DESIGN OF A SPOKEN LANGUAGE DIALOGUE SYSTEM

A Study of the Initial Specification Phase

Niels Ole Bernsen, CCI, Roskilde University and Risø National Laboratory

Summary: The paper presents a study of the initial design specification phase of a spoken natural language dialogue system with particular emphasis on the interaction between the criteria of realism, usability and naturalness, on the one hand, and the criteria of feasibility on the other. The results are generalised into a logic of initial design specification to facilitate comparison with results from other studies of early design processes.

Keywords: Design criteria, realism, usability and naturalness, designer reasoning, early design, spoken language dialogue systems.

1. Introduction

This paper is a study of a four-year, in-house CCI system design process which forms part of a larger national R&D project. The project started in 1991 and the partners are the Speech Technology Centre (STC), Aalborg University, the Centre for Language Technology (CST), Copenhagen University, and CCI. The goal of the project is to develop two experimental human-machine spoken dialogue system prototypes of increasing complexity and performance. Both prototypes should be able to carry out a spoken dialogue with users on specified tasks subject to a number of scientific, technological and other constraints.

In the present paper, the initial specification phase for the first, and smaller, prototype is described and analysed as follows. Section 2 describes the setting-up of the initial design space focusing on the criteria of realism, usability and naturalness. Section 3 describes the interaction between these criteria and the technological requirements and constraints. Section 4 summarises the constraints on, and the tasks of, the subsequent knowledge acquisition phase as derived from initial artifact specification. Section 5 provides an overview of the general logic of initial artifact specification as induced from the design process described in this paper.

In broad terms, the current design process has six, partly iterative, phases:

- initial artifact specification;
- knowledge acquisition;
- detailed architecture specification;
- knowledge representation;
- system implementation;
- system testing.

The term "initial artifact specification" indicates that this is where artifact development begins. However, the specification refinement process will continue throughout the design process and will interact strongly with the other phases of artifact development. Many design decisions made during initial specification may have to be revised later. The initial specification phase is operationally defined as the phase of artifact specification *up to but not including* the

knowledge acquisition phase and the detailed definition of system architecture involving specification of the internal operation of all system modules as well as of the interfaces between the modules. Partly because of the type of artifact being developed, the initial specification phase is fairly well-defined since there are strict limits to the level of detail to which the artifact can be specified prior to knowledge acquisition. In the current design process, in other words, the knowledge acquisition phase is of crucial importance to artifact development beyond a certain point. The knowledge acquisition phase is primarily taking place at CCI and will be the subject of a subsequent paper (Klausen, in preparation). The detailed system architecture specification process is taking place at the time of writing and concurrently with the knowledge acquisition phase. It must be emphasized that the distinction made below between design space delimitation and first design decisions within the delimited design space is analytical, not temporal. In particular, *it is often only when having started to work within an initially vaguely delimited design space that some of the conditions contributing to its firmer delimitation are discovered.* Thus the description provided below represents an analytical approach to initial artifact specification rather than a temporal one.

The first implementations of spoken language dialogue systems were made about 15 years ago. Since then a number of systems have been developed almost all of which are still research systems. The few applications in actual use are quite primitive and typically recognise only isolated words. To be at all possible, today's spoken language dialogue systems require a rather narrow and precisely circumscribed task domain. Furthermore, speech recognition and understanding techniques have still not matured to the point where they can be used in computer systems allowing users to conduct a relatively natural, large-vocabulary conversation pertaining to some task domain. If the vocabulary is large, which in this context means more than a few hundred words, and a relatively complex dialogue and discourse structure is permitted, the system response will take too long and large efforts are required to make the system speaker-independent. If real-time operation and speaker-independence are required, the vocabulary will be small and the dialogue structure simple. Usually, only one-word- or at least very short user utterances will be allowed (see Bernsen et al. 1992). In addition, there are still open research questions about how to improve basic speech recognition capabilities, how to efficiently combine syntactic and semantic processing, how to control the dialogue, how to process various phenomena in discourse, and how to achieve efficient integration within an overall system architecture.

2. Setting up the initial design space: Realism, usability and naturalness

The design task is constrained, firstly, by limitations in time, manpower, available machine power and so on. I will not go further into these *general feasibility constraints* but merely note that the first prototype is to be ready for demonstration after a little more than one year's work by the designer teams. Secondly, the design task is constrained by *scientific and technological feasibility* (see Sect. 3 below). Thirdly, the artifact to be designed is subject to a number of important conditions concerning its *realism, usability and naturalness.* These conditions help shaping (or constraining) the initial design space within which design decisions are to be made. The distinction between realism, usability and naturalness can be viewed as a structured alternative to the standard notion of system "usability" or "habitability".

An important point is that, analytically speaking, realism, usability and naturalness are clearly distinct from *design decisions.* They are *not* design decisions but, rather, conditions which are presupposed by design decisions or, in the QOC (Questions, Options, Criteria) terminology of the Design Rationale (DR) framework, they are general *criteria* to be used throughout the

design process in evaluating different design options prior to the making of design decisions (see, e.g., MacLean et al. 1990, MacLean et al. 1991). The following description of the initial artifact specification phase demonstrates a number of cases in which the criteria of realism, usability and naturalness support the making of early design decisions, sometimes in conjunction with other kinds of constraint on the design process. Clearly, each conclusion drawn from the criteria of realism, usability and naturalness is a design decision just like any other design decision. As such, it is conditional upon both general and technological feasibility. Sometimes, as will become apparent below, this feasibility is *hypothetical* at the time the design decision is being made. Later in the design process, such design decisions may have to be changed simply because the feasibility which was initially assumed turns out not to be there.

Realism

The artifact to be designed should be *realistic*, that is, it should demonstrate the use of the generic technology (i.e., spoken language dialogue systems), in a task domain in which such systems might actually be superior, equivalent or at least acceptably inferior to other known ways of performing similar tasks. Among the preferred *task domains* for spoken language dialogue systems today are time table enquiries for flights and trains. It was decided in the current design project to select one such domain, i.e., that of information on flight travels and reservation of tickets to be obtained over the telephone. Currently there are several artifact development efforts going on internationally in this task domain which has become something of a competitive testing ground for advanced spoken dialogue systems. Needless to say, the available information on related artifact development efforts was carefully studied during the initial design specification phase and constituted an important background for the design decisions made (Bernsen et al. 1992). This background has an influence on all three types of constraints on the design process, i.e., general feasibility, scientific and technological constraints, and realism, usability and naturalness. Moreover, those efforts continue to be monitored during subsequent design phases. The chosen task domain is realistic in the sense that, in all or most countries today, the telephone is already being used extensively for the tasks indicated. In principle, the relevant tasks might be mechanised instead by having users perform long series of telephone keystrokes or through the use of screen, keyboard and mouse. However, the first option implies rapidly decreasing usability as the task-relevant dialogues grow in complexity, and the second option is not (yet) available to most potential users.

It may be added that the chosen task domain differs along at least two dimensions from other major domains of interest to speech recognition and speech understanding technology. Firstly, the task domain is "mono-medium" in the sense that the interface consists entirely of spoken language. There is no mixture of spoken language with written language, graphics or animation. Secondly, the systems to be built will replace humans performing the same tasks. This class of systems seems to be distinct from systems which are being controlled by a human operator through spoken language. In such cases, the human voice is being used for process control and no human operator is being replaced. These two differences may limit the generalisations relevant to HCI to be made from the current design effort.

In summary, the realism of the artifact to be designed comes from the facts that it (a) addresses real and known user needs, (b) is preferable to other technological solutions currently available and (c) is assumed to become tolerably inferior in performance to humans performing the same tasks. It is proposed that *these criteria of realism are valid for all artifacts designed to replace humans on some task or set of tasks.*

Usability

A *usable* spoken language dialogue system should meet at least the following requirements, most of which are self-evident implications of the overall design goal in conjunction with various other assumptions:

- the system should *understand* all or most of the users' utterances in their appropriate task context as evidenced from its responses to users. In cases where the system fails to understand an utterance, it should be able to repair the dialogue through appropriate responses to the user. This feature of the system is often called "robustness". It is self-evident that if a system is not sufficiently robust in this sense it will not be usable;

- the recognised *vocabulary* should be large enough to encompass all or most terms relevant to completing the dialogues necessary for the chosen tasks. If the vocabulary is too limited, then users will have a difficult time getting the system to do what they want;

- the system's *grammar* should be natural to users, i.e., the system should recognise and understand the varieties of syntactic forms users find it natural to use during their dialogue with the system. It is a fact of cognition that it is practically impossible for users to remain within the bounds of an unnaturally restricted syntax during natural dialogue. This point may be less self-evident than others on this list and hence might be overlooked by designers not familiar with the literature on earlier spoken language dialogue systems or with relevant principles of human language processing;

- the system's *semantics* should be appropriate to the chosen vocabulary and grammar. Anyone who could design the linguistic parts of the system would know that;

- the system's handling of *discourse phenomena* should be natural to users during their dialogue with the system. The principle behind this condition is the same as in the case of grammar above. The point about lacking self-evidence made in connection with the grammer holds true here as well;

- the system should *respond to user input* in something not too remote from real time. This is evident from practical considerations even without invoking cognitive theories of attention span;

- the system should preferably do *speaker-independent* recognition of speech as usability in the chosen task domain is seriously affected by the training process which is otherwise needed for the system to adapt to a new user. In other task domains (e.g., voice process control in the cockpit), speaker-independence may be less important;

- the system should preferably do *continuous speech recognition* as constraints on users' sentence pronounciation seriously affect the usability of the system in the task domain. Again, in other task domains this requirement may be less important;

- the system should clearly communicate to users *which tasks* they can accomplish with the system in the chosen task domain and the system should possess the *task domain information* necessary for users to accomplish those tasks. Since the system will be limited in performance, users obviously have to know what it can and cannot do. And when the system announces it capability of doing something, it should be able to do it.

Although this list is (mostly) self-evident and states conditions necessary to usability, there does not seem to exist any procedure for making sure that it is sufficient. Moreover, it is clearly possible for different designers to derive somewhat different sub-sets of the necessary conditions for system usability. Let us take a closer look at how these conditions are derived.

The above usability criteria are *minimal* in the sense that they state conditions necessary for the system to be at all able to replace a human operator in the task domain. They are derived from the overall goal specification of the system to be designed in conjunction with task domain information and information on user needs and users' cognitive capabilities and limitations. An important cue as to how this derivation takes place is the following. In general, when a system is usable, it *can do* the tasks done by the human operator it replaces. This may not be sufficient to ensure usability but it certainly seems to be necessary. In other words, *to determine usability criteria for a system to be designed, one has to identify conditions such that, if they are not met by the designed system then the system cannot do the tasks to be assigned to it.* As long as we do not have any systematic way of identifying such conditions, it will be up to the collective know-how present in the designer team to identify as many usability conditions as possible. There is no way of guaranteeing that the identified set of usability criteria is complete and different designers or designer teams may well identify different sub-sets of the usability criteria for a given artifact. Moreover, as indicated, *since important assumptions about the cognitive capabilities and limitations of users may be involved, the expertise needed for deriving a reasonably complete set of usability criteria during early design is not necessarily present in the designer team* and strong penalties may have to be paid later on if some crucial usability criterion has been overlooked. Some way of testing the cognitive assumptions about users underlying the design of a particular artifact is therefore mandatory.

One possibility of ameliorating the situation just described of unprincipled derivation of usability criteria might be to identify some of the basic structural properties of the design space in which system design takes place (Bernsen, in preparation).

Naturalness

Usability is a very basic standard indeed for the design of interactive computer systems. Arguably, a usable system corresponding to the above usability requirements could be developed by having a dialogue with users which is totally controlled by the system so that the user will only have to respond by "yes" or "no" throughout. However, such a dialogue would not be *natural*, i.e., it would not even remotely correspond to the way in which spoken language dialogues are normally conducted in the task domain. In other words, a natural dialogue requires that users can use language (i.e., vocabulary, grammar, semantics and discourse) to some considerable extent freely during their dialogue with the system. Note that, in the case of “pure” natural language interfaces to computers, we actually do have an objective standard of naturalness. This is certainly not the case with respect to most other types of human-computer innterfaces. On the other hand, a fully natural dialogue cannot be allowed at this stage given the task domain and the state of the relevant science and technology, so there have to be some constraints on naturalness. In other words, *naturalness can be* (and in fact has to be) *traded for system feasibility* (general as well as scientific and technological). But, just as importantly, *those constraints should be principled and natural* in the sense that they can be easily assimilated and practiced by users, possibly based on some initial advice communicated to them by the system itself. In order to develop and test the naturalness of the system, the design decision was that the system should be usable by occasional- or "walk-up-and-use" -users as well as by experienced "travel-bookers" such as company secretaries.

The naturalness criterion also helps generate implications with respect to the definition of possible user tasks. The envisioned users want to book flights or to have flight information. A user wanting to do these things may want to:

- have information on departures and arrivals of flights;

- have information on fares and conditions on obtaining reduced fares;

- book flights and obtain the tickets;

- change flight reservations;

- have information of connecting means of transport;

- have information which allows planning of how to use local transportation (e.g., distance to the nearest city);

- book hotels, rent cars, book connecting trains, book tourist bus tours, theatre tickets, tables at restaurants, etc.;

- have information on general travel conditions ("Can I bring my cat ?");

- make travel insurance arrangements;

- have information on which different airline companies serve a specific destination;

- have information on alternative itineraries;

- etc.

Moreover, the flights and destinations involved could be any flights and destinations. Due to the feasibility constraints on the design process, technological and otherwise, this long list has to be reduced. In this situation, the naturalness condition demands that reduction be made in a principled and easily comprehensible manner. Since the system will deal with walk-up-and-use - users, its limitations should be made clear to users at the start of the dialogue and the limitations should be immediately comprehensible. This implies that the system be able to handle *a natural subset* of the possible tasks within the task domain. Moreover, the system should be able to handle this subset thoroughly and completely (cf. the usability requirements above). The design decisions made on this basis were that the system will deal with flights between two specific cities and will allow users to book flights, change previous bookings and obtain information on fares, arrivals and departures (and nothing else). It might be argued that these decisions go against the realism criterion. The answer has to be that in this case (1) we are merely dealing with a question of scale and that, therefore, (2) subsequent scaling-up to a realistic set of user tasks in a realistic application is feasible. Assumption (1) seems true whereas assumption (2) is hypothetical.

Finally, the naturalness criterion potentially has implications for the structure of the dialogue to be conducted between the user and the artifact. If, for instance, during knowledge acquisition it turns out to be the case that users have specific conceptions of the *order* in which to complete a certain task or sub-task, then the system should respect that order in the sequence of requests

and responses it gives to users. Such conceptions might be conceived of as a kind of script-like knowledge structures.

3. Setting up the initial design space: Technological requirements and constraints

Given the above design space constraints which have been developed using the notions of realism, usability and naturalness, let us now look at the constraints on the design space derived from technological requirements and constraints and the initial design choices based on these in conjunction with the realism, usability and naturalness constraints as well as the general feasibility constraints.

In *speech recognition*, the acoustic signal is decoded into phonetic/linguistic elements. Speech recognition today is dominated by the Hidden Markov Models technology which, although far from perfect, is still unbeaten by alternative approaches. No further justification is needed for the design decision that this technology will be used in the current design effort for speaker-independent recognition of continuous speech in close-to-real-time (cf. the usability requirements above). The training of word models (whether these be full-word- or sub-word-models) for speaker-independent word recognition is time-consuming and hence costly as it is necessary to use from 50 to 100 different speakers to obtain reasonably robust word models. This implies that the addition of new words to the system's vocabulary requires a non-negligible amount of effort. This again imposes stringent demands on the knowledge acquisition process (see below). The close-to-real-time- and naturalness design constraints described above jointly lead to the design decision that a vocabulary of 400-500 recognised words would be necessary and sufficient for the selected user tasks. This design decision of course rests on a *hypothesis* about the nature of the task domain sublanguage which is to be tested during the knowledge acquisition phase. A further design decision based on feasibility was made in this context. Given the amount of time available for completing the prototype, it was agreed that *limited* speaker-independent word recognition might be acceptable in the first prototype. This would save an amount of effort which might preferably be invested in other aspects of artifact development. Although limited speaker-independence adds to the fragility of system performance, it was felt that limitations in this respect were not of a principled nature, speaker-independence being a straightforward function of the amount of work spent on training the system's word models. In fact, “speaker-independence” is an eufemism for “group-dependence” and “speaker-dependent” recognition systems are simply systems capable of recognising voices from a rather small group. So if, due to the limited speaker-independence of the prototype system, a specific speaker cannot make himself or herself understood by the system - then system performance will simply have to be tested on other persons ! I will not go further into the design decisions pertaining to the speech recognition part of the system as such since the decisions needed to meet the usability and naturalness criteria above have now all been accounted for.

Speech recognition accuracy can be improved by incorporating a *grammar* in the speech recogniser, which provides additional constraints on the strings of words acceptable to the system for further processing. Current grammars for this purpose tend to be rather primitive, and it is a major unknown emerging from the initial design specification phase which grammar will be able to meet the usability condition on the speech recognition grammar noted above. The sublanguage corpus collected during the knowledge acquisition phase is expected to provide important input for the design decision to be made on this point. But the uncertainty on this point, which directly affects the expected overall utterance recognition and understanding capability of the system, lead to a further design decision. It is that, during each turn of the user-system dialogue in which users address the system, they should use brief utterances, and

preferably single sentences, in doing so. This will increase the likelihood of utterance recognition and understanding by the system. At the same time, this constraint on the user-system dialogue seems to be *principled* and thus meets the naturalness criterion described above. The assumption is that the brevity constraint can be easily understood, assimilated and respected by users during their dialogue with the system.

The symbolic output from the speech recogniser is passed on to a *natural language parsing module* which performs a syntactic analysis of the output and builds a semantic representation of the spoken utterance. This again involves the use of a grammar which may or may not be identical to the grammar built into the speech recogniser. Although no design decision has yet been made, it is likely that the grammar of the natural language parsing module will be different from and more sophisticated than the grammar used in the speech recognition module. Final design decisions on the grammar and parser to be used in the natural language parsing module will have to wait until the sublanguage corpus collected during the knowledge acquisition phase has been analysed. The same holds true of the relationship between syntactic and semantic processing as well as the amount of processing of discourse phenomena (anaphora, ellipsis, etc.) needed.

The semantic representation of the spoken utterance is interpreted by a *dialogue handling module* which decides what action to take. For instance, the system may have to go back to the user for clarification or additional information, it may put a query to the database or it may decide to terminate the dialogue. The dialogue handling module also has to keep track of the dialogue history in order to know when the information needed for the task at hand has been obtained from or communicated to the user. Some of the lower-level design decisions pertaining to the dialogue handling module depend on the results of the knowledge acquisition phase (see Sect. 4.2 below).

A *database* contains the task domain information relevant to the tasks the user can accomplish with the system. Although much of the relevant task domain information can be identified from standard sources directly on the basis of the chosen user tasks, some of this information will have to be identified during the knowledge acquisition phase. Once this has been satisfactorily done, the actual implementation of the database can be regarded as being a relatively simple and trivial part of the system development effort.

An *answer generation module* generates answers or questions to users. During the initial system specification phase little discussion has gone into the specification of this module.

A *speech synthesis module* generates synthetic speech either through true text-to-speech synthesis or through synthesis of pre-recorded human speech. The design decision concerning this module simply is that off-the-shelf components will be used. Since the current quality of Danish synthesized speech is not very good it is likely that pre-recorded speech will be used.

The *detailed system architecture* and *overall control* of information processing in the system are being discussed at the time of writing and concurrently with the knowledge acquisition phase of the project.

With respect to *system implementation*, the design decision is to make use of the DDL (Dialogue Description Language) tool which so far has not been used for the development of spoken language systems. The effectiveness of this tool for the purpose of the design effort is unknown at the time of initial system specification.

4. Results of the initial system specification phase

The criteria of realism, usability and naturalness have turned out to have had quite strong implications for the system to be designed and these implications have become visible in terms of design decisions already during the initial specification phase. The primary precondition for making those design decisions seems to be feasibility (including scientific and technological feasibility and feasibility in terms of manpower, costs, etc.). In other words, *unless an implication from the conditions of realism, usability and naturalness is not feasible, it should be incorporated into the system being designed.* And it can be added that, if the conditions of realism and usability cannot be met then the system will not be practically usable.

It has become abundantly clear from Sections 2 and 3 above that once the initial specification phase has been completed, the critical step in the artifact development cycle becomes that of knowledge acquisition. This section describes, first, the constraints identified during initial artifact specification which are directly relevant to the knowledge acquisition phase; secondly, a summary is provided of the tasks to be addressed during the knowledge acquisition phase as identified during initial specification.

4.1 Constraints relevant to knowledge acquisition

The following constraints have been derived from conditions of realism, usability and naturalness, from scientific and technological constraints and from general feasibility constraints:

1. The task domain is that of Danish domestic flight travels between Copenhagen and Aalborg.

2. The tasks of the system are to provide users with information on flight arrivals and departures, ticket prices and to permit users to book flights.

3. The corresponding tasks of users are to obtain information on flight arrivals and departures, fares, and to book flights.

4. The system should clearly communicate to users which tasks they can accomplish with the system in the chosen task domain and the system should possess the task domain information necessary for users to accomplish those tasks.

5. The types of users who will be addressing the system are both experienced "travel bookers" such as company secretaries and occasional users.

6. The user-system interface medium will be a telephone. Both user and system will be using spoken natural language.

7. The system should be capable of responding via the interface in approximately real time, i.e., close enough to make a demonstration tolerable.

8. The system should be able to do speaker-independent recognition and understanding of continuous speech.

9. The system should be able to understand users' utterances in their appropriate task context as long as they remain within the confines of the specified tasks. This implies that the

system should have a vocabulary, a grammar, semantics and discourse handling capabilities sufficient for that purpose.

10. As far as possible within the given feasibility constraints, the system should be capable of conducting via the interface a natural dialogue with users within the task domain. The main constraint on the naturalness of the dialogue decided on so far is that users have to use short-, and preferably single-, sentences when speaking to the system.

11. Knowledge acquisition has to be completed in about 6 months.

The above constraints (1)-(11) are relevant to the knowledge acquisition phase of the design effort in a quite specific sense. *The constraints help define the optimal situation of knowledge acquisition.* They provide a best model for knowledge acquisition. The optimal situation of knowledge acquisition is exactly a situation in which the above constraints are respected. If knowledge acquisition can be done in such a situation, there is a maximum likelihood that the observed user behaviour will be identical to the user behaviour which will be exhibited towards the final system. Obviously, at an early stage of design such as the present one in which no part of the system has yet been implemented, one will often have to resort to doing knowledge acquisition in situations which are more or less remote from the optimal situation. In such cases, the optimal model of knowledge acquisition provides a more or less precise standard against which to estimate the likelihood of valid transfer of observations of user behaviour from the non-optimal situation to the actual situation of use of the future system. However, given the importance of the results of the knowledge acquisition phase for subsequent artifact design, it clearly seems worthwhile to go to quite some length to realise a situation of knowledge acquisition which is as close as possible to the optimal situation.

4.2 The knowledge acquisition tasks

Vocabulary to be recognised by the system: For technological reasons, the (first prototype) artifact to be designed will be able to recognise only about 400-500 words. The knowledge acquisition phase has to (a) test the *hypothesis* that a reasonably natural dialogue between system and users can be conducted within the task domain if users are allowed to use a vocabulary this size; and (b) identify the individual words of this vocabulary with a high degree of precision. If the hypothesis turns out to be false and more than 400-500 words are needed, then further constraints will have to be imposed on the naturalness of the dialogue. If it turns out later that important words have not been identified during knowledge acquisition, then additional training of word models will have to be undertaken. This is costly in terms of time and manpower and may cause serious delay in the design process. If, on the other hand, the needed additional word model training is not performed the consequence will be a fragile system with compromised usability and which breaks down at unexpected points during user-system dialogue.

Speech recogniser grammar: A major unknown in the design process after initial specification is which type of grammar can support speech recognition without leading to the rejection of perfectly correct grammatical utterances. If such rejections are made, the system will be fragile and its usability will be seriously compromised. On the other hand, if the grammar used is too liberal with respect to the word-strings it accepts, word recognition will be impaired and the system will be fragile for different reasons. The core risk here is that this latter fragility might become so marked that the system's recognition and understanding capabilities turn out not to be acceptable to users so that usability will be seriously compromised. The knowledge

acquisition phase is expected to provide important input for the design decisions to be made on this point.

Natural language processing module: The analysis of the sublanguage corpus collected during the knowledge acquisition phase is crucial to a number of design decisions concerning this module. This is true of the grammar and parser to be used by the system; its form of semantic representation; the relationship between syntactic and semantic processing; and the amount of processing of discourse phenomena needed. Complex discourse handling may be difficult to achieve because the science base is not yet able to support it. However, the single sentence constraint on user utterances has the dual purpose of facilitating user utterance recognition and reducing the amount and complexity of discourse phenomena to be handled by the system. As to grammar, parsing and semantics the science base can be assumed to be in place, but the processing loads involved might endanger the system's ability to respond in close-to-real-time.

The dialogue handling module: The knowledge acquisition phase should clarify the structure of and the relationships between the tasks of the user and the tasks of the system during dialogue. It should also clarify the structure of actual user behaviour during dialogue. Such clarification is essential to the design of appropriate dialogue handling on the part of the system. For instance, if it turns out to be the case that users have specific conceptions of the order in which to complete a certain task or sub-task, then the system should respect that order in the sequence of requests and responses it gives to the user.

The database of task domain information: Initial system specification has been done without a fully detailed knowledge of the task domain, i.e., the possibility still remains that bits and pieces of relevant task domain knowledge have been overlooked. The knowledge acquisition phase should ensure that all relevant task domain knowledge is collected for subsequent implementation in the system's database and possibly also for the creation of new sub-dialogues between user and system.

Detailed system architecture and overall control: Initial system specification has produced a broad overall system architecture. It is an open question whether this architecture will have to be revised, in addition to being refined, in the light of the results obtained during knowledge acquisition.

In conclusion, the knowledge acquisition phase is critically important in the design of today's spoken language dialogue systems. The basic reasons why this is the case are the state-of-the-art in speech recognition science and technology and to a lesser extent in natural language processing science and technology. If the goal is to implement as natural and *as naturally constrained* a dialogue with users as possible, it becomes critical to carefully adjust the dialogue structure to the limited vocabulary recognised and understood by the system. The basic task of the knowledge acquisition phase, then, is to get the usability and naturalness vs. science, technology and general feasibility trade-off "right". For obvious reasons, this trade-off cannot be experimented with by building a full system which is then tested on real users. If this was not already disallowed due to resource limitations, the knock-down argument is that to build such a system in the first place one would have to define and implement its vocabulary. And there is no way other than empirical investigation to arrive at a vocabulary suitable for the task domain and corresponding to the feasibility constraints. In other words, some kind of empirical user studies is crucial to artifact design in this case. A positive point is that we already have a rough, optimal model of the circumstances in which to conduct empirical user studies (see Section 4.1 above).

5. The logic of initial system specification

The logic of the initial specification phase described above can now be characterised with a view to comparing the process to other early design processes. The characterisation removes the details of the artifact being designed. Removal of domain-specific detail allows structural comparisons to be made between this and other early design processes. Such comparisons might enable us to arrive at stable generalisations at some point.

Overall goal of the design process:

- to develop a state-of-the-art spoken language dialogue system prototype capable of replacing a human operator.

Realism criteria:

- the system should meet real and/or known user needs;

- the system should be preferable to current technological alternatives;

- the system should be tolerably inferior to the human it replaces, i.e., it should be at least usable;

- if the realism criteria cannot be met, the system will be practically useless;

- relationship to design decisions: realism criteria normally are basic to the decision whether to design a system at all.

Usability criteria:

- constitute a basic set of criteria for the design of interactive computer systems;

- serve to ensure that the system *can do* the tasks done by the human it replaces;

- if the usability criteria cannot be met, the system will be practically useless;

- are identified by asking for conditions such that, if they are not met by the system to be designed then the system will not be practically usable;

- many usability criteria can be self-evidently derived; the derivation of others may involve the notion of naturalness and may therefore require (cognitive) science-based expertise;

- relationship to design decisions: usability criteria normally are basic to the decision whether to design a system at all.

Naturalness criteria:

In some cases of systems design, such as mono-medium spoken language dialogue systems, we do have objective naturalness criteria for system performance. These criteria:

- aim at increasing the naturalness of user-interaction with the system;

- unless a naturalness criterion cannot be met for feasibility reasons, it should be incorporated into the system being designed;

- relationship to design decisions: can be traded for system feasibility;

- constraints on system naturalness resulting from trade-offs with system feasibility have to be made in a principled fashion based on knowledge of users in order to be practicable by users; decisions as to whether a trade-off is principled in this sense may require science-based expertise;

- constraints on system naturalness have to be clearly communicated to users;

- constraints on system naturalness may affect at least the tasks the users can perform with the system, the task domain covered by the system, the mode of user-system interaction and the types of users who can operate the system;

- when objective naturalness criteria for system performance are available, general naturalness criteria for system design can be more or less easily derived during early design;

- detailed and specific naturalness criteria may require investigation of users.

Realism, usability and naturalness:

- these criteria pose challenges to the *feasibility* of system design including technological and scientific feasibility as well as feasibility in terms of available manpover, cost, time, etc.;

- the criteria become operative through a process of *interpretation* in which they are interpreted with respect to relevant aspects of the artifact being designed (system aspects, interface aspects, task and task domain aspects, user aspects, etc. (cf. Bernsen in preparation));

- the criteria continue to be operative throughout the artifact design process and can be invoked at any stage during design as criteria for choosing between different design options;

- the criteria may require deep knowledge of users and the effects on performance of users with different degrees of experience;

- only criteria of naturalness may be traded for feasibility in design decisions. Examples from the current design effort are:

    • limits on the types of task users are allowed to perform;

    • limits on the task domain within which users are allowed to operate;

    • limited speaker-independent utterance recognition;

    • users should address the system in short utterances.

The initial design space:

- the shape of the design space resulting from the initial specification phase is a function of an interaction - sometimes involving trade-off decisions - between the following types of constraint:

  • general feasibility in terms of available manpover, cost, time, etc.;

  • technological and scientific feasibility (known or hypothetical);

  • interpretation of the criteria of realism, usability and naturalness;

  • specific choices of technology and tools for system development which may result from either feasibility or designer preferences.

- the shape of the initial design space is discovered through the concurrent development of the above types of constraint and the making of design decisions based on them;

- the initial design space may provide a rough model of the optimal situation of knowledge acquisition for the system to be developed, against which the actual circumstances of knowledge acquisition may be judged as to their similarity to the optimal situation and their consequent likelihood of providing relevant information for further system specification refinement;

- the initial design space may provide a detailed list of tasks for the knowledge acquisition phase of system development.

References

Bernsen, N.O.: The Structure of the Design Space. CO-SITUE Illustrated by a Study in Early Artifact Design. Contribution to Esprit Basic Research Action AMODEUS (to appear 1993).

Klausen, T: Talking to a Wizard. Report from the design of a natural speech understanding system. Contribution to Esprit Basic Research Action AMODEUS (to appear 1993).

Larsen, L.B., Bernsen, N.O., Brøndsted, T., Dybkjær, H., Dybkjær, L., Music, B., Povlsen, C., and Ravnholt, O.: *Spoken Language Dialogue Systems. A Survey of the State-of-the-Art* Report 1.1 from the project: Spoken Language Dialogue Systems. STC, Aalborg University, CLT, Copenhagen University and CCI, Risø National Laboratory and Roskilde University. August 1992.

MacLean, A., Bellotti, V. and Young, R.: What Rationale is There in Design ? In Diaper, D., Gilmore, D., Cockton, G. and Shackel, B. (Eds.): *Proceedings of INTERACT '90: Third IFIP Conference on Human-Computer Interaction*. Amsterdam: Elsevier North-Holland 207-212.

MacLean, A., Young, R., Bellotti, V. and Moran, T.P.: Questions, Options, and Criteria: Elements of Design Space Analysis. *Human-Computer Interaction* Vol. 6, 1991, pp. 201-50.

Acknowledgements. I am grateful to Hans Dybkjaer, Laila Dybkjaer and Tove Klausen for their helpful comments. The work reported here was done in part under a grant from the

Danish Research Council for the Technical Sciences on "Spoken Language Dialogue Systems" and in part under AMODEUS II, Esprit Basic Research Action 7040.