Knowledge acquisition for a constrained speech system using WoZ

Laila Dybkjær & Niels Ole Bernsen & Hans Dybkjær
Centre for Cognitive Informatics (CCI), Roskilde University
PO Box 260, DK-4000 Roskilde, Denmark
emails: laila@ruc.dk, nob@ruc.dk, dybkjaer@ruc.dk

This paper describes the knowledge acquisition phase in a national project[1] aimed at the design of realistic spoken language dialogue system prototypes in the domain of airline ticket reservation and flight information [Dybkjær and Dybkjær, 1993].

The goals of the knowledge acquisition phase were to define a dialogue structure and a sublanguage vocabulary and grammar for subsequent implementation of a first prototype. The development method was the Wizard of Oz simulation technique [Fraser and Gilbert, 1991]. The dialogue model had to satisfy a number of conflicting constraints, most importantly: (1) A maximum user vocabulary of 500 word forms. (2) A maximum user utterance length of 10 words and an average length of 3-4 words. (3) A usable dialogue, including sufficient domain and task coverage, robustness and real-time system performance. (4) A natural form of dialogue and language.

A *usable* system is one which can do the tasks required of it. In principle, it can replace a human operator on those tasks. A *natural* system, on the other hand, is one which allows users to use free and unconstrained spontaneous speech in efficiently achieving their goals. In the development of the first prototype to be described here, the focus was on usability (constraints (1)-(3) above) and on laying the foundations for meeting the naturalness constraint (4) in a second prototype. The real-time requirement of (3) forces the recogniser to handle at most 100 active words at a time, and together with (1) and (2) this obviously pushes the dialogue model towards a rigid system-directed dialogue structure.

Seven iterations of Wizard of Oz experiments were performed involving taped and transcribed dialogues between the wizard and subjects. Voice distorting hardware (equalizer and harmonizer) was only used in the final set of experiments. A wizard's assistant was used in the three last sets of experiments. From iteration 3 onwards, the wizard used a graph structure based on the notion of basic tasks and containing canned phrases in the nodes and contents of possible user answers along the edges. In addition, users were instructed to answer questions briefly and one at a time in order to be understood by the system. Users were given broadly described scenarios the goals of which they had to achieve in dialogue with the system. In the last three iterations 23 subjects performed in all 107 dialogues with 28 different scenarios using a total of 4455 words.

The constraints (1) and (2) above on vocabulary size and maximum and average user utterance length have been met. In the last iteration only 3 user utterances out of 881 contained more than 10 tokens and the average number of tokens per user turn was 1.85. The total number of word types was 165 excluding numbers, weekdays, months, and destinations. Additional inflexions and a complete list of numbers, weekdays, months, and destinations are incorporated in the final sublanguage which includes close to 500 word forms.

In order to evaluate the simulated system's usability and naturalness (3)-(4), users were given a questionaire asking them about their opinion of the system. On average they found the system desirable (62%), efficient (60%), robust (82%), reliable (73%), easy to use (73%), simple (78%), and friendly (82%), but still 81% preferred to talk to a human travel agent! Apart from a general preference for talking to humans this is probably due to the rigid menu-like structure. As for robustness the wizard did not simulate misrecognitions. This may result in lack of robustness in the first prototype. The domain and task coverage was sufficient for the scenarios used and the system would seem adequate for handling the tasks which were found in recordings from a travel agency.

The vocabulary is believed to be usable but its natural limits have not yet been identified. Moreover, subjects tended to model formulations from the scenarios. To improve data reliability, scenarios should be used which only provide an abstract scenario frame and force subjects to be inventive.

The second prototype should demonstrate improved naturalness, including: a less rigid menu structure which allows immediate focused choice; longer average user utterances; well-tested robustness; and an increased amount of information transferred between different tasks and subtasks.

[1]The project is carried out in collaboration with the Speech Technology Centre at Aalborg University (STC) and the Centre for Language Technology at Copenhagen University (CST). We gratefully acknowledge the support of the project by the Danish Government's Informatics Programme.

References

[Dybkjær and Dybkjær, 1993] Laila Dybkjær and Hans Dybkjær. *Wizard of Oz Experiments in the Development of the Dialogue Model for P1.* Report 3, STC, CCI, CST, 1993.

[Fraser and Gilbert, 1991] Norman M. Fraser and G. Nigel Gilbert. Simulating Speech Systems. *Computer Speech and Language*, no. 5, 1991.