Hej Lars Bo

Hermed vores ide til at bringe udkastet ned på ca. tre sider. Der er skåret overalt - ikke bare i vores egne afsnit. Det har været nødvendigt med en grundig omgang med grovfilen dels for at bringe længden af teksten ned og dels for at bringe graden af detaljerethed for de enkelte afsnit på et mere ensartet niveau. Check grundigt at jeres egne afsnit har et ensartet overordnet niveau. Detaljerne skal først komme i det endelige papir.

Der er sat referencer ind for vores afsnit. Du kan udskifte med tal, når du har styr på samtlige referencer. Dette vil reducere tekstlængden. Det vil også reducere tekstlængden, hvis du i stedet for blanklinier mellem afsnit bare laver indrykning for hvert nyt afsnit, der ikke følger efter en overskrift.

De fleste underoverskrifter er nu smidt ud. De har været udmærkede som arbejdsoverskrifter, men de bør ikke med i endelige abstract. Teksten er alt for kort til at bære så mange underoverskrifter, og de æder for meget af pladsen.

Hvis det kan nås, læser vi gerne resultatet af din gennemarbejdning af teksten på baggrund af de kommentarer du får (bl.a. disse).

THE DANISH SPOKEN LANGUAGE DIALOGUE PROJECT - A GENERAL OVERVIEW

Anders Baekgaard*, Niels Ole Bernsen**, Tom Brøndsted*, Paul Dalsgaard*, Hans Dybkjær**, Laila Dybkjær**, Jan Kristiansen*, Lars Bo Larsen*, Børge Lindberg*, Bente Maegaard***, Bradley Music***, Lene Offersgaard***, Claus Povlsen***

* = Center for PersonKommunikation, Aalborg University, Denmark

** = Center for Cognitive Science, Roskilde University, Denmark

*** = Center for Language Technology, Copenhagen, Denmark

1. Introduction

This paper provides a general overview of the Danish Spoken Language Dialogue Project. The project started mid 1991 funded by the Danish Government's Informatics Programme and is scheduled to run until early 1995. The project is carried out with an effort of 28 man/years and combines the expertise of its three partners which are the Centre for Cognitive Science (CCS), Roskilde University, the Centre for Language Technology (CST) in Copenhagen, and the Center for PersonKommunikation (CPK), Aalborg University.

The primary goal of the project has been to investigate, through the development of prototypes, the combination of the research areas: Speech technology, dialogue management, natural language processing, human-computer interaction and dialogue handling. It builds upon previous experiences gained by the partners, in i.a. the EUROTRA- and SUNSTAR European projects.

This paper describes the entire development process from choice of application and domain analysis through to implementation and test of the P1 prototype system.

2. Application development

The overall domain chosen for the project was flight travels. Within this domain it was decided to develop an application for Danish domestic flight ticket reservation and travel information.

2.1 Dialogue model

A dialogue model was developed for both these tasks through the use of the experimental and iterative simulation method called Wizard of Oz (WOZ) [Fraser and Gilbert 1991, Dybkjær and Dybkjær 1993, Dybkjær et al. 1993]. Totally seven iterations were performed to achieve an acceptable and feasible dialogue model.

However, during this process it became clear that only a system-directed dialogue would be feasible given the technological and other constraints which had to be satisfied. As confirmed by recordings of customer calls to a travel agency the reservation task is a well-structured task for which the information to be exchanged to achieve the goal can be prescribed. For this reason the reservation task can be handled acceptably through system-directed dialogue. The information task, however, consists of a number of different information subtasks which can be combined and addressed in almost arbitrary order. This makes the information task ill-suited for system-directed dialogue [Bernsen et al. 1994, Dybkjær et al. 1994] and it was decided only to implement the part of the dialogue model covering the reservation task.

2.2 Sublanguage

The WOZ experiments were not only used to develop a dialogue model but also served as a basis for defining a sublanguage for the application. Based on automatically generated frequency lists and concordances of the words in the transcribed WOZ dialogues, sublanguage vocabulary and grammatical coverage were defined.

Grammars were implemented in the APSG-formalism. In the development of the structural APSG-grammar, conformance to of the constraints set by speech technology was done in two ways:

- the generality of the structural analysis was reduced by application of domain-specific selectional restrictions and categories.

- the overall domain-specific grammar was divided into subgrammars, the union set of which is equivalent to the defined linguistic coverage of the system.

2.3 Speech databases and -models

The APSG implementation of the domain sublanguage was in turn used to automatically generate very large sets of sentences (typically above 100.000). The sets were then subject to a selection process, in order to optimise a training corpus for acoustic models. The optimisation was done to obtain as compact a training set as possible, while ensuring that sufficient training tokens were present to train the desired speech sound models. The combined occurrences of words as well as context dependent phonemes (triphones) were taken into account.

A CDHMM was trained for each word in the corpus, using an iterative approach. In order to achieve real-time performance on a DSP-32, the model architecture was limited to single mixture models based on telephone-quality speech.

3. Platform and implemented system modules

The system architecture of P1 is based on the SUNSTAR DDL/ICM platform which was developed in the Esprit SUNSTAR project and which has been further developed in the Dialogue Project. The architecture is modular and based on events. A bus called the Dialogue Communication Manager carries messages between the modules. These may be other programs or hardware and communicate with the bus through drivers.

The core module is the Interpretation and Control Module (ICM). ICM interprets the dialogue model implemented in DDL (Dialogue Description Language). DDL is an experimental language originally intended for primitive dialogues not involving natural language. DDL has been extended in the Dialogue Project to meet the particular needs of the prototype systems. The dialogue model has been implemented by using the DDL-Tool which is a graphical editor and debugger.

3.1 Implemented system modules

P1 has five main modules implemented on the described platform: a speech recogniser, a parser, a dialogue model, a database and an output module.

The acoustic decoding is carried out by a continuous speaker-independent speech recogniser based on CDHMMs and implements the token parsing recognition algorithm. A lattice N-best algorithm has recently been implemented.

The recogniser version used in P1 runs in real time with an active vocabulary of up to 100 words on a single DSP implementation. The acoustic units used by the recogniser may be whole words, subwords (triphones) or both. In P1 whole words are used for acoustic modelling. The active set of grammars is augmented with a set of garbage models and transitions, enabling the recogniser to spot phrases.

Finite state descriptions of the sublanguage model, automatically derived from the APSG implementation are used as constraint networks by the recogniser. The recogniser version used in P1, however, is based on word pair grammars, which enable a faster acoustic decoding than more equivalent finite state approximations.

The dialogue manager dynamically generates grammar sets depending on the present dialogue focus. These grammars are downloaded to the recogniser. In this way the speech recogniser is optimally constrained according to the current dialogue focus.

The parser has been implemented as a module external to the ICM dynamically linked to the rest of system. It is a left-right bottom-up chart parser, optimized by precompiling a left corner dependency tree of the subgrammar (expressed in APSG) into a table. The parser takes as input a sentence hypothesis from the speech recogniser and delivers a semantic interpretation which is used as input to the dialogue model.

Parsing is done robustly: if a complete syntactic analysis of a sentence hypothesis is not found doing bottom-up parsing, a recovery method is used to build a so-called robust state made up of partial syntactic analyses of the hypothesis. The robust state covers all parts of the input and a semantic interpretation is then derived in the usual way from this result.

The dialogue model for the reservation task was implemented in DDL. The implemented dialogue model is expressed as a graph with dynamic computed conditional branches. There are static links between subtasks within the default task template, and dynamic links to previous subtask and global tasks like help. Simple subtasks concerning a single item are modelled by a template structure for system-user exchanges [Dybkjær and Dybkjær 1994].

The database is implemented in C++. It is divided into three main parts. One part contains the interface to the rest of the system and handles messages sent to and from the database, a second part contains rules and constraints, e.g. how to compute the day of week given a date, and a third part contains facts, i.a. departure times [Dybkjær and Dybkjær 1994].

Output is determined by the dialogue model and produced by putting together and replaying pre-recorded words and (parts of) sentences.

4. The running demonstrator

The P1 system has been demonstrated at CRIM/FORWISS and at the ANLP conference in Stuttgart, October 1994, running on a single 486-PC with a DSP-board, a microphone, and a loudspeaker. In Stuttgart, during two full return-ticket demonstrations only one error occured, caused by a misrecognition and caught by the domain check.

We perform a test of P1 with naive users in November 1994. A report on the results from this user test will be included in the full paper. In particular, there will be a comparison with the results from the WOZ experiments.

5. Conclusion and future work

P1 seems to be a state-of-the-art system at the same level as other current spoken language dialogue systems with speaker independent real-time capabilities. As examples of such systems can be mentioned the Philips train timetable information system [Oerder and Aust 1994] which is currently available via the public telephone line, and the systems developed in the SUNDIAL project [Peckham 1993].

Future work will focus on generalising and extending the platform, on multimodality and on mixed-initiative dialogue handling.

The ICM/DDL platform will be generalised to handle multimodal applications. The current ICM concept is able to handle parallel and distributed input/output devices, but there is no support of the integration of input sources nor of the coordination and distribution of simultaneous output to several devices.

The DDL-Tool provides a programming and debugging environment. This environment will be extended to a full-scale program development environment with support of i.a. high-level specifications of dialogue structure, phrases etc., and rapid prototyping, in particular bionic WOZ experiments.

Mixed-initiative dialogue is necessary to allow an acceptable minimum of naturalness in the handling of ill-structured tasks, i.e. tasks for which the information to be exchanged cannot be determined in advance. The class of ill-structured 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.

References

[Bernsen et al. 1994] Niels Ole Bernsen, Laila Dybkjær, and Hans Dybkjær: *A dedicated task-oriented dialogue theory in support of spoken language dialogue systems design*, proceedings of ICSLP'94, Yokohama, pp 875-878.

[Dybkjær et al. 1993] Dybkjær, H., Bernsen, N.O. and Dybkjær, L.: Wizard-of-Oz and the Trade-off between Naturalness and Recogniser Constraints. *Proceedings of EUROSPEECH '93*, Berlin, September 1993, 947-50.

[Dybkjær and Dybkjær 1994] Hans Dybkjær and Laila Dybkjær: *Representation and Implementation of Spoken Dialogue*, report 6b, Spoken Language Dialogue Systems, CPK Ålborg University, CCS Roskilde University, CST Copenhagen, 1994.

[Dybkjær et al. 1994] Laila Dybkjær, Niels Ole Bernsen, and Hans Dybkjær: Different Spoken Language Dialogues for Different Tasks. A Task-Oriented Dialogue Theory. *Human Comfort and Security,* Springer Research Report 1994 (in print).

[Dybkjær and Dybkjær 1993] Laila Dybkjær and Hans Dybkjær: *Wizard of Oz Experiments in the Development of a Dialogue Model for P1*, report 3, Spoken Language Dialogue Systems, CPK Ålborg University, CCS Roskilde University, CST Copenhagen, 1993.

[Fraser and Gilbert 1991] Fraser, N.M. and Gilbert, G.N.: Simulating Speech Systems. *Computer Speech and Language* 5, 1991, 81-99.

[Oerder and Aust 1994] Martin Oerder and Harald Aust: A Realtime Prototype of an Automatic Inquiry System, Proceedings of ICSLP'94, Yokohama 1994, pp 703-706.

[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, 1993, 33-40.