Towards an Incremental Theory of Task-Oriented Spoken Human-Computer Dialogue

Niels Ole Bernsen, Hans Dybkjær, Laila Dybkjær

Centre for Cognitive Science (CCS), Roskilde University

PO Box 260, DK-4000 Roskilde, Denmark

phone: +45 46 75 77 11; fax: +45 46 75 45 02

emails: nob@cog.ruc.dk, dybkjaer@ruc.dk, laila@cog.ruc.dk

1. On the Need for a Task-Oriented Dialogue Theory

A new generation of task-oriented spoken language dialogue systems (SLDSs) using mixed initiative spontaneous language is becoming ready for use in over-the-telephone applications and elsewhere [Peckham 1993]. This raises the need for an integrated, application-oriented theory which can support the design of increasingly sophisticated, but still restricted spoken natural language dialogues that are usable and reasonable seen from the user's point of view and which can be managed by the system. The theory should have at least four related aims: (1) To optimise the *methodologies* for developing SLDSs in order to reduce resources and cost [Fraser and Gilbert 1991] [Bernsen et al. 1994a]. (2) To support the *functional* aspects of SLDSs by identifying the concepts and distinctions which should be considered when specifying the system and its architecture. (3) To support practical systems design by identifying the constraints which should be observed in order to optimise system *usability* [Bernsen 1993a,b, Bernsen et al. 1994b]. (4) To address SLDS evaluation, benchmarking and standards [Simpson and Fraser 1993]. This paper deals with the second, functional part of such a theory.

The background of the paper is our dialogue model development using the Wizard of Oz method and a corpus of human-human dialogues in the task domain [Dybkjær and Dybkjær 1993], implementation [Dybkjær and Dybkjær 1994] and ongoing testing of a prototype, P1, of a real-time, 500 words vocabulary, speaker-independent SLDS in the domain of Danish domestic airline ticket reservation over the telephone. P1 has been designed as a first prototype of a realistic system. Our partners in the project are the Center for PersonKommunikation, Aalborg University, and the Centre for Language Technology, Copenhagen.

P1 takes as input a speech signal which is recognised and passed as a sentence hypothesis to the linguistic analysis module. This module makes a syntactic and semantic analysis of the sentence and represents the result in a set of frame-like structures called semantic objects. The dialogue handling module performs a task-oriented interpretation of the semantic objects received from the linguistic analysis module and takes action according to this input, e.g. through updates or queries to the application database or decisions on the next output to the user. In P1 the output module uses pre-recorded speech rather than language generation and text-to-speech synthesis. The architecture of P1 is described in [Larsen et al. 1993]. A more advanced version, P2, of the first prototype will be developed on the basis of P1. P2 will have superior output functionality based on language generation and speech synthesis. Improved recognition techniques, an improved parser and extended grammars and vocabulary are expected to allow the design of more natural dialogues than in P1.

The process of developing the P2 dialogue model based on that of P1 through motivated extensions of P1's functionality is analogous to developing an incremental theory of SLDS functionality which may support the design of systems in many different domains of application. The sections below constitute first steps towards the latter aim as they cover elements of a task-oriented human-computer dialogue theory which should be taken into account when designing practical applications. Some of the elements are only rudimentarily present, if at all, in P1 but are needed in more advanced systems such as P2. Further elements will no doubt have to be added to the theory in order to support system capabilities beyond those of P2. The application-oriented nature of the theory means that it aims to directly support the specification of data structures and operations needed for implementation. Sects.2 and 3 below deal with the *what* of a dialogue model, i.e. dialogue level decomposition and task structure. Sects. 4 to 9 deal with the *how* of a dialogue model, i.e. how dialogue naturalness could be optimised within current technological constraints using mixed initiative, feedback, predictions, history, user models and meta-communication. Sect.10 provides a concluding discussion.

2. Dialogue Level Decomposition

~~**There is broad agreement in the literature on the number of hierarchical levels into which**~~ task-oriented dialogues should be decomposed for the purpose of adequate description [cf. Bilange and Magadur 1992]. Task-oriented SLDS dialogue may be decomposed into the following three levels of description and analysis each of which is illustrated by examples from P1:

1. *Task level.* A *dialogue task* N consists of one or more tasks which are referred to as *dialogue sub-tasks* relative to N. Tasks may be embedded in, and hence be sub-tasks relative to, other tasks. Task N is realised through realising its dialogue sub-tasks a, b, c, ..., n. The global unfolded dialogue task structure shows all tasks and their embeddings, i.e. which tasks are sub-tasks relative to a given task. The global unfolded dialogue task structure of P1 illustrating all tasks and their embeddings is shown schematically in Fig. 1. Only the structure of the domain-related dialogue is shown, not the meta-communication (cf. Sect. 9).

2. *Turn-taking or utterance level.* In SLDSs task N is realised through user-system turn-taking involving a sequence of *dialogue turns* S (System)1, U (User)1, S2, U2, ..., Sn, Un. A turn consists of a user or system *utterance*. Each turn can at least be characterised by the dialogue act(s) it contains and by whether the speaker (user or system) *has the initiative* or *responds* to an initiative taken by the interlocutor ([Bilange and Magadur 1992], Sect. 4 below). The following example shows the completion of three sub-tasks during six dialogue turns:

S1: How many persons are going to travel?

U1: One.

S2: What is the id-number of the person?

U2: Fifty-seven.

S3: Id-number fifty-seven Jens Hansen. Where does the travel start?

U3: In Aalborg.

3. *Dialogue act level.* An utterance may contain one or more *dialogue acts*. In the example

above, the system's third turn S3 contains two dialogue acts, the first being an assertion (a declarative act stating a fact) which provides echo feedback on the dialogue act in the preceeding user turn U2, the second being a question to the user. Dialogue acts are similar to speech acts [Searle 1969]. Dialogue acts are dynamic semantic entities, i.e. they occur in a specific dialogue task context and are defined in terms of their modification of that context [Bunt 1989].

3. Task Structure

Many tasks, such as the flight ticket reservation task of P1, have a stereotypical structure. A *task stereotype* prescribes which information must be exchanged between the dialogue partners to complete the task and, possibly, in which order this may be done naturally. The dialogue task structure of Fig. 1 expresses the reservation task stereotype.

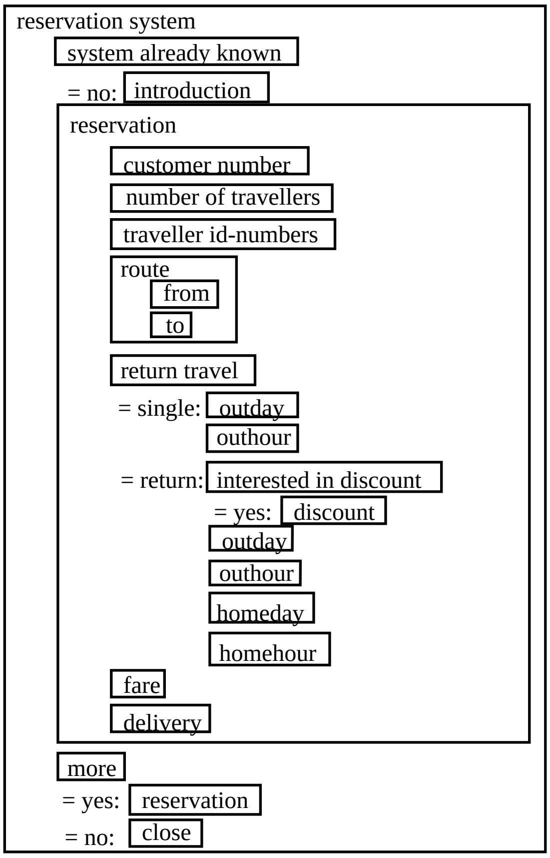

Figure 1: The unfolded domain dialogue task structure for P1. A labelled box indicates a task. If a box A contains another box B then B is a sub-task relative to A. At some points during dialogue the structure to follow depends on the user's answer to the most recent question. In such cases an answer is indicated as '= [answer]:' followed by the tasks to be performed. The dialogue task structure is a cyclic graph with conditional branches.

This structure conforms to the most common structure found in corresponding human-human reservation task dialogues recorded in a travel agency [Dybkjær and Dybkjær 1993].

The work on P1 has suggested that, when shared by user and machine, task stereotypes strongly facilitate dialogue systems design because they allow the computer to direct the dialogue through asking questions of the user without the user feeling this to be a major drawback of the design. Adding observations by [Walker and Whittaker 1990], the following hypothesis emerges: *System-directed dialogue is acceptable to users in cases where there is a directly applicable task stereotype which is assumed by the user to form part of the system's expertise.* Thus, system and user do not even have to share stereotypical task knowledge in cases where (1) the system has sufficient knowledge of the user's situation to embark on the stereotypical task right away and (2) the user has sufficient confidence in the system's task knowledge to let it do so. This would bring a considerable number of task types within the scope of system-directed spoken language dialogue, including many tasks in which the user is novice or apprentice and the system acts as an expert instructing the user on what to do or say.

A travel information task including flight schedules, fares and travel conditions has been specified for P1 using the Wizard of Oz. The task was not implemented, however, and will only be so in P2, for the following reason. Whereas ticket reservation tasks conform to a single, basic stereotype, travel information tasks do not. Knowing that a user wants travel information does not help the system know what to offer and in which order. This means that travel information tasks are not well-suited for system-directed dialogue. The corresponding hypothesis is that *if a task has no stereotypical structure but contains a large number of optional sub-tasks, then the system cannot take and preserve the initiative during dialogue without unacceptable loss of dialogue naturalness.* In such cases, mixed initiative dialogue is necessarily called for to allow an acceptable minimum of naturalness. In the task stereotype case, although always preferable to rigid, system-directed dialogue, mixed initiative dialogue is not strictly required. The class of non-stereotypical tasks seems to be quite large including, i.a., tasks in which users seek information, advice, or support, or otherwise want to selectively benefit from a system's pool of knowledge or expertise.

4. Dialogue Initiative

The interlocutor who controls the dialogue at a certain point has the *initiative* at this point and may decide what to talk about next, such as asking questions which the dialogue partner is expected to answer.

As only the stereotypically structured reservation task has been implemented in P1, it seems acceptable that the system, with two exceptions to be mentioned shortly, takes and preserves the initiative throughout the dialogue. The distinction between user and system initiative, therefore, has not been explicitly represented in the implementation. The system takes and preserves the initiative by concluding all its turns (except when closing the dialogue) by a question to the user. The questions serve to implicitly indicate that initiative belongs to the system rather than the user. Only in meta-communication is the user allowed to take the initiative by using keywords (cf. Sect. 9) which enable the system to immediately identify both the user initiative and the task the user intends to perform.

Even if the described solution may work for stereotypical tasks, keywords-to-be-remembe-red are unnatural and systems for non-stereotypical tasks need user initiative. P2 will have mixed-initiative dialogue for improved naturalness of meta-communication and in order to solve the problem posed by the non-stereotypical information task (cf. Sect. 3). *If* an explicit system representation of who has the initiative throughout a dialogue will be needed in order to achieve those aims, one way for the system to establish who has or takes the initiative, might be to use control rules based on dialogue context and a simple taxonomy of user dialogue acts [Walker and Whittaker 1990].

The correlated distinctions between stereotypical tasks/system initiative and unstructured tasks/mixed initiative dialogue provides a rough guideline for determining where the emphasis should lie given a certain type of task to be performed interactively between user and system. In fact, there seems to be a continuum between full system control through use of questions, declarative statements or commands, and mixed initiative dialogue in which the system only assumes control when this is natural. Even SLDSs for stereotypical tasks need some measure of mixed initiative dialogue to be fully natural [Seneff et al. 1991, Peckham 1993]. And sys–tems performing non-stereotypical tasks, such as large numbers of unrelated sub-tasks, are often able to go into system-directed mode once a stereotypically structured sub-task which the user wants performed has been identified [Guyomard et al. 1991].

5. System Feedback

The provision of sufficient feedback to users on their interactions with the system is particularly crucial in speaker-independent SLDSs because of the frequent occurrence of misunderstandings of user input. The user needs to know whether or not a task has been successfully completed and hence whether repair or clarification is needed.

P1 provides *continuous feedback* on the user commitments made during a task. When the system decides that it has sufficient information to complete a sub-task, the user receives feedback on that information. Users who accept the feedback information do not have to reconfirm their commitment as the system will carry on with the next sub-task in the same utterance. Two such cases of feedback can be seen in the example dialogue in Sect. 2 above: In S2, the term ‘person’ confirms that only one person will be travelling. In S3, the Id-number provided by the user is echoed and the name of the person added for extra confirmation. This type of *echo* or *masked echo* feedback is obviously more parsimonious than, and hence preferable to, *explicit* feedback which requires the system to repeat what the user just said with an added request for confirmation from the user. An even more sophisticated solution may be to use acoustic scores, or acoustic scores combined with perplexity as a basis for determining which type of feedback to give to the user in a particular case, as proposed by [Bilange and Magadur 1992]. If the score drops below a certain threshold indicating considerable uncertainty about the input, explicit feedback might be offered. If the user does not accept the feedback information, meta-communication is needed (cf. Sect. 9).

In addition to continuous feedback, P1 offers *summarising feedback.* On closing the reservation task, the system summarises the information provided by the user. Summarising feedback provides the user with an overview of the commitments made and thus has a role which is distinctly different from that of continuous feedback. Users should be able to initiate meta-communication in cases where the summarised commitments are no longer viewed as satisfactory.

6. Predictions and System Focus

Predictions are expectations as to what the user will say next and help identifying the sub-vocabulary and sub-grammars to be used by the recogniser. Predictions constrain the search space and express the sub-tasks which the user is expected to address in the next utterance. If the user chooses to address other sub-tasks, system understanding will fail unless some prediction-relaxation strategy has been adopted. The more stereotypical structure a task has, the easier it is to make good predictions provided the user is cooperative. One key reason why practical mixed initiative systems are hard to realise is that they make user input prediction more difficult, especially in non-stereotypical tasks [Guyomard et al. 1991]. In mixed initiative dialogue in general, and in non-stereotypical task dialogue in particular, the first challenge the system faces on receiving a user utterance, is to identify the sub-task the user intends to perform.

Predictions are based on the set of sub-tasks currently in *system focus.* The set of sub-tasks in system focus are the tasks which the user is allowed to refer to in the next utterance. A useful heuristics for stereotypical task systems seems to be that the set of sub-tasks in system focus always include the preceeding sub-task (if any), the current sub-task, the possible succeeding sub-task(s) according to the default dialogue task structure and the meta-communicative tasks which might be initiated by the user. Ideally, the system focus should correspond to the *common dialogue focus* shared by the interlocutors. The heuristics just mentioned should make the correpondence achievable in many types of task-oriented dialogue based on task stereotypes, provided that the needed prediction sets are technologically feasible. In such cases, the heuristics may ensure correspondence between system focus and the set of sub-tasks which the user will find it natural to address at a given point during dialogue. In general, of course, the more overlap there is between system focus and user focus, the more likely it is that the dialogue will proceed smoothly. This field is one in which practical systems design expects to benefit from basic research on discourse.

In P1, the dialogue handler predicts the next possible user utterances and tells the speech recogniser and the parser to download the relevant sub-vocabulary and sub-grammars. To obtain both real-time performance and acceptable recognition accuracy it has been necessary to restrict sub-vocabularies to contain at most 100 words [Dybkjær et al. 1993]. The system's predictions include the current sub-task and the meta-communicative possibilities of the user saying 'correct' or 'repeat'. In some cases P1's predictions include more than the current sub-task. For instance, when the system expects an arrival airport, the departure airport is also included in its predictions and may therefore be provided by the user in the same turn as the arrival airport.

Information on the sub-tasks in system focus is hardwired in P1. For each point in the dialogue structure it has been decided which sub-grammars should be active and how the system's utterances should be expressed. The decision on sub-grammars depends on the number of active words required. This approach will not work for mixed initiative dialogue where the user has the opportunity to change task context (or topic) by taking the initiative. When part of the initiative is left to the user, deviations from the default domain task structure may be expected to occur from time to time and in such situations the system has to be able to determine the set of sub-tasks in system focus at run-time. Mixed initiative dialogue therefore requires a dynamically determined set of sub-tasks in system focus.

7. Dialogue History

A *dialogue history* is a log of information which has been exchanged so far in the dialogue. We distinguish between four different kinds of dialogue history each of which has its own specific purpose. Further distinctions among dialogue histories are likely to be needed at some stage [cf. Eckert and McGlashan 1993]. Firstly, the *linguistic dialogue history* logs the surface language of the exchanges (i.e. the exact wording) and the order in which it occurred. Linguistic dialogue history is primarily used to support the resolution of anaphora and ellipses and has to do its work before producing semantic frames for the dialogue handler. Therefore linguistic dialogue history has a closer relation to the linguistic module than to the dialogue model. It is an open question if SLDSs will ever need access to the entire linguistic dialogue history or whether a window of, say, the four most recent user-system turns is sufficient. P1 has no linguistic dialogue history because it only accepts an average user utterance length of max 4 words. With P2's longer user utterances and increased user initiative, a linguistic dialogue history will be necessary to allow, i.a, anaphora resolution.

User input surface language is not needed for dialogue handling, only input order and semantics. We call a history which records the order of dialogue acts and their semantic contents a *dialogue act history*. In P1 the dialogue act history logs

1. an identification of the previous sub-task in order to be able to make corrections on user request;
2. the logical contents of the latest system question. It is, e.g., important to the interpretation of a yes/no answer from the user to know how the question was phrased. It makes a difference if the question was "Is it a one-way travel?" or "Is it a return travel?";
3. the semantic contents of the user's latest utterance.

The dialogue act history is used for correcting the most recent user input. Corrections to information exchanged prior to the most recent user input cannot be made in P1. A larger dialogue act history would probably not help in this case. As in human-human dialogue, the most convenient solution for the user will be to explicitly indicate the piece of information to be corrected. This requires i.a. a task record (see below), maintenance of inter-dependencies between task values and an implementational strategy for revisiting earlier parts of the dialogue structure.

The third kind of dialogue history is the *task record* which logs task-relevant information that has been exchanged during a dialogue, either all of it or that coming from the user or the system, depending on the application. All task-oriented dialogue systems would seem to need a task record because they have to keep track of task progress during dialogue. However, a task record does not keep track of the order in which information has been exchanged and ignores insignificant exchanges relative to the task. The task record also logs which tasks are pending and which ones have been completed. The system may have to suspend the current task if it discovers that it needs some value in order to proceed, which can only be obtained by performing a task which is prior in terms of the task structure. For instance, to determine whether a certain departure hour is acceptable it is necessary to know the date of departure.

In P1 all values obtained from the user concerning the reservation and the extent to which the values have been checked by the system is recorded. Pending sub-tasks are not allowed.

The fourth kind of dialogue history is the *user model record*. This record updates a model of how well the user performs during the dialogue and may be used to influence the way the system addresses the user. P1 does not have a user model record. The next section discusses user modelling in more detail.

8. User Modelling

In human-human dialogue, a participant is normally prepared to change the way the dialogue is being conducted in response to special needs of the interlocutor. During dialogue each participant builds a model of the interlocutor to guide adaptation of dialogue behaviour (cf. the user model record, Sect. 7). In other cases, a participant already has a model of the interlocutor prior to the dialogue, upon which to base dialogue behaviour. The participant knows, for instance, that the interlocutor is a domain expert who only needs update information. A reservation system might do the same by, e.g., using the user's previous reservation-making record as a guide to how to handle the dialogue - or by simply asking the user.

P1 incorporates a small amount of user modelling. In the dialogue opening task phase, the user is asked: "Do you know this system?" (cf. Fig. 1). If the answer is "No", the user is presented with an introduction on how to use the system. If the answer is "Yes", the introduction is by-passed. In P2, we will try to extend system adaptivity by experimenting with a user model record which helps the system determine how to address the user, i.e. whether, for instance, increased use of spelling requests, explicit yes/no questions or multiple choice questions might be helpful to allow the dialogue to succeed. Otherwise, the sky is the limit in how adaptive user models may be created and used in future generations of SLDSs.

9. Meta-Communication

Today's SLDSs require that users provide cooperative utterances so that it is possible to make valid predictions [Bilange 1991, Eckert and McGlashan 1993]. Cooperative utterances are utterances which a user has a right to expect the system to be able to understand. It is up to the system to inform users on the system's understanding capabilities and limitations. Cooperative utterances must conform to this information. However, even when users are cooperative the system may fail to understand, or misunderstand, them. Meta-communication serves as a means of resolving misunderstandings and lacks in understanding between the dialogue partners during dialogue. In current SLDSs, meta-communication for *dialogue repair* is essential because of the sub-optimal quality of the systems' recognition of spontaneous spoken language. Similarly, meta-communication for *dialogue clarification* in order to resolve cases of ambiguous or incomplete information, is common in human-human dialogue and the ability to perform clarification dialogues is generally needed in SLDSs. We shall look at dialogue repair in what follows.

If understanding failure is due to difficulties in recognising a user's pronunciation of certain words, a first reaction could be to ask the user to repeat the utterance. This is the least possible step in the direction of trying to repair an understanding problem. However, if understanding failure is due to, e.g., an overly complicated utterance, simple repetition will not help. In this case it is necessary to make the user express the information more simply. As it is probably impossible for the system to always detect exactly why understanding has failed, a general method for repairing system understanding problems is needed. System-prompted graceful degradation appears to be a promising method (for a combination of graceful degradation and feedback, see [Heisterkamp 1993]).

When using graceful degradation, the system will explicitly ask the user to provide the missing information in increasingly simple terms. The degradation in *user input level* will continue until either the system has understood the input or no further degradation is possible. In P2, distinction will be made between the following five different, system-prompted user input levels, roughly listed in the order of increasing input complexity: (1) spelling question,

(2) yes/no question, (3) multiple choice question, (4) focused question and (5) open, mixed initiative. It is not always a solution to degrade to the level immediately below the current one. For instance, when asking for an arrival airport it would not make sense to use a multiple choice question if there were, e.g., ten possible destinations. In this case the next relevant level would be to ask the user to spell. So the problem is how to decide the next relevant level. This may be done as follows: For each piece of information to be obtained, all five levels are indicated together with a grammar telling how to ask for that information. If it does not make sense to use a certain level, no grammar is indicated, and if it only sometimes makes sense the grammar is conditioned. When the system has understood the user the dialogue returns to the user input level used immediately before degradation. A tentative method for carrying out graceful degradation when system understanding fails, involves the following three steps:

1. Initialisation: If the system does not yet have the initiative it takes the initiative and asks the user a question concerning what it believes to be the topic of the user's utterance. The topic may be determined on the basis of the system focus set. If understanding fails again the system proceeds to step 2, otherwise degradation stops. For instance, the system believes that the user wants to make a reservation but has no information on the reservation yet, and therefore asks "Where does the travel start?".

2. Either repeat or do explicitness iteration: In explicitness iteration, the system makes explicit to the user what was implicit in its original question. If understanding still fails, the system proceeds to step 3, otherwise degradation stops. The question "Where does the travel start?" can be repeated but hardly made more explicit. An example of explicitness iteration is when the system stresses that the user's answer should mention one out of three options offered by the system.

3. Level iteration: The system asks an equivalent question which can be answered in a different and simpler way, i.e. degrades to the next relevant level and then proceeds to step 2 if understanding still fails, otherwise degradation stops. When no lower level exists a bottom stop condition is activated, such as asking the user to address a human travel agent. In the "Where does the travel start?" case, for instance, the system asks the user to spell the name of

the departure airport.

P1 does not have the spelling level and only asks a single open question ("Do you want more?"). P1 initiates repair meta-communication by telling the user that it did not understand what was said or, in case the user signals, using the *correct* command, that the system has mis–understood something, by repeating its penultimate question.

We would like to end this section by pointing out that, in practical applications, the term 'meta-communication' must be taken in a rather wide sense. Our P1 work suggests that, in addition to repair and clarification functionality, the following functions will be needed in practice:

- a 'wait' function for use when the user needs time, e.g. to think or to talk to somebody;
- a 'dialogue help' function for use when the user needs help from the system to get on with the dialogue. Actions such as unexpected, confusing or irrelevant answers or repeated use of the repeat function may be taken to indicate a need for help;
- a 'restart' function for use when the user needs to start all over again, e.g. because too many things have gone wrong during the dialogue.

10. Conclusion

As the development of SLDSs moves from science towards craftsmanship, and from the laboratory towards commercial applications the need arises to address, in a systematic and integrated fashion, the different aspects of the design process which must be mastered to build usable systems efficiently. We have identified four such aspects: methodology, functionality, usability and evaluation/standards. Steps towards an incremental theory of SLDS functionality have been presented, amounting to a 'toolbox' of functionalities which system designers might consider in order to select those which are appropriate for their particular application. Given the practical aim of the paper, we have had little to say about future research aspects. It is quite clear, however, that SLDS theory still needs a lot of medium-term and long-term research. Systems are only now becoming able to perform anything resembling real mixed initiative dialogue. We believe that the more integrated, SLDS theory becomes, the easier it will be to get the priorities right for the research which is yet to be done.

References

[Bernsen 1993a] Bernsen, N.O.: The structure of the design space. In Byerley, P.F., Barnard, P.J. and May, J. (Eds.): *Computers, Communication and Usability: Design issues, research and methods for integrated services.* Amsterdam, North-Holland, 221-244.

[Bernsen 1993b] Bernsen, N.O.: Types of User Problems in Design. A Study of Knowledge Acquisition Using the Wizard of Oz. *Esprit Basic Research project AMODEUS Working Paper* UM/WP 14, 1993. In Deliverable D2: *Extending the User Modelling Techniques.*

[Bernsen et al. 1994a] Bernsen, N.O., Dybkjær, H. and Dybkjær, L.: Wizard of Oz prototyping: How and when? *CCI Working Papers in Cognitive Science and HCI*, WPCS-94-1.

[Bernsen et al. 1994b] Bernsen, N.O., Dybkjær, L. and Dybkjær, H.: Task-Oriented Spoken Human-Computer Dialogue. *Report 6a, Spoken Language Dialogue Systems, CPK Aalborg University, CCS Roskilde University, CST Copenhagen.*

[Bilange 1991] Bilange, E.: A Task Independent Oral Dialogue Model. In *Proceedings of the 5th EACL*, Berlin, April, 83-88.

[Bilange and Magadur 1992] Bilange, E. and Magadur, J.-Y.: A Robust Approach for Handling Oral Dialogues. *Actes de COLING-92*, Nantes, August, 799-805.

[Bunt 1989] Bunt, H.C.: Information dialogue as communicative action in relation to partner modelling and information processing. In Bouwhuis, D., Taylor, M. and Néel, F. (Eds.): *The Structure of Multimodal Dialogues Including Voice.* Amsterdam: North-Holland, 1-19.

[Dybkjær and Dybkjær 1993] Dybkjær, L. and Dybkjær, H.: Wizard of Oz Experiments in the Development of a Dialogue Model for P1. *Report 3, Spoken Language Dialogue Systems, STC Aalborg University, CCS Roskilde University, CST Copenhagen.*

[Dybkjær et al. 1993] Dybkjær, H., Bernsen, N.O. and Dybkjær, L.: Wizard-of-Oz and the Trade-off between Naturalness and Recogniser Constraints. In *Proceedings of Eurospeech '93*, Berlin 21-23 September, pp. 947-950.

[Dybkjær and Dybkjær 1994] Dybkjær, H. and Dybkjær, L.: Representation and

Implementation of Spoken Dialogues. *Report 6b, Spoken Language Dialogue Systems, CPK Aalborg University, CCS Roskilde University, CST Copenhagen.*

[Eckert and McGlashan 1993] Eckert, W. and McGlashan, S.: Managing Spoken Dialogues for Information Services. In *Proceedings of Eurospeech '93*, Berlin 21-23 September, pp. 1653-1656.

[Fraser and Gilbert 1991] Fraser, N.M. and Gilbert, G.N.: Simulating Speech Systems. *Computer Speech and Language*, no.5, pp.81-99, 1991.

[Guyomard et al. 1991] Guyomard, M., Siroux, J. and Cozannet, A.: The Role of Dialogue in Speech Recognition. The Case of the Yellow Pages System. *Proceedings of Eurospeech '91*, Genova, Italy, September, 1051-1054.

[Heisterkamp 1993] Heisterkamp, P.: Ambiguity and Uncertainty in Spoken Dialogue. In *Proceedings of Eurospeech '93*, Berlin 21-23 September, pp. 1657-1660.

[Larsen et al. 1993] Larsen, L.B., Brøndsted, T., Dybkjær, H., Dybkjær, L. and Music, B.: Overall Specification and Architecture of P1. *Report 2, Spoken Language Dialogue Systems, STC Aalborg University, CCS Roskilde University, CST Copenhagen*.

[Peckham 1993] Peckham, J.: A New Generation of Spoken Dialogue Systems: Results and Lessons from the SUNDIAL Project. In *Proceedings of Eurospeech '93*, Berlin 21-23 September, pp. 33-40.

[Searle 1969] Searle, J.R.: *Speech Acts*. Cambridge: Cambridge University Press.

[Seneff et al. 1991] Seneff, E., Hirschman, L. and Zue, V.W.: Interactive Problem Solving and Dialogue in the ATIS Domain. *Proceedings of the Pacific Grove Workshop,* CA, 354-359, February.

[Simpson and Fraser 1993] Simpson, A. and Fraser, N.M.: Black Box and Glass Box Evaluation of the SUNDIAL System. In *Proceedings of Eurospeech '93*, Berlin 21-23 September, pp. 1423-1426.

[Walker and Whittaker 1990] Walker, M. and Whittaker, S.: Mixed Initiative in Dialogue: An Investigation into Discourse Segmentation. *Proceedings of the ACL*, 70-79.