# Notes on Dialogue Structure

David Traum

May 8, 1998

**Abstract**

Here are some musings on different things in discourse structure of dialogue that should be considered in formulating a DRI coding scheme. The general focus is on viewing individual coding scheme efforts with an eye towards generalization and unification, where possible. This draft just presents the organizing principles, future drafts will actually apply this to extant schemes, with some speculations about which ones are actually in conflict or could be fruitfully combined.

## 1 Introduction

There have been many, often very different, approaches to discourse structure. The following document is an attempt to synthesize some characteristics that unify and distinguish these various approaches, in order to select or construct a general-purpose coding scheme for discourse structure in dialogue. This is not intended to be an exhaustive classification, merely a first attempt to set forth some of the dimensions along which schemes can be compared. In particular, by trying to classify what is being structured and how, it opens the possibilities of using various parts of different schemes together in one multi-component scheme for dialogue.

[Halliday, 1961], distinguishes two types of distinctions of linguistic structures: *levels* represent different kinds of structure (e.g., phonology vs. syntax), and *rank* is used to denote a hierarchical point within a level, such as between words and phrases within syntax. A basic question is whether there one or many levels of discourse structure? [Sinclair and Coulthard, 1975] say one level with a number of ranks for discourse. Others, e.g, [Grosz and Sidner, 1986, Lambert and Carberry, 1991, LuperFoy, 1991] talk about a tri-partite model (though with different levels), most of these latter models don't have different ranks within the levels. Others [Bunt, 1994, Allwood, 1995] have even more. I present several ways to classify discourse structure schemes, based on answers to the following questions:

- Granularity: how much stuff (time, text, turns, etc.) is covered by the units (minimum, maximum, and average)?
- Content: what is this a structure *of* (e.g., intentions, accessibility, effects, etc.)?
- What kinds of units and structuring principles are used (e.g., flat, set inclusion, hierarchical/CFG structuring, relational)?
- How many types of units are allowed (one basic unit type, 2 or three types of units, or several types)?
- How many different ranks or levels are given within the scheme as a whole?

# 2 Granularity

There seems to be roughly three ranges of granularity of discourse structures. The difficult thing is that most coding schemes seem to cover more than one (though rarely all three) ranges, and the boundaries are not quite clear. I name these three *micro*, *meso*, and *macro*, to avoid scheme-specific concepts.

## 2.1 Micro-Range Discourse Structure

This is roughly everything within a single speaker's turn, perhaps extending to a few things in multiple turns. On the low end, it's unclear where to draw the boundaries; many of the same phenomena at this range, e.g., rhetorical relations such as "purpose", "explanation" occur sub-sententially. Arguably intra-sentential syntactico-semantic relations also occur at this range. The main types of structure at this range are coherence relations linking multiple things, usually with a nucleus and satellite (head and adjunct according to syntax). Discourse relations like ellipsis and centering also probably occur at this range. Arguably, a few relations between contributions of different speakers, such as syntactic completions and second turn repairs also occur at this range, though most inter-speaker phenomena would fall at the next range.

## 2.2 Meso-Range Dialogue Structure

This range can roughly be glossed as a "subdialogue". Its minimal unit is an "exchange" which includes connected speech between two speakers. It can also extend for several turns on a coherent topic. This range is called "Game" by the HCRC folk. Discourse Units (Traum and Allen) and most Contributions (Clark And Schaefer) also occur at this range, as do adjacency pairs from conversational analysis, and IR pairs in the Linda scheme, and clarification subdialogues (Litman and Allen). A key question is whether the units at this range should be flat (all constituents are below the meso-range, such as grounding acts in a DU) or hierarchical (some constituents may also be meso-range, such as (I R) units in LINDA, or Clark and Schaefer's Contributions).

## 2.3 Macro-Range Dialogue Structure

This is a range related to phases of a dialogue, such as *Opening*, *Body*, and *Closing*, for Conversational Analysis, or *Transactions* in the HCRC scheme. Each phase might take several meso-range structures to complete. For the most part, macro-range units would be related to stereotypical dialogue types, e.g., in TRAINS, *get-goal*, *specify-plan*, *evaluate-plan*, or more elaborate ones for info-seeking, I think, for the most part, intentional discourse segments would also map to this range. Most of the intention at this range is related to the activity going on, whereas meso-range structure is mainly related to the dialogue participation itself - specific linguistic realizations, and coherence of dialogue acts. While a meso-range sub-dialogue might be related to the elicitation or expression of a single piece of information, the macro-range structure would be more related to how that piece of information fits in with the overall dialogue agenda.

# 3 Content

There are wide divergences in what a unit of structure is supposed to capture. Some rough categories are presented here.

## 3.1 Accessibility

One broad class of structuring principles is whether elements of the structure are accessible for one or more types of processing at other points in time. Such accessibility is meant to be calculated on the basis of the current structure, according to some rules or guidelines. There are several principles which can block once possible accessibility. These include:

- deletion - the item is no longer accessible, because it has been deleted from the structure (e.g., popping from a stack).
- blocking - the item remains in the structure, but in a part that is "hidden" by a more prominent part, (e.g., by a higher element on the stack).
- rule-based - processing is blocked by rules forbidding traversing a sub-structure in a particular configuration (e.g., C-command), Sub-boxes in DRT.
- hidden - the structure is not immediately accessible, but must be searched for, or brought back into a cache, and so might not be used in processing that requires immediate access.

There are also several different kinds of structure, depending on what things are supposed to be accessible, and what for. Some of these, include:

### 3.1.1 Memory Structure

Here, the main organizing principle is deleting (i.e., forgetting), though also some distinction between short-term (focussed) memory, and longer-term memory that is not immediately accessible, but can be located and brought in. There has not been much work on memory structures, probably because it seems more a phenomenon of individual agents, rather than the dialogue itself, and different agents are likely to follow their own strategies. The most explicit work on memory in dialogue is Lyn Walker's thesis work [Walker, 1993], using an AWM model of memory, which had some properties about cost of retrieval. She also suggested using a cache [Walker, 1996]. Luperfoy [LuperFoy, 1991] also relied on a kind of memory decay in the *linguistic* structure to disallow accessibility of the actual words, etc. beyond a certain point.

#### 3.1.2 Attention/Reference

Which items are available to be referred to using:

- pronouns
- deictics
- definite reference
- ellipsis

Some structure of attention is usually posited to control these phenomena. Attention can be linguistic - related to discourse structure, or also to some aspects of the non-linguistic context - visual, intensional, or background. Grosz and Sidner suggested using stacks for attentional structure, though some have been dissatisfied with some of the restrictions this causes. Are there compelling examples that can't be handled with stacks, or intermediate restrictions that better capture accessibility?

### 3.2 Intention/Purpose

Another way of characterizing dialogue structure is by what the speakers are trying to accomplish in the conversation. Thus one can link bits of content that strive to achieve the same purpose, and segment these bits by the different sub-tasks they are working at, as is done by [Nakatani *et al.*, 1995]. One could also look at different types of intentions, e.g., whether they operate at the 'domain' level, problem solving level, or discourse level [Lambert and Carberry, 1991]. Another complexity in dialogue is that there is more than one speaker. Thus one can talk about intention of each individual speaker, as well as intentions of the other conversational participants, and also joint intentions. Another questions is that, since there may be many intentions behind a particular utterance, how many of these should be coded? Also, should there be any particular structural mapping (e.g., strict nesting), or should overlapping purposes be marked? Should one only mark speaker's intention (or perhaps joint intention), or all the utterances, even by another speaker, that contribute to satisfying an intention?

## 3.3 Turn-taking/Initiative

This kind of structure is with respect to who has *control* over that part of the dialogue. It is similar to intentional structure, to some degree, but here the idea is to label which participant had control rather than what the purpose that participant might have. *Turn* is roughly who is speaking, while *initiative* is roughly who controls the topic of the speech.

*Turn* is used with at least two different meanings. As used by some (e.g., [Sacks *et al.*, 1974] and the various efforts in the TRAINS project [Gross *et al.*, 1993, Traum and Hinkelman, 1992, Heeman and Allen, 1995, Traum and Heeman, 1997], the structure is fairly easy to determine, just by observing who is speaking. Others, however, (e.g., [Orestrom, 1983, Allwood, 1995], use the term rather to describe who has the *right* to speak, and don't count backchannels or unsuccessful interruptions as changing the turn. [Sacks *et al.*, 1974] use the term *floor* to refer to the rights that are shared out.

It's still not completely universally agreed as to what *Initiative* means (see several definitions in the recent AAAI Spring symposium and upcoming UMAI special issue). E.g., Chu-Carroll [Chu-Carroll and Brown, 1997] talks of discourse vs. domain initiative, while Novick et. al [Novick and Sutton, 1997] talk of three kinds of initiative, including turn-taking as one of these. Others [Smith and Hipp, 1994] talk of varying levels of initiative, often related to dialogue moves or syntactic types of utterances used to realize them [Walker and Whittaker, 1990]. Still, at a broadest level, we could classify who holds the initiative for a given segment, assuming a mixed-initiative dialogue, where it actually changes throughout the dialogue.

## 3.4 Grounding/Mutual Understanding

Another type of structure is the way that information is shared in dialogue, leading to mutuality. Clark and Schaefer call this process of adding to the common ground, *grounding*, and propose a hierarchical model of *Contributions*, which contain *presentation* and *acceptance phases*, each of which could also contain embedded contributions [Clark and Schaefer, 1989]. Traum [Traum and Allen, 1992, Traum, 1994] proposed a model in terms of *Discourse Units*, which were finite state sequences of grounding acts, with a bounded stack to relate the accessible units.

[Allwood *et al.*, 1992] point out that the same kind of mutuality can apply at other levels than perception and understanding. They also include levels of contact and agreement. As well as the grounding structures, one might also want to track the negotiation process, leading to shared agreement [Baker, 1994, Sidner, 1994].

### 3.5 Information

Another way of structuring dialogue is according to the information expressed or described. First, (as a principle for whether content forms a coherent unit) are multiple sub-units talking about the same general thing, and next, what is the relationship between them? Subject-matter relations [Mann and Thompson, 1987] also describe the information structure [Moore and Pollack, 1992].

## 4 Types of Dialogue Structure

As well as there being many ways in which bits of dialogue can be coherent, there are also several different way to label/conceptualize that kind of coherence. I divide this consideration into two related sub issues:

1. how is the structure organized from substructures

2. what is the nature of the sub-units

### 4.1 Type of Structure

- span of text
  the simplest kind of structure is just a region of contiguous text or time. Here all that is necessary is to set the boundary points.

- set/sequence of subunits (not necessarily contiguous)
  here one must also divide the span into sub-units, and one builds up the structure by concatenating sub-units. One may also allow non-contiguous subunits (i.e., a medial gap which is not part of a segment). A further complication is whether to allow sub-units to overlap in span, or have multiple defining notions (e.g., purpose).

- stack
  This is used mainly for accessibility structures, to organize a recency preference, while maintaining relatively simple computational properties, and allow sub-goals.

- cache/queue
  this has higher complexity, but may be necessary or more faithful to human processing for some phenomena.

- unordered relation
  here a structure is built not by just including a sequence of sub-units, but defined by a particular relationship holding between subunits. Thus there is some semantic claim about this relationship of units.

- ordered (e.g., nucleus/satellite) relation
  this is even more strict, since one must point out the particular roles that the sub-units play to the relation/larger unit as a whole. There are attempts made to generalize these relations so that one can point out types of roles, e.g., head/specifier/adjunct in syntax, and nucleus/satellite, for rhetorical relations. Here, one might run into problems if one would like to designate a single unit by multiple relations, where the different relations might have different "heads", as shown by [Moore and Pollack, 1992].

- structured sequence (i.e., grammar)
  Dialogue grammars (e.g., [Reichman, 1985]) indicate possible next dialogue moves/acts, given a current state. They had been declining in popularity in favor of plan/intentional models, because of the wide range of possible move sequences, but have found more popularity now that people are trying to build (spoken) dialogue systems.

- other/haphazard (e.g., AWM)
  Some kinds of structure are hard to categorize, e.g., the AWM structure used for memory in [Walker, 1993], in which successive bits of information was put one (random) step away in a three-dimensional lattice structure.

## 4.2 Cardinality of Units

Another issue for units is whether there is only a single type of unit, in which case one would only mark the presence (or absence) and extent of the unit, or whether there are multiple types and one must mark the type as well. If there are multiple types, then one must distinguish whether the types belong to some fixed set, from which a tagger must choose, or whether the space is potentially unbounded (e.g., topics in [Nakatani *et al.*, 1995]).

## 4.3 Nature of the sub-Units

A related question is what type of sub-units build up the unit. For the *span*, this question is not important, as one does not actually need sub-units. For the other types, though, there are several possibilities:

- hierarchical
  here, the subunits are (can be) the same as the unit itself. Usually this is associated with a stack-like structure for managing the hierarchical structure.

- compositional
  the subunits are (possibly) other kinds of units, but only a fixed "grammatical" structure of sub-units will constitute a unit. E.g., Clark and Schaefer's Contributions are composed of two phases, *presentation* and *acceptance*, with the first preceding the second, and definitions for the purposes of each phase.

- multiple sub-units
  allowing a unit to be composed of some arbitrary number and ordering of subunits.

# 5 Coding Schemes

We consider the following analyses of discourse structure and coding schemes, in light of the above organizing principles.

## 5.1 Birmingham

[Sinclair and Coulthard, 1975, Coulthard *et al.*, 1981]

## 5.2 Conversation Analysis

There isn't a specific coding scheme, as far as I'm aware, but several organizing principles, as described in places like, [Schegloff and Sacks, 1973, Sacks *et al.*, 1974, Scheglff *et al.*, 1977], and summarized in [Levinson, 1983]

## 5.3 HCRC

[Carletta *et al.*, 1997]

## 5.4 Grosz and Sidner

### 5.4.1 Nakatani et. al

### 5.4.2 Flammia

## 5.5 Rhetorical Structure Theory

### 5.5.1 Pitt

### 5.5.2 Marcu

## 5.6 Clark

## 5.7 Rochester

## 5.8 Linköping

# References

[Allwood *et al.*, 1992] Jens Allwood, Joakim Nivre, and Elisabrth Ahlsen.

On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9, 1992.

[Allwood, 1995] Jens Allwood. An activity based approach to pragmatics. Technical Report (GPTL) 75, Gothenburg Papers in Theoretical Linguistics, University of Göteborg, 1995.

[Baker, 1994] Michael Baker. A model for negotiation in teaching-learning dialogues. *Journal of Artificial Intelligence in Education*, 5(2):199–254, 1994.

[Bunt, 1994] Harry Bunt. Interaction management functions and context representation requirements. In *Proceedings of the Twente Workshop on Langauge Technology: Dialogue Management in Natural Language Systems (TWLT 11)*, pages 187–198, 1994.

[Carletta *et al.*, 1997] Jean Carletta, Amy Isard, Stephen Isard, Jacqueline C. Kowtko, Gwynewth Doherty-Sneddon, and Anne H. Anderson. The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1):13–31, 1997.

[Chu-Carroll and Brown, 1997] Jennifer Chu-Carroll and Michael K. Brown. Tracking initiative in collaborative dialogeu interactions. In Hans Uszkoreit, editor, *Proceedings of the Thirty-Fifth Meeting of the Association for Computational Linguistics*, pages 262–270. Association for Computational Linguistics, 1997.

[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.

[Coulthard *et al.*, 1981] M. Coulthard, M. Montgomery, and D. Brazil. Developing a description of spoken discourse. In M. Coulthard and M. Montgomery, editors, *Studies in Discourse Analysis*, pages 1–50. Routledge & Kegan Paul, 1981.

[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.

[Halliday, 1961] M. A. K. Halliday. Categories of the theory of grammar. *Word*, 17:241–92, 1961.

[Heeman and Allen, 1995] Peter A. Heeman and James F. Allen. The Trains spoken dialog corpus. CD-ROM, Linguistics Data Consortium, April 1995.

[Lambert and Carberry, 1991] Lynn Lambert and Sandra Carberry. A tri-parite plan-based model of discourse. In *Proceedings of the 29$^{th}$ Annual Meeting of the Association for Computational Linguistics*, pages 47–544, 1991.

[Levinson, 1983] Stephen C. Levinson. *Pragmatics*. Cambridge University Press, 1983.

[LuperFoy, 1991] Susann LuperFoy. *Discourse Pegs: A computational Analysis of Content-Dependent Referring Expressions*. PhD thesis, University of Texas at Austin, 1991.

[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.

[Moore and Pollack, 1992] Johanna D. Moore and Martha E. Pollack. A problem for RST: The need for multi-level discourse analysis. *Computational Linguistics*, 18(4):537–544, 1992.

[Nakatani *et al.*, 1995] Christine H. Nakatani, J. Grosz Barbara, David D. Ahn, and Julia Hirschberg. Instructions for annotating discourses. Technical Report 21-95, Harvard University, 1995.

[Novick and Sutton, 1997] David Novick and Stephen Sutton. What is mixed-initiative interaction. In *Working Notes AAAI Spring Symposium* *on Computational Models for Mixed Initiative Interaction*, pages 114–116, march 1997.

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

[Reichman, 1985] Rachel Reichman. *Getting Computers to Talk like You and ME.* MIT Press, Cambridge, MA, 1985.

[Sacks *et al.*, 1974] H. Sacks, E. A. Schegloff, and G. Jefferson. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50:696–735, 1974.

[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.

[Sidner, 1994] Candace L. Sidner. An artificial discourse language for collaborative negotiation. In *Proceedings of the forteenth National Conference of the American Association for Artificial Intelligence (AAAI-94)*, pages 814–819, 1994.

[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.

[Smith and Hipp, 1994] Ronnie W. Smith and D. Richard Hipp. *Spoken Natural Language Dialog Systems: A Practical Approach.* Oxford University Press, 1994.

[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 Heeman, 1997] David R. Traum and Peter Heeman. Utterance units in spoken dialogue. In Elisabeth Maier, Marion Mast, and Susann

Luperfoy, editors, *Dialogue Processing in Spoken Language Systems — ECAI-96 Workshop*, Lecture Notes in Artificial Intelligence, pages 125–140. Springer-Verlag, Heidelberg, 1997.

[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.

[Walker and Whittaker, 1990] Marilyn A. Walker and Steve Whittaker. Mixed initiative in dialogue: An investigation into discourse segmentation. In *Proceedings ACL-90*, pages 70–78, 1990.

[Walker, 1993] Marilyn A. Walker. *Informational Redundancy and Resource Bounds in Dialogue.* PhD thesis, University of Pennsylvania, 1993.

[Walker, 1996] Marilyn A. Walker. Limited attention and discourse structure. *Computational Linguistics*, 22(2):255–264, 1996.