Modelling dialogue: Challenges and Approaches

David Schlangen

Modelling dialogue, that is, designing formal systems that reproduce aspects of natural conversation, is a challenging task. In this overview paper we describe some of the challenges, and review extant approaches to dealing with them.

1 Introduction

Modelling dialogue, that is, designing formal systems that reproduce aspects of natural conversation, is a challenging task. Not only must a dialogue model be able to handle most (if not all) of the linguistic phenomena that make monological discourse hard to model (e.g., anaphora, presuppositions, implicatures), there are also additional requirements: unlike (written) text, dialogue unfolds in time, and timing matters; and unlike monologue, a dialogue is an interaction, and is shaped by the interests and capabilities of all dialogue participants (DPs), while it unfolds.

Describing these challenges is one aim of this overview paper; the other is reviewing some of the extant approaches to adressing them in dialogue models. In the very general sense of the term as used above, all dialogue systems, i.e., computer systems that interact via natural language with users, are models of dialogue. Traditionally, however, the term has been mostly employed for those modules within dialogue systems that handle what may be called *content-flow* aspects of dialogue, organising the *coherence* of utterances and *cooperation* on the task. Other, more *interactional* aspects such as the timing of utterances or cooperation on the dialogue itself have been dealt with by other modules, often in a less principled manner. As we argue, observable interactions between these phenomena indicate that a better strategy might be to combine models of these aspects.

This, then, is the structure of the paper: in Section 2 we review some of the phenomena observable in human-human dialogue that make it a challenging subject for study, such as the organisiation of whose turn it is to speak, strategies for recovery from communication problems, and regularities in the "flow" of the dialogue. In Section 3 we review approaches that have been taken to modelling these phenomena. We close with a brief discussion of shortcomings of current models and implementations and possible ways to overcome them.

2 Challenges

We take a descriptive stance in this section, simply describing some phenomena that can be observed in natural dialogue, proceeding from the smallest units, utterances, to the overall organisation of dialogues. Whether they need to be modelled or not is a decision that will be discussed in the next section—models may choose to structure the interaction in such a way that the occurence of some phenomena can be avoided (e.g. by keeping strict control over the develeopement of the dialogue). Moreover, implemented models may want to make a distinction between the range of phenomena the model must be able to *interpret* and those it must be able to *generate*; e.g., disfluencies (see below) will likely occur in input from human users, but whether there is any value in letting the model also produce them depends on the purpose of the model.

2.1 Utterances

Spoken dialogue does not come pre-segmented into units in the same way as written language does, where segmentation is provided by typographic means. This opens the question what the basic units of analysis are in spoken dialogue. The notion of *sentence* or *clause*, as used in the analysis of written monologue, can only be transferred with modifications, for two reasons.

First, there are artefact of the spontaneous nature of the language that are not normally present in written language, and which render the products non-sentential. E.g., the string shown (together with some annotation) in (1) below is not grammatical due to the abortion of one element and a subsequent correction. Such artefacts are called *disfluencies* or *self-repairs*, and seem to follow a fairly regular pattern [26, 18], so that they can be relatively easily detected and removed [16]. (1) is glossed with the names introduced by [18, 30] for the different parts of self-repairs.

(1) *until you're* | *at the le-* || *I mean* ||
start reparandum editing terms
at the right-hand | *edge of the quarry*
alteration continuation

Second, turns in dialogue can be "shorter" than full sentences, and consist for example just of nominal phrases, as in (2)–B.

(2) A: *Who came to the party?*
B: *Sandy.*

Such sub-sentential utterances are called *fragments*, and are related to elliptical sentences in that their form is, unlike that of disfluencies, not the result of "accidental" processes but seems to be intended. Models of their contextual resolution are provided by [13, 28], *inter alia*.

This is the first observation, then: the *utterance* as the basic unit of analysis is not congruent with the syntactic notion of sentencehood and needs to be given a disjunctive, intention-oriented definition [1]: it can consist of a sentence (possibly containing disfluencies), or smaller syntactic elements, as long as they form an *intentional unit*.

2.2 Dialogue Acts and Coherence

This brings us to the next observation. Utterances in dialogue are produced in order to achieve something, they can be analysed as *acts*, just like other intentional behaviour. The effects of these acts can be usefully bundled and labelled with *dialogue act types* (a notion that generalises Searle's concept of *speech acts* [29]). Several such schemes have been developed, of which we mention only DAMSL [2] here. (3) gives an example of utterances annotated with DAMSL-acts.

(3) A: Please open the door. *[action-directive]*
B: OK. *[committ; accept; signal-und./acknowledge]*

In DAMSL, a single utterance can be analysed as constituting more than one act. This is an important feature, because it allows a principled analysis of the various functions an utterance can have. DAMSL distinguishes the *forward-looking function (FF)*, which is individuated with respect to the way an utterance constrains future discourse (e.g., a promise commits the promiser to intend the promised action), and the *backward-looking function (BF)*, which describes how the utterance relates to previous discourse.

This analysis integrates two important observations: One is that there is a certain sense of "connectedness" to discourse, where new utterances relate to previous discourse. Some explanatory theories of discourse (e.g., [15, 21, 8]) go further than the descriptive DAMSL schema and take connectedness of all utterances (*coherence*) to be a necessary feature of "good" discourse. In any case, reproducing this impression of coherence in discourse is a major task of all dialogue models.

The second, related observation is that the choice of an utterance constrains the future discourse. The FF in DAMSL is meant only to record the effect of utterances on the public beliefs and obligations of the utterer, but some researchers have pointed out that there are regular patterns of utterance types, and that in this sense an utterance more concretely has an influence on further discourse. In the field of *conversation analysis* for example ([25], [19] for a comprehensive review), it has long been noted that for certain utterance types (e.g., questions) there are preferred replies (e.g., answers) and dispreferred replies (e.g., a refusal to answer), the latter typically being more elaborate linguistically (e.g., specifying reasons). Dialogue models might chose to model these patterns as well.[1]

Lastly, DAMSL also offers an additional dimension of analysis, namely the *information level* at which an act operates, where the scheme distinguishes between acts that pertain to the task, to task-management, or to communication-management. The observation that utterances can contribute to organising the interaction of which they are part will be further detailed in the next section.

To summarise: utterances in dialogue are intentional acts that have to be interpreted with reference to the context they occur in (backward-looking function) and the changes they make to that context (forward-looking f.).

2.3 Interaction Management

In dialogue, participants take turn in speaking. This is normally accomplished very smoothly: *overlaps* of speech *are rare* (less than 5% of turns in one study, see review in [19]), and *pauses* between turns *are short* (often even shorter than motor-planning the next utterance takes, see review of evidence in [11]). Taking or handing over the floor seems to be organised cooperatively between the DPs through the use of certain *turn-taking devices* [24], such as using particular intonational contours, producing certain dialogue acts that "select" the next speaker (questions, requests), or making visible signs that one wants to begin speaking—more on models of turn-taking below in Section 3.3.

Controlling who gets the next turn via the devices mentioned above is one aspect of *initiative* in dialogue, which [10] calls the *dialogue initiative* aspect. *Task initiative*, on the other hand, is held by the dialogue participant "driving" the task forward. These factors can vary independently; e.g., asking a clarification question gives you dialogue, but not necessarily task initiative. There are dialogue genres where initiative is regularly distributed unevenly between the DPs (e.g., interviews, tutorials, exams), but in free conversation it is normally held in equal measures by all DPs. Implemented dialogue models might choose to restrict initiative for technical reasons, however.

As [11] points out, there is a class of utterances that seems to be systematically exempt from the preference to minimise overlaps, the so-called *feedback utterances* ("uhu, yeah", etc.; [7]). To account for the function of such utterances, [11] introduces the metaphor of parallel *tracks* on which conversation proceeds, where one track is devoted to the "official business" of the conversation, and the other to managing the interaction; the observations about turn-taking then can be restricted to utterances that contribute to the main track. (This distinction has occured before above, in the DAMSL *information level* and the distinction between kinds of initiative.)

The particular aspect that feedback utterances manage is that of *ensuring mutual understanding*, a process that [11] called *grounding*. Recipients in dialogue produce these signals to indicate their success with understanding the speaker's contribution, and speakers appear to use them

[1] A question that will arise again is whether such patterns are the result of other, for example intentional or politeness-related constraints, or whether they should have an independent status in a theory. Practical models, in any case, might simply chose to directly produce these patterns rather than try to explain them.

to determine whether their communicative goals have been reached.[2] A sub-class of FBs that is particular prominent in dialogue is that of *clarification requests (CRs)* as in (4). They make understanding problems explicit and ask for repair (*other-repair* compared to the self-repair discussed above).

(4) A: *I saw Peter.*
B: *Peter? / Who? / You did what? / Pardon?*

Given the frequency of CRs in human-human dialogue (4–5% of utterances in task-oriented dialogue, [27] and references therein) and the fact that implemented systems will likely only have more understanding problems, modelling grounding behaviour is an important part of modelling dialogue.

To summarise, besides having to ensure coherence of the dialogue, the participants are also responsible for organising their interaction—they cooperate on taking turns so as to avoid ineffective overlaps, and on reaching mutual understanding on what was said.

2.4 Conversation Structure

Clarification requests as discussed above can be seen as opening sub-dialogues that are "inserted" into the main dialogue; similarly, sub-dialogues can concern corrections, knowledge preconditions ("do you know what this is?"), etc. At an even higher level of abstraction, one might want to distinguish *phases* in conversations (a notion coming from *conversation analysis* again, see [19, 11]), like *openings, main business, closings*, with different conventions. Dialogue models might use such phases to restrict the range of dialogue acts possible; but again the question arises whether these phases have independent explanatory value or are results of other constraints.

3 Approaches

We now turn to the approaches to modelling dialogue. This overview is organised into sections on models of *what was said* (recognising dialogue acts), of *what to say next*, and of *when to say it*. It closes with an overview of approaches to combining such models.

3.1 Models of what was said

The first task here is to identify utterance boundaries and remove disfluencies. This is often factored out in implemented models to the speech recognition component, and so will not be further discussed here (see [16] for one approach).

We have said above that a useful abstraction of the intentions connected with utterances is the concept of dialogue acts. But how can the connection between an actual utterance token and the dialogue act types it instantiates (i.e., the intentions they convey) be made? (If it is desired; not all dialogue models make it, in some the input directly determines the reaction of the system—see next section.)

There are two general strategies that are used in practical dialogue systems (which can be seen a complimentary and whose results can be combined, [17]), using either *symbolic* or *statistic methods to evaluate cues* in the input. The latter work directly on surface information (words, lexical features, prosodic information, etc.) and make their classification decision based on probabilistic models induced from annotated corpora (see for example [31]). Such models are robust, but prone to miss subtle contextual nuances [6].

A common symbolic strategy is to use a two-stage approach, where a set of *surface speech acts* is computed through syntactic and semantic analysis of the utterance, among which the intended speech act is identified through contextual reasoning (a textbook description of this method is [5], different implementations are [6, 17]). In the context of a travel information setting for example, such an approach would derive for an utterance of "Can you tell me the direction?" the surface acts *request* and *yes-no-question*, among which it would identify the *request*-reading, based on contextual reasoning, the extent of which varies between approaches.

A more principled symbolic model is offered by a theory called SDRT [8], which uses a non-monotonic logic to combine in a tractable manner information coming from various sources (including lexical and compositional semantics, discourse structure, and cognitive states of DPs) and makes detailed predictions about what makes dialogues coherent. However, it has so far only been implemented for a carefully restricted domain [28] and it is unclear whether it can be used in large-scale models.

3.2 Models of what to say next

The models we discuss in this section can all be described as specifying *states* and *transitions* between them; where they differ is in what they assume the states are, i.e., which kind of information they assume must be represented, and in how the transitions are specified.[3]

3.2.1 Structured Dialogue Models

The first class of models we discuss has the most reduced view of dialogue states. Among these approaches the oldest are those that use *finite-state automata* to specify a set of legal dialogues (for historical overview and references see [22]). The states in these models are atomic, that is, they carry information only in virtue of their position in the network. The states are associated with topic-specific elements such as prompts to be played (e.g. "where do you want to fly to?") and speech recognition grammars to load (e.g., specialised for city names). Transitions between states are

[2] It is useful to distinguish layers of understanding here, e.g., acoustic, semantic, and pragmatic understanding, see [7, 11, 23, 27]; we gloss over this for reasons of space.

[3] This common framework for describing dialogue models was proposed by the TRINDI consortium, who introduced the general term *information state update* for it [33]. It is important to distinguish, as intended by the authors, this way of *talking* about models from the particular models that were developed within the TRINDI project (see below).

triggered by specific events, such as recognition results ("To Berlin." vs. `not-recognised`, or, if they use the additional abstraction, recognition of certain dialogue acts).

Such models *impose* a structure on the dialogue rather than explain it; by keeping (both dialogue- and task-) initiative to the system, they fully control the flow of the dialogue. This has technical advantages because the search space for interpretation is reduced, but also obvious disadvantages: deviations from the pre-scripted path are not tolerated and may result in unforeseen behaviour. Moreover, they make no distinction between task and dialogue control, and so recovery-mechanisms (as discussed under *grounding* above) can not be specified abstractly but must be integrated into the dialogue-script.

There are various attempts to make such approaches more flexible (see e.g. [9, 12]), for example through allowing information to be passed between states, making transitions dependent on the execution of programs (raising the formal complexity of the underlying automaton), or, more substantially, through using *forms* as the structuring device, where the dialogue is controlled by which information is missing in a form (this allows overanswering of questions as in "To Berlin on Monday" as an answer to the question above). The fact that the task has to be pre-structured remains, however.

To summarise, the models sketched here mostly are justified on practical grounds (robustness). They are good for pre-structured tasks such as constructing data-structures (e.g., collecting information for a database lookup), and professional development environments exists (see [22]). However, the feature that gives them their practical advantage—the strict control over the dialogue—also makes them less appropriate for more dynamic and self-organised tasks.

3.2.2 Plan-Based Dialogue Models

Plan- or agent-based models can be seen as being positioned at the other end of the flexibility scale. In these models the flow of the dialogue is based on local inferences over rich contextual representations (recording beliefs, desires and intentions of agents), and develops out of principles that are seen to be general for intentional behaviour.

These approaches have been developed mostly within the AI-commmunity, and started with attempts to connect the notion of speech acts (see above) with AI planning concepts. In one influential early system [4], responses as in (5) were modelled.

(5) Patron: *When does the Montreal train leave?*
Clerk: *At 3:15 at gate 7.*

The system, in the role of the clerk, recognised the *plan* of the customer (to go to Montreal, by train), and identified possible *obstacles* (missing knowledge of time *and place* of departure), which it adressed. This was an attempt to model the inferences which seem to lie behind helpful (strictly speaking, overinformative) behaviour as in the example. In contrast to the approaches discussed in the previous section, this response is not pre-scripted.

Because of their principled, declarative nature, such rule systems are theoretically very appealing. However, the high computational cost of the required reasoning made it necessary already in the first systems to introduce task-specific heuristics that restrict the search space, reducing the declarativity of the models. Moreover, these early systems only had available plans that concerned the domain level plan, which made it difficult to deal with utterances that have interaction management functions (see above). To adress these problems, [20] introduced a distinction between *domain plans* and *discourse plans*, the latter handling for example clarification sub-dialogues. (This distinction echoes that into "tracks" discussed above.)

The still quite severe computational problems, however, meant that such models were never employed in practical systems, and for a while there was not much further developement. James Allen's TRAINS/TRIPS research project [6] then revived some of the central ideas while aiming to maintain real-time behaviour through adding domain restrictions to the reasoning. It models sophisticated practical problem solving, using concepts such as *objectives* (goals, constraints), *solutions* (plans) and *resources* (for use in solutions), which can become the topic of discussion, being available for operations like *evaluate, modify, repair* and *abandon*. These operations are described abstractly and are only connected to respective domain operations by separate rules, allowing for adaptation to different domains. Besides tracking beliefs and intentions, the system also keeps note of the *grounding status* of utterances (confirmed, unconfirmed, to be clarified), and adds as a new theoretical concept *obligations* that arise during the dialogue; this notion is used to explain why people bother giving negative replies, which was a problem for earlier plan-based approaches.

The resulting system is impressively flexible while also being robust (see overview of evaluation in [22]); a possible criticism is that applications appear to be rather resource-intensive to build, and large-scale evaluations of attempts by other groups to adapt the model to different domains have still to be made.

3.2.3 Information-State Update: The QUD Model

To bring out the similarities and differences between the models, we have used the general terminology of information-states and transitions throughout this section. The approaches developed under the label ISU, however, define information states more narrowly as records (typically displayed as attribute-value structures) of the information required by the model, and give transitions by defining *update rules* that use this information to determine the new state (i.e., to update it). Various models have been implemented within this framework (see [33]); we describe Larsson's QUD-model here.

The IS in this model distinguishes between information that is deemed *private* to the system (e.g., its plans and agenda) and such which is *shared* (i.e., has been grounded between the DPs). The dialogue is structured around the concept of the *question under discussion (QUD)* [14]. This QUD might be something like "Where do you want to travel?"; if the user replies "To Berlin, on Saturday", the update rules remove this question from the QUD (“downdate” it) and also search for a question on the plan that matches the unused bit “on Saturday” and put it on the QUD (accommodate it). While this is in effect not much more powerful than forms-based models, the strength of this approach lies in the declarative way in which this is formulated, which for example allows grounding-behaviour to be integrated. For this, domain-independent update rules look at the quality of the input-recognition and track possible understanding problems, and control generic protocols to deal with them (e.g., through generation of feedback or clarification questions).

Independent from the details of the models implemented within it, the ISU approach is attractive because it encourages a declarative formulation of the required information sources and the rules for computing transitions. The TRINDI project also provides a toolkit for realising models in this framework [33].

3.3 Models of when to say it

Most implemented dialogue systems use fairly strict turn-taking models. TRAINS for example uses a push-to-talk system, where taking and releasing the turn is explicitly announced. Most models based on structured-dialogue approaches also enforce strict turn-taking, allowing the user only to react at pre-determined states. More advanced systems sometimes allow “barge in”, that is, allow users to interrupt a system turn.

Natural turn-taking behaviour is more flexible than this, as discussed above. The evidence cited there indicates that a model of natural turn-taking must be *projective* rather than *reactive*, that is, must be able to predict points where turns might end *before* the actual event. [24] present such a model, based on the notion of *turn constructional units* and *transition relevance points (TRPs)*. One cue they use to project the position of TRPs is syntax (when is the utterance likely to be completed?), other cues have been identified as well (see [32] for a recent overview). At TRPs, either the speaker selected to speak next (see above) is obliged to react, or, if none was selected, the floor is open for anyone, including the previous speaker. Models like this have been implemented, interestingly mostly in the context of systems that use modalities other than speech as well (e.g., user's gaze), see [32]. One problem holding back the use of such models will be discussed in the next section.

3.4 Perspectives

In Section 2 we noted interactions between certain phenomena, for example that certain utterance types (feedbacks) underly different turn-taking constraints. In the previous section, we noted that a natural turn-taking model must be projective, that is, must determine whether a TRP is upcoming while the utterance is processed. This means that a good model of these phenomena must be *incremental* and *parallel*, that is, must be able to decide on grounding and turn-taking while it is still determining what was said, unlike traditional systems which use a sequential processing strategy. Interesting first steps have been taken in this direction [3], but major research issues remain (e.g., incremental parsing, incremental contextual reasoning, etc.) However, if progress is to be made towards *natural, flexible conversational systems*, then such models will be required.

4 Conclusions

In this paper, we have reviewed some phenomena found in natural dialogue, and discussed extant approaches to modelling essential parts of a full dialogue model, namely of what was said, what to say next, and when to say it. What should have become clear during the discussion is that there is “correct” approach—which one chooses will depend on the purpose of the modelling. However, the constraints on models that want to achieve natural-sounding dialogue should also have become clear.

Due to the complexity of the subject, this review has been rather cursory. Further information can be found in the following monographs: [22], from a more practical perspective; [5] for the foundations of plan-based approaches; and [11] for a comprehensive, non-computational model of interaction. Current work at the interface between theoretical and practical models of dialogue is presented mostly at the workshop series of the special interest group on discourse and dialogue (SIGdial) of the Association for Computational Linguistics and at the “Semantics and Pragmatics of Dialogue” (SemDail) series.

Kontakt

Dr. David Schlangen
Universität Potsdam, Institut für Linguistik
Postfach 601553, D-14415 Potsdam
Tel.: +49 (0)331 977 2956
e-mail: das@ling.uni-potsdam.de
WWW: http://www.ling.uni-potsdam.de/ das

Bild

David Schlangen is currently a Post-Doc at the University of Potsdam, working on computational models of conversational competence. Before that he completed a PhD at the University of Edinburgh, developing a logical model of the interpretation of non-sentential utterances in dialogue.

References

[1] Jan Alexandersson, Bianka Buschbeck-Wolf, Tsutomu Fujinami, Michael Kipp, Stephan Koch, Elisabeth Maier, Norbert Reithinger, Birte Schmitz, and Melanie Siegel. Dialogue acts in Verbmobil-2: Second edi-

tion. Verbmobil Report 226, Verbmobil Consortium, Saarbrücken, July 1998.

[2] James Allen and Mark Core. Draft of damsl: Dialog act markup in several layers. Discourse Research Initiative, October 1997.

[3] James Allen, George Ferguson, and Amanda Stent. An architecture for more realistic conversational systems. In *Proceedings of the conference on intelligent user interfaces*, Santa Fe, USA, June 2001.

[4] James Allen and C. R. Perrault. Analyzing intention in utterances. *Artificial Intelligence*, 15:143–178, 1980.

[5] James C. Allen. *Natural Language Understanding*. Benjamin/Cummings, Redwood City, USA, 2nd edition, 1995.

[6] James F. Allen, Lenhart K. Shubert, George Ferguson, Peter Heeman, Chung Hee Hwang, Tsuneaki Kato, Marc Light, Nathaniel G. Martin, Bradford W. Miller, Massimo Poesio, and David R. Traum. The TRAINS project: A case study in building a conversational planning agent. *Journal of Experimental and Theoretical AI*, 7:7–48, 1995.

[7] Jens Allwood, Joakim Nivre, and Elisabeth Ahlsén. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9(1), 1993.

[8] Nicholas Asher and Alex Lascarides. *Logics of Conversation*. Cambridge University Press, 2003.

[9] Harald Aust, Martin Oerder, Frank Seide, and Volker Steinbiss. The philips automatic train timetable information system. *Speech Communication*, 17:249–262, 1995.

[10] Jennifer Chu-Carroll and Michael K. Brown. An evidential model for tracking initiative in collaborative dialogue interactions. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics (ACL97)*, pages 262–270, Madrid, Spain, 1997.

[11] Herbert H. Clark. *Using Language*. Cambridge University Press, Cambridge, 1996.

[12] Paul C. Constantinides, Scott Hansma, and Chris Tchouand Alexander I. Rudnicky. A schema based approach to dialog control. In *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP1998)*, Sydney, Australia, December 1998.

[13] Raquel Fernández and Jonathan Ginzburg. Non-sentential utterances in dialogue: A corpus-based study. In Kristiina Jokinen and Susan McRoy, editors, *Proceedings of the Third SIGdial Workshop on Discourse and Dialogue*, pages 15–26, Philadelphia, USA, July 2002. ACL Special Interest Group on Dialog.

[14] Jonathan Ginzburg. Dynamics and the semantics of dialogue. In J. Seligman, editor, *Language, Logic and Computation: The 1994 Moraga Proceedings*. CSLI Press, 1996.

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

[16] Peter A. Heeman and James F. Allen. Speech repairs, intonational phrases and discourse markers: Modeling speakers' utternaces in spoken dialogue. *Computational Linguistics*, 25(4):527–571, 1999.

[17] Stephan Koch, Uwe Küssner, Manfred Stede, and Dan Tidhar. Contextual reasoning in speech-to-speech translation. In *Proceedings of 2nd International Conference on Natural Language Processing (NLP2000)*, Springer Lecture Notes in Artificial Intelligence, 2000.

[18] Willem J. Levelt. Monitoring and self-repair in speech. *Cognition*, 14(4):41–104, 1983.

[19] Stephen C. Levinson. *Pragmatics*. Cambridge Textbooks in Linguistics. Cambridge University Press, 1983.

[20] Diane J. Litman and James F. Allen. A plan recognition model for subdialogues in conversations. *Cognitive Science*, 11:163–200, 1987.

[21] William C. Mann and Sandra A. Thompson. Rhetorical structure theory: A theory of text organization. In Livia Polanyi, editor, *The Structure of Discourse*. Ablex Publishing Corporation, Norwood, N.J., 1987.

[22] Michael F. McTear. *Spoken Dialogue Technology*. Springer Verlag, London, Berlin, 2004.

[23] Matthew Purver. *The Theory and Use of Clarification Requests in Dialogue*. PhD thesis, King's College, Unversity of London, London, UK, August 2004.

[24] H. Sacks, E. A. Schegloff, and G. A. Jefferson. A simplest systematic for the organization of turn-taking in conversation. *Language*, 50:735–996, 1974.

[25] Emanuel A. Schegloff. Sequencing in conversational openings. In J.J. Gumperz and D. H. Hymes, editors, *Directions in Sociolinguistics*, pages 346–380. Holt, Rhinehart & Winston, New York, 1972.

[26] Emanuel A. Schegloff, Gail Jefferson, and Harvey Sacks. The preference for self-correction in the organisation of repair in conversation. *Language*, 53(2):361–382, 1977.

[27] David Schlangen. Causes and strategies for requesting clarification in dialogue. In *Proceedings of the 5th Workshop of the ACL SIG on Discourse and Dialogue*, Boston, USA, April 2004.

[28] David Schlangen and Alex Lascarides. Resolving fragments using discourse information. In *Proceedings of EDILOG 2002*, pages 161–168, Edinburgh, September 2002.

[29] J. Searle. *Speech Acts*. CUP, 1967.

[30] Elizabeth E. Shriberg. *Preliminaries to a theory of speech disfluencies*. PhD thesis, University of California at Berkeley, Berkeley, USA, 1994.

[31] Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol van Ess-Dykema, and Marie Meeter. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3), 2000.

[32] Kristinn R. Thórisson. Natural turn-taking needs no manual: computational theory and model, from perception to action. In Björn Granström, David House, and Inger Karlsson, editors, *Multimodality in Language and Speech Systems*, pages 173–207. Kluwer, Dordrecht, The Netherlands, 2002.

[33] David Traum and Staffan Larsson. The information state approach to dialogue management. In Ronnie Smith and Jan van Kuppevelt, editors, *Current and New Directions in Discourse and Dialogue*, pages 325–353. Kluwer, Dordrecht, The Netherlands, 2003.