# Natural Interaction in Spoken Dialogue Systems

*Kristiina Jokinen*

University of Helsinki and University of Art and Design Helsinki
Finland
Kristiina.Jokinen@helsinki.fi

## Abstract

This paper discusses challenges for building dialogue systems that would exhibit intelligent and natural interaction capabilities. While technological development leads us to construct systems that will be embedded in our environment and allow ubiquitous computing, there is simultaneously a need for more intelligent software which takes into account requirements for complex interaction: the users' knowledge and intentions, variation in their viewpoints and interests, the context in which interaction takes place. Representation and processing of such knowledge is one of the important issues in dialogue systems research.

## 1 Introduction

The world around us is rapidly changing and we are surrounded by electronic devices which provide us with more and more information. It is likely that in a few years' time intelligent home robotics will cater for many everyday needs for us, and we are required to interact with a complex environment which will not consist of only people, but of computers embedded in our daily environment. E.g. Sjöberg and Backlund (2000) envisage the future information and communication systems contain computers that are built into products such as clothes, books, beds, and sporting gear, and which communicate easily with other objects. Computers will also have senses and they can interpret human expressions, can smell, feel, hear, see and taste, and there will be intuitive human-computer interfaces that mimic human communication.

One of the main problems in furnishing computers with intelligent communicative capabilities is the type and representation of knowledge that is needed in order to enable interaction and rational initiative-taking behaviour in the designed artefacts. For humans, the world is huge and exciting, full of meaning that is inferable through their interaction with the environment, but this is not the case with computers which, although being able to store and crunch large amounts of data, still lack adequate means to represent and use this data in a meaningful way.

The problem is not new. Knowledge acquisition has been one of the main topics in AI-research, ranging from collecting and representing common sense information in an explicit way like in the CYC-project, to Brooks' anti-representational robotics where meanings emerge through interaction of the processors. Early dialogue research concerned plan-based and BDI-systems (e.g. Cohen et al. 1990, Jennings 1993, and references therein), and emphasized the separation of static task knowledge from the processing knowledge. Task knowledge was represented in concept hierarchies and knowledge bases were manually built for the demonstration purposes of the systems. One of the main innovations of the PLUS project (Black et al. 1991) was to produce a more generic system by separating application specific knowledge from more general world knowledge, thus aiming at portability to different applications.

However, manual construction of concept hierarchies and knowledge bases that were large enough for practical (and commercial) purposes is both time and resource consuming, and automatic construction of the appropriate conceptual classes is still an on-going research topic. For instance, Jokinen (2000) identified knowledge acquisition and learning as the main challenges for developing dialogue systems. The adding of speech to dialogue systems also directed research towards the problems of speech processing, and the higher-level interaction and knowledge management were largely ignored.

Recently, three reasons seem to have brought ontologies, knowledge acquisition and representation into research focus again:

- Technical development in constructing computers which can sense our presence and react in a meaningful way,
- Complex application tasks such as negotiation in eCommerce, free navigation in the web, interaction with a complex environment in the ubiquitous computing visions,
- Initiatives to support participation for all in the knowledge-based economy, so as to enable accessibility to information technology for as many citizens as possible.

In this paper I discuss challenges for building dialogue systems that would exhibit intelligent and natural interaction capabilities. Section 2 concerns issues in dialogue modelling, Section 3 presents a dialogue system, and Section 4 summarises the challenges.

# 2 Dialogue Management

## 2.1 Computer as an Agent

The computer is usually regarded as a tool which supports human goals and its role has been a passive and transparent 'slave' under human control. As more sophisticated systems become available, and also connected with each other into networks of interacting processors, human control over their behaviour is not so straightforward and easily managed as before. In recent years another metaphor has thus been brought out: computer as an agent capable of mediating interaction between human users and the application. For instance, in the Smartkom project (Wahlster et al. 2001), this metaphor is realised in the Smartakus -interface agent which takes care of the user's requests with respect to different applications and application scenarios. Furthermore, information within Internet and electronic databases has grown into such dimensions that meaningful, accurate and truthful information is more difficult to get mined, retrieved and extracted from them. There appears a need for a mediating agent that would understand the logic behind natural language requests and master the complicated structure of the network, thus making the interaction with digital information easier and more accessible.

Although there is a fundamental asymmetry between the user and the system, I wish to put forward an approach which takes into account rich interaction possibilities for the user, and also accords with the best-practice design principles in order to provide a transparent and robust interface. The system should be intelligent enough to understand the user and her incomplete, imprecise, vague and fragmental input, but it should not aim at producing responses on the same level of inferential richness. Rather, it should restrict its responses to those that have clear communicative goals.

## 2.2 Scripts, Frames and Conversations

By a spoken dialogue system I refer to a speech-based system that has special models for representing and manipulating information at multiple levels of abstraction. A speech interface, on the other hand, is a software component that enables the user to use speech input when interacting with the system (and possibly also the system to use speech when giving information back to the user), but it does not generate abstract concepts for meaning analysis.

In interface design, the system's responses are geared towards clear, unambiguous prompts whereby the user will be provided with explicit information about the task and the system's capabilities. For instance, the 20 Laws of Interface Design (Weinschenk and Barker, 2000) consist of such aspects as linguistic clarity, simplicity, predictability, accuracy, suitable tempo, consistency, precision, forgiveness, and responsiveness, which make the interface easy and transparent to use. The aim is to take human cognitive limitations into consideration, to fit the user's work and thinking together, and to help the user to feel in control of the application.

On the other hand, Sadek (1999) advocates user-friendly interaction with the help of an intelligent and rational system: it is the system's intelligence that provides the good interface and ergonomy of the service. The users' perception of the system depends on the system's communication capabilities and affects satisfaction and trust on the system: natural communication supports user satisfaction even over some obviously undesirable features like long waiting times and mere errors. Dialogue capabilities thus allow the system to function in a wider range of situations, and provide services in an enjoyable and acceptable manner.

The simplest way to build a dialogue system is to use scripts which define possible actions at each dialogue point. Task and dialogue knowledge are evident through the designer's plan that has been employed in building the system. Scripting languages like VoiceXML can be used to write descriptions which can also include subroutines so that it is possible to produce fairly sophisticated dialogues. The approach is feasible only for limited domains, however, since the intertwining of static task knowledge with procedural interaction knowledge makes it difficult to reason over abstract events and relations, which is necessary in richer interaction situations.

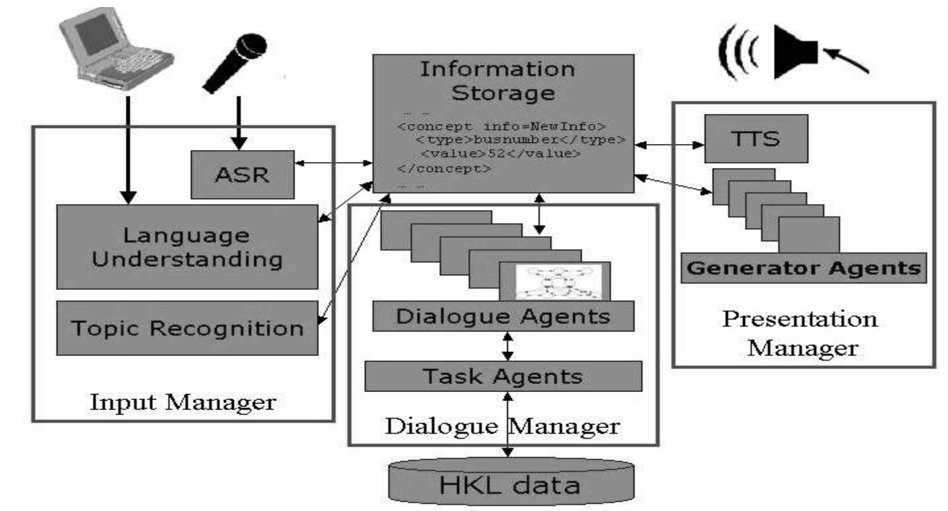
To allow more flexible dialogues, task knowledge can be separated from the possible dialogue actions by using forms or frames which define the information needed to complete the underlying task. Form-based dialogue management is suitable for tasks where the actions can be executed in various orders and the dialogue be driven by the information needed (e.g. Larson et al. 2001, Jokinen et al. 2002). The form also provides a dialogue context in which the actions can be interpreted and planned so as to allow varied utterances.

To enable truly natural interaction, in Sadek's sense of the system possessing capabilities for intelligent and rational reasoning, it is necessary to equip the system with knowledge about the world, and with effective means to reason about the information. Various statistical and machine-learning techniques can be applied to conversational phenomena, and it is important to take into account communicative principles concerning cooperation, obligations, rights, and trust (Allwood et al. 2001). Conversational dialogue management thus aims at improving computer interaction by taking into account human conversational capabilities and by building models for their computational treatment. On the other hand, natural interaction patterns among humans need not be natural when interacting with a computer and vice versa (e.g. spoken/signed language is the most natural means of communication among humans, whereas keyboard and mouse are the most natural means to interact with computers). Extension to multimodality and a better understanding of how multimodal conversation takes place among humans is thus necessary for designing conversational dialogue systems. More research is also needed to evaluate what input and output modalities are possible for the user and the system and which areas of dialogue management will benefit most from the sophisticated models.

## 3 Intelligent systems

Figure 1 shows the architecture of the Interact-system (Jokinen et al. 2002). The system consists of managers which handle general coordination between functional modules and the physical system components (Input Manager, Dialogue Manager and Presentation Manager), and of several agents which within each manager take care of various tasks typical for the functional domain. The agents can vary from simple counting procedures to complex software agents. The modules and agents communicate via a shared knowledge base called Information Storage. The system exhibits distributed dialogue management style (Kerminen and Jokinen 2003) whereby the dialogue situations are handled by several dialogue agents dealing with such subtasks as recognition of dialogue acts and topics. The DM also contains separate task agents which take care of application-related goals and the system's interaction with the database. Dialogue states are characterized by the agent's perception of the context (dialogue acts, topic, task goals, the speaker), but the knowledge itself is encoded in the system of agents and their processes. Information Storage contains dynamic knowledge produced during the interaction.

**Figure 1:** System Architecture




## 4 Challenges

The main challenge in the 21st century Communication Technology Society is to integrate engineering with language technology research, and to equip electronic devices with natural interaction capabilities. The following seem to be the main issues in this respect:

- Type and representation of knowledge, and the nature and number of reasoning rules. Data-driven learning techniques are used for concept and rule-extraction but have not yet been incorporated into dialogue systems as clustering and classification algorithms to extract necessary information to build knowledge bases.
- Flexibility. The system should be capable of learning and adapting itself to new tasks, situations, users, and language expressions. Instead of operating in a strict sequential manner, having their task and user model fixed, the systems should have dynamic control of the interaction and take the user's personalised needs into consideration. Adaptive methods are needed to identify and classify relevant features, and also to enable systems evolve through time to meet users who speak several languages, have varied knowledge and mixed cultural backgrounds.
- Presentation techniques. The systems often need to combine and present complicated information to the user on-line, and thus sophisticated generation models are needed to provide information in an accessible and clear form (Wilcock 2001). QA-systems that combine text mining with reasoning techniques also refer to the Gricean maxims that have been the basis for cooperative dialogue systems (Harabagiu 2000).

- Robustness, handling of errors and incomplete information. The system should be capable of overcoming processing errors and lack of knowledge on one level by exploiting expectations and information available on the next level. Besides context modelling, this requires that processing units and their interaction is specified in a way that allows ambiguous information exchange between the modules.
- Multimodal interaction. Sophisticated models of communication take multimodality into account, and extend the system's communicative repertoire to cater various new interface possibilities and also user groups with special needs.
- Interaction between computer architectures and human processing models. Advances in hardware development have mostly been engineering research with no connection to human cognitive processing. Research efforts directed towards connecting speech and language technology with cognitive processing are needed.
- A general theory/formalisation of human communication capabilities. This is related to constraints on representations, architectures and processing through the emergence of repetitive patterns on processing units, and must be pursued together with the above mentioned efforts combining evolving and learning systems.

## 5 References


Allwood, J., Traum, D., & Jokinen K (2001). Cooperation, dialogue and ethics. *International Journal of Human-Computer Studies, 53,* 871-914.

Black, W.J., H.C. Bunt, F.J.H. Dols, C. Donzella, G. Ferrari, R. Haidan, W.G. Imlah, K. Jokinen, T. Lager, J-M. Lancel, J. Nivre, G. Sabah, & T. Wachtel (1991). *A Pragmatics-Based Language Understanding System*. PLUS Deliverable D1.2.

Cohen, P.R., Morgan, J. & Pollack, M. (Eds.) (1990). *Intentions in Communication*. Cambridge: MIT Press.

Harabagiu, S., Pasca, M., and Maiorano, S. (2000). Experiments with open-domain textual question answering. *Proceedings of COLING-2000*, 292-298.

Jennings, N. R. (1993). Specification and Implementation of a Belief- Desire-Joint-Intention Architecture for Collaborative Problem Solving. *Int. Journal of Intelligent and Cooperative Information Systems*, 2 (3), 289-318.

Jokinen, K. (2000). Learning Dialogue Systems. In L. Dybkjaer (Ed.) *LREC 2000 Workshop: From Spoken Dialogue to Full Natural Interactive Dialogue – Theory, Empirical Analysis and Evaluation*, Athens, pp. 13-17.

Jokinen, K., Kerminen, A., Kaipainen, M., Jauhiainen, T., Wilcock, G., Turunen, M., Hakulinen, J., Kuusisto, J., Lagus, K. (2002). Adaptive Dialogue Systems - Interaction with Interact. *Procs of the 3rd SIGdial Workshop on Discourse and Dialogue*, Philadelphia, 64-73.

Kerminen, A. & Jokinen K. (2003). Distributed Dialogue Management. *Procs of the EACL Workshop Dialogue Systems: interaction, adaptation and styles of management,* Budapest.

Larsson, S., Cooper, R. & Ericsson, S. (2001): menu2dialog. In K. Jokinen (Ed.): *Knowledge and Reasoning in Practical Dialogue Systems*, Workshop Program, IJCAI-2001. pp. 41-45

Sadek, D. (1999). Design considerations on dialogue systems: from theory to technology. *Procs of the ESCA workshop on Interactive Dialogue Systems*, Kloster Irsee, Germany.

Sjöberg, C. & Backlund, A. (2000). Technology Foresight: Visions of Future Developments in Information and Communication Systems. NUTEK, Swedish National Board for Industrial and Technical Development. http://www.jrc.es/pages/iptsreport/vol49/english/MET1E496.htm

Weinschenk, S. & Barker, D. (2000). *Designing Effective Speech Interfaces*. London:Wiley.

Wilcock, G. (2001). Pipelines, templates and transformations: XML for natural language generation. *Procs of the 1st NLP and XML Workshop*, Tokyo, 1-8.