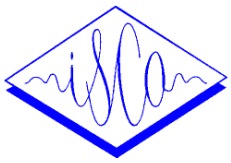

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

A DEDICATED TASK-ORIENTED DIALOGUE THEORY IN SUPPORT OF SPOKEN LANGUAGE DIALOGUE SYSTEMS DESIGN

Niels Ole Bernsen, Laila Dybkjær and Hans 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, laila@cog.ruc.dk, dybkjaer@ruc.dk

ABSTRACT

This paper presents steps towards an incremental dialogue theory in support of functional design of successive generations of spoken language dialogue systems. Dialogue functionality theory departs from a simple task taxonomy and develops a systematic set of concepts or dialogue elements and implementation strategies important to dialogue management. Increasingly complex tasks require the introduction of an increasing number of dialogue elements to ensure acceptable user-system interaction.

1. INTRODUCTION

As spoken language dialogue systems (SLDSs) technology is taking off in the market place and the next generations of SLDSs are being prototyped, a need arises for theory which may adequately support the development of such more and more sophisticated but still restricted SLDSs. A complete dialogue theory would support efficient SLDS development from initial requirements capture through to the testing phase. It would include support of methodology optimisation for dialogue model development and implementation, appropriate functionality design, usability optimisation and SLDS evaluation [1, 8].

This paper presents steps towards a dialogue theory in support of functionality design. Dialogue functionality theory departs from a simple task taxonomy and develops a systematic set of concepts or *dialogue elements* and implementation strategies important to dialogue management. The theory draws upon results from human-human dialogue theories when needed and takes the form of an incremental task-oriented dialogue theory which attempts to anticipate the problems to be addressed in developing successive system generations. Incrementality should ensure that new elements may easily be added without the rest of the theory having to be revised. The theory to be presented primarily draws on lessons from the development of dialogue models for SLDS prototypes in the Danish Dialogue project.

The Dialogue project began in 1991 and involves an effort of 28 man/years by the Center for Person-Kommunikation (CPK), Aalborg University, the Centre for Language Technology (CST), Copenhagen, and the Centre for Cognitive Science (CCS), Roskilde University. The first prototype, P1, is currently being tested. A dialogue model in the domain of Danish domestic airline ticket reservation and travel information was developed through Wizard of Oz simulations. This involved a series of difficult adjustments of natural spoken dialogue to meet technological constraints [5]. On discovering that information tasks impose requirements different from reservation tasks, only the reservation part of the system was implemented [6]. The testing of P1 serves as a basis for a more advanced system, P2, which is now being specified. P2 will differ from P1 by, i.a., leaving more initiative to the user and allowing longer and more complex user utterances.

Sect. 2 below addresses the concept of task which is the basic element in dialogue modelling, and how increasing task complexity leads to increased dialogue complexity. Sect. 3 discusses how increased dialogue complexity requires an increasing number of dialogue elements. Sect. 4 concludes the paper.

2. A TASK TAXONOMY

For the purpose of this paper, SLDSs are defined as requiring (input) speech understanding, thus excluding, e.g., voice response systems and 'speech typewriters'.

In broad agreement with the literature, e.g. [3], task-oriented dialogue may be decomposed into three hierarchical levels.

A *task* to be performed interactively between user and system consists of one or more *sub-tasks*. Tasks may be embedded in, and hence be sub-tasks relative to, other tasks. A task is done by performing its sub-tasks. In SLDSs every task involves at least one *dialogue turn* and tasks often consist in a sequence of turns. A turn is a user or system utterance. An utterance may contain one or more *dialogue acts* such as, e.g., assertions or questions.

Based on task size and structure, tasks may be roughly divided into three types that require increasingly complex dialogue handling to support the functionality needed. Fig. 1 presents this division and lists the demands imposed on dialogue type as well as the minimum demands on the technology needed to handle increasingly complex task types *in a way which is acceptable to users* [7].

Current SLDSs are becoming able to acceptably handle tasks belonging to the first two columns of Fig. 1. However, it is desirable to be able to handle the class of tasks belonging to the third column of Fig. 1. This requires significant improvements of SLDS technologies. Each time the technology has improved to allow progression from one column to the next, this not only means that a new class of tasks can be managed but also that the management of less complex task types can be improved beyond what is minimally acceptable to users.

Tasks which may be acceptably managed by system-directed dialogue have a stereotypical structure that prescribes which information needs to be exchanged between the dialogue partners to complete the task and, possibly, roughly in which order this may be done naturally. The P1 reservation dialogue task structure conforms to the most common structure found in similar human-human reservation task dialogues recorded in a travel agency [9].

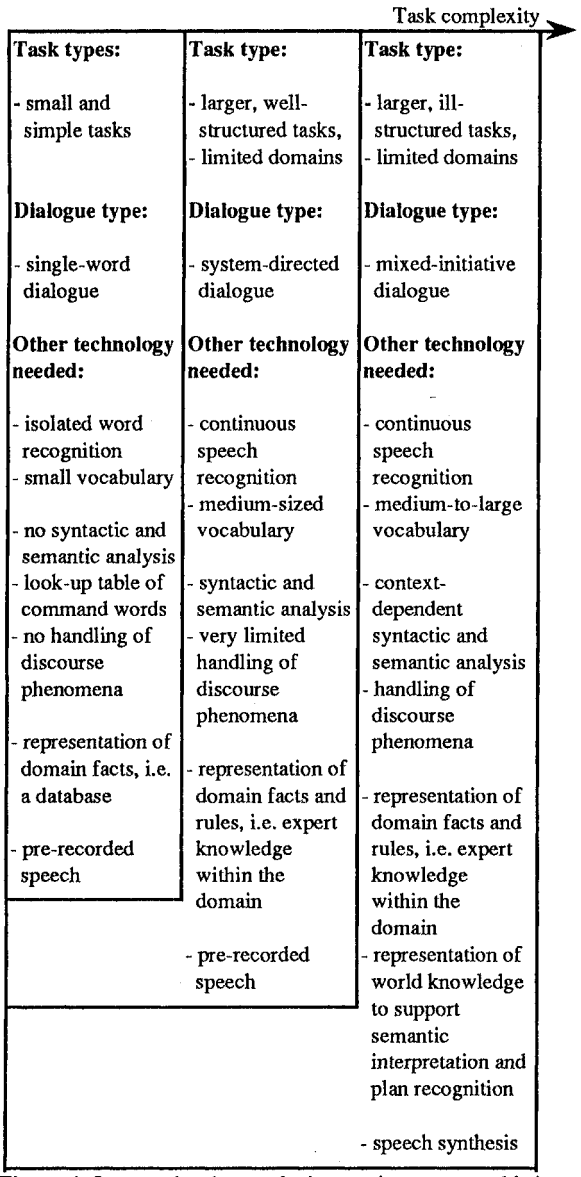

Figure 1. Increased task complexity requires more sophisticated dialogues to maintain an acceptable level of habitability. This again requires more and better technologies.

The class of complex stereotypical tasks is large. It even appears that system and user do not have to *share* stereotypical task knowledge, namely in cases where (1) the system has sufficient knowledge of the user's situation and (2) the user has sufficient confidence in the system's task knowledge [17]. This brings a considerable number of task sub-types within the scope of system-directed spoken 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.

Ill-structured tasks such as the information task specified but not implemented for P1, typically contain a large number of optional sub-tasks and hence are ill-suited for system-directed dialogue. Knowing, e.g., that a user wants travel information does not help the system know what to offer and in which order. In such cases, an amount of mixed initiative dialogue is necessary to allow an acceptable minimum of naturalness. The class of non-stereotypical tasks seems to be 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.

3. DIALOGUE ELEMENTS

The tasks which can be managed in dialogues based on single-word user utterances are small and simple tasks requiring few dialogue turns. Tasks which can be handled by system-directed dialogue may require many turns but are well-structured. Tasks requiring mixed-initiative dialogue typically contain a large number of sub-tasks each of which may involve several turns. Each sub-task may be well-structured but their inter-relationship is ill-defined and a typical dialogue will include an unpredictable sub-set of sub-tasks.

The number and complexity of the dialogue elements needed in an application depend on the type of dialogue they are to support (cf. Fig. 2). We shall briefly define each of the dialogue elements of Fig. 2 and then explain when an element is needed for a specific dialogue type. For a more detailed discussion of dialogue elements see [1].

3.1 Definition of dialogue elements

The interlocutor who controls the dialogue at a certain point has the *initiative* at this point and decides on what to talk about next.

(Explicit, spoken) *system feedback* is a repetition by the system of the key information provided by the user in previous turn(s).

Predictions are expectations as to what the user will say next, and help identify the sub-vocabulary and sub-grammars to be used by the recogniser.

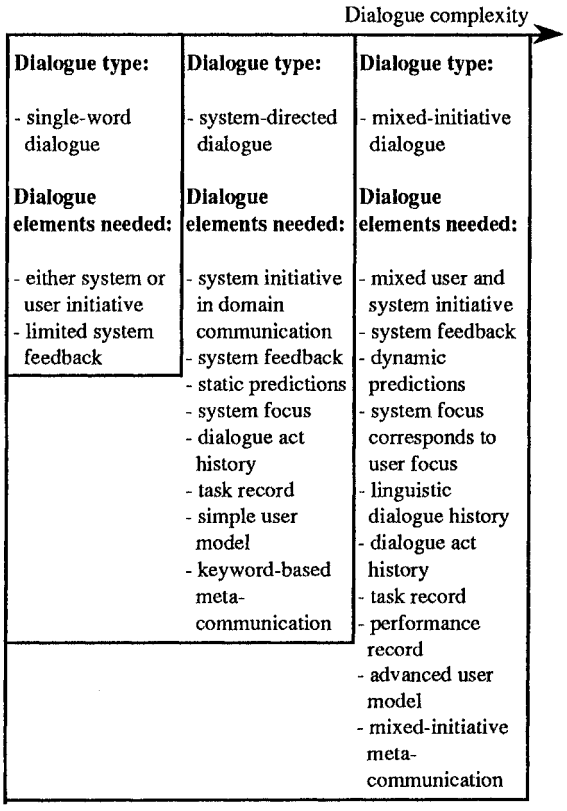

Figure 2. The more sophisticated the dialogue, the larger the demands on dialogue theory and the elements supporting the dialogue model.

The *system focus* is the set of sub-tasks which the user is expected to refer to in the next utterance. Predictions are based on the set of sub-tasks currently in system focus.

A *dialogue history* is a log of the information which has been exchanged in the dialogue. *Linguistic dialogue history* logs the surface language of the exchanges (i.e. the exact wording) and the order in which it occurred. *Dialogue act history* records the order of dialogue acts and their semantic contents. A *task record* logs the 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. A *performance record* updates a model of how well the dialogue with the user proceeds and may be used to influence the way the system addresses the user.

In human-human dialogue each participant builds a model of the interlocutor to guide adaptation of dialogue behaviour. Participants sometimes have a model of the interlocutor prior to the dialogue. SLDSs may need to have or build a *user model* to guide their dialogue behaviour.

Meta-communication is distinct from *domain communication* and serves as a means of resolving misunderstandings and lacks in understanding between the dialogue partners during dialogue.

3.2 Single word dialogue

Systems with dialogues based on single word user utterances are typically either medium/large vocabulary speaker adaptive [4] or small vocabulary speaker independent [14]. *Initiative* is typically either fully with the system or fully with the user. When the user has the initiative, the dialogue structure is usually flat, e.g. there may be a selection of command words for having files opened, saved, printed, etc., cf. [4]. When the system has the initiative, the dialogue structure is typically not entirely flat nor does it have any deep levels [14].

System feedback may be needed. If, e.g., the system has understood that the user wants to transfer money to another account, it may check this with the user before executing. Often, however, explicit feedback is superfluous as the system's next action makes it clear which choice it has recognised. In most cases it will be harmless if the system performs a wrong task such as giving today's weather report instead of tomorrow's.

Especially in small vocabulary systems there is no need for *predictions*. *System focus* will correspond to the possible user tasks and is hardwired in the dialogue structure.

Dialogue history is not needed as reference to previous parts of the dialogue rarely are required. Even explicit *meta-communication* facilities can often be left out. Error correction then happens by uttering the same or a different command.

Many systems in this group are speaker adaptive. However, their dialogue does not otherwise adapt to the user and a more elaborate *user model* is not required.

3.3 System-directed dialogue

What has been termed system-directed dialogue in Fig. 2 typically is much more complex than single word dialogue. Increased complexity is primarily due to task complexity, user utterance length and vocabulary size. The P1 system has system-directed dialogue and will serve as example below.

In system-directed dialogue the system by definition has the *initiative* during domain communication. P1 takes and preserves the initiative by concluding all its turns (except when closing the dialogue) by a question to the user. The questions implicitly indicate that initiative belongs to the system.

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. *Continuous feedback* may vary from an echo or masked echo through to an explicit request for user confirmation. *Echo* and *masked echo* feedback are direct and indirect ways, respectively, of repeating the key information in what the user just said. P1 provides continuous (both echo and masked echo) feedback on the user commitments made. Users who accept the feedback information do not have to reconfirm their commitment as the system will carry on with the next subtask in the same utterance. A sophisticated way of combining echo feedback with explicit requests for user confirmation is to use acoustic scores, or acoustic scores and perplexity, as a basis for determining which type of feedback to offer, as proposed by [3]. If the score drops below a certain threshold indicating considerable uncertainty about the input, feedback may be offered which the user explicitly has to confirm. If the user does not accept the feedback information, meta-communication is needed.

In addition to continuous feedback, P1 offers *summarising feedback* on closing the reservation task to summarise the commitments made. Users should be able to initiate meta-communication in cases where these commitments are unsatisfactory.

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. Information on the sub-tasks in *system focus* is hardwired.

The *dialogue act history* in P1 logs information for use in correcting the most recent user input. Corrections to earlier input cannot be made.

All task-oriented dialogue systems appear to need a *task record* as they have to keep track of task progress during dialogue. P1 has a task record.

P1 incorporates a small amount of *user modelling* in that the system introduction can be avoided by users who already know the system.

P1 has limited *meta-communication* facilities. P1 initiates repair meta-communication by telling the user that it did not understand what was said. Users initiate meta-communication through one of the keywords *correct* and *repeat*. The use of keywords enables the system to simultaneously establish that the user takes the initiative and which task the user intends to perform.

3.4 Mixed-initiative dialogue

P1 has the dialogue elements needed to support effective system-directed dialogue. However, the system clearly might benefit from the addition of mixed-initiative dialogue elements.

To support mixed-initiative dialogue a system must establish and explicitly represent who has or takes the *initiative*. One way to do this might be to use control rules based on dialogue context and a simple taxonomy of user dialogue acts [17]. In practice, there is a continuum between full system control through 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 [15, 16]. And systems performing non-stereotypical tasks, such as large numbers of unrelated sub-tasks, are often able to go into system-directed mode once a stereotypical sub-task which the user wants performed has been identified [11].

We have observed no need for additional *system feedback* compared to what is already in P1 but refinements are possible.

The P1 approach to *predictions* and *system focus* 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 (if any) can be expected and the system must then be able to generate the set of sub-tasks in system focus at run-time. Mixed-initiative dialogue thus requires a dynamically determined set of sub-tasks in system focus. Mixed-initiative systems make user input prediction more difficult, especially in non-stereotypical tasks [11]. 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.

Discourse phenomena such as anaphora occur more frequently in mixed-initiative dialogues. *Linguistic dialogue history* [13] is primarily used to support the resolution of anaphora and ellipses and has to do its work before producing semantic input to the dialogue handler.

Dialogue act history may be used to support correction of input other than the most recent input. As in human-human dialogue, the most convenient solution for the user is to state the piece of information to be corrected. This requires, i.a., maintenance of inter-dependencies between task values, an implementational strategy for revisiting earlier parts of the dialogue structure, and a task record. In addition, the *task record* may also log pending tasks. 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 task structure.

System adaptivity can be extended by introducing a *user model* which helps the system determine how to address the user, whether, e.g., increased use of spelling requests, explicit yes/no questions or multiple choice questions might be helpful to allow the dialogue to succeed.

Although current SLDSs assume a cooperative user [2, 10], *meta-communication* for dialogue repair is essential because spontaneous speech recognition is still fragile. System-prompted *graceful degradation* of user input level is a promising method for dialogue repair [12]. When using graceful degradation, the system takes the initiative and explicitly asks the user to provide the missing information in increasingly simple terms. Degradation continues until either the system has understood the input or no further degradation is possible. When the input has been understood, the dialogue returns to the user input level used immediately before degradation.

On the user side, work on P1 suggests a need for functionality in addition to *correct* and *repeat*, such as asking for *help* and asking the system to *wait*. As walk-up-and-use users cannot be expected to remember large numbers of keywords, improved meta-communication requires mixed-initiative dialogue.

4. CONCLUSION

We have discussed how increasingly complex tasks require the introduction of an increasing number of dialogue elements to ensure acceptable user-system interaction in SLDSs. Steps towards an incremental theory of SLDS functionality have been presented in terms of the dialogue elements needed in increasingly complex SLDS types. As mixed-initiative SLDSs and incorporation of SLDS technologies in multimodal systems are becoming feasible, the functional dialogue theory presented here will evidently need to be augmented in future work.

Acknowledgements. The work described is being funded on a grant from the Danish Government's Informatics Programme whose support is gratefully acknowledged.

REFERENCES

[1] 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*, 1994.

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

[3] Bilange, E. and Magadur, J.-Y.: A Robust Approach for Handling Oral Dialogues. *Actes de COLING-92*, Nantes, August, 799-805, 1991.

[4] Christ, B.: På talefod med PC'en. In *Alt om Data*, vol.3, pp.114-118, March 1992.

[5] 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, 1993.

[6] Dybkjær, H., Bernsen, N.O. og Dybkjær, L.: Dialogue Development and Implementation in the Danish Dialogue Project. To appear in book on *European Speech Projects.*, Springer Verlag, 1994, (in print).

[7] Dybkjær, L.., Bernsen, N.O. og Dybkjær, H.: Roles of Spoken Language in User-System Interaction. In *Human Comfort and Security*, Springer Research Report, 1994 (to appear).

[8] 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*, 1994.

[9] 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*, 1993.

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

[11] 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, 1991.

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

[13] Heisterkamp, P., McGlashan, S. and Youd, N.: Dialogue Semantics for an Oral Dialogue System. In Proceedings of ICSLP '92, Banff 12-16 October, pp.643-646, 1992.

[14] MAX: Reference Card for MAX. ECHO, European Commission Host Organisation, B.P. 2373, L-1023 Luxembourg G.D., 1991.

[15] 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, 1993.

[16] 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, 1991.

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