Levels/Structure of Dialogue Notes

In general:
Dialogue researchers posit a multi-leveled analysis of the structure of dialogue. These levels relate signals towards some discourse purpose or task. The number of levels may be few or many… but commonly center on two or three levels. Active research areas still center on the question of a fundamental unit of dialogue and how such a level relates to an intentional analysis. The Discourse Resource Initiative (DRI) use the common ground unit (CGU) as the basic level for intentional analysis but concedes they aren't sure how these levels align given that grounding and intentional task structure are somewhat independent (Core et al., 1998). In part, this accounts for why there may be some lumping and splitting in regards to task, dialogue and domain-related acts among various models of dialogue. Middle-level posited to be the basic level of analysis for dialogue (Pilkington, 1999; Mushin et al., 1999; Nakatani and Traum, 1999). Reflects some notion of joint action and mutual understanding.

There is at least two inter-twining thread: joint action involves the coordination of both content (what people intend to express) and processes (how people coordinate information exchange).

1. *Packaging*. How much information is presented at one time? There is a trade-off associated with efficiency of communication. You can present a large amount of information all at once and run the risk of misunderstanding of some part of this information. Or you could present tiny chunks of information and be sure that there is little or no misunderstanding but take a long time communicating.
2. *Coordination*. In order to account for the process of coordination in joint action, Clark (1996) distinguishes between two communicative tracks. Track one is concerned with the presentation of information – the communicative acts. Track two is concerned with the communication itself –the metacommunicative acts. Information in track one is the primary presentation. Information in track two may be more backgrounded. For example, participants may be presenting positive evidence for understanding by nodding the head. Any utterance may be communicative on one or both tracks. Because, conversation is a joint activity, C&S posit a basic structure in dialogue that has two phases: one concerned with *presentation* of information and the second with an *acceptance* of that information. Through these phases, participants each closure for each signal.

Level	**Type (Researcher)**	**Comments**
	Utterance	Has been considered a basic unit of analysis for spoken language analysis. Nakatani and Traum consider it a *minimal* unit of analysis. "Act of speech… Spans of speech terminated by prosodic cues: boundary tones and pauses." (Traum and Heeman, 1997).
Micro-	*Move* (Carletta 1997) *Act* (Nakatani & Traum, 1999)	"People add to discourses through communicative acts or signals" (Clark, 1994). Generally, considered co-inciding with an utterance or turn unit for simplicity. Associated with discourse phenonemena (e.g., ellipsis, referent resolution) and coherence relations (e.g., sentential rhetorical relations). Also, centering, self-repair, etc.
	Turn (Mushin et al. 1999)	A turn is generally associated with who is speaking. May also be associated with who has the right to speak. (Allwood, 1995).

		A turn may be composed of multiple moves or acts. The amount of information in a turn depends on how speakers chunk presentations. It becomes more difficult to distinguish turns when you factor in collateral signals such as backchanneling.
Meso- (sub-dialogue)	*Conversational games* (Carletta et al. 1997; composed of moves) *Dialogue games* (Carlson 1983; Power 1979). *Interactions* (Houghton 1986). *Exchanges* (Sinclair & Coulthard, 1975; composed of moves and acts). *Adjacency pairs* (Conversational analysis). *Clarification subdialogues* (Litman and Allen, 1987). *Discourse units* (Traum and Allen, 1992). *Common ground units* (Nakatani and Traum 1999). *Discourse segments* (Grosz and Sidner 1986). *Contribution* (Clark and Schaeffer 1989).	Can span several turns. DRI 1998 Report asks whether this unit should be flat or (all constituents are below this level) or whether it can be hierarchical (e.g., contributions of C&S). How about compositional where there is a fixed type of structure as described in Pilkington, 1999? Corresponds to: illocutionary acts (accord. to Traum in Core et al. 1998; Pilkington 1999); discourse purpose (Carletta 1997); patterns of turn-taking or initiative (Mushin et al. 1999). This level is concerned with both content and meta-communicative acts.
Macro- (Intentional / Informational)	*Transaction* (Carletta et al. 1997). *Episode* (Pilkington 1999). *Phases of dialogue* (Conversational analysis). *Intentional discourse segments* (Grosz and Sidner 1986).	Topic or common focus space (Pilkington 1999). How information fits with overall dialogue purpose.

Cahn, Janet 1992 (and Cahn and Brennan, 1999)
The contribution model illustrates collaborative processes in both domain and repair interactions, and so, can be expected to provide the following benefits in human computer interaction[Bre91b]:

- An ongoing and current account of discourse state and progress;
- Support for local repair strategies in the context of larger domain-driven strategies;
- A collaborative rationale for maintaining context over more than one adjacency pair, and therefore, a basis for discourse segmentation;
- A representation that can be applied to human computer interaction in non{linguistic modes.

The contribution model provides a collaborative rationale for maintaining context over more than one adjacency pair. The minimal collaborative requirement is that both conversants understand what each other has said in one exchange. Context, then, is maintained at least long enough to model a presentation, all its repairs and the partner's response. A more collaborative rationale would also model the contribution hierarchies as they re reflect hierarchies of domain and conversational tasks. Such extensions would begin to address the discourse segmentation required by the focus space model of discourse[GS86]. However, as this seems to require a theory of knowledge organization, it is outside the scope of the contribution model and (currently) its computational adaptation.

Criticisms of the Contribution Model.
Although the contribution model is a process model, its explication is illustrated only with completed structures[CS87, CS89]. Thus, the main challenge is that a step by step description of building the model is missing and must be invented. The main task, then, is to design the step by step processes and the structures

that would support them. All the modifications to the original contribution model arise from the needs of this task. … For example, task structuring is missing and, in addition, the operation of the grounding criteria has no formal representation.
As the basis for computation in a user interface, the contribution model presents specific difficulties:

- Underspecification about whose view is represented in the model and when;
- The inclusion in one utterance of more than one presentation;
- The inclusion in one presentation of many utterances, presented in separate turns;
- The indeterminacy of the acceptance phase structure;
- The lack of explicit task structuring;
- The lack of a formal proposal for the grounding criteria.

For computational purposes, the contribution model is underspecified not only in form but in content. Questions remain about the proper treatment of: utterances in which more than one presentation is presented; the installment presentations, in which one presentation is completed over several speaking turns; ambiguities in the structural representation of the acceptance phase. The contribution model also lacks an explicit representation of the common purpose that binds the contributions of two conversants; and formal mechanism with which to represent the grounding criteria (and the grounding threshold) in the model. [Cahn's adaption to the model, for current purposes, does not deal with] utterances that contain more than one presentation, or conversely, presentations that require more than one utterance (turn) to complete. Instead, they require only one presentation per utterance. The resolution of other problems includes: the use of only unambiguous structures to model acceptance phase processes and the addition of a task structure that binds the contributions of each conversant.

Reasoning for Exchange structure.
The exchange replaces the contribution as the paradigmatic structure of the model. In addition, a new structure, the gist, is proposed to model the use of the exchange after its completion. The gist summarizes the collaborative work accomplished in an exchange. Together, the exchange and gist provide the foundations of an architecture that links collaborative conversational work to the establishment of mutual belief in common ground.

The exchange model incorporates a revised definition of the acceptance phase, such that only one structural configuration (not two) can signal its completion. Thus, The changes to the contribution model are, chiefly:

- Exclusion of ambiguous forms;
- Inclusion of the exchange to represent explicitly the task link between two contributions;
- The imposition of a more deterministic form on the acceptance phase.

The **exchange, a task structure**, plays a role in redefining the acceptance phase. It differs from the contribution model in the following ways:

- The highest level structure within a dialog is an exchange instead of a contribution. (For now the dialog is trivially defined as the root node.)
- The task motivation for an interaction is explicitly represented by the exchange -- two contributions, one per conversational partner.
- The acceptance is formulated as a sequence of two distinct parts: 1) the exchange, representing the interim work toward mutual understanding and 2) the final utterance that enable an agent to conclude that mutual understanding has been achieved. The interim work towards grounding, represented by one or more exchanges, can only end felicitously with the presentation of the ratification utterance, the positive evidence that uniquely signals grounding.

The exchange is introduced to model explicitly the requirements that a dialog consist of at least two contributions and that each contribution become grounded. Two contributions are linked at the root to represent the requirements of the task, while connections to the leaf nodes continue to represent the progression of understanding. The conceptual justification for the exchange is that, by linking two contributions at the root, it represents explicitly a shared conversational or domain task. The task supported by an exchange corresponds to the Discourse Segment Purpose in the focus space model[GS86]. *That is, **two contributions are linked structurally because each contributes to the satisfaction of a common purpose***.

The purpose or task motivation applies to exchanges at every level of discourse structure**. Thus, anchored below a dialog, an exchange models a task in the domain. Anchored below an acceptance, an exchange models a repair or clarification.** The model is the system's current hypothesis about two states - the state of the task and the state of mutual understanding. The state of the task describes the relevance of the current contribution to the current domain or conversational task. It is represented by links at and above the root node of the current contribution. The state of mutual understanding describes only whether one contribution appears to be evidence for grounding a prior contribution. It is represented by links to the utterance at the leaf node.

The contribution model explains why partners seek evidence about whether their presentation is understood, and how - they apply their grounding criteria to determine the state of mutual understanding. Introduction of task structuring adds another metric by which understanding is measured - the state of the task. Conversational work on either phase of a task ends only when both partners mutually understand that their collaborative conversational work has successfully defined or executed the current task. Conversational evidence is evaluated against the grounding criteria of each speaker. These may vary according to the discourse context, speaker preferences and the domain task. They are not formalized for either the contribution or exchange models. However, the exchange, a task representation, is the structural foundation for applying task based criteria, and therefore, a step toward a formal representation of the grounding criteria.

Carletta, Jean et al., 1997

The coding distinguishes three levels of dialogue structure, similar to the three middle levels in Sinclair and Coulthard's analysis of classroom discourse. At the highest level, dialogues are divided into transactions, which are subdialogues that accomplish one major step in the participants' plan for achieving the task. The size and shape of transactions is largely dependent on the task. In the Map Task, two participants have slightly different versions of a simple map with approximately fifteen landmarks on it. One participant's map has a route printed on it; the task is for the other participant to duplicate the route. A typical transaction is a subdialogue which gets the route follower to draw one route segment on the map.

Transactions are made up of conversational games, which are often also called dialogue games [Carlson1983,Power1979], interactions [Houghton1986], or exchanges [Sinclair and Coulthard1975], and show the same structure as Grosz and Sidner's discourse segments Grosz&Sidner-CL when applied to task-oriented dialogue. All forms of conversational games embody the observation that, by and large, questions are followed by answers, statements by acceptance or denial, and so on. Game analysis makes use of this regularity to differentiate between initiations which set up a discourse expectation about what will follow, and responses which fulfill those expectations. In addition, games are often differentiated by the kind of discourse purpose which they have -- for example, getting information from the partner or providing information. A conversational game is a set of utterances starting with an initiation and encompassing all utterances up until the purpose of the game has been either fulfilled (e.g., the requested information has been transferred) or abandoned. Games can nest within each other if one game is initiated to serve the larger goal of a game which has already been initiated (for instance, if a question is on the floor but the hearer needs to ask for clarification before answering). Games are themselves made up of conversational moves, which are simply different kinds of initiations and responses classified according to their purposes.

Mushin et al., 1999

Traum (1998) and Nakatani & Traum (1999) have recently proposed taking grounding as the basic principle behind the structuring of dialogue at levels higher than the dialogue act. Minimal units of acknowledged common ground have been considered as the building blocks of higher level dialogue structures based on intentional or informational content (eg. ‘Common Ground Units’, or ‘CGUs’ Nakatani & Traum (1999)). CGUs, which represent grounding at the ‘illocutionary level’ (Clark 1996), have been proposed as a meso- level dialogue structure - roughly the same level that dialogue games (Carletta et al, 1997) or adjacency pairs (eg. Sinclair & Coulthard 1975) occupy in their dialogue structure frameworks. The appeal of taking units based on grounding as the level of dialogue structure above the microlevel of ‘act’ (as argued in Nakatani & Traum 1999) lies in its prioritization of mutual understanding as a central component of dialogue, regardless of the type of initiation and response. In the ‘CGU’ framework, some responses themselves get grounded so that the result is a complex configuration of overlapping and

embedded units of information entered into the common ground of the participants. This approach thus acknowledges importance of the contributions by both participants in the grounding process. It highlights the 'joint action' aspect of dialogic communication.

Pilkington, 1999

In Sinclair and Coulthard's (op cit.), *transactional analysis*, the *exchange* is viewed as the smallest unit which can stand alone and still make sense. For example, an exchange might consist of statement and counter-statement, question and answer or offer and acceptance. The notion of achieving some goal or completing some business through dialogue lies behind the term "transactional". In DISCOUNT we extend the work of Sinclair and Coulthard in a similar way to that suggested by Stubbs (1983). Sinclair and Coulthard (1975) describe educational classroom exchange structure as consisting of [I R <F>] where I is the initiating turn that opens a new exchange and predicts a responding turn, which may be followed by an (optional) feedback turn. In DISCOUNT the notation F for feedback is replaced with Rc for Response-Complement to avoid confusion with the move level term {feedback}. In DISCOUNT moves are a category of description at the level of Communicative Act and below the level of the turn. A single turn, or even a single word, may consist of none, one or more such moves. In contrast mark-up at the Exchange Structure level is aimed at describing patterns of turn taking or *initiative* as it is shared between participants according to their roles. Feedback suggests the evaluation of one participant's contribution by another.

Limitations of marking up *Exchange Structure* alone include that it does not describe these *outcomes* and *functions* of the transaction in ways which allow us to evaluate the success of the exchange relative to such goals (the intentions or purposes) of the participants. Furthermore, the detailed *rhetorical strategies* and *tactics* - the selection of particular moves - are not recorded, nor is it tracking of *topic-focus* or *issue spaces.* The tracking of topic-focus is necessary in order to determine the successful completion of transactions (outcomes) and/or the likelihood of participants needing to return to elaborate a topic further.

An Episode consists of one or more (at least one but optionally more) exchange(s) on a developing theme or topic (a common focus-space):

(EXCHANGE <EXCHANGE>)

An Exchange consists of a minimum of one Initiating (I) turn and one Responding (R) turn but may also include optional Reinitiating turns (RI), Response-Complement (Rc) and Stand-Alone (SA) sequences. Where anything in < > brackets is optional and can occur as many times as wanted, the minimal structure (excluding possible embedding) is:

[I <RI Rn Rcn> <SA> R <Rc>]

A topic can develop over a series of exchanges to form a line of argument or multiple parallel threads of argument. The links between moves sharing this topic will be marked by the use of co-ordinating conjunctions - such as *and, or, but, not, "as well"* - or similar devices which maintain reference such as pronoun substitution. A new topic, on the other hand is typically accompanied by a longer pause between turns and the use of more definite description to introduce it. The tracking of focus within the current topic, is thus, based upon the notion of coherence, theme and rheme and the work of Halliday (1967). Although this will often suffice to determine episode boundaries there remains a strong element of judgement in determining a true topic shift from a developing sub-topic or daughter. An episode is, thus, defined as a topic focus-space within which the topic is developing in parent-daughter relationship.

Within a turn we can identify *moves*, tactics that serve a dialogue goal. The identification of moves proceeds, in part, bottom-up from an identification of rhetorical predicates.

References

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

Allen, J.F. and D.J. Litman. 1987. A Plan Recognition Model for Subdialogues in Conversations. Cognitive Science 11, 2, 163-200.

Cahn, Janet. 1992. A Computational Architecture for the Progression of Mutual Understanding in Dialog. Technical Report 92-4, Music & Cognition Group, Media Laboratory. Massachusetts Institute of Technology.

Cahn, Janet and Susan E. Brennan. 1999. A Psychological Model of Grounding and Repair in Dialog. Proceedings of the Fall 1999 AAAI Symposium on Psychological Models of Communication in Collaborative Systems. Sea Cliff, Massachusetts. November 5-7, Pages 25-33.

Carletta, Jean, Amy Isard, Stephen Isard, Jacqueline C. Kowtko, Gwyneth Doherty-Sneddon, and Anne H. Anderson. 1997. The Reliability of a Dialogue Structure Coding Scheme, Computational Linguistics 23(1):13-31.

Carlson, Lauri. 1983. Dialogue Games: An Approach to Discourse Analysis. D. Reidel.

Clark, Herbert H. 1996. Using Language. Cambridge University Press.

Core, Mark, Masato Ishizaki, Johanna Moore, Christine Nakatani, Nobert Reithinger, David Traum, and Syun Tutiya (ed.). 1999. The Report of The Third Workshop of the Discourse Resource Initiative, Chiba Univeristy and Kazusa Academia Hall, Technical Report No.3 CC-TR-99-1, Chiba Corpus Project.

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

Houghton, George. 1986. The Production of Language in Dialogue: A Computational Model. Ph.D. thesis, University of Sussex, April.

Mushin, Ilana, Lesley Stirling, Janet Fletcher and Roger Wales. 2000. Identifying Prosodic Indicators of Dialogue Structure: Some Methodological and Theoretical Considerations. 1st SIGDial Workshop on Discourse and Dialogue. October 7-8, 2000. Hong Kong. In conjunction with ACL-2000: The 38th Annual Meeting of the Association for Computational Linguistics.

Nakatani, Christine and David Traum, Coding discourse structure in dialogue (version 1.0). Technical Report UMIACS-TR-99-03, University of Maryland.

Pilkington. Rachel. 1999. Analysing Educational Discourse: The DISCOUNT Scheme. Technical Report No. 99/2.

Power, Richard J. D. 1979. The organisation of purposeful dialogues. Linguistics, 17:107-152.

Sinclair, John M. and R. Malcolm Coulthard. 1975. Towards an Analysis of Discourse: The English used by teachers and pupils. Oxford University Press.

Stirling, Leslie, Ilana Mushin, Janet Fletcher, Roger Wales. 2000. The Nature of Common Ground Units: an Empirical Analysis Using Map Task Dialogues. Gothenburg Papers in Computational Linguistics 00-5.

Traum, David 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, David and Peter Heeman. Utterance Units in Spoken Dialogue. In Processing in Spoken Language Systems, Ed. Elisabeth Maier, Marion Mast, and Susann LuperFoy, Lecture Notes in Artificial Intelligence, Springer-Verlag, Heidelberg, pages 125--140, 1997.