Bridging the Gap
Between Dialogue Management and Dialogue Models

Weiqun Xu and **Bo Xu** and **Taiyi Huang** and **Hairong Xia**
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
Beijing, 100080, P. R. China
{wqxu, xubo, huang, hrxia}@nlpr.ia.ac.cn

Abstract

Why do few working spoken dialogue systems make use of dialogue models in their dialogue management? We find out the causes and propose a generic dialogue model. It promises to bridge the gap between practical dialogue management and (pattern-based) dialogue model through integrating interaction patterns with the underling tasks and modeling interaction patterns via utterance groups using a high level construct different from dialogue act.

1 Introduction

Due to the rapid progress of speech and language processing technologies (Cole et al., 1998; Juang and Furui, 2000), ever-increasing computing power, and vast quantity of social requirements, spoken dialogue systems (SDSs), which promise to provide natural and ubiquitous access to online information and service, have become the focus of many research groups (both academic and industrial) with many projects sponsored by EU, US (D)ARPA and others in the past few years (Zue and Glass, 2000; McTear, 2002; Xu, 2001). The last decade saw the emergence of a great deal of SDSs.

Despite so much progress, some problems still remain, prominent among which are usability and reusability (or portability across domains and languages). Through a survey of typical working spoken (or natural language) dialogue systems in the nineties (Xu, 2001), we find their central control-

ling component – dialogue management – is relatively less well-established than other components. In most working SDSs, the design of dialogue management is usually guided by some principles (den Os et al., 1999), strategies (Souvignier et al., 2000), or objectives (Lamel et al., 2000). In some even these guidelines are implicit. The problem is more outstanding in those SDSs developed by the speech recognition community, in which most working SDSs come into being. Among many causes, we think, the most important is that dialogue management is short of solid theoretical support from dialogue models (the distinction between dialogue management and dialogue model will be explicated in section 2), in addition to the design of SDSs being a real world problem.

The approach we adopt in building dialogue management model for SDSs is to study human-human dialogues solving the same or similar problem. Though human-computer dialogues may be different in some aspects from human-human dialogues, the design of human-computer dialogue will benefit a lot from the study of human-human dialogues. It will not be clear whether those that characterize human-human dialogues are applicable to human-computer dialogues until they are well studied. Applicable or not, they are sure to contribute some insights to the design of dialogue management.

In what follows we first inspect main approaches to dialogue modeling and dialogue management and find two deep causes behind the gap between them (section 2). Against the causes we propose a generic dialogue model which distinguishes five ranks of discourse units and three levels of dialogue dynam-

ics (section 3). Then we apply it to information-seeking (one of the most common tasks adopted in the study of SDSs) dialogues and elaborate interaction patterns as utterance groups, which are classified along two dimensions (initiative and direction of information flow) into four basic types with some variations (section 4). We also experiment on segmenting utterance groups in our corpus with a subject and three algorithms.

2 The Gap

Why do most working SDSs make little use of dialogue models in their dialogue management? Or, why is there a gap between dialogue management and dialogue models?

To make it clear, we first distinguish between dialogue models and dialogue management models[1], or equivalently, between dialogue modeling and dialogue management modeling.The goal of dialogue modeling is to develop general theories of (cooperative task-oriented) dialogues and to uncover the universals in dialogues and, if appropriate, to provide dialogue management with theoretical support. It takes an analyzer's point of view. While the goal of dialogue management modeling is to integrate dialogue model with task model in some specific domain to "develop algorithms and procedures to support a computer's participation in a cooperative dialogue" (Cohen, 1998, p.204). It takes the viewpoint of a dialogue system designer.

Next, we briefly overview main approaches to dialogue modeling and dialogue management, then point out the causes behind the gap.

2.1 Dialogue Models

There are mainly two approaches to dialogue modeling: pattern-based and plan-based.[2]

[1]The distinction between dialogue models and dialogue management models is close to what Cohen (1998) makes in dialogue modeling. He distinguishes "two related, but at times conflicting, research goals ... often adopted by researchers of dialogue". Roughly speaking, one is theoretical and the other is practical.

[2]Cohen (1998) gives a more detailed discussion on dialogue modeling. Below we draw a lot from there. He mentions "three approaches to modeling dialogue – dialogue grammars, plan-based models of dialogue, and joint action theories of dialogue". We treat joint action theories as further development of original plan-based approach. So his latter two correspond to our plan-based approach in general.

Patten-based approach models recurrent interaction patterns or regularities in dialogues at the illocutionary force level of speech acts (Austin, 1962; Searle, 1969) in terms of *dialogue grammar* (Sinclair and Coulthard, 1975), *dialogue/conversational game* (Carlson, 1983; Kowtko et al., 1992; Mann, 2001), or *adjacency pairs* (Sacks et al., 1974). It benefits a lot from the insights of discourse analysis (Sinclair and Coulthard, 1975; Coulthard, 1992; Brown and Yule, 1983) and conversation analysis (Levinson, 1983).

Plan-based approach relates speech acts performed in utterances to plans and complex mental states (Cohen and Perrault, 1979; Allen and Perrault, 1980; Lochbaum et al., 2000) and uses AI planning techniques (Fikes and Nilsson, 1971). Later developments of plan-based dialogue models include multilevel plan extension (Litman and Allen, 1987; Litman and Allen, 1990; Carberry, 1990; Lambert and Carberry, 1991), theories of joint action (Cohen and Levesque, 1991) and SharedPlan (Grosz and Sidner, 1990; Grosz and Kraus, 1996).

Pattern-based dialogue model describes what happens in dialogues at the speech act level and cares little about why. Plan-based dialogue model explains why agents act in dialogues, but at the expense of complex representation and reasoning. In other words, the former is shallow and descriptive and the latter is deep and explanatory. Hulstijn (2000) argues for the complementary aspects of the two approaches and claims that "dialogue games are recipes for joint action".

Since, on the one hand, our target tasks belong to the class of simple service, like information-seeking and simple transactions, which are relatively well-structured and well-defined and not too complex for pattern-based dialogue models, on the other hand, there are some significant problems in using plan-based models in practical SDSs – those of "knowledge representation, knowledge engineering, computational complexity, and noisy input" (Allen et al., 2000), we will choose pattern-based instead of plan-based dialogue model as our theoretical basis for practical dialogue management at present.

2.2 Dialogue Management Models

We view dialogue management as an organic combination of dialogue model with task model in some

specific domain. Its basic functionalities include interpretation in context, generation in context, task management, interaction management, choice of dialogue strategies, and context management. All of them require contextual (linguistic and/or world) knowledge.

According to how task model and dialogue model are used, approaches to dialogue management can be classified into four categories[3] in Table 1.

Table 1: Classifying dialogue management models

		Task Model	
		implicit	explicit
Dialogue	implicit	DITI	DITE
Model	explicit	DETI	DETE

DITI or graph-based, both dialogue model and task model are implicit. Dialogue is controlled via finite state transitions (McTear, 1998). Topic flow is predetermined. It is neither flexible nor natural, but simple and efficient. It's suitable for simple and well-structured tasks similar to automated services over ATMs or telephones with DTMF input.

DITE or frame-based, with no explicit dialogue model, but task is explicitly represented as a frame or a form (Goddeau et al., 1996), a task description table (Lin et al., 1998), a topic forest (Wu et al., 2000), or an agenda (Xu and Rudnicky, 2000), etc. Both system and user may take the initiative. Topic flow is not predetermined. It's more flexible than that of DITI, but still far from naturalness and friendliness, since it makes no explicit use of dialogue models. Most working SDSs adopt this way of dialogue management.

DETI there is no practical dialogue management

using such a combination of task model and dialogue model.

DETE both dialogue model and task model are explicit. This type of dialogue management shares the advantages of frame-based one. At the same time it is potential to allow of more natural interactions according to the dialogue model used. This is what we are after here.

2.3 The Causes behind the Gap

From the analysis above we can see the surface gap between (DITE) dialogue management in most working SDSs and (pattern-based) dialogue models is mainly due to a deep one, i.e., the one between dialogue models and the underlying tasks.

There is another important cause – the interaction patterns are described at the level of speech act or dialogue act.[4] To link dialogue acts to utterances, three problems[5] must be addressed at the same time:

- Dialogue act classification scheme and its reliability in coding corpus, (Carletta et al., 1997; Allen and Core, 1997; Traum, 1999);

- Choice of features/cues that can support automatic dialogue act identification, including lexical, syntactic, prosodic, collocational, and discourse cues;

- A model that correlates dialogue acts with those features.

Some of the problems are discussed in (Jurafsky et al., 1998; Stolcke et al., 2000; Jurafsky and Martin, 2000; Jurafsky, 2002). The empirical work on dialogue act classification and recognition did not begin until some dialogue corpora (like Map Task, Verbmobil, TRAINS, and our NLPR-TI) were available.

But how could dialogue act recognition be successfully applied to practical dialogue management remains to be seen. So we choose a higher level

---

[3]For a more comprehensive discussion on dialogue management (and SDSs), see (McTear, 2002). He identifies two aspects of dialogue control (i.e., dialogue management) – initiative and flow of dialogue, and three strategies for dialogue control – finite-state-based, frame-based, and agent-based. The first two are similar to DITI and DITE respectively and the third is a collection of some other approaches which are now hardly applicable for practical dialogue management, among which is plan-based. Our classification below is more clear.

[4]Following Jurafsky (2002), we will adopt the term *dialogue act*, which captures the illocutionary force or commucative function of speech act. Though there are some arguments in (Levinson, 1983) and others against using dialogue act to model dialogues, and there are indeed some unresolved problems in linking dialogue acts to utterances, it will be our choice for the time being.

[5]We extend Webber's (2001) idea by splitting feature choice out.

construct (UT-3, see section 3.1.3) to describe inter-action patterns instead. We are by no means denying the important role dialogue act plays in dialogue modeling, but try to incorporate high level knowledge into dialogue modeling.

3 The Bridge – GDM

Against the above gap and its causes we propose a generic dialogue model (GDM) for task-oriented dialogues, which consists of five ranks of discourse units and three levels of dialogue dynamics. It captures two important aspects of task-oriented dialogue – interaction patterns at the low level and underlying task at the high level.

3.1 Discourse Units

We distinguish five ranks of discourse units in describing task-oriented dialogues: dialogue, phase, transaction, utterance group, and utterance.

3.1.1 Dialogue, Phase, and Transaction

The overall organization of a typical task-oriented *dialogue* can be divided into three *phases*, namely, an opening phase, a closing phase, and between them a problem-solving phase, which can be subdivided into *transactions* depending on how the underlying task is divided into subtasks. Each subtask corresponds to a transaction. If a task is atomic, there will be only one transaction in the problem-solving phase, just like the task of tourism information-seeking.

3.1.2 Utterance Group

In performing a subtask (or task, if atomic), some interaction patterns will recur. We name the interaction patterns *utterance groups* (or groups, for short). It's also called exchanges or conversational games (see section 2.1). The unit at this level involves complex grounding process towards common ground or mutual knowledge (Clark and Schaefer, 1989; Clark, 1996; Traum, 1994).

3.1.3 Utterance

The elementary unit in our model is *utterance*. Every utterance either initiates a new group, continues, or ends an old one. Usually it is what a speaker utters in his/her turn (for simplification, overlaps will not be considered here). But there are some turns with two or more utterances. These multi-utterance turns usually end an old group with their first utterance and begin a new one with their last utterance. Similar observations are found in Verbmobil corpus (Alexandersson and Reithinger, 1997).

Each utterance can be analyzed at three levels and assigned a type correspondingly (utterance type, UT):

UT-1 sentence type or mood, i.e., declarative, imperative, and interrogative (including yes-no question (ynq), wh-question (whq), alternative question (atq), disjunctive question (djq), which can be identified using surface lexical and prosodic features).

UT-2 dialogue act, see section 2.3.

UT-3 a more general communicative function, relative to a group, of a small number, including initiative (I), response/reply (R), feedback (F), acknowledgement (A) (typical in information-seeking dialogues), and others. It can be identified using UT-1 and semantic content (or utterance topic) and preceding UT-3s, *It is at this level that interaction patterns are more obvious.* What's more, it can be recognized without UT-2 (dialogue act) but contribute to dialogue act recognition.

3.2 Dialogue Dynamics

By dialogue dynamics, we mean the dynamic process within dialogues, i.e., how dialogues flow from one partner's utterance to another's all the way till the closing. The dynamic process includes that of intra-utterance (micro-dynamics) and that of inter-utterance. Inter-utterance dynamics is further divided into intra-group dynamics (meso-dynamics) and inter-group dynamics (macro-dynamics).

3.2.1 Micro-dynamics

Micro-dynamics deals with how discourse phenomena (like anaphora, ellipsis, etc.) within one utterance are decoded (interpretation) or encoded (generation) in discourse context and how utterance level intention (dialogue act) is recognized using lexical, prosodic, and other cues and discourse structure (see section 2.3). Discourse phenomena contain much discourse-level context information. It is those

that contribute partly to the naturalness and coherence in human-human dialogues. But it's very difficult for computers to make full use of them, either in interpretation or in generation. They are implemented in few of present SDSs, though much effort has been put on the study of computational models of discourse phenomena (see (Webber, 2001) for an overview and references therein for further details).

3.2.2 Meso-dynamics

Meso-dynamics explains utterance-to-utterance moves within one group which present recurrent interaction patterns. Our corpus study shows that those patterns in information-seeking dialogues are closely related to two factors – initiative and direction of information flow between user and server (see section 4.1).

3.2.3 Macro-dynamics

Macro-dynamics describes inter-group moves, which may take place intra-transactionally within one subtask or inter-transactionally between subtasks. Inter-group moves are subject to the underlying task. It's difficult to give an account like intra-group moves, because they reflect the process how a problem is solved. The account depends on how tasks are represented and reasoned. We may gain some hints from the study of general problem solving in AI (Bonet and Geffner, 2001).

3.3 Discussion

GDM as we propose above is a DETE dialogue management framework with fine-grained patterns. We discuss related work and its implication for dialogue management below.

3.3.1 Discourse Unit

Different discourse units are used by different researchers in studying discourse. In (Sinclair and Coulthard, 1975), five ranks of units are used to analyze classroom interactions: lesson, transaction, exchange, move, and act. The first four roughly correspond to our dialogue, transaction, group, utterance. We add the unit phase and omit act, which is a sub-utterance unit. In (Alexandersson and Reithinger, 1997), four ranks of units are used to analyze meeting scheduling dialogues: dialogue, phase, turn, and dialogue act. Turn is a natural unit that appears in dialogues, but is it an basic unit? Four

units with conversation acts (Traum and Hinkelman, 1992; Traum, 1994), are used to analyze TRAINS (freight scheduling) dialogues: multiple discourse unit (argumentation act), discourse unit (core speech act), utterance unit (grounding act), sub-utterance unit (turn-taking act). Theirs differ a lot from ours partly because they pay more attention to grounding.

3.3.2 Discourse Structure

In GDM the structure of discourse[6] is accounted for from two aspects: local structure is reflected in utterance groups and shaped by meso-dynamics; global structure is determined by the underlying task and shaped by macro-dynamics. This is obvious to task-oriented dialogues in view of GDM.

3.3.3 Dialogue Strategies

In most working SDSs dialogue strategies are handcrafted by system developers. Recently there are some efforts in applying machine learning approaches to the acquisition of dialogue strategies (Walker, 2000; Levin et al., 2000). We hope to find out what strategies are used in human-human dialogue and how they could be applied to human-computer dialogue. We first refine the concept of dialogue strategies. From the view of GDM, the strategies a dialogue agent may choose can also be classified into three levels, i.e.,

Micro-level strategies how to realize information structure, anaphora, ellipsis, and others, in utterances,

Meso-level strategies what to say regarding current group status, so as to complete ongoing group more friendly,

Macro-level strategies how to choose discourse topic regarding current task status, so as to complete the underlying task more efficiently.

---

[6]Grosz and Sidner (1986) proposed a tripartite discourse model consisting of attentional state, intentional structure, and linguistic structure. It is influential and covers both dialogue and text. But their intentional structure fails to capture the distinction between global level and local level structure. Their discourse unit – *discourse segment* – is used without noticing that there are different ranks of discourse unit in dialogues. This is partly due to that they looked more at the similarities between dialogue and text and less at the differences between them. Dialogue and text, as two types of discourse, share something in common, but there is also something that makes them different.

3.3.4 The Complexity of Dialogue Management

Since dialogue management is closely related to dialogue model and underlying task and domain, the complexity of dialogue management can be decomposed into three parts, i.e., the complexity of dialogue model, the complexity of task, and the complexity of domain. The complexity of dialogue model is affected by what kind of initiative and dialogue phenomena are allowed. The task complexity is affected by the number of its possible actions and by whether it is well-structured and well-defined. The domain complexity is affected by domain entities and their relations and by the volume of information. The three are not independent but intertwined.

4 Utterance Groups in GDM-IS

We now apply GDM to information-seeking dialogues (GDM-IS) and search for interaction patterns in the NLPR-TI corpus. We first try to classify and segment utterance groups. This is a preliminary step toward group pattern recognition. Details of the recognition process and results will be given in (Xu, 2002).

4.1 Group Classification

Group patterns are recurrent, but how many? Or, is there a limited number? In our NLPR-TI corpus information-seeking dialogues (see section 4.2.1), we find four basic groups with some variations.

4.1.1 Basic Groups

The recurrent patterns, according to our observation, can be classified into one of the four types in Table 2 along two dimensions – initiative and the direction of information flow (determined using world knowledge in the domain).

Table 2: Basic utterance groups

		Information Flow	
		S=>U	U=>S
Group	User	UISU	UIUS
Initiative	Server	SISU	SIUS

Direction of information flow In the dialogues of information-seeking, there are two directions of information flow: one from user to server (U=>S) and the other from server to user (S=>U). In the tourism domain, the former includes intended route (or sight-spot, or a rough area, obligatory), intended start time, number of tourists (optional); the latter includes start time, duration, vehicle, price, accommodation, meal, schedule, and more. Server must know the information about user's intended route before providing user with other information.

Initiative[7] In GDM, initiative always starts a new utterance group. It is one of utterance's general communicative functions relative to a group, together with reply, feedback, acknowledgement, as we mention in section 3.1.3. Regarding one group topic there are user initiatives (UI) and there are server initiatives (SI). Group patterns depend heavily on who initiates the group regarding some specific topic. This is due to the role asymmetry of the dialogue partners.

4.1.2 Complex Groups

Though most groups can be covered by the above basic patterns, there are some exceptions which are more complex. They are usually embedded ones. When one partner signals non-understanding or non-hearing, or a normal group is suspended, one or two more utterances will be inserted, either to repeat previous utterance or resume suspended group. The embedded groups may also be precondition groups. Precondition groups occur when some obligatory information is missing before the salient issue could be addressed. Once the missing is provided, the outer group will continue. Complex groups can also occur when one partner lists more than one items or does some repairing.

4.2 Group Segmentation

Given the above group classification, how to recognize them? We have to segment and classify groups, and determine UT-3 of every utterance within groups. This is a big problem. Only the experiment on group segmentation is reported in this paper.

---

[7] We note that there are task initiative and dialogue initiative (Chu-Carroll and Brown, 1998) and there are local initiative and global initiative (Rich and Sidner, 1998). Our initiative-in-group is more task-related and global. For a comprehensive discussion on mixed initiative interaction, see (Haller and McRoy, 1998, 1999).

To segment a dialogue into groups is first to determine the beginning of a group, i.e., to determine if an utterance is an initiative or not. (Multi-utterance turns are manually segmented beforehand for simplification.)

4.2.1 NLPR-TI Corpus

We use NLPR-TI corpus (Xu et al., 1999) in the experiment. It consists of 60 spontaneous human-human dialogues (about 5.5 hours) over telephones on tourism information-seeking. There are total 2716 turns (1346 by the user and 1370 by the server). The average length of user's turns is about 7 Chinese characters and server's about 9. The first 20 dialogues (transcript) are used for current group segmentation.

4.2.2 Manual Segmentation

A subject was given the basic ideas about GDM and utterance groups in GDM-IS and segmented two dialogues with an expert's guide before starting the work.

To test the reliability of group segmentation within GDM-IS, we calculate the kappa coefficient (K)[8] (Carletta, 1996; Carletta et al., 1997; Flammia, 1998) to measure pairwise agreement between the subject and the expert. Two coders segmented the first 20 dialogues (totally 845 utterances). They reached $K = .85 > .8$, which shows a high reliability. Using the expert's segmentation as reference, we also measure the subject's segmentation using information retrieval metrics – precision (P), recall (R), and F-measure[9] (see Table 3 for the result).

4.2.3 Automated Segmentation

Three simple algorithms in Figure 1 are used to perform the same task on the 20 dialogues. The input is a semantic tag sequence produced by a statistical parser (Deng et al., 2000)[10].

---

[8] $K = (P(A) - P(E))/(1 - P(E))$, where $P(A)$ is the proportion of times that the coders agree and $P(E)$ is the proportion of times that one would expect them to agree by chance. – From (Carletta, 1996)

[9] Combined metric $F = (\beta^2 + 1)PR/(\beta^2 P + R)$, from (Jurafsky and Martin, 2000, p.578), $\beta = 1$.

[10] That we adopt such deep features in discourse segmentation is mainly due to our target application – dialogue management. This makes it different from others using surface features like (Passonneau and Litman, 1997).

I. Using topic only for segmentation
if topic is new
then UT-3 = *initiative*
else UT-3 = *non-initiative*

II. Using UT-1 only for segmentation
if UT-1 $\in$ interrogatives
then UT-3 = *initiative*
else UT-3 = *non-initiative*

III. Using both for segmentation
if topic is new $\land$ UT-1 $\in$ interrogatives
then UT-3 = *initiative*
else UT-3 = *non-initiative*

Figure 1: Group segmentation algorithms

Given the semantic tag sequence of an utterance, we determine its topic[11] and UT-1 (what we are most interested in is interrogatives (ynq, whq, atq, and djq)). Since the parser performs with an error rate of 26.4%, there will be some wrong semantic tags, which lead to errors in assigning UT-1 and topic. Then we use the three simple algorithms to segment groups in the 20 dialogues. Their performance (also using the expert's segmentation as reference) is given in Table 3.

Table 3: Group segmentation results

	subject	I	II	III
Precision	.88	.59	.67	.83
Recall	.92	.82	.62	.56
F-measure	.90	.69	.64	.67

4.3 Discussion

Table 3 shows the results of group segmentation, both manual and automated. Though none of the three algorithms outperforms the subject, they all show that topic change and UT-1 as interrogative are acceptable and also good indicators of utterance group beginning, esp. when topic and UT-1 are the

---

[11] We presume that the topic of an utterance is the last one in the candidate tags. This seems problematic but is true to most of the utterances according to our observation. How to determine the topic of an utterance needs further study.

only information sources and when discourse markers (Schiffrin, 1987) in spontaneous speech are unavailable in current deep analysis.

There is no obvious performance difference in segmenting dialogue into groups with the three algorithms. The performance of algorithm I may be improved if the noises brought by the parser and our simple topic identification algorithm are cleared. This implies that topic change is a potentially better indicator of the beginning of new groups. The result using UT-1 only is the worst. This is partly because not all groups begin with interrogatives and that interrogatives do not always occur at the beginning of a group. When using both topic and UT-1, the performance changes little, though seemly more constraints are used. This possibly is because topic change and UT-1 as interrogative overlap a lot.

5 Conclusions

After a survey of the main approaches to dialogue modeling and dialogue management in working SDSs, we find the causes behind the gap between practical dialogue management and dialogue models and propose GDM, which consists of five ranks of discourse units and three levels of dialogue dynamics. It promises to bridge the gap through integrating meso-dynamics at the group level with macro-dynamics at the task level, and modeling interaction patterns via utterance groups using UT-3.

Then we apply it to information-seeking dialogues and elaborate utterance groups (or interaction patterns) in the model. We also classify and segment utterance groups in our information-seeking corpus, which takes a preliminary step toward better dialogue modeling for practical dialogue management with empirical justification. A more challenging task – group pattern recognition – is under way (Xu, 2002). After that we will investigate how local discourse structure in terms of utterance group structure could contribute to the recognition of dialogue act (UT-2).

GDM takes a step further toward better dialogue modeling for practical dialogue management with empirical justification. It is expected to be used in practical dialogue management in the near future for better usability and portability.

References

Jan Alexandersson and Norbert Reithinger. 1997. Learning dialogue structures from a corpus. In *Proceedings of the 5th European Conference on Speech Communication and Technology*, volume 4, pages 2231–2234.

James Allen and Mark Core. 1997. Draft of damsl: Dialog act markup in several layers. Available from `http://www.cs.rochester.edu/research/cisd/resources/damsl/`.

James F. Allen and C. Raymond Perrault. 1980. Analyzing intention in utterances. *Artificial Intelligence*, 15(3):143–178.

James Allen, George Ferguson, Bradford W. Miller, Eric K. Ringger, and Teresa Sikorski Zollo, 2000. *Dialogue Systems: From Theory to Practice in TRAINS-96*, chapter 14, pages 347–376. In Dale et al. (Dale et al., 2000).

J. L. Austin. 1962. *How to do Things with Words*. Clarendon Press, Oxford.

Blai Bonet and Héctor Geffner. 2001. Planning and Control in Artificial Intelligence: A Unifying Perspective. *Applied Intelligence*, 14(3):237–252.

Gillian Brown and George Yule. 1983. *Discourse Analysis*. Cambridge University Press.

Sandra Carberry. 1990. *Plan Recognition in Natural Language Dialogue*. ACL-MIT Press Series in Natural Language Processing. A Bradford book, MIT Press, Cambridge, Massachusetts.

J. Carletta, A. Isard, S. Isard, J. C. Kowtko, G. Doherty-Sneddon, and A. H. Anderson. 1997. The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1):13–31.

Jean Carletta. 1996. Assessing agreement on classification tasks: The Kappa statistic. *Computational Linguistics*, 22(2):249–254.

Lari Carlson. 1983. *Dialogue Games: An Approach to Discourse Analysis*. D. Reidel, Dordrecht, Holland.

Jennifer Chu-Carroll and Michael K. Brown. 1998. An evidential model for tracking initiative in collaborative dialogue interactions. *User Modeling and User-Adapted Interaction*, 8(3-4):215–253.

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

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

Philip Cohen, Jerry Morgan, and Martha Pollack , editors. 1990. *Intentions in Communication*. MIT Press.

P. R. Cohen and H. J. Levesque. 1991. Teamwork. *No ûs*, 25(4):487–512.

P R. Cohen and C. R. Perrault 1979. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3(3):177–212.

Phil Cohen, 1998. *Dialogue Modeling*, chapter 6.3. In Cole et al. (Cole et al., 1998).

Ronald Cole, Joseph Mariani, Hans Uszkoreit, Giovanni Varile, Annie Zaenen, Antonio Zampolli, and Victor Zue, editors. 1998. *Survey of the State of the Art in Human Language Technology*. Cambridge University Press, Cambridge.

Malcolm Coulthard, editor. 1992. *Advances in Spoken Discourse Analysis*. Routledge. London.

Robert Dale, Hermann Moisl, and Harold Somers, editors. 2000. *Handbook of Natural Language Processing*. Marcel Dekker. New York.

Yunbin Deng, Bo Xu, and Taiyi Huang. 2000. Chinese spoken language understanding across domain. In *Proceedings of the 6th International Conference on Spoken Language Processing*, volume 1, pages 230–233.

Richard Fikes and Nils J. Nilsson. 1971. STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2(3-4):189–208.

Giovanni Flammia. 1998. *Discourse segmentation of spoken dialogue: an empirical approach*. Ph.D. thesis, MIT.

D. Goddeau, H. Meng, J. Polifroni, S. Seneff, and S. Busayapongchai. 1996. A form-based dialogue manager for spoken language applications. In *Proceedings of the 4th International Conference on Spoken Language Processing*, volume 2, pages 701–704.

Barbara J. Grosz and Sarit Kraus. 1996. Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269–357.

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

Barbara J. Grosz and Candace L. Sidner. 1990. *Plans for Discourse*. In Cohen et al. (Cohen et al., 1990).

Susan Haller and Susan McRoy, editors. 1998, 1999. *User Modeling and User-Adapted Interaction, Special Issue on Computational Models for Mixed Initiative Interaction*, 8(3-4),9(1-2).

Joris Hulstijn. 2000. Dialogue games are recipes for joint action. In *Proceedings of the Forth Workshop on the Semantics and Pragmatics of Dialogue (Gotalog'00)*.

Biing-Hwang Juang and Sadaoki Furui, editors. 2000. *Proceedings of the IEEE , Special Issue on Spoken Language Processing*, 88(8).

Daniel Jurafsky, Rebecca Bates, Noah Coccaro, Rachel Martin, Marie Meteer, Klaus Ries, Elizabeth Shriberg, Andreas Stolcke, Paul Taylor, and Carol Van Ess-Dykema. 1998. Switchboard discourse language modeling project report. Technical Report Research Note No. 30, Center for Speech and Language Processing, Johns Hopkins University, Baltimore, MD.

Daniel Jurafsky and James H. Martin. 2000. *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. Prentice-Hall.

Daniel Jurafsky, 2002. Pragmatics and Computational Linguistics. To appear in Laurence R. Horn and Gregory Ward, editors. *Handbook of Pragmatics*. Blackwell, Oxford.

J. Kowtko, S. Isard, and G. M. Doherty. 1992. Conversational games within dialogue. Research Paper 31, Human Communication Research Centre, Edinburgh University, Edinburgh.

Lynn Lambert and Sandra Carberry. 1991. A tripartite plan-based model of dialogue. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pages 47–54, Berkeley, CA.

L. Lamel, S. Rosset, J.L. Gauvain, S. Bennacef, M. Garnier-Rizet, and B. Prouts. 2000. The LIMSI ARISE system. *Speech Communication*, 31(4):339–354.

Esther Levin, Roberto Pieraccini, and Wieland Eckert. 2000. A stochastic model of human-machine interaction for learning dialog strategies. *IEEE Transactions on Speech and Audio Processing*, 8(1):11–24.

Stephen C. Levinson. 1983. *Pragmatics*. Cambridge University Press.

Y. Lin, T. Chiang, H. Wang, C. Peng, and C. Chang. 1998. The design of a multi-domain mandarin Chinese spoken dialogue system. In *Proceedings of the 5th International Conference on Spoken Language Processing*, volume 1, pages 230–233.

Diane J. Litman and James F. Allen. 1987. A plan recognition model for subdialogues in conversation. *Cognitive Science*, 11(2):163–200.

Diane J. Litman and James F. Allen. 1990. *Discourse Processing and Commonsense Plans*. In Cohen et al. (Cohen et al., 1990).

Karen E. Lochbaum, Barbara J. Grosz, and Candace L. Sidner, 2000. *Discourse Structure and Intention Recognition*, chapter 6, pages 123–146. In Dale et al. (Dale et al., 2000).

William C. Mann. 2001. The genre diversity of dialogue game theory. Available from `http://www-rcf.usc.edu/~billmann/memos.htm`.

Michael F McTear. 1998. Modelling spoken dialogues with state transition diagrams: experiences with the CSLU toolkit. In *Proceedings of the 5th International Conference on Spoken Language Processing*, volume 2, pages 1223–1226.

Michael F. McTear. 2002. Spoken dialogue technology: Enabling the conversational user interface. *ACM Computing Surveys*,34(1):90–169.

Els den Os, Lou Boves, Lori Lamel, and Paolo Baggia. 1999. Overview of the ARISE project. In *Proceedings of the 6th European Conference on Speech Communication and Technology*, volume 4, pages 1527–1530.

Rebecca Passonneau and Diane Litman. 1997. Discourse segmentation by human and automated means. *Computational Linguistics*, 23(1):103–140.

Charles Rich and Candace L. Sidner. 1998. Collagen: A collaboration manager for software interface agents. *User Modeling and User-Adapted Interaction*, 8(3-4):315–350.

H. Sacks, E. A. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.

Deborah Schiffrin. 1987. *Discourse Markers*. Cambridge University Press.

J. R. Searle. 1969. *Speech Acts*. Cambridge University Press.

John M. Sinclair and Malcolm Coulthard. 1975. *Towards an Analysis of Discourse: The English Used by Teachers and Pupils*. Oxford University Press.

Bernd Souvignier, Andreas Kellner, Bernhard Rueber, Hauke Schramm, and Frank Seide. 2000. The thoughtful elephant: Strategies for spoken dialog systems. *IEEE Transactions on Speech and Audio Processing*, 8(1):51–62.

Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3):339–371.

David R. Traum and Elizabeth A. Hinkelman. 1992. Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8(3):575–599.

David R. Traum. 1994. *A Computational Theory of Grounding in Natural Language Conversation*. Ph.D. thesis, University of Rochester.

David R. Traum. 1999. 20 questions for dialogue act taxonomies. In *Proceedings of the Third Workshop on the Semantics and Pragmatics of Dialogue (Amstelog'99)*.

Marilyn A. Walker. 2000. An application of reinforcement learning to dialogue strategy selection in a spoken dialogue system for email. *Journal of Artificial Intelligence Research*, 12:387–416.

Bonnie Webber. 2001. Computational Perspectives on Discourse and Dialogue. In Deborah Schiffrin, Deborah Tannen, and Heidi Hamilton, editors. *Handbook of Discourse Analysis*. Blackwell, Oxford.

Xiaojun Wu, Fang Zheng, and Mingxing Xu. 2000. Topic forest: A plan-based dialog management structure. In *Proceedings of ICASSP*, volume 1, pages 617–620.

Wei Xu and Alexander I. Rudnicky. 2000. Task-based dialog management using an agenda. In *Proceedings of ANLP/NAACL 2000 Workshop on Conversational Systems*, pages 42–47.

B. Xu, T.Y. Huang, X. Zhang, and C. Huang. 1999. A Chinese spoken dialogue database and its application for travel routine information retrieval. In *Proceedings of the Second International Workshop on East-Asia Language Resources and Evaluation*, Taipei.

Weiqun Xu. 2001. Survey of the state of the art in spoken dialogue systems. Manuscript.

Weiqun Xu. 2002. Grouping utterances in information-seeking dialogues. In preparation.

Victor Zue and Jim Glass. 2000. Conversational interfaces: Advances and challenges. *Proceedings of the IEEE*, 88(8):1166–1180.