Degrees of Grounding Based on Evidence of Understanding

Antonio Roque
USC Institute for Creative Technologies
Marina del Rey, CA, USA
roque@ict.usc.edu

David Traum
USC Institute for Creative Technologies
Marina del Rey, CA, USA
traum@ict.usc.edu

Abstract

We introduce the Degrees of Grounding model, which defines the extent to which material being discussed in a dialogue has been grounded. This model has been developed and evaluated by a corpus analysis, and includes a set of types of evidence of understanding, a set of degrees of groundedness, a set of grounding criteria, and methods for identifying each of these. We describe how this model can be used for dialogue management.

1 Introduction

Dialogue system researchers are active in investigating ways of detecting and recovering from error, including determining when to provide confirmations or rejections, or how to handle cases of complete non-understanding (Bohus and Rudnicky, 2005a; Bohus and Rudnicky, 2005b; Skantze, 2005).

Studying the strategies that humans use when speaking amongst themselves can be helpful (Swerts et al., 2000; Paek, 2003; Litman et al., 2006). One approach to studying how humans manage errors of understanding is to view conversation as a joint activity, in which **grounding**, or the process of adding material to the common ground between speakers, plays a central role (Clark and Schaefer, 1989). From this perspective, conversations are highly coordinated efforts in which participants work together to ensure that knowledge is properly understood by all participants. There is a wide variety of grounding behavior that is determined by the communication medium, among other things (Clark and Brennan, 1991).

This approach is developed computationally by Traum, who presents a model of grounding which adapts Clark and Schaefer's contributions model to make it usable in an online dialogue system (Traum, 1994). Other computational approaches to grounding use decision theory (Paek and Horvitz, 2000a) or focus on modeling belief (Bunt et al., 2007).

Grounding models generally consider material to be in one of three states: ungrounded, in the process of becoming sufficiently grounded, or sufficiently grounded. (An exception is (Paek and Horvitz, 2000b), who use a continuous model of groundedness.) We are developing a model of grounding that is attentive to a larger set of types of evidence of understanding than is typical, and use this to define a model of **Degrees of Grounding**, which tracks the extent to which material has become a part of the common ground.

This model includes a set of types of **Evidence of Understanding** that describes the kinds of cues that the dialogue gives about the state of grounding. A set of **Degrees of Groundedness** describes the extent to which material has achieved mutual belief while being discussed. A set of **Grounding Criteria** describes the degree to which material needs to be grounded. Finally, the model provides algorithms to assist dialogue management.

The next section describes the radio domain which we used to begin developing this model. The dialogues in this domain contain a large amount of confirmation behavior, which make it a good testbed for the initial development of the model. However, because these radio dialogues are highly structured we are not yet able to make strong claims about the

Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue, pages 54–63,
Columbus, June 2008.

generality of this model.

In following sections we describe the components of the model, annotation evaluations, and ongoing development of the model.

2 Domain

The domain for this corpus analysis involves a radio-based military training application. This corpus was developed while building the Radiobot-CFF system (Roque et al., 2006) in which soldiers are trained to perform artillery strike requests over a simulated radio in an immersive virtual environment.

Calls for Fire (CFFs) are coordinated artillery attacks on an enemy. Several teams work together to execute a CFF. A **Forward Observer** (FO) team locates an enemy target and initiates the call. The FO team is made up of two or more soldiers, usually with one soldier dedicated to spotting the enemy and another soldier dedicated to operating the radio. The FO radio operator communicates with the **Fire Direction Center** (FDC) team, which decides whether to execute the attack, and if so, which of the available fire assets to use. An example CFF is given in the Appendix.

3 Evidence of Understanding

An influential description of evidence of understanding was presented in (Clark and Schaefer, 1989), as shown in Table 1. This set of types of evidence was described as being "graded roughly from weakest to strongest" and was part of the acceptance phase of a two-phase grounding process. (Clark and Brennan, 1991) further develop Clark's notion of evidence, describing "the three most common forms of positive evidence" as being acknowledgments, initiation of the relevant next turn, and continued attention.

The Degrees of Grounding model exchanges Clark and Schaefer's two-phase model for an approach that tracks grounding acts in a way similar to (Traum, 1994). Also, rather than concerning itself with the strength of a given type of evidence, the current model tracks the strength of material based on its degree of groundedness, which is derived from sequences of evidence as described in Section 4.

Evidence in the Degrees of Grounding model is tracked per **Common Ground Unit** (CGU) in an information state, as in (Traum and Rickel, 2002). An example of such a CGU is given in Figure 1. Material under discussion is disambiguated by several identifying components of the CGU: in this domain this is the dialogue move, the parameter, the mission number, and the adjust number. Note that parameter value is not used as an identifying component; this allows for reference to the material by participants who may not yet agree on its value.

Evidence	Description
Continued Attention	B shows he is continuing to attend and therefore remains satisfied with As presentation.
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.
Acknowledgement	B nods or says "uh huh," "yeah," or the like.
Demonstration	B demonstrates all or part of what he has understood A to mean.
Display	B displays verbatim all or part of As presentation.

Table 1: (Clark and Schaefer, 1989)'s Evidence of Understanding between speakers A and B

```
information:
  dialogue move: target location
  parameter: direction
  value: 5940
  mission number: to be determined
  adjust number: 0
evidence history:
  submit-G91, repeat_back-S19
degree of groundedness: agreed-content
grounding criteria met: true
```

Figure 1: Example Common Ground Unit

The remainder of this section describes the kinds of evidence of understanding found in the corpus. Section 6 describes inter-annotator agreement studies that determine that humans can reliably identify these types of evidence.

3.1 Submit

A **Submit** type of evidence is provided when material is introduced into the common ground for the first time. The Submit type of evidence is derived from the Presentation phase of (Clark and Schaefer, 1989).

An example of a Submit is given in line 1 of Table 2: "direction 6120" is information that had not yet been mentioned and has no assumed values.

Line	ID	Utterance	Evidence
1	G91	direction 6120 over	Submit
2	S19	direction 6120 out	Repeat Back
3	G91	correction direction 6210 over	Resubmit

Table 2: Example Dialogue

Dialogue systems that do not specifically model grounding generally assume that material is grounded when it is first Submitted unless there is evidence to the contrary.

3.2 Repeat Back

A **Repeat Back** type of evidence is provided when material that was Submitted by another dialogue participant is presented back to them, often as part of an explicit confirmation.

The Repeat Back evidence is related to the "Display" evidence of (Clark and Schaefer, 1989) and described in Table 1, however here it is renamed to indicate that it pertains to verbal repetitions, rather than general displays which may be in other modalities, such as visual. In fact, there is evidence that grounding behavior related to visual feedback is different from that related to auditory feedback (Clark and Brennan, 1991; Thompson and Gergle, 2008).

An example is given in line 2 of Table 2: the "direction 6120" information given in line 1 is Repeated Back as part of a confirmation.

3.3 Resubmit

A **Resubmit** type of evidence is provided when material that has already been Submitted by a dialogue participant is presented again as part of a self- or other-correction. This is an example of what (Clark and Brennan, 1991) call negative evidence, which indicate a lack of mutual belief.

An example is shown in Table 2; the direction information which was Submitted in turn 1 and Repeated Back in turn 2 is Resubmitted in turn 3.

In this domain, follow-up presentations of material were almost always corrections, usually of information that has been repeated back by the other participant, or based on new occurences in the virtual world (for example, the lifting of smoke that was previously obscuring a target.) Due to the nature of the task, this corpus had few instances of non-correction follow-up behavior, where material was presented a second time for the purposes of further discussion. Such follow-ups are an evidence of understanding whose behavior is probably different from that of the Resubmit type of evidence as described here, and will be examined in future work as described in Section 7.

3.4 Acknowledge

An **Acknowledge** type of evidence is a general statement of agreement that does not specifically address the content of the material. Acknowledges are identified by semantic interpretation. Acknowledges are a part of (Clark and Schaefer, 1989)'s set of types of evidence of understanding.

Table 3 contains an example: in line 1 the speaker G91 Submits information about the target's status, which is then Acknowledged by speaker S19 in turn line 2.

Line	ID	Utterance	Evidence
1	G91	end of mission target destroyed over	Submit
2	S19	roger	Acknowledge

Table 3: Example of an Acknowledgment

3.5 Request Repair

A **Request Repair** type of evidence is a statement that indicates that the speaker needs to have the material Resubmitted by the other participant. Request Repairs are identified by semantic interpretation. Request Repairs are another example of negative evidence (Clark and Brennan, 1991).

Table 4 gives an example: in line 1 G91 submits a map grid coordinate, and in line 2 S19 asks that the other speaker "say again" that grid coordinate, which is a Request for Repair.

Line	ID	Utterance	Evidence
1	G91	grid 5843948	Submit
2	S19	say again grid over	Request Repair

Table 4: Example of a Request Repair

3.6 Move On

A **Move On** type of evidence is provided when a participant decides to proceed to the next step of the task at hand. This requires that the given task have a set of well-defined steps, and that the step being Moved On from needs to be grounded before the next step can be discussed. Move Ons are identified based on a model of the task at hand. Move Ons are related to (Clark and Schaefer, 1989)'s "Initiation of the relevant next contribution," although Clark and Schaefer do not specify that "next contributions" should be dependent on sufficiently grounding the previous step.

A Move On provides evidence because a cooperative dialogue participant would typically not move on to the next step of the task under such conditions unless they felt that the previous step was sufficiently grounded.

Table 5 shows an example of a Move On. In line 1, G91 indicates that the kind of artillery fire they want is a "fire for effect"; this is Repeated Back in line 2. G91 then Submits grid information related to the target location. The task specification of Calls for Fire indicates that fire requests should proceed in several steps: after a Warning Order is established, a Target Location should be given, followed by a Target Description. By moving on to the step in which a Target Location is provided, G91 tacitly indicates that the step in which a Warning Order is established has been dealt with to their satisfaction.

Line	ID	Utterance	Evidence
1	G91	fire for effect over	Submit
2	S19	fire for effect out	Repeat Back
3	G91	grid 45183658	Submit, Move On

Table 5: Example of a Move On

Line	ID	Utterance	Evidence
1	S19	message to observer kilo 2 rounds AB0001 over	Submit
2	G91	mike tango oscar kilo 2 rounds target number AB0001 out	Repeat Back
3	S19	shot over	Submit

Table 6: Example of a non-Move On

Not all typical sequences provide Move On evidence. In the example in Table 6, in line 1 S91 submits a "message to observer" indicating the kind of fire that is being delivered, which is followed in line 2 by a confirmation by G91. S19 then proceeds to the next step of the task by indicating in line 3 that the artillery has been fired. Line 3, however, is not a Move On because although it is typically the next step in the task, providing that information is not dependent on fully grounding the material being discussed in line 2 - in fact, line 3 will be provided when the artillery has been fired, and not based on any other decision by S19.

3.7 Use

A **Use** type of evidence is provided when a participant presents an utterance that indicates, through its semantics, that a previous utterance was understood. Uses are related to (Clark and Schaefer, 1989)'s "Demonstration".

In the Radiobot-CFF corpus, most Uses are replies to a request for information, such as in Table 7, where S19's request for a target description in line 1 is answered with a target description, in line 2.

Line	ID	Utterance	Evidence
1	S19	s2 wants to know whats the target description over	Submit
2	G91	zsu over	Submit, Use

Table 7: Example of a Use

Another example of Use is shown in Table 8, in which S19 is providing an intelligence report in line 1 regarding an enemy target, and line 2 replies with a statement asking whether the target is a vehicle. The utterance in line 2 uses information provided in line 1.

3.8 Lack of Response

A **Lack of Response** type of evidence is provided when neither participant speaks for a given length of time. Identifying a Lack of Response type of evidence involves determining how much silence will be significant for signalling understanding or lack of understanding.

Line	ID	Utterance	Evidence
1	S19	again it should have rather large antennas affixed to it uh they are still sending out signals at the time	Submit
2	G91	this is some kind of vehicle over	Submit, Use

Table 8: Example of a Use

In the example shown in Table 9, G91 submits an identifying utterance to see if S19 is available. After 12 seconds, G91 has heard nothing back; this is negative evidence of grounding, so in line 3 G91 resubmits the utterance.

Line	ID	Utterance	Evidence
1	G91	S 1 9 this is G 9 1	Submit
2		(12 seconds of silence)	Lack of Response
3	G91	S 1 9 this is G 9 1	Resubmit

Table 9: Example of a Lack of Response

A Lack of Response can also be an indication of positive grounding, as in Table 10. In line 1, G91 submits information about a target, which in line 2 is repeated back. Line 3 indicates a period of silence, in which neither speaker took the opportunity to request a repair or otherwise indicate their disapproval with the state of the groundedness of the material. In that sense, the silence of line 3 is positive evidence of understanding.

Line	ID	Utterance	Evidence
1	G91	b m p in the open over	Submit
2	S19	b m p in the open out	Repeat Back
3		(10 seconds of silence)	Lack of Response

Table 10: Example of a Lack of Response

4 Degrees of Groundedness

Degrees of groundedness are defined such that material has a given degree before and after any sequence of evidence given. For example, in Table 10 the target description given in line 1 has a certain degree of groundedness before it is Submitted, another degree after it is Submitted, another degree after it is Repeated Back, and another degree after the Lack of Response.

Degreee	Pattern/Identifier
Unknown	not yet introduced
Misunderstood	(anything,Request Repair)
Unacknowledged	(Submit, Lack of Response)
Accessible	(Submit) or (anything,Resubmit)
Agreed-Signal	(Submit, Acknowledgment)
Agreed-Signal+	(Submit, Acknowledgment, other)
Agreed-Content	(Submit, Repeat Back)
Agreed-Content+	(Submit, Repeat Back, other)
Assumed	grounded by other means

Table 11: Degrees of Groundedness

A key part of defining these degrees is to determine which of these degrees is worth modeling. For example, Table 3 shows a CGU further grounded by a single Acknowledgment. In this domain, for the purposes of determining grounding criteria and dialogue management algorithms, it is not worth distinguishing between the case in which it had been followed by one more Acknowledgment and the case in which it had been followed by two or more Acknowledgments.

Table 11 shows the significant degrees identified during the corpus study, as well as the definition or identifying pattern of evidence. These degrees are shown from Unknown, which is least grounded, to Assumed, which is grounded by other means, such as written information given during a scenario briefing. Most degrees are identified by patterns of evidence. For example, a CGU is misunderstood if the latest item of evidence provided is a Request Repair, and CGU is Unacknowledged if it is Submitted followed by a Lack of Response.

The degree of groundedness is used to compute how much (if any) additional evidence is needed to reach the grounding criterion, or "criterion sufficient for current purposes" as defined by (Clark and Schaefer, 1989). This computation can be used in dialogue management to help select a next utterance.

In this domain, information such as target numbers have high grounding criteria, such as Agreed-Content+; they would need to be Repeated Back, and followed at least by a Lack of Response, giving the other participant an opportunity to correct.

Other information might have a grounding criterion of Agreed-Signal, needing only an Acknowledgment to be grounded, as in Table 3. Future work will address the fact that grounding criteria are variable: for example, in noisy conditions where errors are more probable, the grounding criteria may increase.

5 Dialogue Management

Exploiting this model of grounding for dialogue management involves several steps. Evidence of understanding must be identified given a semantic interpretation and the history of evidence provided so far. Given an utterance's new evidence and a CGU's current degree of groundedness, the CGU's new degree of groundedness must be determined.

Once a CGU's current degree is determined, it can be compared to its grounding criterion to determine whether or not it has been sufficiently grounded, and if not, a new item of evidence may be suggested to help further ground the material.

All of these can be put together in one algorithm, as shown in Figure 2.

```
for each dialogue act parameter,
 identify the relevant CGU
 identify evidence of understanding
 compute the CGU's degree of groundedness

for each CGU not sufficiently grounded
 determine evidence to be given
 compute the CGU's degree of groundedness

if Lack of Response detected
 compute the CGU's degree of groundedness
```

Figure 2: Dialogue Management Algorithm

The specifics of how this algorithm is integrated into a system and how it influences task decisions will vary based on the system being used. To explore the domain-independence of this model, we are currently integrating it into a dialogue manager in a domain unrelated to the CFF task.

6 Evaluation

The validity of this model has been evaluated in several corpus tests to measure inter-annotator agreement in identifying evidence, to ensure that identifying evidence can reliably be done by an algorithm, to measure inter-annotator agreement in identifying the increase or decrease of the degree of groundedness, and to ensure that identifying the increase or decrease of a degree of groundedness can reliably be done by an algorithm.

Human transcribers produced transcriptions of several sessions between two sets of humans acting as Forward Observer and Fire Direction Center radio operators in the training simulation. A subset of the corpus was used for close analysis: this subset was made up of 4 training sessions, composed of 17 fire missions, totaling 456 utterances; this provided a total of 1222 possible indicators of evidence of understanding made up of 886 dialogue move parameters and 336 period of silence.

We automatically performed a dialogue act interpretation on the dialogue move parameters, which were then manually corrected. We then manually annotated the evidence of understanding identified in each dialogue move parameter and period of silence. An example of the data produced from this process is given in the Appendix.

6.1 Inter-Annotator Agreement - Identifying Evidence

An inter-annotator agreement study was performed in which two annotators tagged a subset of the corpus (318 dialogue move parameters and 74 silences) to identify the evidence of understanding, given an utterance and dialogue act interpretation. One annotator was the first author of this paper, and the other was a computer professional who had no previous experience with the domain or with tagging data.

Table 12 shows the results, broken down by the Standalone types of evidence, which could occur by themselves (Submit, Repeat Back, Resubmit, Acknowledge, and Request Repair), the Additional types of evidence, which only occurred with other types of evidence (Move On and Use), and the Silence-Related Lack of Understanding type of evidence. Each of these showed acceptable levels of agreement, with the exception of the Kappa for the additional evidence. The low score on the additional evidence is probably due to the fact that Move On judgments depend on a strong understanding of the domain-specific task structure, as described in section 3.6; to a lesser extent Use judgments tend to rely on an understanding of the scenario as well.

Evidence Type	P(A)	Kappa
Standalone	0.95	0.91
Additional	0.87	0.53
Silence-Related	0.92	0.84

Table 12: Inter-Annotator Agreement - Evidence

Evidence Type	P(A)	Kappa
Standalone	0.88	0.81
Additional	0.98	0.92
Silence-Related	1.0	1.0

Table 13: Algorithm Agreement - Evidence

This highlights the fact that for most of the evidence of understanding (all except for Move On and Use), agreement can be reached with a non-expert annotator.

6.2 Algorithm Agreement - Identifying Evidence

The results of the inter-annotator agreement test were merged into the larger 1222-markable corpus, to create a consensus human-annotated corpus. This was used in the next test, to identify whether an algorithm can automatically identify evidence.

We authored a set of rules to identify evidence of understanding based on the order in which CGUs were introduced into the common ground, the identity of the speaker who introduced them, and the semantic interpretations. The rules were then applied to the 1222-markable corpus, and the resulting identifications were then compared to the identifications made by the human annotators. The results are shown in Table 13. The respectable agreement and kappa values indicate that it is possible for an algorithm to reliably identify evidence.

6.3 Degree Increase/Decrease Agreements

Finally, we explored whether humans could reliably agree on whether a given material's groundedness had increased or decreased after a given turn.

We studied this because we are not here claiming that humans explicitly model degrees of groundedness or perform a computation to compare a given material with something they had grounded previously. It is more likely that humans track evidence, determine whether material is more or less grounded than it was before, and check whether it has reached a grounding criterion. A dialogue system need not be tied to human behavior to be effective, so given these human behaviors, we are interested in whether computer algorithms can be built to produce useful results in terms of task completion and human-realistic behavior. For this reason we evaluate the model of degrees of grounding based on how human-realistic its ability to identify whether a CGU's degree of groundedness has increased or decreased, and in future work study whether a system implementation performs acceptably in terms of task completion and managing human-realistic grounding behavior.

Agreement Type	P(A)	Kappa
Human-Human	0.97	0.94
Human-Algorithm	0.87	0.73

Table 14: Degree Increase/Decrease Agreements

To perform the test of whether degree increase or decrease could be reliable detected, we annotated a subset of the corpus with a non-domain expert. For a set of CGUs, we tracked the sequence of evidence that was provided to ground that CGU. Before and after each item of evidence, we asked the annotators to determine whether the CGU was more or less grounded than it was the turn before.

We also developed a set of rules based on the definition of the degrees of groundedness defined in section 4 to determine after each utterance whether a CGU's degree of groundedness had increased or decreased from the utterance before. We then compared the results of that set of rules with human-consensus judgments about degree increase and decrease.

The results are shown in Table 14, indicating that humans could reliably agree among themselves, and a rule-based algorithm could reliably agree with the human consensus judgments.

7 Discussion and Future Work

In this paper we describe the initial development of the Degrees of Grounding model, which tracks the extent to which material has been grounded in a dialogue. The Degrees of Grounding model contains a richer variety of evidence of understanding than most models of grounding, which allows us to define a full set of degrees of groundedness.

We recognize that the initial domain, although rich in grounding behavior, is not typical of most human conversation. Besides the structured dialogues and the domain-specific word use, the types of evidence of understanding presented in Section 3 does not cover all possible types of evidence. For example, (Clark and Schaefer, 1989) describe "continued attention" as another possibility, which was not available with the radio modality used in this study. Furthermore, it is a feature of this domain that Resubmit evidence generally indicates lack of understanding; in general conversation, it is not true that the repeated mention of material indicates that it is not understood, so a "Follow-Up" evidence is likely, as are variations of "Use."

To explore these questions, we are extending work to other domains, and are currently focusing on one in which virtual humans are used for a questioning task. Also, we plan to run evaluations in implemented systems, exploring performance in terms of task completion and believable human behavior.

Acknowledgments

This work has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

The authors would like to thank Kevin Knight and the anonymous reviewers for feedback about the evaluation.

References

Dan Bohus and Alexander Rudnicky. 2005a. Error handling in the RavenClaw dialog management architecture. In *Proceedings of HLT-EMNLP-2005*.

Dan Bohus and Alexander Rudnicky. 2005b. Sorry, I didn't catch that! - an investigation of non-understanding errors and recovery strategies. In *Proceedings of SIGdial-2005*. Lisbon, Portugal.

Harry Bunt, Roser Morante, and Simon Keizer. 2007. An empirically based computational model of grounding in dialogue. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*.

Herbert H. Clark and Susan E. Brennan. 1991. Grounding in communication. In *Perspectives on Socially Shared Cognition*, pages 127–149. APA Books.

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

Diane Litman, Julia Hirschberg, and Marc Swerts. 2006. Characterizing and predicting corrections in spoken dialogue systems. *Computational linguistics*, pages 417–438.

Tim Paek and Eric Horvitz. 2000a. Conversation as action under uncertainty. In *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 455–464.

Tim Paek and Eric Horvitz. 2000b. Grounding criterion: Toward a formal theory of grounding. Technical report, Microsoft Research, April. Microsoft Technical Report, MSR-TR-2000-40.

Tim Paek. 2003. Toward a taxonomy of communication errors. In *Proceedings of the ISCA Tutorial and Research Workshop on Error Handling in Spoken Dialogue Systems*, pages 53–58, August 28-31. Chateau d'Oex, Vaud, Switzerland.

Antonio Roque, Anton Leuski, Vivek Rangarajan, Susan Robinson, Ashish Vaswani, Shri Narayanan, and David Traum. 2006. Radiobot-CFF: A spoken dialogue system for military training. In *9th International Conference on Spoken Language Processing (Interspeech 2006 - ICSLP)*, September.

Gabriel Skantze. 2005. Galatea: a discourse modeller supporting concept-level error handling in spoken dialogue systems. In *Proceedings of SigDial*, pages 178–189). Lisbon, Portugal.

Marc Swerts, Diane Litman, and Julia Hirschberg. 2000. Corrections in spoken dialogue systems. In *Proceedings of the 6th International Conference of Spoken Language Processing (ICSLP-2000)*, October.

Will Thompson and Darren Gergle. 2008. Modeling situated conversational agents as partially observable markov decision processes. In *Proceedings of Intelligent User Interfaces (IUI)*.

David Traum and Jeff Rickel. 2002. Embodied agents for multi-party dialogue in immersive virtual world. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-agent Systems (AAMAS 2002)*, pages 766–773, July.

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

Appendix

Line	ID	Utterance	Semantic Interpretation	Evidence: Standalone	Evidence: Additional
1	G91	fire for effect over	WO-MOF: fire for effect	Submit	
2	S19	fire for effect out	WO-MOF: fire for effect	Repeat Back	
3		Silence: 0.7 seconds			
4	G91	ah roger	ROGER	Acknowledge	
		grid four five four two ah three six three eight	TL-GR: 45423638	Submit	Move On
5		Silence: 2.3 seconds			
6	S19	grid four five four two three six three eight out	TL-GR: 45423638	Repeat Back	
7		Silence: 0.7 seconds			
8	G91	ah roger	ROGER	Acknowledge	
		b r d m	TD-TYPE: b r d m	Submit	Move On
		in the open over	TD-DESC: in the open	Submit	
9		Silence: 1.3 seconds			
10	S19	b r d m	TD-TYPE: b r d m	Repeat Back	
		in the open out	TD-DESC: in the open	Repeat Back	
11		Silence: 9.9 seconds		Lack of Response	

Comments:

This dialogue is between G91 as a Forward Observer identifying a target, and S19 as a Fire Direction Center who will send the artillery fire when given the appropriate information.

In line 1, G19's utterance is interpreted as a Warning Order - Method of Fire (WO-MOF), describing the kind of artillery fire requested, whose value is "fire for effect." This is the first mention of a WO-MOF for this particular CFF, so it is identified as a Submit type of evidence related to a new CGU, which now has an Accessible degree of groundedness.

In line 2, a WO-MOF is again given. The WO-MOF is identified as referring to the CGU introduced in line 1, and a Repeat Back type of evidence is added to that CGU's evidence history, which gives it an Agreed-Content degree of groundedness.

In line 3 there follows a silence that is not long enough to be a Lack of Response.

In line 4, G91 provides an Acknowledge type of evidence, and Moves On to the next task item: identifying the Target Location - Grid (TL-GR) of the CFF. The Acknowledge and Move On, referring to the CGU created in line 1, raise that CGU's degree of groundedness to its grounding criterion of Agreed-Content+, at which point it becomes grounded. At the same time, the introduction of the TL-GR information creates a new CGU, whose degree is Accessible.

In line 6 the TL-GR CGU is Repeated Back, thereby raising its degree of groundedness to Agreed-Content.

In line 8 an Acknowledge is provided and a set of information related to the Target Description (TD-) is given, providing a Move On, thereby grounding the TL-GR CGU. So by line 8, two CGUs (WO-MOF and TL-GR) have been added to the common ground, and two more CGUs (TD-TYPE and TD-DESC) have Accessible degrees and are in the process of being grounded.

In line 10 the TD CGUs are Repeated Back, raising their degree of groundedness to Agreed-Content.

In line 11 the Lack of Response raises the TD CGUs to Agreed-Content+ thereby grounding them. At this point there is enough information in the common ground for S19 to send the artillery fire.

Line	ID	Utterance	Semantic Interpretation	Evidence: Standalone	Evidence: Additional
12	S19	message to observer kilo	MTO-BAT: kilo	Submit	Move On
		two rounds	MTO-NUM: two	Submit	
		target number alpha bravo zero zero one over	TN: AB001	Submit	
13		Silence: 3.1 seconds			
14	G91	a roger mike tango alpha ah alpha	ROGER	Acknowledge	
		target number alpha bravo zero zero zero one	TN: AB0001	Repeat Back	
		a kilo	MTO-BAT: kilo	Repeat Back	
		two rounds out	MTO-NUM: two	Repeat Back	
11		Silence: 11.4 seconds		Lack of Response	
16	S19	shot	SHOT	Submit	
		rounds complete over	RC	Submit	
17		Silence: 0.8 seconds			
18	G91	shot	SHOT	Repeat Back	
		rounds complete out	RC	Repeat Back	
19	S19	splash over	SPLASH	Submit	
20		Silence: 1.5 seconds			
21	G91	splash out	SPLASH	Repeat Back	
22		Silence: 30.4 seconds		Lack of Response	
...					
23	G91	ah end of mission a target number alpha bravo zero zero one	TN: AB001	Submit	
		one	EOM-NUM: one	Submit	
		b r d m	EOM-TYPE: b r d m	Submit	
		destroyed over	EOM-BDA: destroyed	Submit	
24	S19	end of mission b r d des m d correction b r d m	EOM-TYPE: b r d m	Repeat Back	
		destroyed out	EOM-BDA: destroyed	Repeat Back	

Comments:

In line 12, S19 provides information about the artillery fire that is going to be sent. This includes the battery that will be firing (MTO-BAT), the number of rounds to be fired (MTO-NUM) and the target number that will be used to refer to this particular fire mission from that point on (TN).

In line 14, G91 Repeats Back the information presented in line 12 along with an Acknowledge.

In line 16, S19 notifies that the mission has been fired; in line 18 this is confirmed. Likewese, in line 19 S19 notifies that the mission is about the land; in line 21 this is confirmed.

Between lines 22 and 23 several turns have been removed for space reasons. These turns were related to an adjustment of the artillery fire: after the initial bombardment, the Forward Observer requested that the same artillery be fired 100 meters to the left of the original bombardment. This was confirmed and delivered.

In line 23, G91 sends a description of the amount of damage suffered by the target: the number of enemy affected (EOM-NUM), the type of enemy (EOM-TYPE) and the extent of the damage (EOM-BDA). These are Repeated Back by S19, thereby ending the CFF. Note that S19 does not Repeat Back the EOM-NUM. In this particular instance the number of enemies is implied by the EOM-TYPE being singular, but throughout the corpus EOMs are seen to have a low grounding criteria.