

A “SPEECH ACTS” APPROACH TO GROUNDING IN CONVERSATION*

David R. Traum and James F. Allen
Computer Science Department
University of Rochester
Rochester, New York 14627-0226 USA
{traum, james}@cs.rochester.edu

Abstract

We propose that *Grounding*, the process of achieving mutual understanding between participants in a conversation, be analyzed in terms of the actions performed by the conversants which contribute to achieving this mutual understanding. We propose a set of *Grounding Acts* which facilitate this analysis. This paper describes Grounding Acts, and a “grammar” stipulating which series of performance of grounding acts result in grounded content.

1 Introduction

Speech act theory has been used in AI to provide an account of how language can be used by agents to achieve communicative goals (e.g. [4; 7]). One of the main effects in most accounts of speech acts is a change in the mutual beliefs, or *common ground* of the agents, but most accounts of speech acts have made overly strong predictions of when mutual beliefs are expected to result. This has been a result of assuming that utterances are heard and understood correctly by the listener as they are uttered and that speech acts

*Appears in Proceedings International Conference on Spoken Language Processing (ICSLP-92), pages 137-40, October 1992

are single agent plans executed by the speaker with the listener only passively present.

Theories incorporating these assumptions can not account for the frequent acknowledgements and occasional repairs that occur in spoken conversation. Mutual belief seems to require positive effort by both parties in a conversation, and thus traditional speech acts are inherently multi-agent actions, requiring several utterances in order to be realized. This process of reaching common ground is called *grounding*. Clark & Schaefer [2] present a structural model of grounding based on hierarchical “contributions”, however this model presents a number of difficulties for on-line processing of discourse, since the functions of a particular utterance can often only be determined after analysis of later utterances, and there is no straightforward mapping between their “acts” and the intentions of performing agents.

2 Grounding Acts

Our solution to the problem is to introduce a level of action below that of traditional speech acts to model the grounding process. **Grounding Acts** are performed at the level of a single utterance unit (UU) - roughly an intonation phrase, and build up to a level of common ground that is necessary for communication of beliefs, intentions, and obligations. We call this level at which common ground is built up a **Discourse Unit (DU)**. A DU consists of a series of grounding acts performed by different conversational participants. A DU may contain one or more (partial) traditional speech acts. Table 1 summarizes the multi-level theory of conversation acts described in [14; 15]. Turn-taking acts, which coordinate who’s turn it is to talk, occur within the scope of a single utterance unit. Grounding acts occur at the level of a single utterance unit, and coordinate the grounding of content. Core speech acts, the analogue of the traditional speech acts - though in our formulation requiring grounding acts by both parties to be successfully performed, are realized at the DU level. Argumentation acts range over multiple DUs, meeting higher level conversational goals than a single core speech act.

We use the following set of acts to model the grounding process:

Initiate An initial utterance component of a Discourse unit - traditionally this utterance alone has been considered sufficient to accomplish the core speech act. An **initiate** usually corresponds to what Clark &

Level	Act Type	Sample Acts
<UU	Turn-taking	take-turn keep-turn
UU	Grounding	Initiate Continue Ack Repair
DU	Core Speech Acts	Inform YNQ Accept Request
>DU	Argumentation	Elaborate Q&A Clarify Repair

Table 1: Conversation Act Types

Schaeffer would consider the (first utterance in the) presentation phase of a top level Contribution [2].

Continue A continuation of a previous act performed by the same speaker. Part of a separate intonation phrase, but syntactically and conceptually part of the same act. This category also includes **restart-continue**, which is where some part of the previous utterance is repeated before continuing on.

Acknowledge Shows understanding of a previous utterance. It may be either a repetition or paraphrase of all or part of the utterance, a *backchannel response* (e.g. “okay”, “uh-huh”), or implicit signalling of understanding, such as by proceeding with the initiation of a new DU which would naturally follow the current one in the lowest level argumentation act. Typical cases of implicit acknowledgement are answers to questions. Acknowledgements are also referred to by some as

confirmations (e.g. [3]) or *acceptances* (e.g. [2]). We prefer the term *acknowledgement* as unambiguously signalling understanding, reserving the term *acceptance* for a higher level speech act signalling agreement with a proposed suggestion or request.

Repair Changes the content of the current DU. This may be either a correction of previously uttered material, or the addition of omitted material which will change the interpretation of the speaker’s intention. A **repair** can change either the content or illocutionary force of acts in the current DU (e.g. a tag question can change the interpretation of a declarative utterance from an **Inform** to a **YNQ**). **Repair** actions should not be confused with domain clarifications, e.g. **CORRECT-PLAN** and other members of the *Clarification Class* of Discourse Plans described by Litman & Allen [7]. **Repairs** are concerned merely with the grounding of content and not changes to previously grounded content.

ReqRepair A request for repair. Asks for a repair by the other party. This is roughly equivalent to what Schegloff, et. al. term a *Next Turn Repair Initiator* [9]. Often a **ReqRepair** can be distinguished from a **repair** or **acknowledge** only by intonation. A **ReqRepair** invokes a discourse obligation on the listener to respond with either the requested repair, or an explicit refusal or postponement (e.g a followup request).

ReqAck Attempt to get the other agent to acknowledge the previous utterance. This invokes a discourse obligation on the listener to respond with either the requested acknowledgement, or an explicit refusal or postponement (e.g a followup repair or repair request).

Cancel Closes off the current DU as ungrounded. Rather than repairing the current DU, a **cancel** abandons it; the underlying intention, if still held, must be expressed in a new DU.

3 From Grounding Acts to DUs

We name the agents taking part in constructing a DU as follows: the *Initiator* is the one who performs the **initiate** act to start off the DU. The other participant is called the *Responder*. Agents may take different roles in different

DUs in a mixed initiative conversation. A completed Discourse Unit is one in which the intent of the Initiator becomes mutually understood (or *grounded*) by the conversants. While there may be some confusion among the parties as to what role a particular utterance plays in a unit, whether a discourse unit has been completed, or just what it would take to complete one, only certain patterns of actions are allowed. For instance, a speaker cannot acknowledge his own immediately prior utterance. He may utter something (e.g. “ok”) which is often used to convey an acknowledgement, but this cannot be seen as an acknowledgement in this case. Often it will be seen as a request for acknowledgement by the other party, a signal that the speaker is shifting to a new topic, or a pre-closing device [10]. Similarly, a speaker cannot **continue** an utterance begun by another speaker. Depending on context, this will be interpreted as either an acknowledgement (e.g. if one is just completing the other’s thought), a **repair** (if one is correcting to what *should* have been said), or an **initiate** of a new DU (if this is new information).

We can identify at least seven different possible states for a DU to be in. These can be distinguished by their relevant context: what acts have been performed and what is preferred to follow, as shown in Table 2.

State	Entering Act	Preferred Exiting Act
S	—	Initiate ^I
1	Initiate ^I	Ack ^R
2	ReqRepair ^R	Repair ^I
3	Repair ^R	Ack ^I
4	ReqRepair ^I	Repair ^R
F	Ack ^{I,R}	Initiate ^{I,R} (next DU)
D	Cancel ^{I,R}	Initiate ^{I,R} (next DU)

Table 2: Meanings of Discourse Unit States

Acts in this table are superscripted with the initial of the agent who performs them, *I* for the *Initiator*, and *R* for the *Responder*. State S represents a DU that has not been initiated yet. State F represents one that has been grounded, though we can always add on more, as in a further acknowledgement or some sort of repair. State D represents an abandoned DU, ungrounded and ungroundable. The other states represent DUs which

still need one or more utterance acts to be grounded. State 1 represents the state in which all that is needed is an acknowledgement by the Responder. This is also the state the results immediately after an initiation. However, the Responder may also request a repair, in which case we need a repair by the Initiator before the responder acknowledges, this is state 2. The Responder may also repair directly (state 3), in which case the Initiator needs to acknowledge this repair. Similarly the Initiator may have problems with the Responder’s utterance, and may request that the Responder repair, this would be state 4.

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	

*repair request is ignored

Table 3: DU Transition Diagram

Although these states have acts which are in some sense *preferred*, any of a number of acts can follow at any given state. Table 3 shows a finite state machine which gives the possible transitions from state to state and tracks the progress of Discourse Units.

The entries in the table signal which state to go into next given the current state and the utterance act. A Discourse Unit starts with the utterance of an initiator (state S), and is considered completed when it reaches the final state

Speaker	UU#	Utterance	UU Act
S	141	that's four loads of orange juice	Initiate ₁
M	142	oka.. we..	Ack ₁
M	143	of oranges	Repair ₁ (141)
S	144	oranges, yes	Ack ₁
M	145*	and we have tanker T-two with us, right	Initiate ₂
S	146	that's right	Ack ₂ Initiate ₃
M	147	okay	Ack ₃

Table 4: Example Conversation Fragment with Grounding Acts

(state F). As can be seen, however, it may continue beyond this point, either because one partner is not sure that it has finished, or if it gets reopened with a further repair. At each state, there are only a limited number of possible next actions by either party. Impossible actions are represented in the table by blanks. If one is in a state and recognizes an impossible action by the other agent, there are two possibilities, the action interpretation is incorrect, or the other agent does not believe that the current DU is in the same state (through either not processing a previous utterance or interpreting its action type differently). Either way, this is a cue that repair is needed and should be performed. One also always has the option of initiating a new DU, and it may be the case that more than one is open at a time. If a DU is left in one of the non-final states, then its contents should not be seen as grounded.

4 Examples of Grounding Acts

Table 4 presents a small spoken conversation fragment from the TRAINS domain [1], annotated with examples of grounding acts. The table can be read as follows: the first column shows the speaker, *M*, or *S*. The second column gives the number of the utterance unit, the third column the transcription of the utterance, and the last column the grounding act(s) performed by the utterance, with the number of the Discourse Unit of which it is a part subscripted. `Repair`, `ReqRepair`, and `Continue` also take a UU argument in parentheses, as in UU#143 which is a repair of UU#141. The asterisk in utterance #145 indicates that its production overlaps in time with the previous utterance; the text is lined up directly under the point in the previous utterance at which the overlap begins.

With utterance #141, S initiates an inform DU, which (at the argumentation level) also serves to summarize the previous plan suggested by M. M (starts to) acknowledge, with utterance #142 but M realizes that S said “orange juice” instead of oranges. He corrects with utterance #143, which is then acknowledged by S with utterance #144. M then begins a new DU with utterance 145, a YNQ, which is (implicitly) acknowledged by S with utterance #146. Utterance #146 realizes both an **Acknowledge** act in DU #2, and an **Initiate** act for DU #3, an inform which serves to answer the question and complete the Q&A argumentation act.

5 Comparisons With Related Work

Cohen & Levesque [3] remark on the frequent occurrence of acknowledgements and requests for acknowledgement, and point out the inability of previous speech act theories of dialogue to account for them. They provide an account of the goals that lead to acknowledgements, based on conversants having a joint intention to perform a task which commits each agent to achieving mutual knowledge of the success or impossibility of the task, in this case communication.

Although they state that mutual knowledge is achieved by acknowledgements, and that the same goals account for repair requests by the responder and acknowledgement requests by the initiator, they do not give an account of just what role particular types of utterances play in leading to this mutual belief. Thus, [3] can be seen as providing an account of *why* agents utter acknowledgements, while section 3 gives the beginnings of an account of *how* these utterances contribute towards mutual belief.

Clark & Schaefer [2] provide a similar account of grounding to the one given here. Their account of grounding is by *Contributions*, which consist of a presentation part and an acceptance part, each of which may be composed of nested contributions as well. For Clark & Schaefer, each “signal” is a presentation of some contribution (how precisely to divide up a signal into units is not specified; most units seem to correspond to turns at talk, though there are sometimes more than one unit per turn). In addition, each signal that a previous presentation has been accepted is an acceptance, creating a complete contribution. Contributions are thus dovetailed, with every contribution having as its acceptance part the presentation part of another

contribution. Contributions may also be recursive, in which the presentation or acceptance phase contains one or more complete contributions. This happens in the case of repairs and more complicated side sequences, as well as when contributions are presented in installments, with each turn being only part of a sentence.

We take a more local approach to grounding, assigning to each utterance a grounding function as it is uttered. It is often very difficult to assign the role that an utterance plays in Clark & Schaefer’s *Contribution* system until subsequent utterances are analyzed, limiting the utility of contributions in a processing model. We treat acknowledgements and repairs as atomic actions in support of a DU, not as contributions in their own right which require acknowledgement as well. We also do not treat the higher level coherence of grounded units in the same manner. While Clark & Schaefer treat linked speech such as presentation by installments as contributions containing contributions, we treat this at a higher level of Conversation Act, occurring at the **Argumentation** level, along with other elaborations and the relationship between questions and answers.

6 Future Work

This classification scheme has been based on the analysis of a corpus of collected conversations in the TRAINS domain, and difficulties encountered in trying to label the corpus using previous schemes. Currently, we have marked 4 complete conversations and several small fragments (totalling about 1 hour of speech, about 1000 utterances). We intend in the near future to try to evaluate the scheme by testing whether it can be taught to others who can independently code dialogues reliably. Our hope is that the only conflicts will be either different paths leading to the same state of beliefs and commitments about the conversation (in which case it didn’t matter which labelling was used) or cases in which there is a genuine confusion about the state of knowledge of the conversational participants.

Another avenue we are pursuing is providing a formal semantics for grounding acts based on general sub-action relationships [13]. One of the difficulties with DUs and grounding acts as presented here is that the “content” of DUs is often not stipulated in advance, but negotiated between an initiator and responder. Occasionally a responder will understand and ac-

knowledge one part of an open DU, while requesting repair on another part (e.g. acknowledging a proposition while questioning the referent of one of the referring expressions used therein). In this case, the only option is to “split” the DU into two different DUs, now in different states. While this is somewhat awkward using the analysis given in Section 3, the semantics makes this relationship fairly clear.

We are also making attempts to implement grounding ideas in the context of the TRAINS dialogue manager [11]. In this system, the state of grounding of information is tracked using a set of belief contexts, and recognized acts will move information from one context to another. The relationship of information in these spaces will also serve as a guide to system produced utterances (e.g. when to produce a repair or acknowledgement). Further details of this proposal may be found in [12].

Finally, we would like to use this classification scheme to conduct studies on the surface cues which tend to signal particular acts, along the lines of [8; 5; 6]. It is our hope that a finer grained notion of “action” will provide a sounder basis for such studies.

Acknowledgements

This material is based upon work supported in part by the NSF under research grant no. IRI-9003841, by ONR under research grant no. N00014-90-J-1811, and by DARPA/ONR under contract N00014-92-J-1512. We would like to thank Derek Gross and Shin’ya Nakajima for collecting the TRAINS dialogues and providing the initial transcriptions which have been used as data for this analysis.

References

- [1] James F. Allen and Lenhart K. Schubert. The trains project. TRAINS Technical Note 91-1, Computer Science Dept. University of Rochester, 1991.
- [2] Herbert H. Clark and Edward F. Schaefer. Contributing to discourse. *Cognitive Science*, 13:259 – 94, 1989.

- [3] Phillip R. Cohen and Hector J. Levesque. Confirmations and joint action. In *Proceedings IJCAI-91*, pages 951–957, 1991.
- [4] Phillip R. Cohen and C. R. Perrault. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3(3):177–212, 1979.
- [5] Beth Ann Hockey. Prosody and the interpretation of “okay”. In *Working Notes AAAI Fall Symposium on Discourse Structure in Natural Language Understanding and Generation*, November 1991.
- [6] Jacqueline C. Kowtko. On the intonation of mono- and di-syllabic words within the discourse framework of conversational games. In *Proceedings ACL-92*, pages 282–284, 1992.
- [7] Diane J. Litman and James F. Allen. Discourse processing and common sense plans. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*. MIT Press, 1990.
- [8] Shin’ya Nakajima and James F. Allen. A study on the roles of prosody in the cooperative dialogue. In *Research on Modelling of Speech Dialogue and its Computational Processing: Final Reports*, pages 144–152. National Science Foundation, December 1991.
- [9] E. A. Schegloff, G. Jefferson, and H. Sacks. The preference for self correction in the organization of repair in conversation. *Language*, 53:361–382, 1977.
- [10] D. Schiffrin. *Discourse Markers*. Cambridge University Press, Cambridge, 1987.
- [11] David R. Traum. The discourse reasoner in trains-90. TRAINS Technical Note 91-5, Computer Science Dept. University of Rochester, 1991.
- [12] David R. Traum. Towards a computational theory of grounding in natural language conversation. Technical Report 401, Computer Science Dept. University of Rochester, October 1991.
- [13] David R. Traum and James F. Allen. Abstract plan execution relations: An initial report. Unpublished manuscript.

- [14] David R. Traum and James F. Allen. Conversation actions. In *Working Notes AAAI Fall Symposium on Discourse Structure in Natural Language Understanding and Generation*, November 1991.
- [15] David R. Traum and Elizabeth A. Hinkelman. Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8(3), forthcoming 1992. Special Issue on Non-literal language.