Coding Dialogs with the DAMSL Annotation Scheme

Mark G. Core and James F. Allen
Department of Computer Science
University of Rochester
Rochester, NY 14627
mcore, james@cs.rochester.edu

Abstract

This paper describes the DAMSL annotation scheme for communicative acts in dialog. The scheme has three layers: Forward Communicative Functions, Backward Communicative Functions, and Utterance Features. Each layer allows multiple communicative functions of an utterance to be labeled. The Forward Communicative Functions consist of a taxonomy in a similar style as the actions of traditional speech act theory. The Backward Communicative Functions indicate how the current utterance relates to the previous dialog, such as accepting a proposal, confirming understanding, or answering a question. The Utterance Features include information about an utterance's form and content, such as whether an utterance concerns the communication process itself or deals with the subject at hand. The kappa inter-annotator reliability scores for the first test of DAMSL with human annotators show promise, but are on average 0.15 lower than the accepted kappa scores for such annotations. However, the slight revisions to DAMSL discussed here should increase accuracy on the next set of tests and produce a reliable, flexible, and comprehensive utterance annotation scheme.

Introduction

There are two classes of applications that require the automatic analysis of dialogs: a computer system may act as a participant in a dialog with users, or it may act as an observer attempting to interpret human-human dialogs. In both cases, the system must keep track of how each utterance changes the commonly agreed upon knowledge (common ground (CS89)) including the conversational agents' obligations and plans. Dialog text annotated with the communicative actions of each utterance would aid in training and testing such systems. In addition, linguists studying dialog would greatly benefit from annotated corpora that could be used to reveal the underlying structures of dialogs.

DAMSL (Dialog Act Markup in Several Layers) defines a set of primitive communicative actions that can be used to analyze dialogs. For the purposes of this paper, we will define communicative actions as referring to explicit manipulations of the common ground, and not include more subtle phenomena such as listeners forming opinions about speakers based on the tone and style of their speech.

Speech act theory (Sea75) was one of the first attempts at developing a set of communicative actions. Searle's action classification included Representatives, that introduce information into the common ground; Directives, that attempt to create an obligation on the listener; and Commissives, that involve speakers attempting to introduce an obligation on themselves. Over the years, many researchers (All95; CL90; Han79) have noticed that a major problem with speech act theory is that it attempts to capture an utterance's purpose(s) with one label. DAMSL addresses this problem by allowing multiple labels in multiple layers to be applied to an utterance. Thus an utterance might simultaneously perform actions such as responding to a question, confirming understanding, promising to perform an action, and informing.

The classes of communicative actions discussed here are high-level and designed to be applicable to various types of dialogs. The idea is that for a particular domain, these classes could be further subdivided into acts that are relevant to the domain. The common level of abstraction across domains, however, would allow researchers to share data in a way that would not be possible if everyone developed their own scheme.

The overall structure of DAMSL has been developed by the Multiparty Discourse Group in Discourse Research Initiative (DRI) meetings.[1] The DAMSL annotation manual and annotation tools have been developed at Rochester. The annotation manual describing each action class in DAMSL and when it applies is available at "ftp://ftp.cs.rochester.edu/pub/packages/dialog-

[1] See the DRI home page for more details: `http://www.georgetown.edu/luperfoy/Discourse-Treebank/dri-home.html`

annotation/manual.ps.gz". It is important to note that this is a working document rather than a completed project, and the scheme is sure to be refined and extended in subsequent meetings once we have more experience with using DAMSL. In addition, the focus of DAMSL has primarily been on task-oriented dialogs, where the participants are focused on accomplishing a specific task. While we believe the taxonomy is applicable to all dialogs, the distinctions made here are the ones most prevalent and important to task-oriented situations.

The following sections of this paper will give a short description of the DAMSL scheme and discuss some preliminary inter-annotator reliability scores.

The DAMSL Annotation Scheme

Speech act theories generally only allow an utterance to have one speech act and maybe an additional indirect speech act. This is a problem because utterances can simultaneous respond, promise, request, and inform. To handle responses, researchers have created subclasses of Representative/Inform such as Accept and Reject (ASF^{+}94). However, consider the two dialogs below. Note, the labels *u* and *s* are used to refer to different speakers.

```
u: let's finish the report today
s: okay

u: it is raining
s: oh no
```

In contexts above, it seems strange that the utterance "okay" would be labeled with the same category as "it is raining" (both would be Informs and "okay" would be an Accept to be more specific). The accepting and rejecting character of an utterance seems to belong in a separate action class dealing with a speaker's reactions to previous utterances. You can find other types of phenomena that fit into this class such as signaling understanding with acknowledgments and answering questions. These phenomena will be called Backward Communicative Functions while speech act categories not related to responses will be called Forward Communicative Functions since they affect the future portions of the dialog. For example, a request for information will cause you to give an answer. A third set of labels, Utterance Features, includes features that characterize the content and structure of utterances.

Forward Communicative Function

The Forward Communicative Functions include the speech act categories: Representatives, Directives, and Commissives. However, the categories are now independent so an utterance can simultaneously give information, make a request, and make a promise (although it is unlikely one utterance will do all of these).

All the Forward Communicative Functions are shown below. Representatives, utterances making claims about the world, are now called Statements. This class is further subdivided based on whether the speaker is trying to affect the beliefs of the hearer, or is repeating information for emphasis or acknowledgment. Directives fit under the more general category, Influencing-Addressee-Future-Action, which includes all utterances that discuss potential actions of the addressee. Directives are subdivided into two categories: Info-Request, which consists of questions and requests such as "tell me the time", and Action-Directive, which covers requests for action such as "please take out the trash" and "close the door". Influencing-Addressee-Future-Action also includes Open-Option where a speaker gives a potential course of action but does not show preference toward it, "how about going to Joey's Pizza". Commissives are given the more descriptive name, Committing-Speaker-Future-Action, and are subdivided into Offers and Commit(ments). The Performative category includes utterances that make a fact true in virtue of their content, such as your boss firing you by saying "you are fired"). Since the Performative category is an independent component of the Forward Function, such utterances can be marked in other categories (such as Statement) as well. The Other Forward Function category is a default choice for communicative actions that influence the future of the dialog in a way not captured by the other categories. Sentence initial words such as "okay" are often separated into separate utterances and marked as Other Forward Function. These words may have Forward Communicative Functions such as signaling a repair or change in topic or holding the turn (while the person is thinking) as well as Backward Communicative Functions such as Accepting and Acknowledging. Future work in this annotation effort will include developing classes of Other Forward Functions.

- Statement
  - Assert
  - Reassert
  - Other-Statement
- Influencing Addressee Future Action
  - Open-option
  - Directive
    - Info-Request
    - Action-Directive
- Committing Speaker Future Action
  - Offer
  - Commit
- Performative
- Other Forward Function

Backward Communicative Function

The Backward Communicative Functions in the DAMSL scheme are shown below. The classes Agreement, Understanding, Answer, and Information-Relation are independent so an utterance may simultaneously accept information and acknowledge that the information was understood as well as answer a question.

Agreement has several subclasses; Accept and Reject refer to fully accepting or rejecting an utterance or set of utterances. Accept-Part and Reject-Part refer to partially accepting or rejecting a proposal. In the next version of DAMSL, a label such as Accept-and-Reject will be added to deal with utterances such as “I’ll take everything except the curtains”, that both accept and reject parts of an offer (assume that this is a response to an offer such as “what would you like to take to school”). Note, it is difficult to break this into accepting and rejecting pieces since separating “I’ll take everything” from the rest changes its meaning. Hold refers to utterances such as clarification questions that delay the listener’s reaction to a proposal or question. Maybe refers to cases where the listener refuses to make a judgment at this point. The examples in figure 1 illustrate each type of agreement in response to the offer “Would you like the book and its review?”.

The Understanding dimension concerns whether the listener understood the speaker. The listener may signal understanding or non-understanding or attempt to correct the speaker (showing that they either did not understand or that they did understand but that the speaker misspoke). Non-understanding can be indicated by utterances such as “huh?”, clarification questions (“To Dansville?”) and by explicit questions about what the speaker said or meant. Understanding can be indicated by acknowledgments such as “right” or “okay”, by repeating some of the speaker’s utterance, or by continuing or completing the speaker’s sentence.

```
Context:     A: Would you like the book and
                its review?
Accept       B: Yes please.
Accept-Part  B: I’d like the book.
Maybe        B: I’ll have to think about it
                (intended literally)
Reject-Part  B: I don’t want the review.
Reject       B: No thank you.
Hold         B: Do I have to pay for them?
```

Figure 1: Example annotations using the Agreement Label

The Answer dimension indicates that an utterance is supplying information explicitly requested by a previous Info-Request act. This is a highly specific function that you might expect could be generalized into some other form of response, but we have not as yet been able to identify what the generalization would be.

Information-Relations are intended to be like the Rhetorical Relations of (MT87) and describe how the information in the current utterance relates to previous utterances in the dialog: “does the utterance provide evidence for a claim in a previous utterance”, “is it giving an example of something mentioned previously?”. So an utterance can certainly have Information-Relations as well as answering a question, accepting a proposal, and acknowledging understanding. A set of information relations for DAMSL has not been constructed yet.

- Agreement
  - Accept
  - Accept-Part
  - Maybe
  - Reject-Part
  - Reject
  - Hold
- Understanding
  - Signal-Non-Understanding
  - Signal-Understanding
    - Acknowledge
    - Repeat-Rephrase
    - Completion
  - Correct-Misspeaking

- Answer
- Information-Relation

Utterance Features

The third part of DAMSL consists of the Utterance Features, which capture features of the content and form of utterances. The Information Level dimension encodes whether the utterance deals with the dialog task, the communication process, or metalevel discussion about the task. This dimension eliminates the need to have tags such as Communication-Info-Request, for utterances such as "What did you say?", and Task-Info-Request for utterances such as "What times are available?". With this information, we can identify three independent subdialogs within a single dialog. The topic motivating the dialog is developed and discussed in the Task part of the dialog. The Task-Management part of a dialog involves explicit planning and monitoring of how well the task is being accomplished. The physical requirements of the dialog (such as being able to hear one another) are maintained in the Communication-Management part of the dialog. Note that in some sense all utterances have a Communication-Management component. It is only marked, however, when the utterance has no Task or Task Management component.

Communicative Status and Syntactic Features are hints about the possible communicative acts of an utterance. Communicative Status labels of Abandoned and Uninterpretable suggest that an utterance has little effect on the dialog because it was broken off or garbled beyond recognition. Syntactic Features currently only flag conventional sentences such as "hello", "may I help you" and exclamations such as "wow". Conventional utterances are often at the Communication Management level and Exclamations are usually Statements about the speaker's feelings.

- Information Level

  Task

  Task Management

  Communication Management

  Other

- Communicative Status

  Abandoned

  Uninterpretable

- Syntactic Features

  Conventional Form

  Exclamatory Form

Utterance Segmentation

This paper assumes an utterance is a set of words by one speaker that is homogeneous with respect to Information Level and Forward and Backward Communicative Functions. This means in a case like the one below when the set of communicative acts being conveyed changes, a new utterance begins:

```
utt1 u: we'll get that couch
utt2    how about that end table?
```

Utterances are not required to be single clauses, and if the set of communicative acts being conveyed stays the same, several clauses may form one utterance:

```
utt1 u: we'll take the train to Corning
|       then we'll pick up boxcars in Avon
|       and go on to Dansville to pick up oranges
```

Usually the only utterances shorter than a clause are sentence initial words such as "okay". Words such as "um" and "er" and phrases such as "I mean" have communicative functions separate from the clauses in which they appear. However, utterances are not hierarchical so labeling "I mean" as a separate utterance below would mean cutting off "Friday" from "we'll go Tuesday". DAMSL is not designed for annotating speech repairs, reference, or other intra-clause relations so we decided to use a simple definition of utterance that leaves out such phenomena.

```
utt1 u: we'll go Tuesday I mean Friday
```

Short interruptions by another speaker do not break up an incomplete utterance (incomplete meaning an interruption in the syntax). In the example below, "take the product to to Corning" is treated as one utterance. So this is a functional notion of utterance as opposed to a definition based on prosody.

```
u: take the product to
s: yes?
u: to Corning
```

Experiments

One of the key requirements for any annotation scheme is that the scheme can be used reliably by trained annotators. To explore this, we performed a reliability experiment on the current DAMSL scheme using test dialogs from the TRAINS 91-93 dialogs (GAT93; HA95), a corpus of discussions between humans on transportation problems involving trains. One person (the user) was given a problem to solve such as shipping boxcars to a city and the other person was instructed to act as a problem solving system. In addition, this system had information (the times to travel various paths) that the manager did not. An excerpt from a TRAINS dialog is shown in figure 2.

```
u: _we_ have to ship a boxcar of oranges to
    Bath by 8 AM
 : and it is now midnight
s: okay
u: okay all right so there are two boxcars
    at Bath and one at Dansville and there's
s: and there's
u: wait I've forgotten where the oranges are
    where are the oranges
s: the oranges are in the warehouse at
    Corning
u: okay so we need to get a boxcar to
    Corning
s: right
u: alright so why don't we take one of the
    ones from Bath
```

Figure 2: An excerpt from a TRAINS 91 dialog (d91-7.1)

Three undergraduates and a graduate student were given informal training consisting of annotating some dialogs and having their results compared against canonical annotations as well as comparing their results against one another. A GUI-based annotation tool, DAT[2] was developed to test the DAMSL scheme. This tool displays the dialogs and can play audio for individual utterances so annotators can listen to the actual dialogs as well as studying the transcripts. DAT also gives warnings to users when a suspicious pattern of inputs is entered and allows them to correct the annotation if desired. Here is a list of what the tool defined as suspicious.

- Question and answer have different Info Levels
- An acceptance that is not an acknowledgment
- An acknowledgment (but not acceptance) that is not at the Communication Management Information Level
- Answers that are not Asserts.
- A check question[3] whose answer does not have an Agreement label.
- A response to an Action-Directive or Open-Option that does not have an an Agreement label.
- A response to a question that is not an answer.

After training, the students independently annotated a series of dialogs as shown in table 1[4]:

Dialog	Utts	Annotators	Total Annotations/Tag
d1	133	2 UG	266
d2	72	2 UG	144
d3	40	2 UG 1 GR	120
d4	41	1 UG 1 GR	82
d5	19	1 UG 1 GR	38
d6	88	1 UG 1 GR	176
d7	159	1 UG 1 GR	318
d8	52	1 UG 1 GR	104
total	604		1248

UG = undergraduate
GR = graduate student

Table 1: Experimental Setup

Results

The statistics used to measure interannotator reliability are percent pairwise agreement (PA), expected pairwise agreement (PE), and kappa (PA adjusted by PE): $K = \frac{PA - PE}{1 - PE}$. These are defined formally in (SJ88). Statistics were collected for each tag over each dialog. Then an average PA, PE, and kappa for each tag were computed as follows: average = $\sum(d_i * TAPT_i)/\sum TAPT_i$ where TAPT is total annotations per tag and d_i is the PA, PE, or kappa for a tag over dialog i.

According to (Car96) even for tentative conclusions to be drawn, kappas must be above 0.67 with above 0.8 being considered reliable. The results suggest that with revisions to the annotation manual, annotators should be able to produce labelings of at least usable quality (between 0.67 and 0.8). The results are shown in tables 2, 3, and 4 (note, IAF is Influence on Addressee Future Action and CSF is Committing Speaker Future Action). The Resp-to abbreviation refers to Response-to, an annotation of which utterances a response responds to. Note, Exclamation was only labeled yes three times in the test set and Performative was never labeled yes in the test set, so both labels are left out of consideration.

Two of the lowest kappa scores of the annotations occur in the Committing-Speaker-Future-Action and

[2] Available at
http://www.cs.rochester.edu/research/trains/annotation/

[3] A check question is defined in the annotation manual as a statement about the world made for the purposes of confirmation, as in "We're using the blue sofa, right?". Check questions are labeled as both Asserts and Info-Requests and their answers are both Asserts and Accepts (or possibly Rejects).

[4] d1-d8 correspond to TRAINS dialogs d92a-2.1, d92a-2.2, d92a-3.1, d92a-4.1, d92a-4.3, d93-13.2, d93-13.3, and d93-16.1.

Measure	Statement	IAF	CSF	Other For Funct
PA	0.82	0.88	0.88	0.93
PE	0.49	0.60	0.87	0.85
Kappa	0.66	0.70	0.15	0.48

Table 2: Reliability for Main Forward Function Labels

Measure	Understand	Agree	Ans	Resp-to
PA	0.83	0.78	0.95	0.84
PE	0.60	0.62	0.73	0.29
Kappa	0.57	0.42	0.76	0.77

Table 3: Reliability for Backward Function Labels

Agreement dimensions. The major reason for disagreements in these dimensions is that annotators have a hard time deciding whether a response is an acceptance (labeled under the Agreement dimension) or just an acknowledgment. In the example below, it is unclear whether *u* thinks going through Corning is a good idea or is waiting to hear more before making a judgment.

```
s: so we'll take the train through Corning
u: okay
s: and on to Elmira.
```

Hearing the audio sometimes helps, but there are many cases where the annotator would have to be able to read the speaker's mind in order to make the distinction. To make matters worse, this one decision also affects two other dimensions: the Committing-Speaker-Future-Action dimension because acceptances many times mean commitment but acknowledgments do not, and the Information Level dimension since acknowledgments are at the Communication Management level while agreements are at the Task level. Thus, we have differences in at least three dimensions based on a single subtle distinction that often cannot be made. The two interpretations are summarized in table 5.

This problem, where a slight change in interpretation causes major changes in the annotation, clearly indicates a need for revision. One possibility would be to introduce some labels that capture the ambiguity, but this would have to be done in each dimension and might serve to aggravate the problem by introducing additional choices. The other possibility is to force an agreement reading based on how the proposal/request is eventually treated in the dialog. Thus in the example above, unless the speaker goes on to reject or question the proposal, the response would count as an implicit accept, Interpretation 1 would not be allowed, and the response would have to be labeled with some Agreement tag. Following this rule could be encouraged by having DAT give the user a warning every time an utterance is tagged an Acknowledgment but no Agreement tag is specified.

Measure	Info level	Abandoned	Unintelligible
PA	0.83	0.98	0.99
PE	0.57	0.94	0.98
Kappa	0.60	0.64	0.14

Table 4: Reliability for Utterance Features

Dimension	Interp 1	Interp 2
Understanding	ACK	ACK
Agreement	N/A	ACCEPT
CSF	N/A	COMMIT
Info Level	COMM-MANAGE	TASK

Table 5: Two interpretations of an utterance such as "okay".

The Other-Forward-Function category also has a low kappa score; this is partially due to the fact that the expected agreement for it is high since its value is usually Not-Present. This category applies most often to words such as "okay" that are very ambiguous in their meaning even when heard in context. It will be interesting to develop subcategories of Other Forward Function such as "turn holding" and "signaling a repair" to give us a better idea of what phenomena annotators are having trouble labeling.

Most of the other labels have kappas around 0.6 meaning the annotations are fairly reliable but that some problems still remain. One problem that affects several labels involves check questions. Check questions are statements about the world made for the purposes of confirmation, as in "We're using the blue sofa, right?". Check questions are labeled as both Asserts and Info-Requests and their answers are both Asserts and Accepts (or possibly Rejects). However, it is difficult for annotators to consistently recognize a check question, leading to disagreements in the Statement and Influencing Addressee Future Action dimensions (is it an assert, is it a question?), and disagreements about whether the next utterance is an Answer and Assert or simply an Accept (or Reject).

Another problem arises with indirect speech acts such as requests made by statements such as "it would be nice to have some light". There is a continuum of interpretations for such an utterance, ranging from a pure Assert act through to a pure Action-Directive act depending on the annotator's view of what the speaker intended and how the utterance was taken in the dialog. The DAMSL scheme alleviates this problem somewhat by not forcing an annotator to choose between the two options. They can mark an utterance as both acts. In practice, however, we still see a fair amount of inconsistency and some more specific guidance appears to be needed. This may have to be done on a domain-by-domain basis, however. For instance, in the TRAINS domain, the users often state their goals, as in "I have to get trains there by noon". We have been taking these utterances simply as Asserts, but this is somewhat arbitrary as there is a sense in which such utterances influence the hearer's future action as with Action Directives.

Another difficult example in TRAINS occurs when the speaker summarizes a plan that has already been developed, as in:

```
utt1: s: we'll go through Corning
utt2: u: mm-hm
utt3: s: pick up the oranges, and
          unload at Dansville
```

If utt1 and utt3 are really just descriptions of what has been agreed upon, they would be Reasserts, but annotators often want to add an Action-Directive interpretation as well because of their surface form. Such cases may be resolved with domain-specific instruction, but it is unclear whether unambiguous generic instructions can be found.

Another problem with the Statement dimension is the label, Reassert. When information is asserted that has been discussed previously, the annotators have to decide whether the information was forgotten by the hearer (and thus constitutes an Assert) or whether the speaker is trying to reintroduce the information to make a point (and hence it would be a Reassert). A similar confusion occurs with the Repeat-Rephrase tag of the Understanding level where annotators have to decide how far back a Repeat-Rephrase utterance can refer and how close the paraphrase must be. Annotators also get confused if a speaker simultaneously makes a repetition and goes on to make a correction or completion. Some work needs to be done to clarify the definitions of these labels.

Another label that confuses annotators is the Task Management label of the Information Level dimension. In TRAINS, the domain is planning so an utterance such as "we can't do that because there is a train already on that track" is Task level but something like "we could do that another way. do you want to change the plan?" would be considered Task-Management since it explicitly discusses the course of the dialog while the first only implicitly signals a possible change in the course of the dialog. The difference is very subtle and hard to annotate.

Conclusions

For the interpretation of a dialog, it is critical to have a primitive abstraction of the purpose of each utterance. The general strategies of a system trying to participate in a dialog or understand a dialog will be tied to these primitives. For example, a system might have a rule such as "a statement is something to add to the database". The system will then use a more detailed representation of the utterance in its processing. As another example, if an utterance is an Information Request, the system will process the semantic interpretation of the sentence to determine what information is being asked for. As the system adds utterances to its data structures, it will create higher level forms such as hierarchical multi-agent plans and discourse structures analogous to paragraphs and chapters.

The representation driving the creation of such data structures needs to be extremely flexible. Speech act theory is currently the most popular representation used; however, it is a set of mutually exclusive categories and does not allow utterances to perform multiple actions simultaneously. Unfortunately it is common in dialogs, especially problem solving dialogs, for an utterance to perform several actions such as signaling understanding and accepting a task. The DAMSL annotation scheme has many independent layers that allow the labeling of all these actions. The annotation scheme also separates utterances into those that deal with the communication process, those that deal with the task at hand, and utterances that deal with how to solve the task. This type of annotation is not typically seen in speech act theory but it is critical to interpreting dialogs since the utterances at these levels must be processed using different strategies. Dealing with the communication process might mean repeating a previous utterance or changing the volume of the speech output. Utterances discussing how to solve the task can be viewed as direct messages to a system's planner, "let's solve this subgoal first" or "is that the best solution".

The DAMSL annotation scheme makes reference to linguistic phenomena such as "check questions" and "acknowledgments by repetition". A serious question is whether these phenomena can be defined precisely enough for humans to recognize them and annotate them reliably in a corpus of dialogs. A corpus reliably annotated with DAMSL labels would provide a valuable resource in the study of discourse as well as a source of training and testing for a dialog system using DAMSL labels in its utterance representation. The experiments in this paper show reliability results close to those considered usable for drawing scientific conclusions. Given that this is the first major test of

DAMSL, it seems likely that the revisions mentioned in the Results section will allow reliable annotation with DAMSL.

Acknowledgments

This work was supported in part by National Science Foundation grants IRI-95-03312 and IRI-95-28998, the latter of which was under a subcontract from Columbia University. Thanks to Teresa Sikorski for her help in running the experiments and interpreting the results. Thanks also to Lenhart Schubert for his helpful comments, and George Ferguson for his annotation tool.

References

J. Allwood. An activity based approach to pragmatics. Technical report, Dept of Linguistics, Univ. of Goteborg, 1995.

J. F. Allen, L. K. Schubert, G. M. Ferguson, P. A. Heeman, C. H. Hwang, T. Kato, M. N. Light, N. G. Martin, B. W. Miller, M. Poesio, and D. R. Traum. the TRAINS project: A case study in building a conversational planning agent. Technical Report 532, Department of Computer Science, University of Rochester, Rochester, NY 14627-0226, September 1994.

J. Carletta. Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, 22(2), 1996.

Philip R. Cohen and Hector J. Levesque. Rational interaction as the basis for communication. In Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors, *Intentions in Communication*, SDF Benchmark Series, pages 221–255. MIT Press, 1990.

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

D. Gross, J. Allen, and D. Traum. the TRAINS 91 dialogues. TRAINS Technical Note 92-1, Department of Computer Science, University of Rochester, Rochester, NY 14627-0226, 1993.

P. Heeman and J. Allen. the TRAINS 93 dialogues. TRAINS Technical Note 94-2, Department of Computer Science, University of Rochester, Rochester, NY 14627-0226, 1995.

M. Hancher. The classification of cooperative illocutionary acts. *Language in Society*, 8(1):1–14, 1979.

W. C. Mann and S. A. Thompson. Rhetorical structure theory: a theory of text organization. Technical Report ISI/RS-87-190, Univ. of Southern CA - Information Sciences Institute, 1987.

J. R. Searle. *Language, Mind, and Knowledge. Minnesota Studies in the Philosophy of Science*, chapter A Taxonomy of Illocutionary Acts. University of Minnesota Press, 1975.

S. Siegel and N. J. Castellan Jr. *Nonparametric Statistics for the Behavioral Sciences.* McGraw-Hill, 1988.