the task goals or collaborative plans, the grounding criterion is high for agreement, and thus worth an extensive debate, if necessary. On the other hand, lots of other potential points of disagreement need not be resolved if they have only a minor influence on the overall objectives (e.g, specific sub-actions, or reasons for doing or believing something). Likewise, in a collaborative situation, the *ability* to communicate and receive messages is very important, in case something comes up requiring urgent attention, but in some phases of a task involving individual actions, it may not be so important to actually have access to individual communicative actions (and certainly not as important to perceive or understand them). Likewise, the higher level demands can set the criterion for a lower level function (e.g., one needs to have access and perceive in order to understand).

We have tried to apply this analysis to understand the behavior of agents collaborating in a multi-modal computer mediated problem solving environment. In the BOOTNAP project, [Dillenbourg *et al.*, 1996, Traum and Dillenbourg, 1996, Dillenbourg *et al.*, 1997a], two participants used a MOO and a shared electronic whiteboard to communicate and solve a murder mystery embedded in the MOO.[3] In this environment, the collaborators had a range of possible action types that they

[3] See [Dillenbourg *et al.*, 1996, Dillenbourg *et al.*, 1997a] for more discussion of how the MOO was used in this task, and [Curtis, 1993] for more on MOOS in general.

could use to communicate, including: drawing on the whiteboard, putting text messages on the whiteboard, acting in the moo (e.g., moving locations or asking suspects questions), and two types of text message sending in the MOO ("SAY" which requires being in the same virtual room, and "PAGE" which doesn't, but incurs extra reading costs).

We can see that these different media have different costs and grounding-related benefits at the various levels. For instance, collaborators are always accessible by PAGE, but not by SAY. When the collaborators need to have extended discussions, they often act to reach a higher level of access, so that the lower cost SAY command can be used. Also, information posted on the whiteboard is always accessible, and is generally more perceivable and has a higher degree of groundedness, than MOO messages, since they persist on the whiteboard until erased, while MOO messages scroll up the screen and can be hard to notice when superseded by others. Another important factor for us what the nature of the information, itself. Even for propositional information about the domain, we made a distinction between *facts*, which meant information directly provided by the MOO, as clues or answers to questions, and *inferences*, which the collaborators had to infer from facts on their way to solving the mystery. At the level of understanding, there was no particular difference between these two, since they both had the same degree of complexity and ambiguity. On the other hand, at the level of agreement, there was much less prior grounding of inferences, since these could easily be disputed by alternate theories, and might depend on facts (or other inferences) which were not available to both parties. Given this combination of factors, we would expect the following relationships of acknowledgments (extra grounding acts beyond the initial one mentioned): more acknowledgments for MOO than whiteboard, more acknowledgments for inferences than facts.

This is precisely what occurred: in out 20 pairs of collaborators, we observed the following rates of acknowledgment (The interaction effect is significant: $F=6.09$, $df=4$, $p=0.001$):

	MOO Talk	Whiteboard
Facts	0.37	0.06
Inferences	0.50	0.38

Further experiments have also qualitatively validated other aspects of (1). For instance [Montandon, 1996, Dillenbourg *et al.*, 1997b] performed a controlled experiment with tasks in which the grounding criterion was high for a particular type of information, and in which the degree of information provided by particular acts was controlled. As expected, more relevant actions were performed in the information poor setting. What remains is a more quantitative account of the aspects shown in (1). Preliminary work relies on the degree of availability of the information (e.g, including factors like persistence of media presentation, e.g., whiteboard vs. MOO) and quantity and type of explicit feedback actions (was something acknowledged or not, and if so, how (e.g., some of the different styles mentioned by [Clark and Schaefer, 1989]. We intend to test various formulations both by further experiments to evaluate grounding related behavior under various conditions, as well as by building such a model into the performance of artificial collaborating agents.

Acknowledgments

This work was funded by the Swiss National Science Foundation (Research contract 1114-040711.94). The first author was also supported in part during the writing of this paper by NSA Contract MDA904-96-C-1250. We would like to thank members of the TECFA Unit (especially Daniel Schneider, Patrick Jermann, Lydia Montandon and Patrick Mendelsohn) for helpful discussions on the issues presented in this paper.

References

[Allwood *et al.*, 1992] Jens Allwood, Joakim Nivre, and Elisabrth Ahlsen. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9, 1992.

[Baker, 1994] Michael Baker. A model for negotiation in teaching-learning dialogues. *Journal of Artificial Intelligence in Education*, 5(2):199–254, 1994.

[Clark and Brennan, 1991] Herbert H. Clark and Susan E. Brennan. Grounding in communication. In L. B. Resnick, J. Levine, and S. D. Teasley, editors, *Perspectives on Socially Shared Cognition*. APA, 1991.

[Clark and Marshall, 1981] Herbert H. Clark and Catherine R. Marshall. Definite reference and mutual knowledge. In Aravind K. Joshi, Bonnie L. Webber, and Ivan A. Sag, editors, *Elements of Discourse Understanding*. Cambridge University Press, 1981. Also appears as Chapter 1 in [Clark, 1992].

[Clark and Schaefer, 1989] Herbert H. Clark and Edward F. Schaefer. Contributing to discourse. *Cognitive Science*, 13:259–294, 1989. Also appears as Chapter 5 in [Clark, 1992].

[Clark and Wilkes-Gibbs, 1986] Herbert H. Clark and Deanna Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22:1–39, 1986. Also appears as Chapter 4 in [Clark, 1992].

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

[Clark, 1994] Herbert H. Clark. Managing problems in speaking. *Speech Communication*, 15:243 – 250, 1994.

[Curtis, 1993] Pavel Curtis. LambdaMOO programmer's manual. Xerox Parc, 1993.

[Dillenbourg *et al.*, 1996] Pierre Dillenbourg, David Traum, and Daniel Schneider. Grounding in multi-modal task-oriented collaboration. In *Proceedings of the European Conference on AI in Education*, 1996.

[Dillenbourg *et al.*, 1997a] P. Dillenbourg, P. Jermann, D. Schneider, D. Traum, and C. Buiu. 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)*, pages 15–22, 1997.

[Dillenbourg *et al.*, 1997b] P. Dillenbourg, L. Montandon, D. Traum, and C. Buiu. Does text-based virtual space impact on collaboration? In *roceedings of the* Workshop on Collaborative Learning/Working Support System with Networking*, 8th World Conference on Artificial Intelligence in Education, Kobe*, 1997.

[Montandon, 1996] Lydia Montandon. Etude des méchanisms de coordination spatiale dans un environnement virtuel de collaboration. Diplôme Thesis, TECFA, FPSE, Université de Genève, 1996.

[Sidner, 1994] Candace L. Sidner. An artificial discourse language for collaborative negotiation. In *Proceedings of the forteenth National Conference of the American Association for Artificial Intelligence (AAAI-94)*, pages 814–819, 1994.

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

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

[Traum and Heeman, 1997] David R. Traum and Peter Heeman. Utterance units in spoken dialogue. In Elisabeth Maier, Marion Mast, and Susann Luperfoy, editors, *Dialogue Processing in Spoken Language Systems — ECAI-96 Workshop*, Lecture Notes in Artificial Intelligence, pages 125–140. Springer-Verlag, Heidelberg, 1997.