References

[Allen and Schubert, 1991] James F. Allen and Lenhart K. Schubert, "The TRAINS Project," TRAINS Technical Note 91-1, Computer Science Dept. University of Rochester, 1991.

[Hartman, 1990] Leo B. Hartman, *Decision Theory and the Cost of Planning*, PhD thesis, University of Rochester, Rochester, NY 14627, 1990, available as URCS TR # 372.

[Martin and Miller, 1991] Nat Martin and Brad Miller, "The TRAINS-90 Simulator," TRAINS Technical Note 91-4, Computer Science Dept. University of Rochester, 1991.

[Poesio, 1992] M. Poesio, "Conversational Events and Discourse State Change: A Preliminary Report," In *Proc. KR-92*, Boston, MA, October 1992.

[Traum and Hinkelman, 1992] David R. Traum and Elizabeth A. Hinkelman, "Conversation Acts in Task-oriented Spoken Dialogue," *Computational Intelligence*, 8(3):575–599, 1992, Special Issue on Non-literal language.

5 The Dialogues

The collected dialogues are given below. There are 8 subjects, each of whom participated in two dialogues. For each subject, first is given the warmup dialogue followed by the test dialogue. Here is a summary of the lengths of the dialogues:

Dialogue	**time**	**# turns**	**# utterances**
91-1.1	1'38"	20	63
91-1.2	9'51"	95	382
91-2.1	2' 2"	28	77
91-2.2	8'57"	99	333
91-3.1	58"	15	36
91-3.2	4'37"	45	166
91-4.1	1'56"	34	95
91-4.2	7'58"	89	313
91-5.1	39"	11	36
91-5.2	7'50"	102	351
91-6.1	1'18"	20	54
91-6.2	12'27"	162	559
91-7.1	2'42"	33	113
91-7.2	7'47"	76	330
91-8.1	1'26"	36	81
91-8.2	10'19"	135	430

Table 1: Dialogue Lengths

A compressed ascii version of the dialogues is available by anonymous ftp from ftp.cs.rochester.edu, in the directory pub/papers/ai, file 92.tn1.trains.91.dialogues.txt.Z (file 92.tn1.trains.91.dialogues.ps.Z is a compressed postscript version of this document). For the audio signal, please write the second author at the address listed on the title page.

```
56.5  : goes back
56.6  : makes OJ
56.7  : and then takes
56.8  : the
57.1 M: puts the OJs  in the   tanker
58.1 S:       +southern route+
58.2  : right
58.3  : load the tanker
58.4  : then go back
58.5  : to Corning and take the southern route to Avon
```

Even though thematically 58.1 belongs in the same turn with 56.5-8 and not with 58.2 (the response to 57.1), that would involve a higher-order notion of coherence. Since 57.1 started before 58.1, 58.1 starts a new turn. Similarly, 58.2 does not start a new turn, because no new utterance by M started after the start of 58.1.

4.2 Other Markings

/ is used to signal a discontinuity in the flow of the utterance unit. At the end of an utterance, it signals an abandoned intonation phrase, one that does not have a boundary tone, but is not continued with subsequent speech, although there is no significant pause in the speech signal. In the middle of an utterance, this signals a repair marker - generally speech is clipped and speeded up, but remains in the same intonation phrase.

.. is used to signal a pause of one beat within an utterance.

... signals a silence of 1-2 second duration. The marking is indicated at the beginning of the utterance following the silence when the silence occurs within a turn and on its own line when the silence falls at a turn boundary.

[nsec] indicates silences of more than 2 seconds, rounded to nearest second (n). This occurs on it's own line between the utterances.

[] brackets are also used to indicate non- or meta-linguistic sounds. Of these, [snirk] stands for the audible component of a smirk

oranges underlining signals a word with extra stress.

%% signals a comment by the analyst

4 Conventions for Marking Transcripts

Our aim in these transcriptions is to provide a maximum of useful information with a minimum of theoretical overhead and controversy. We have tried as much as possible to use only acoustic features rather than syntactic or pragmatic analyses in segmentation and non-word markings. Each utterance is labeled with the utterance, the speaker ("S" for System, "M" for Manager, but labeled only for the first utterance in each turn) followed by the text of the utterance.

4.1 Segmentation

The text is segmented into *turns* and within these turns, *utterance units*, roughly corresponding to intonational phrases. By *turn*, we mean any stretch of speech by one speaker without intervening speech by the other. We do not imply any coherence of purpose within a turn; turns will often contain unrelated bits of speech with a long intervening silence. Also, many researchers do not count so-called *backchannel responses* such as "ok" or "uh-huh" as turns in their own right or as dividing turns of another speaker. Since it may often be difficult to distinguish a backchannel response from other speech, we choose to count all speech the same.

An utterance unit is terminated by either

- a boundary tone
- a pause in speech longer than a single beat
- a resetting of the pitch level, starting a new intonation phrase

The utterances are numbered by turn and utterance. As an example, in the following segment, taken from dialogue 91-1.1,

```
7.5      : okay
7.6      : we have to get engines to the boxcars
8.1     S: right
```

the first utterance 7.5 is the fifth utterance in the seventh turn, the next is the sixth utterance in the seventh turn, and then, with the speaker change, the first utterance of the eigth turn.

A difficult case is the occurrence of simultaneous speech. We have marked simultaneous speech with the symbol "+". Speech which temporally overlaps speech by the previous speaker is set off with "+" markers signalling the points of commencement and completion of overlap. A new turn is started by every utterance by one conversant in which an utterance by the other conversant was started after the start of the previous utterance by the first conversant. Thus, in dialogue 91-4.2, we have the following sequence:

what problem you have to solve. You may use the layout in any way you wish to help you plan. The time is 12 midnight.

Remember that the system does not know what problem you need to solve, so you must tell it.

Once you have tried to warmup problem, please solve the problem described on the next page.

Have fun!

<< on a separate page >>

Test Problem

The situation is the same as before the last problem. The time is 12 midnight.

You have to get one tanker car of orange juice to Avon, and a boxcar of bananas to Corning by 3 PM today.

3.2 System Instructions

The "system" was given a map similar to that in Figure 1, but which also included timing information about loading and manufacturing commodities and transit between connected cities. In addition, for the "test" problem, after completion of the plan the system introduced a problem with one of the boxcars, requiring plan revision.

3 Dialogue Collection

The dialogues in this study were collected on a single day (June 6, 1991) under the circumstances described in the previous section. All of the participants were native speakers of North American English in their 20's and 30's, studying Computer Science or Linguistics at the University of Rochester. The "System" was played by the same person for all dialogues. Participants in the dialogues were known to each other. For each "Manager", two dialogues were collected in succession: first a short "warmup" dialogue, to introduce the "manager" to the domain, then by a longer "test" dialogue.

3.1 User Instructions

These are the instructions that were given to the user before the dialogue collection started. The user was also presented a copy of the map in Figure 1.

The TRAINS World

We are interested in building a computer system that can help people in simple problem solving tasks. We are collecting some conversations with a person simulating the system in order to better understand the task.

You will be making plans to manufacture and/or ship various goods between cities as shown on the map on the other page. This map shows the layout consisting of several different cities connected by different rail lines. Some cities have warehouses for storage of goods. In this situation, the goods are either oranges or bananas. The rail lines are single track so trains cannot pass each other on the same track.

The "system" will have reasonably up-to-date information about the state of the world and you can ask it questions at any time. You, however, have the final say in what actions are to be performed. The system will keep you informed of the situation when asked, and try to anticipate problems that arise in the plan and help you when possible.

In looking at the map, you will see the following:

Engines - labeled E1, E2 and E3
Boxcars for shipping - as indicated on the map
A Tanker car, at Corning, usable for shipping OJ
Two warehouses - one for oranges at Corning, and one for bananas at Avon.
An OJ Factory at Elmira - which can convert a boxcar of oranges into a tanker of OJ

While it may look complicated, it really is quite simple once you get started. The "system" is available to help you at any time. The best way to learn is by trying a sample problem. Here's one to get started with:

WarmUp Problem

You need to ship a boxcar of oranges to Bath by 8AM today. Here are some specific facts about the current situation, which are fairly reliable. You may obtain further information and confirm facts by asking the system. Remember that the system doesn't initially know

2.1 Uses of the Data

There are a wide range of problems that we intend to use this data to help solve. These dialogue have been extensively used in our work on plan-based dialogue models and speech acts (e.g., [Traum and Hinkelman, 1992], on discourse models for handling co-reference and scoping (e.g., [Poesio, 1992] and other projects in the TRAINS project. For an overview of the project, see [Allen and Schubert, 1991]. Here are a few examples of work in progress or work we are planning to start once additional data is available.

- parsing spontaneous dialogue. It is well known that spoken language is quite different in structure than written text. In addition, spontaneous dialogue seems very different than spoken language in a setting where speakers have time to plan and prepare their utterances in advance. This database will allow one to study the structure of interactive dialogue.

- speech repairs. From our pilot study, it is clear that in-line speech repairs are very common in interactive dialogue. We find almost one third of our turns contain repairs: repetitions, corrections, and cancellations.

- prosody in spontaneous dialogue. While the study of intonation and prosody is an area of growing interest, there is little work done so far on spontaneous dialogue. More frequently, studies involve read speech, or utterances that are planned and prepared in advance of the utterance. The database will provide a rich source of data on repairs;

- turn-taking strategies. With the precise timing information available, more detailed studies of turn-taking will be possible than previously could be done using only analyses based on timing by hand. Effects of intonation, use of certain phrases, and so on, can be studied in relation to the turn-taking behavior observed;

- the use of referring expressions in dialogue. With a large database of dialogue, and some additional annotation on co-reference phenomena, various theories of reference resolution, such as centering, will be able to be empirically tested. In addition, the effect of segmentation into intonational phrases on reference phenomena can be studied;

- global discourse structure. Several theories of the global structure of discourse have been proposed, but none have been the subject of systematic study because of a lack of enough data in a well-behaved domain.

- speech act analysis. It is believed that common ways of phrasing utterances conventionally signal certain (indirect) speech acts. The database will allows us to study such issues quantitatively.

- the effect of general world knowledge on dialogue structure. A number of theories of topic flow and change have been proposed using plan-based models, but there has never been a systematic study of such theories based on a large corpora of dialogues.

- accounting for the discourse behavior present in natural dialogue, *including* local coherence issues of reference and scope ambiguity, and global discourse issues such as topic flow and structure;

- representing the reasoning and control of the discourse agent, including reasoning about plans in the trains domain as well as the reasoning processes that drive the system's behavior in the dialogue itself; and

- providing the knowledge representation and reasoning tools that are required to support planning and scheduling in realistic size domains, including reasoning about time, events, actions and plans.

2 Data Collection in the TRAINS Domain

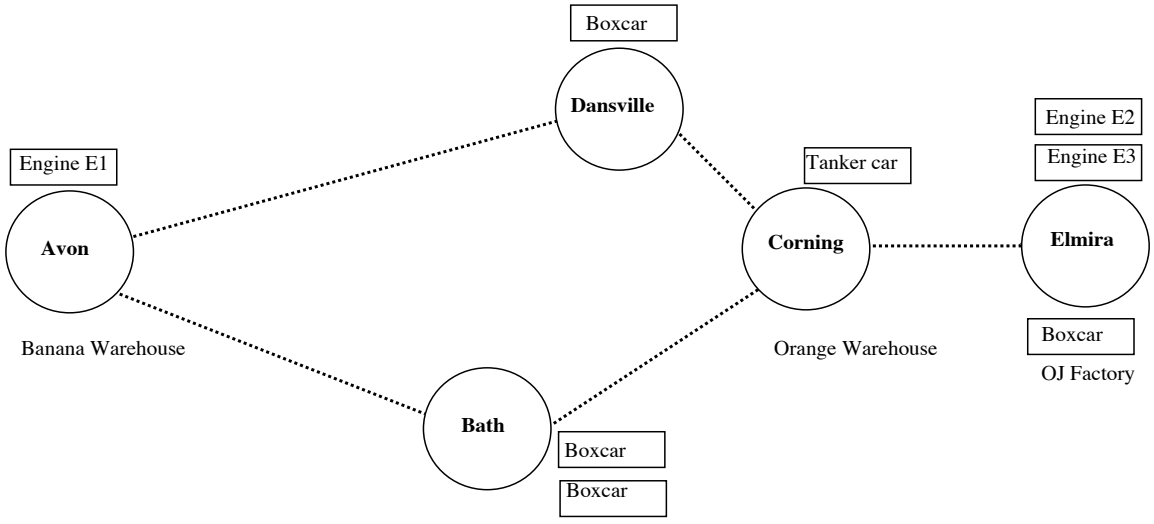

Figure 1: Trains World Set-up for Dialogue Collection

To provide empirical data to drive the research, we are collecting a corpus of dialogues in this domain using a person to play the role of the system. These dialogues are collected in a normal office setting but the two conversants cannot see each other and can only communicate using microphones and headphones. The only common domain knowledge they share is a map of a TRAINS world. The world used for the dialogues in this report is shown in Figure 1. We have found that even a world this size provides a complex enough setting for extended dialogue. The dialogues were recorded on audio tape and have been transcribed and annotated by at least two annotators. The annotations were then compared and the discrepancies were resolved. This report contains transcripts of the initial dialogues collected June 1991. Additional dialogues have been collected since and are now being transcribed using the WAVES system. These dialogues will be made available as transcripts, and as data files containing the actual speech signals, at a later date.

1 Overview of the TRAINS Project

The TRAINS project is a long-term research project to develop an intelligent planning assistant that is conversationally proficient in natural language. In particular, the TRAINS agent helps a person (the manager) construct and monitor plans about a railroad freight system. The manager is responsible for assigning cargo to trains and scheduling shipments, scheduling various simple manufacturing tasks, and for revising the original plans when unexpected situations arise during plan execution. The system aids the manager in all aspects of this task by interacting in natural language and (eventually) through a graphical interface. In particular, the system typically will perform the following tasks:

- evaluating suggested courses of action, such as calculating expected completion times, detecting conflicts that might interfere with the actions, and so on;
- filling in details of the proposed plan that do not require the manager's attention;
- suggesting ways to solve particular subproblems as they arise;
- presenting and describing the current state of the world and how the proposed plan may affect it, including answering questions from the manager;
- dispatching the plan to the different agents in the world, including the train engineers and factory managers;
- interpreting reports back from the engineers and factory managers in order to monitor the progress of the plan and to anticipate problems before they arise; and
- coordinating the correction and modification of plans with the manager.

To test this system, we have constructed a simulator for the TRAINS world, which also simulates the individual train engineers and factory managers as autonomous agents [Martin and Miller, 1991]. The simulator allows us to actually "execute" the plans that are constructed and to provide realistic tests for our work in plan execution monitoring.

While we aim to produce a functional system, the system itself is not really the goal of the effort. Rather, the domain is a tool for forcing our research to address the problems that arise in building a complete dialogue system. For instance, the dialogue module must be able to handle a wide range of everyday discourse phenomena rather than handling a few selected problems of theoretical interest. While this approach focuses our research directions, the solutions that we seek are general solutions to the phenomena rather than specific solutions that happen to work in the TRAINS domain. The TRAINS project currently has several main foci of research:

- parsing and semantically interpreting utterances as they arise in spoken language, including sentence fragments, repairs and corrections, and long sequences of utterances that incrementally provide information that eventually combines to form a complete interpretation;

Contents

The Trains 91 Dialogues

Derek Gross
derek@psych.rochester.edu

James F. Allen
james@cs.rochester.edu

David R. Traum
traum@cs.rochester.edu

The University of Rochester
Computer Science Department
Rochester, New York 14627

TRAINS Technical Note 92-1

July 1993

Abstract

This report contains a small corpus of transcriptions of task oriented spoken conversations in the TRAINS domain. Included are 16 conversations, amounting to over 80 minutes of speech. Also included are a description of the task and collection situation and the conventions used in transcription and utterance segmentation.

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 Peggy Meeker for assistance in formatting the postscript version of the dialogues.