# Coding Schemes for Spoken Dialogue Structure

David Traum*
TECFA, FPSE
Université de Genève
9 Rte de Drize, Bat D
1227 Carouge, Switzerland.
Email: David.Traum@tecfa.unige.ch

**Abstract**

This document contains a series of rough guidelines for coding several kinds of discourse-related phenomena in spoken dialogues. Different quasi-independent schemes are proposed to handle different levels of structure.

## Contents



*These ideas were developed while coding dialogues from the TRAINS project at the University of Rochester, in association with James Allen, Derek Gross, Peter Heeman, and ShinYa Nakajima, with additional input from Elizabeth Hinkelman and Massimo Poesio. Some of the ideas here belong to them, though the author takes full responsibilities for any misstatements or outlandish suggestions. Peter Heeman and Elisabeth Maier also provided very helpful comments on previous versions of this document.

## List of Tables



# 1 Introduction

Below is presented a set of conventions for marking dialogue phenomena. These are somewhat independent of each other, as they are mostly covering different phenomena. It should be possible to use any of the schemes in the sections below without much regard for the others, although individual marking will sometimes be sensitive to things like utterance segmentation. Examples are presented from dialogues collected in the TRAINS domain [Allen *et al.*, 1995], in particular those described in [Gross *et al.*, 1993] (hereafter called the TRAINS-91 corpus) and [Heeman and Allen, 1994b] (hereafter called the TRAINS-93 corpus). The dialogue phenomena covered are:

**Transcription** Word transcriptions should be done in a manner that is fairly readable (thus a general preference for conventional rather than phonetic spelling), yet still true to the original input rather than the conventions of written discourse, which generally abstract away from important features of spoken language, (such as disfluencies) and require higher level analysis (such as for insertion of punctuation marks). The guidelines used for the TRAINS-93 corpus [Heeman and Allen, 1994b] seem fairly reasonable. Some of the ideas below require identification of prosodic features as well as the text itself. I have no special recommendations for how this should be done, but presumably something along the lines of [Beckman and Hirschberg, 1994].

**Segmentation** How should a dialogue be segmented so that functional phenomena can be labeled for particular spans of text, and relations between spans can be conveniently recorded? We suggest division into turns and utterances within turns, where utterance boundaries are distinguished by prosodic boundaries.

**Inter-turn Coherence** How are individual utterances connected by relevance, grounding, topical coherence, etc. We have labeled the first utterance of each turn as to how it relates to utterances of the previous turn. This provides a rough estimation of inter-turn coordination and relevance.

**Grounding** How is the mutual understanding of the dialog achieved through individual utterance-level contributions? We suggest marking *Discourse Units* (spans of text that are grounded together) and *grounding acts*, which incrementally construct these DUs.

**Surface Form** What locutionary acts are being performed? Is the utterance a statement, question, command, fragment, cue-word, or non-linguistic utterance?

**Illocutionary function** What direct actions are the speakers performing by using language?

**Argumentation structure** how are illocutionary contributions tied together to form higher level *discourse segments*, arguments, subdialogues, etc.?

Some of the ideas for some of these levels are rougher or more theoretical (i.e., less tested) than others. The segmentation ideas in Section 2 have been coded by three coders and used to present the 16 dialogues of the TRAINS-91 corpus [Gross *et al.*, 1993]. The inter-turn coherence coding in Section 3 has been used by two coders on 26 dialogues from the TRAINS-93 corpus. The grounding coding has been used by the author to code about 10 dialogues from the TRAINS-91 corpus, and by the author and several others on parts of other dialogues from the TRAINS-91 and -93 corpora. In addition, this scheme was used by the TRAINS-93 system [Allen *et al.*, 1995] to automatically label and participate in dialogue. The coding schemes in Sections 5, 6, and 7, were not part of any systematic coding efforts, but were developed by at least three researchers in the TRAINS effort, used to label a few dialogues and implemented in the TRAINS-93 system.

### 1.1 General Principles

I think for coding to be true to the on-line nature of dialogue, it must be performed using only the information available to the participants at the time of utterance/interpretation. Thus prior context should be available (and will be crucial), but decisions about what category a particular span of text falls into cannot be based primarily on what comes after (i.e, how one of the participants reacted to the span in question). The text following the decision point can be used only as a guide to see whether a speaker seemed to share the opinion of the coder as to an interpretation, but should not be a deciding factor.

As Shober and Clark [1989] demonstrate, overhearers have less facilty for understanding than the particicipants an utterance was addressed to, even if the adressee and overhearer have equal knowledge. This has implications as well for the ability of an offline coder to identify dialogue functions as well as a participant could understand them in conversation.[1] While one might be tempted to use subsequent information to make up for this deficiency, this is also not without peril. As conversation is a constructive process, one must take care not to analyze the initial utterance as the subsequent reanalysis[MAKE THIS MORE COHERENT].

## 2 Segmentation

These ideas were used to transcribe the TRAINS-91 corpus. The description below is a slightly revised version of the contents of that document.

[1] this observation due to Peter Heeman, p.c.

## 2.1 Turns

The turn is the most obvious textual segmentation in dialogue. The most basic distinction is between spans where one speaker is speaking vs. spans in which another is speaking. The complications and differences in labeling come in during speaker overlaps or short utterances by one speaker in the midst of a longer contribution. 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 sometimes 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 segmenting speech by another speaker into different turns (i.e., before and after the *ok*). Since it may often be difficult to distinguish a backchannel response from other speech (such as a confirmation), we choose to count all speech the same as regards to turn segmentation.

## 2.2 Utterances

While segmentation into turns are sufficient for some purposes, such as constrained task-oriented dialogues in which each speaker is only allowed to utter one sentence before yielding to the other speaker, most analysis of discourse in unconstrained spoken dialogue has required looking at sub-turn level utterances. The principles for dividing turns has varied widely, from intonational (tone units [Pierrehumbert, 1980]) to syntactic (complete utterances or clauses), to pragmatic (speech acts or *moves* [Sinclair and Coulthard, 1975]. We favor using only the prosodic signals for segmentation, as being the least theory-laden. We use any of the following signals to terminate utterance units:

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

The middle one, silence is perhaps the most controversial. While a sufficiently long silence is an opportunity for another partner to take a turn, perhaps with some sort of repair or completion of an ongoing utterance, breaking an utterance at all pauses will interpret many interruption points of speech repairs as boundary points. It also makes a lot of dysfluencies (mid-'utterance' pauses) boundary points as well. This may cause problems for trying to code all speech repairs as intra-utterance phenomenona [Traum and Heeman, 1996].

## 2.3 Utterance labeling

Utterances are numbered by turn and utterance. As an example, in the following segment, taken from dialogue 91-1.1, in [Gross *et al.*, 1993].

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

The first utterance in this fragment, 7.5, is the fifth utterance in the seventh turn of the dialogue, the next is the sixth utterance in the seventh turn, and then, with the speaker change, the first utterance of the eighth 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 signaling the points of commencement and completion of overlap. A new turn is initiated by each utterance by a 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, in which M says "the" (56.8), after which S said "puts" (beginning of 57.1), at which point "the OJs in the" and "southern route" are said at the same time

```
56.5    S: 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 initiates 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.

# 3 Coherence

Perhaps the most basic kind of dialogue structure is relevance of an utterance to what has come before, i.e., whether or not an utterance coheres with other utterances [Sperber and Wilson, 1986]. This section describes an attempt at a very general notion of relatedness between utterances. These ideas were used to mark about 20 dialogues of the TRAINS-93 corpus, as described in [Traum and Heeman, 1996]. In particular, we used this scheme to label the relation of turn-initial utterances to utterances (by the other speaker) in the previous turn. This coding assumes some division of dialogues into turns and utterance units, as described in the previous section, but it does not rely on any particular assumptions of *how* to divide the turns into utterances, and is thus can be used independently of the segmentation principles (although, of course, different segmentation may produce different codings of a text).

## 3.1 Relatedness coding

We used single character annotations to indicate relatedness of utterances. These characters are shown on the line above the text which they label. The coding scheme is explained below, including some examples.

### 3.1.1 Category: Explicit Ack

**Label:** e

**Description:** This category includes all utterances by a new speaker which begin with what I call an **explicit** acknowledgment [Traum, 1994] – the first word of the turn by the new speaker is a word like "okay", "right", "yeah", "well", or "mm-hm" (also "no" and "nope" in a negative polarity context. These markers are also called *backchannel items* [Orestrom, 1983], or simply acknowledgments [Clark and Schaefer, 1989, Novick and Sutton, 1994]. This category shows relatedness to *something*, although it is usually not really possible to tell what exactly (if anything) it is related *to*. This marking is used both for true *backchannels*, which do not claim a turn but overlap speech by the other, isolated acknowledgments between turns by the other speaker, and longer turns beginning with one of these items. This category is also used for repair requests for which it is not clear exactly what should be repaired (e.g., an utterance consisting solely of "what?").

**Example:** d93-18.3 utt11-12

```
u:  so we have to start in Avon
     e
s:  okay
```

### 3.1.2 Category: related

**Description:** This category is for those utterances that *demonstrate* a relationship to a previous utterance. This can be a demonstration style acknowledgment (i.e., repetition or paraphrase of a previous utterance by the other speaker), a repair or repair request, or kind of continuation. Any second part to an *adjacency pair* [Schegloff and Sacks, 1973] (such as an answer to a question) would fit in this category. In addition, the recency of the relationship is marked as follows:

**Subcategories:**

| Label | Description |
|---|---|
| 0 | utterances **related to** the **most recent** utterance unit by the previous speaker. |
| 1 | **related** to the utterance unit one **previous** to the most recent but *not* related to the most recent. |
| 2 | related to utterance unit two previous to the last one (and not to anything more recent). |
| etc. | higher numbers for related to utterances further back. |
| , | **related** to previous material by the other speaker, but it is **unclear** to the coder whether they are related to the immediately previous utterance unit (which would be marked **0**), or to an utterance unit further back (which would be marked **1**, or **2**, etc.). |

### 3.1.3 Category: unrelated

**Label:** u

**Description:** these utterances are **unrelated** to previous speech by the old speaker. This could be either the introduction of a new topic, or could show a relation to previous speech by the same speaker, while ignoring speech by the other speaker in the prior turn.

**Example:** d93-18.3 utt13-18

```
U:  how long does it take to bring engine one to Dansville
     0
S:  three hours
     e
U:  okay <sil> and then <sil> back to Avon to get the bananas
     0
S:  three more hours si(x)- six in all
     u
U:  how long does it take to load the bananas
```

#### 3.1.4 Category: uncertain

**Label:** ?

**Description:** The coder is **uncertain** whether or not these utterances relate to previous speech by the other speaker.

#### 3.1.5 Combination Categories

In addition, we have a special subclass of markings for those cases in which the main category labeling is misleading or uninteresting.

**Category: After explicit**

**Label:** X-e

**Description:** A suffix of **-e** is used for those utterances which are **after** turns which consist *only* of **explicit** acknowledgments (or backchannels). Since these utterances do not have any content that new material *could* be related to, the categories for unrelated to last (**u**, **1**, etc.), are not very significant. Therefore, we append the marking **-e** to the other category in these cases.

**Subcategories:**

| Label | Description |
|---|---|
| **u-e** | material which is unrelated to previous material by the other speaker, but for which the last utterance unit by the other speaker was just an explicit acknowledgment. |
| **1-e** | material which is related to the penultimate utterance unit by the other speaker, but for which the last utterance unit by the other speaker contained just an explicit acknowledgment. This signal is often used for popping sub-dialogues, such as in the following example, in which the last turn by speaker **S** refers to the first utterance by speaker **U**. |

**Example:** d93-18.3: utt51-54

```
U:  how long does that take <sil> again
     0
S:  that ta(ke)- <sil> so just to go from Bath to Corning
     e
U:  mm-hm
     1-e
S:  two hours
```

### 3.2 More Examples

The following examples show labeling of discourse fragments, indicating as well the prosodic boundary tones [Pierrehumbert, 1980] identified from the speech input (labeled with %). The first example shows a simple sequence of **0** and **e**. The first response by S shows a relationship to the first contribution by U. Then the last two start with explicit acknowledgments.

**Example:** d93-13.2: utt18-22

```
                                           %
U:  how long is it from Elmira to Dansville
    0                   %          %
S:  Elmira to Dansville is three hours
    e  %  %             %                                            %
U:  okay um so why don't uh I send engine two with two boxcars to Corning
    e  %
S:  okay
```

The next example shows some of the other categories. The second turn by U shows the **u-e** category, since the previous turn was just "okay". This turn also ends without a boundary tone. S's second turn seems to be some sort of clarification attempt, but it is not clear if it is at all related to the content of U's previous utterance. U's final utterance is merely a continuation of his own previous utterance and is unrelated to the last installment by S (which also has no final boundary tone).

**Example:** d93-16.2 utt29-33

```
               %
U:  then do that
    e  %
S:  okay
     %   u-e
U:  um what is tank- what is engine okay engine two is
    ?
S:  you have you told
    u                        %
U:  it's picking up the oranges
```

### 3.3 Summary of Relatedness Markings

Here, again, are the choices for coding relatedness:

| Label | Description |
|---|---|
| e | **explicit** acknowledgment (e.g., "okay", "right", "yeah", "well", or "mm-hm") |
| 0 | **related to** the **most recent** utterance by the previous speaker |
| 1 | **related** to the utterance one **previous** to the most recent but *not* related to the most recent |
| 2 | related to utterance two previous to the last one (and not to anything more recent) |
| etc. | higher numbers for related to utterances further back. |
| , | **related** to previous material by the other speaker, but it is **unclear** to the coder whether they are related to the immediately previous utterance unit (which would be marked **0**), or to an utterance unit further back (which would be marked **1**, or **2**, etc.) |
| u | **unrelated** to previous speech by the old speaker |
| ? | **uncertain** whether these utterances relate to previous speech by the other speaker |
| u-e | **unrelated following** an **explicit** acknowledgment |
| 1-e | related to the penultimate utterance unit by the other speaker, when last utterance contained just an explicit acknowledgment |

Table 1: Relatedness Markings

# 4 Grounding

Grounding is the process by which participants in conversations add to their common ground or (assumed) mutual knowledge [Clark and Schaefer, 1989]. This section describes the coding scheme for grounding that is developed and used in [Traum and Allen, 1992, Traum and Hinkelman, 1992, Traum, 1994].

While largely ignored in most accounts before the mid-80's (either by making simplifying assumptions about things like the sufficiency of *copresence*, or through an aversion to mentalistic concepts like *mutual belief*), it is a fairly common assumption in many current accounts that some sort of positive feedback is required to assume common ground. While there is a variety of feedback (both positive and negative) in dialogue, we distinguish that feedback that has a specific effect on the immediate mutual understanding, while leaving those features which deal with agreement and negotiation, and clarification of issues that go beyond immediate understanding of what is expressed to other levels of analysis.

Unlike [Clark and Schaefer, 1989], who posit a recursive structure of *contributions* built up by *proposal* and *acceptance* phases (which may contain contributions as constituents), we use a 2-level flat structure to model the grounding process. Spans of text which are grounded together are clustered into sub-dialogues called *discourse units* (DUs). DUs are roughly comparable to the top-level contributions of Clark and Shaefer. Although they also consist of a span of related discourse, and can be coded in a similar way, DUs are different from the discourse segments proposed by Grosz and Sidner [Grosz and Sidner, 1986] and others. Discourse segments relate to the *intentional* structure of discourse — the purpose for uttering everything within a discourse segment must be subordinate to the discourse segment purpose. There is no such requirement for DUs, the analogous requirement is that everything in a DU could be said to be understood (or not understood) and acknowledged together. Thus, while discourse segments are units of intentional structure, DUs are units of (mutual) understanding structure.

## 4.1 Grounding Acts

DUs are constructed from individual utterance level *grounding acts*. There are grounding acts for initiating, continuing, repairing and acknowledging (thus successfully grounding), and cancelling (discontinuing when not grounded) a discourse unit. In addition, there are acts for requests for the other speaker to acknowledge or repair. Each utterance corresponds to one *grounding act* for each DU it is a part of. It is possible for an utterance to be part of more than one DU, contributing possibly different acts to each one. The main example of this is an utterance which acknowledges an open DU initiated by the previous speaker, while initiating a new one (in which the new content must be acknowledged by the other speaker). The grounding acts are summarized below and then individually described and exemplified.

| **Label** | **Description** |
|---:|---|
| initiate | begins a new DU, content separate from previous uncompleted DUs |
| Continue | continuation of previous material by the same speaker |
| Acknowledge | demonstrates or claims understanding of previous material by other speaker |
| Repair | Correct (potential) misunderstanding of DU content |
| Request Repair | Signal of lack of understanding |
| Request Ack | Signal for other to acknowledge |
| Cancel | Stop work on DU, leaving it ungrounded and ungroundable |

Table 2: Grounding Act Markings

### 4.1.1 Initiate

**Label:** init

**Description:** An initial utterance component of a Discourse unit. An `initiate` usually corresponds to the (first utterance in the) presentation phase of a top level Contribution in [Clark and Schaefer, 1989]. Any utterance which introduces new material which is not related to an *ungrounded* DU will be an `initiate`.

In the example below, utterance 32.1 is treated as both an `ack` of the previous DU, as well as an `init` of a new DU, with the content of the answer.

**Example:** d91-6.2

| **UU#** | **Speaker:Utterance** | **grounding act label** |
|---|---|---|
| 31.9 | M:It would get there at 3, | init |
| 31.10 | :is that what you're saying? | |
| 32.1 | S:It would get there at um, 4. | init |
| 33.1 | M:It would get there at 4. | |

**Example:** d91-3.2

| UU# | Speaker:Utterance | grounding act label |
|---|---|---|
| 25.1 | M:go to Corning | init |
| 26.1 | S:then on to Corning | |
| 27.1 | M:grab the tanker | init |
| 27.2 | :pick up oranges | |
| 27.3 | :go to Elmira | |
| 27.4 | :make em into orange juice | |
| 28.1 | S:okay | |

In the above example, although all of M's utterances are part of one large suggestion, 25.1 and 27.1 are `init`'s, since only they are unrelated to *ungrounded* DUs (since 25.1 is grounded by 26.1).

#### 4.1.2 Continue

**Label:** cont

**Description:** A continuation of a previous act performed by the same speaker. A `continue` is expressed in a separate utterance unit, but can be seen syntactically and conceptually as part of a whole with material in an ungrounded DU.

In the examples below, 27.2-4 are all `conts`. 27.1 is not, because the previous related material has already been grounded.

**Example:** d91-3.2

| UU# | Speaker:Utterance | grounding act label |
|---|---|---|
| 25.1 | M:go to Corning | |
| 26.1 | S:then on to Corning | |
| 27.1 | M:grab the tanker | |
| 27.2 | :pick up oranges | cont |
| 27.3 | :go to Elmira | cont |
| 27.4 | :make em into orange juice | cont |
| 28.1 | S:okay | |

#### 4.1.3 Acknowledgment

**Label:** Ack

**Description:** An acknowledgment claiming or demonstrating understanding of a previous utterance. It may be either a repetition or paraphrase of all or part of the utterance, an explicit signal of comprehension such as "ok" or "uh huh", or an implicit signaling 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 acknowledgment are answers to questions. Acknowledgments are also referred to by some as *confirmations* (e.g. [Cohen and Levesque, 1991]) or *acceptances* (e.g. [Clark and Schaefer, 1989]). We prefer the term *acknowledgment* as unambiguously signaling understanding.

There are three types of surface realizations for an acknowledgment act, all having the same effect (though perhaps with different strength) of grounding a DU. These are:

- Explicit – as in Section 3.1.1, a cue word that claims acknowledgment without demonstrating what was acknowledged

- Paraphrase or Repetition – demonstrating *what* was understood.
- Implicit – implicating understanding by doing what would be expected *if* the utterance were understood.

These three types are demonstrated in the examples below. In the second example, 32.1 counts as an implicit ack in virtue of answering a question. This utterance is also marked as an `init`, for the content of the answer itself.

**Example:** d91-1.1

| UU# | Speaker:Utterance | grounding act label |
|---|---|---|
| 4.1 | S:the oranges are in the warehouse at Corning | |
| 5.1 | M:oh okay | ack (explicit) |

**Example:** d91-6.2

| UU# | Speaker:Utterance | grounding act label |
|---|---|---|
| 31.9 | M:It would get there at 3, | |
| 31.10 | :is that what you're saying? | |
| 32.1 | S:It would get there at 4. | ack (implicit) |
| 33.1 | M:It would get there at 4. | ack (repetition) |

#### 4.1.4 Repair

**Label:** repair

**Description:** 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 (or might) change the hearer's interpretation of the speaker's intention. A `repair` can change either the content or core speech act type of material in the current DU (e.g. a tag question can change an interpretation 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 from [Litman and Allen, 1990]. `Repairs` are concerned merely with the grounding of content. Domain clarifications, which modify grounded content, are argumentation-level acts.

It is also important to realize that this `repair` category is a relationship between grounding acts. Speech repairs [Heeman and Allen, 1994a] that occur completely within a single utterance unit should *not* be marked as repair grounding acts, unless they also repair material presented in a previous utterance unit. In addition, simple matching repairs in which one utterance merely repeats some words from the end of a previous utterance (perhaps with an intervening disfluency) should be marked as a continue rather than a repair).

The `repair` category subsumes the various types of repair noted by [Scheglff *et al.*, 1977]. The first two examples show self-initiated self-repair, the former, a repair of a referring expression, and the latter the speech act type from a (potential) inform to a check. The third example shows other-initiated self-repair, where the repair request in 21.1 prompts the repair in 22.1. The fourth example shows other repair, in which t S produces a repair himself in 10.1. All of these examples are treated as repair grounding acts.

**Example:** d91-7.2

| UU# | Speaker:Utterance | grounding act label |
|---|---|---|
| 21.2 | M:fill the tanker car | |
| 21.3 | :with that | |
| 21.4 | :with th' orange juice | repair |

**Example:** d91-6.2

| UU# | Speaker:Utterance | grounding act label |
|---|---|---|
| 31.9 | M:It would get there at 3, | |
| 31.10 | :is that what you're saying? | repair |
| 32.1 | S:It would get there at 4. | |
| 33.1 | M:It would get there at 4. | |

**Example:** d91-8.1

| UU# | Speaker:Utterance | grounding act label |
|---|---|---|
| 20.2 | S:it doesn't take any time to | |
| 21.1 | M:to couple? | |
| 22.1 | S:to couple | repair |
| 23.1 | M:okay | |

**Example:** d91-6.1

| UU# | Speaker:Utterance | grounding act label |
|---|---|---|
| 9.1 | M:so we should | |
| 9.2 | :move the engine | |
| 9.3 | :at Avon | |
| 9.3 | :engine E | |
| 9.5 | :to | |
| 10.1 | S:engine E1 | repair |
| 11.1 | M:E1 | |

#### 4.1.5 Request Repair

**Label:** reqRepair

**Description:** A request for a repair by the other party. This is roughly equivalent to a *next turn repair initiator* [Scheglff *et al.*, 1977]. Often a `reqRepair` can be distinguished from a `repair` or `ack` only by intonation. A `reqRepair` invokes a discourse obligation on the listener [Traum and Allen, 1994] to respond with either the requested repair, or an explicit refusal or postponement (e.g. a followup request). Sometimes a suggestion of appropriate content is presented, as in the first example below, but in questioning rather than asserting form.

**Example:** d91-8.1

| UU# | Speaker:Utterance | grounding act label |
|---|---|---|
| 20.2 | S:it doesn't take any time to | |
| 21.1 | M:to couple? | reqRepair |
| 22.1 | S:to couple | |
| 23.1 | M:okay | |

**Example:** d93-15.2

| UU# | Speaker:Utterance | grounding act label |
|---|---|---|
| 46.4 | U:but I'll have to go to Bath | |
| 46.5 | :and I'm not gonna have much time left I guess | |
| 46.6 | :and I'll pick up the two boxcars | |
| 46.7 | :but after the other two are already in | |
| 46.8 | :I'll go to Corning then | |
| 46.9 | :and get the oranges | |
| 46.10 | :oranges | |
| 47.1 | S:can you say that again I'm sorry | reqRepair |

### 4.1.6 Request Acknowledgment

**Label:** ReqAck

**Description:** 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 acknowledgment, or an explicit refusal or postponement (e.g. a followup repair or repair request).

**Example:** d91-6.1

| UU# | Speaker:Utterance | grounding act label |
|---|---|---|
| 3.7 | M:there are oranges at Corning | |
| 3.8 | :right? | ReqAck |

**Example:** d91-6.2

| UU# | Speaker:Utterance | grounding act label |
|---|---|---|
| 50.7 | S:and you wanna have it pick up | |
| | :.. the boxcar full of oranges | |
| 50.8 | :and pick up the tanker | |
| 50.9 | :and go back to Elmira | |
| 50.10 | :is that right? | ReqAck |

**Example:** d93-16.2

| UU# | Speaker:Utterance | grounding act label |
|---|---|---|
| 48.2 | U:your orange juice is now | |
| 48.3 | :at . Coring | |
| 48.4 | :do you understand what I mean by that | ReqAck |
| 49.1 | S:um yes | |

### 4.1.7 Cancelation

**Label:** cancel

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

**Example:** d91-8.2

| UU# | Speaker:Utterance | grounding act label |
|---|---|---|
| 37.3 | M:send E2 | |
| 37.4 | :send / E2's already going there | cancel |
| 37.5 | :and as soon as it gets there | |
| 37.6 | :couple the cars and take it to uh | |

**Example:** d91-8.2

| UU# | Speaker:Utterance | grounding act label |
|---|---|---|
| 73.6 | M:how long will it take | |
| 73.7 | :to go to Bath | |
| 73.8 | :... pick up a boxcar come back | |
| 73.9 | :okay | cancel |
| 73.10 | :so | |
| 73.11 | :while / while the uh | |
| 73.12 | :[3sec] hmm [2sec] | |
| 73.13 | :yeah | |
| 73.14 | :so go / | |
| 73.15 | :go to Bath and pick up the boxcar | |

### 4.2 Grounding Coding

Minimal coding involves just identifying DUs as spans of dialogue and labeling utterances with grounding acts. A more detailed coding also attributes each grounding act to a DU. An even more detailed version specifically marks each continue, repair, request repair, and demonstration or implicit style acknowledgments with the utterance unit(s) that the act most directly relates to.

## 5 Surface Form

While this category may or may not be very interesting, it is certainly easy to code and will often be clues for determining the coding of other act-types (e.g., an interrogative sentence generally expresses a question, while requests are often expressed with an interrogative. The descriptions below are mostly taken from [Quirk *et al.*, 1985]

| | |
|---|---|
| Declarative | A sentence with subject present and preceding the verb. |
| Interrogative | A sentence with an operator (e.g, "do") or wh-element in initial position. |
| Imperative | A sentence with no grammatical subject. |
| Fragment | Only part of a sentence, an isolated word, phrase, or clause. |
| Cue word | discourse connectives appearing alone or as sentential connectors. This class includes words like the following: *ok, and, or, but, so, well* and *now*, when used in the cue word sense. |
| Non-linguistic | disfluencies, filled pauses, and other non-word sounds such as laughter. |

Table 3: Locutionary Act (Surface Form) Markings

## 6 Illocutionary Function

The illocutionary act types used in the TRAINS-93 system are shown below, with their meanings. One distinction from some speech act taxonomies is that we allow multiple interpretations for a single span of text. Thus a certain utterance might be *both* an `inform` *and* a `suggest`, or both a `YNQ` about ability to perform an action, and a `request` to perform the action.

| | |
|---|---|
| INFORM | The speaker provides new information (including providing requested information when answering a question). |
| YNQ | The speaker asks a yes-no question, trying to determine the polarity of a proposition. |
| CHECK | The speaker is attempting to verify that a certain (suspected) proposition is true. |
| WHQ | The speaker asks a wh-question, trying to determine the value of some term in a proposition. |
| SUGGEST | The speaker proposes a new item (action, proposition, plan constraint). |
| REQUEST | The speaker aims to get the hearer to perform some action. |
| ACCEPT | The speaker agrees to a prior proposal by the hearer. |
| REJECT | The speaker rejects a prior proposal by the hearer. |
| SUPP-INF | The speaker provides additional information (perhaps already known) that augments, or helps the hearer interpret some other accompanying speech act. This is usually performed by a subordinate clause or appositive phrase. |
| SUPP-SUG | The speaker makes a supplementary suggestion of content, which is presupposed to be part of the plan by other accompanying suggestion or request. This is often performed by a purpose clause. |
| EVAL | The speaker provides an evaluation of some item. This includes both *factual* evaluations,( e.g., of the likelyhood of success of an action), and *personal* evaluations, describing how the speaker feels about something. |
| GREET | The speaker seeks to establish connection, e.g., by saying "hello", or naming the hearer at the beginning of a dialogue. |
| APOLOGIZE | The speaker apologizes for some action (e.g., speaking out of turn) or mistaken interpretation. |

Table 4: Illocutionary (Core Speech Act) Markings

# 7 Argumentation Structure

This level, described in [Traum and Hinkelman, 1992, Traum, 1994], concerned the connections of multiple speech acts (produced by one or more agents) into coherent patterns of dialogue. It thus fulfills both the *rhetorical relations* described by [Mann and Thompson, 1987], *adjacency pairs* [Schegloff and Sacks, 1973] which are not specifically concerned with grounding, and the exchange and transaction levels from [Sinclair and Coulthard, 1975]. To date, there has not been an adequate attempt to specify or label these acts, but the table below shows the ones that were used in the TRAINS-93 implementation:

- `(:so` $act_1$ $act_2$`)`
  $act_1$ is relevant to the interpretation of $act_2$. If $act_2$ is an informational act, then the truth of its content should be partially supported by $act_1$. If $act_2$ is a suggestion, then the suggestion should be about (a part of) the plan dominated by $act_1$.

- `(:and` $act_1$ $act_2$`)`
  the interpretation of $act_2$ is connected to $act_1$ in some way to form a coherent whole. If $act_2$ is a suggestion, then it should be part of the same plan as $act_1$.

- `(:and-then` $act_1$ $act_2$`)`
  the interpretation of $act_2$ is connected to $act_1$ in some way to form a coherent whole. If $act_2$ is a

suggestion, then it should be part of the same plan as $act_1$. In addition, $act_2$ should temporally follow $act_1$.

- `(:purpose` *act form*`)`
  *act* is to be done for the purpose of achieving *form*.
- `(:background` $act_1$ $act_2$`)`
  $act_1$ is performed for the purpose of making $act_2$ more clear to the hearer.

# References

[Allen *et al.*, 1995] James. F. Allen, L. K. Schubert, G. Ferguson, P. Heeman, C. H. Hwang, T. Kato, M. Light, N. Martin, B. Miller, M. Poesio, and D. R. Traum. The TRAINS project: a case study in building a conversational planning agent. *Journal of Experimental and Theoretical Artificial Intelligence*, 7:7–48, 1995.

[Beckman and Hirschberg, 1994] Mary E. Beckman and Julia Hirschberg. The ToBI annotation conventions. Manuscript, Ohio State University, (obtain by writing to tobi@ling.ohio-state.edu), 1994.

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

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

[Cohen and Levesque, 1991] Phillip R. Cohen and Hector J. Levesque. Confirmations and joint action. In *Proceedings IJCAI-91*, pages 951–957, 1991.

[Gross *et al.*, 1993] Derek Gross, James Allen, and David Traum. The TRAINS 91 dialogues. TRAINS Technical Note 92-1, Department of Computer Science, University of Rochester, July 1993.

[Grosz and Sidner, 1986] Barbara J. Grosz and Candace L. Sidner. Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.

[Heeman and Allen, 1994a] Peter Heeman and James Allen. Detecting and correcting speech repairs. In *Proceedings of the $32^{th}$ Annual Meeting of the Association for Computational Linguistics*, pages 295–302, Las Cruces, New Mexico, June 1994.

[Heeman and Allen, 1994b] Peter A. Heeman and James Allen. The TRAINS 93 dialogues. TRAINS Technical Note 94-2, Department of Computer Science, University of Rochester, 1994.

[Litman and Allen, 1990] 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.

[Mann and Thompson, 1987] William C. Mann and Sandra A. Thompson. Rhetorical structure theory: A theory of text organization. Technical Report ISI/RS-87-190, USC, Information Sciences Institute, June 1987.

[Novick and Sutton, 1994] David Novick and Stephen Sutton. An empirical model of acknowledgement for spoken-language systems. In *Proceedings ACL-94*, pages 96–101, Las Cruces, New Mexico, June 1994.

[Orestrom, 1983] Bengt Orestrom. *Turn-Taking in English Conversation.* Lund Studies in English: Number 66. CWK Gleerup, 1983.

[Pierrehumbert, 1980] J. B. Pierrehumbert. The phonology and phonetics of english intonation. Doctoral dissertation, Massachusetts Institute of Technology, 1980.

[Quirk *et al.*, 1985] Randolph Quirk, Sifney Greenbaum, Geoffrey Leech, and Jan Svartvik. *A Comprehensive Grammar of the English Language.* Longman, London, 1985.

[Scheglff *et al.*, 1977] Emmanuel A. Scheglff, G. Jefferson, and H. Sacks. The preference for self correction in the organization of repair in conversation. *Language*, 53:361–382, 1977.

[Schegloff and Sacks, 1973] Emmanuel A. Schegloff and H. Sacks. Opening up closings. *Semiotica*, 7:289–327, 1973.

[Schober and Clark, 1989] Michael Schober and Herbert H. Clark. Understanding by addressees and overhearers. *Cognitive Psychology*, 21:211–232, 1989. Also appears as Chapter 6 in [Clark, 1992].

[Sinclair and Coulthard, 1975] J. M. Sinclair and R. M. Coulthard. *Towards an analysis of Discourse: The English used by teachers and pupils.* Oxford University Press, 1975.

[Sperber and Wilson, 1986] Dan Sperber and Deirdre Wilson. *Relevence: Communication and Cognition.* Harvard University Press, 1986.

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

[Traum and Allen, 1994] David R. Traum and James F. Allen. Discourse obligations in dialogue processing. In *Proceedings of the 32$^{th}$ Annual Meeting of the Association for Computational Linguistics*, pages 1–8, 1994.

[Traum and Heeman, 1996] David R. Traum and Peter Heeman. Utterance units and grounding in spoken dialogue. to be presented at *4th International Conference on Spoken Language Processing (ICSLP-96)*, October, 1996.

[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.

[Traum, 1994] David R. Traum. *A Computational Theory of Grounding in Natural Language Conversation.* PhD thesis, Department of Computer Science, University of Rochester, 1994. Also available as TR 545, Department of Computer Science, University of Rochester.