

# EMPLOYING SITUATIONAL-FACTORS IN DIALOGUE PROCESSING

Botond Pakucs

*Centre for Speech Technology (CTT)*

*KTH, Stockholm, Sweden*

botte@speech.kth.se

**Abstract** In this paper, a generic solution is proposed for capturing, representing and employing situational-factors in discourse and dialogue processing. Furthermore, the implementation of this solution within the framework of the SesAME dialogue manager and the Butler demonstrator is described.

**Keywords:** Dialogue management, speech interfaces, situational-factors, context, knowledge representation, context sensitivity.

## 1. Introduction

In natural human-to-human communication, speakers are able to use implicit contextual information to increase the conversational bandwidth. The contextual information is relevant knowledge about the actual situation, conversational partners, domain, topic etc. This type of knowledge is not necessarily part of the linguistic-context (has not been uttered earlier).

The term 'context' is often used in the research field of spoken dialogue systems for referring to different aspects of the communicative act (linguistic, acoustic, discourse etc.) For avoiding misconceptions, the term context is avoided in the rest of this paper and for referring to implicit knowledge about the domain, topic, conversational partners etc. related to the actual situation the term situational-factors is used.

Current spoken dialogue systems are, not able to take full advantage of the context of human-computer dialogue. In this paper, a generic solution is proposed for capturing, representing and employing situational-factors in discourse and dialogue processing. Furthermore, the implementation of this solution within the framework of the SesAME dialogue manager and the Butler demonstrator is described.

## 2. Background

One of the most influential discourse theories was presented by Grosz and Sidner (Grosz and Sidner, 1986). According to this theory, at least three different kinds of information are necessary for determining the discourse structure: linguistic markers, utterance-level intentions and general knowledge about actions and objects in the domain of discourse. The situational-factors appear to be part of the general knowledge. According to Grosz & Sidner, this general knowledge is especially important when the linguistic markers and the utterance-level intentions are insufficient for determining the discourse structure. How this general knowledge fits in Grosz & Sidner's discourse theory was however never elaborated.

A knowledge-based theory of discourse interpretation was presented by (Hobbs, 1985). According to Hobbs the coherence of an utterance is a combination of the utterances local and global coherence. The relation of an utterance with its surrounding discourse is the local coherence, while the global coherence of the utterance is the relation between the utterance and the surrounding environment. For employing the situational-factors in dialogue processing, the handling of the global coherence appears to be central. According to Hobbs, the knowledge base necessary for handling the global coherence "will necessarily be huge, and the project of determining what needs to be represented, how to encode and organize it, and whether or to what extent is consistent is correspondingly huge."

In dialogue management, the term 'context' has according to Bunt (Bunt, 1994) the following dimensions:

- *Linguistic context* consist of the surrounding linguistic material and is usually modeled in the dialogue history.
- *Semantic context*, the state of the underlying task, including facts in the task domain is described in task and domain knowledge models (Flycht-Eriksson, 1999).
- *Cognitive context* is the participants' state of beliefs, intentions, plans, attitudes and attentional states. Plan and intention recognition is supported for instance in the TRIPS system (Allen et al., 2000).
- *Physical and perceptual context* consists of the physical circumstances in which the interactions take place.
- *Social context* is related to the type of interactive situation and the roles of participants in the specific situation.

All of these dimensions have *global* and *local* aspects. The global aspects of these context types are given in the beginning of the interaction and do not change during the dialogue, while the local aspects of the context may be changed during the interaction.

Situational-factors can be part of any of these dimensions of the context. Most experimental and commercial dialogue systems are single-domain dialogue systems. Therefore, especially with regard to the physical and perceptual context a more or less static situation is usually assumed. However, the growing interest for using speech interfaces in mobile and ubiquitous computing environments makes it necessary to develop better solutions for capturing, modeling and adapting to the dynamic variations in the mobile users' physical and perceptual context.

Speech-based interaction with ubiquitous services in mobile environments differs from interacting with desktop or telephony-based speech interfaces. Being on the run, and with hands, eyes and sometimes even the mind busy, the users' requirements on the speech interfaces can be expected to increase. Accordingly, in mobile environments it is even more important to fine-tune the speech based communicative interaction to the specific user and to the user's current situation.

Contextual information can be encoded in several different knowledge models employed in dialogue systems (Dahlbäck and Jönsson, 1999), such as domain knowledge models and user models. These knowledge models are, however most often manually built and tailor made for the actual task and domain. Accordingly, the maintenance, porting and reusing of these knowledge models is not facilitated. For making speech interfaces more natural, automatic and generic solutions are desired.

### 3. Situational-factors and knowledge models

Apparently, the role of contextual information and such, the situational-factors are considered important in discourse theories and in dialogue processing. However, in the above theories it is never elaborated how the situational-factors should be integrated in the discourse and dialogue processing. Furthermore, the challenge of allowing different kinds of non-linguistic knowledge and reasoning to play part in discourse and dialogue processing is considered to be an AI-complete problem (Jurafsky and Martin, 2001) (p 738)<sup>1</sup>. For allowing more natural and flexible interaction in spoken dialogue systems, these issues should and can be addressed. The following three scenarios are real life examples collected

---

<sup>1</sup>This means, that the integration of NLP and non-linguistic knowledge processing cannot be addressed before all the research question in the field of AI are solved.

for illustrating some of the major characteristics of situational-factors employed in dialogue processing.

*Scenario 1 - leaving the office*

Q: When is the next train leaving?

A: At 19:30.

*Scenario 2 - in the elevator*

Q: The second floor?

A: Yes please!

*Scenario 3 - ordering pizza*

Q: Would you like one with mozzarella or the vegetarian?

A: Vegetarian please!

All three scenarios appear somewhat strange, incomplete and even incomprehensible to others than the participants. In spite of this, all dialogues are complete and successful interactions. No other kind of related information was uttered previously. They all exhibit the use of implicit knowledge about the situation such as the identity of the participants, time and place of the interaction etc. All three dialogues appear to be repetitions of similar interactions encountered previously. All three answers contain predictions or beliefs about the individual goals, intentions and preferences of the dialogue partner.

It is desirable to enable spoken dialogue systems to support similar interactions and make use of the same features. It is possible to achieve such a functionality by employing better user modeling and more elaborated domain knowledge models. However such a solution requires a computationally demanding reasoning with the use of explicit, manually built and tailor-made knowledge models. A major challenge is to automatically maintain the correctness of the knowledge models in dynamically changing environments.

Usually the knowledge models used for discourse and dialogue processing utilise the ontological aspects<sup>2</sup> of the relevant knowledge. On the other hand, the type knowledge employed in the above scenarios can be regarded as experiences, the sum of the previous interactions in similar situations, albeit, the epistemological aspects<sup>3</sup> of the knowledge. By employing these characteristics of the knowledge it is possible to capture, represent and employ situational-factors in a generic way.

For providing more natural and conversational speech interfaces in dynamically changing environments, another major challenge is the *simultaneous adaptation to the user and to the context*. Variations in the context may affect the user's behavior and induce different user related

---

<sup>2</sup>How knowledge is related and organised.

<sup>3</sup>How knowledge is created.

variations. For instance, time pressure and increased cognitive load on the user may cause variations in the user's speech (Müller et al., 2001). As the user characteristics and the parameters of the situation may change during the same interaction, both the characteristics of the context and the user's properties have to be considered simultaneously. The usability of a system can be expected to increase if both the user properties and the user's situation are taken into account (Jameson, 2001).

It is feasible to achieve simultaneous adaptation to the users and the users' context by employing context tracking and advanced knowledge models. However this solution requires complex and computationally demanding reasoning. A more promising solution is to integrate the contextual data in to the employed knowledge models. If the knowledge base relevant to the discourse is organized with the help of the context, the use of a combination of smaller knowledge models is possible. While using context based indexing and context aware computing, the problem of finding the relevant pieces of information in these knowledge models seems also somewhat simpler.

Using knowledge models based on the epistemological aspects of the relevant knowledge and integrating situational factors into the knowledge models are the two major features of the solution proposed in this paper. How this is implemented and employed in the SesaME dialogue manager is presented in the next sections.

#### 4. Employing situational-factors in SesaME

SesaME (Pakucs, 2003), see Figure 1, is a generic, task-oriented dialogue manager specially designed for a human-centered approach and for mobile environments. SesaME features an event based dialogue management and a central blackboard based architecture. SesaME relies on the Atlas generic speech technology platform (Melin, 2001). The Atlas platform provides high-level primitives for basic speech I/O, but access to low-level data is also facilitated. The ATLAS platform includes support for the ACE speech recogniser (Seward, 2000), which features dynamical adaptation of the language model.

SesaME features a blackboard and agent based architecture. The central blackboard stores the representation of the *information state* (Larsson and Traum, 2000) of the dialogue. However, this representation is not formalised; the information state is merely a collection of all data available to the dialogue system. The update of the information state is event-based, where events can be dialogue moves, internal events, or changes in the user's external context. The event-based functionality enables an asynchronous information processing (Blaylock et al., 2002).

The SesaME architecture and the theoretical considerations behind it are in some aspects comparable to other agent-based architectures such as the Jaspis (Turunen and Hakulinen, 2000) and the TRIPS (Allen et al., 2000) architectures.

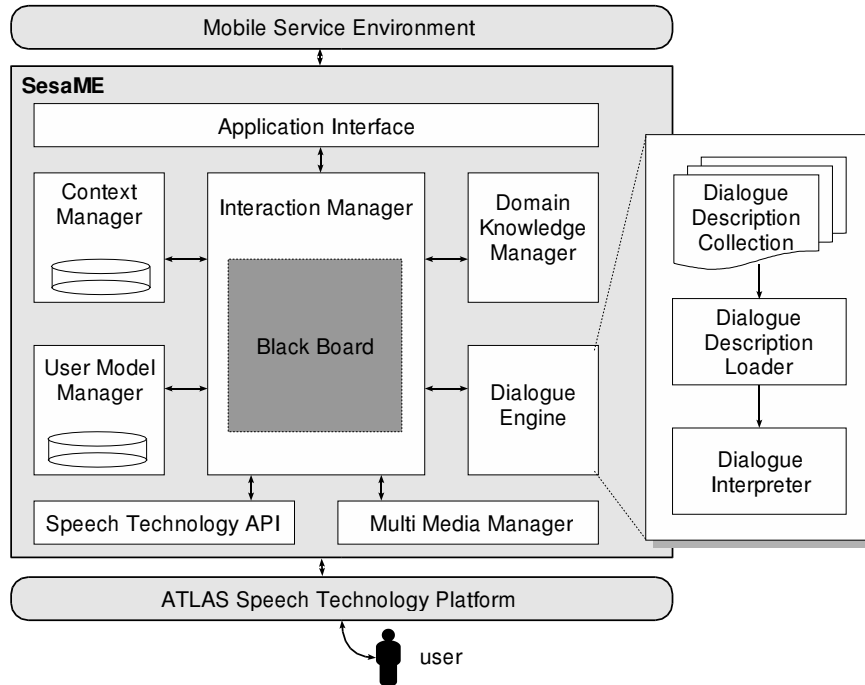


Figure 1. SesaME system architecture.

A major goal in SesaME is to make full use of the potentials of the individual user models and to achieve context-based adaptation. The adopted solution is inspired by attempts to achieve context-aware computing in the research field of ubiquitous computing (Dey, 2001).

In SesaME, the major components involved with employing the knowledge related to situational-factors are:

- *User Model Manager* (UMM) which holds the individual user models which contain even situation specific information. The UMM and its functionality is in detail explained in the next subsection.
- *Context Manager* (CM) The CM keeps track of the current context/situation and employs this information for retrieving relevant aspects of the interaction from the UMM. The functionality of the CM is described in subsection 4.2.

- *Interaction Manager* (IM) Generic, application independent features of the dialogue management are handled by the *Interaction Manager*. The IM also handles error detection, planning, keeps track of dialogue history, and coordinates the different system components and knowledge sources. A central, shared information storage, a blackboard, and a collection of autonomous software agents are the main components of the IM. The detailed description of the IM's functionality is, however, beyond the scope of this paper.

#### 4.1 Knowledge representation in the UMM

In SesaME, after each interaction with the user every utterance is represented as a feature vector containing feature-value pairs of all relevant information (such as topic, start time of the utterance, length of the utterance, user choices etc.). The only common property of the features in the feature vector is the co-occurrence. All the feature-value pairs - containing high or low level information - in the feature vector characterizes in some way the utterance. The feature vectors are indexed and stored in the user model.

```
[user--botte, time--morning, domain--elevator, task--elevator,
destLevel--2, startLevel--0, utteranceDuration--200, time--0832]
```

The user models are individual user models. They are represented as a vector-space model, a common data structure used in information retrieval applications. For manipulating the user model, common information retrieval solutions are used. The weight function used for indexing the user model is the  $tf*idf$  function where  $tf$  is the term frequency and  $idf$  is the inverse document frequency.

Accordingly, the user model is domain and task independent and is automatically built and updated after every interaction. In this way it is no problem to keep the knowledge model up to date and to maintain the correctness in spite of dynamically changing environments. The user model is not formalised in some specific knowledge-based structure. However, it is still possible to apply machine learning solutions such as memory-based learning or similarity-based reasoning.

#### 4.2 Employing situational-factors

The *Context Manager* keeps track of the current context/situation. During a new interaction, based on available contextual information a *query-vector* is built. This query vector is used for retrieving similar interactions from the user model. During this retrieval process straightforward information retrieval methods are used. The employed similarity function is the *cosine-metric function*.

The retrieval results can be used to predict specific features of the ongoing interaction and to achieve adaptation to the current context. For example, based on earlier interactions with a voice controlled elevator it may be possible to detect that the user’s most frequent choice of semantic object was the “fifth floor” when answering to the standardised prompt: “*Which floor would you prefer?*”. Thus, it is possible to predict that the user may want to take the elevator to the fifth floor. By using the additional prompt supported in SesAME, it is possible to ask the user a more natural question: “*Fifth floor, as usual?*” instead of the impersonal standardised prompt.

During the dialogue management process, a slightly modified VoiceXML-based domain and dialogue descriptions are used. For enabling the use of alternative questions, additional system prompts are used

```
<prompt> Which floor would you prefer? </prompt>
<alt-prompt>
<value expr='predicted-floor'> floor as usual?
</alt-prompt>
```

If there are no similar interactions, or no obvious patterns are present in the previous interactions (such as a CD purchasing task), then the default standardised prompt is used.

It is, possible, that the systems prediction is wrong. This is however not a major issue. This kind of mistakes occur naturally in human-to-human communication and are regarded as natural. If the corrections can be handled by the dialogue manager then the interaction breakdown can be avoided. Another major challenge for such a solution is how to detect the users’ changing habits - to allow the system to detect new patterns and forget old ones. For this kind of problems there are several different solutions proposed by the user-modeling research community. For instance it is possible to weight new interactions higher than old ones and in this way facilitate the detection of new patterns.

While combining a user model built and maintained dynamically and containing situation specific knowledge with an situation based retrieval of the relevant knowledge it is possible to capture and employ situational-factors for more flexible and natural interaction.

## 5. Application and Evaluation

For evaluating the support for situational-factors in the SesAME dialogue manager a very carefully designed and implemented application is necessary. Real users and real services are required for building realistic user models. Repetitive use of the services is also necessary for detecting and employing patterns related to user behaviors. In the next section such an application, the Butler, is described.

## **5.1 The Butler**

The Butler is a specially developed telephony based spoken dialogue system specially built for demonstrating and evaluating the use of situational-factors in the SesaME dialogue manager. Butler features a number of different services for the students and employees of our department. The services are related to local services available at the KTH-campus:

- lunch menu information
- subway train information
- commuter train information
- information on meetings at the department
- and possibly personal calendar informations.

All these services are employing on the Internet available services. The relevant knowledge will be automatically extracted from the available documents and transformed into dialogue descriptions.

The identification of the users will be based on the used A-numbers. In this way it will be possible to build, maintain and use individual user models. For achieving adaptation based on situational-factors to individual users a long term usage is necessary. At least 10-20 individual interactions with every service. For proving that detection of individual patterns is possible, at least 10 active users are necessary.

## **6. Conclusions**

This paper presents a generic solution for capturing, representing and employing situational-factors in discourse and dialogue processing. A discussion and analysis of why and how the situational-factors are important for achieving a flexible and natural interaction was also presented.

The described solution relies on using knowledge models based on the epistemological aspects of the relevant knowledge. Another major feature of the presented solution is the integration of the situational factors into the knowledge models. Furthermore, the implementation of this solution within the framework of the SesaME dialogue manager and the Butler application was also described. The Butler is specially designed application which will be used for the evaluation of the presented solution and implementation.

## **Acknowledgments**

This research was carried out at the CTT, Centre for Speech Technology, a competence center at KTH, supported by VINNOVA (The

Swedish Agency for Innovation Systems), KTH and participating Swedish companies and organizations. This work was also supported by GSLT, The Swedish National Graduate School of Language Technology.

## References

- Allen, J., Byron, D., Dzikovska, M., Ferguson, G., Galescu, L., and Stent, A. (2000). An architecture for a generic dialogue shell. *Natural Language Engineering*, 6(3-4):213–228.
- Blaylock, N., Allen, J., and Ferguson, G. (2002). Synchronization in an asynchronous agent-based architecture for dialogue systems. In *Proceedings of 3rd SIGdial Workshop on Discourse and Dialogue*.
- Bunt, H. (1994). Context and dialogue control. *THINK Quarterly*, 3(1):19–31.
- Dahlbäck, N. and Jönsson, A. (1999). Knowledge sources in spoken dialogue systems. In *Proceedings of Eurospeech'99*, pages 1523–1526, Budapest, Hungary.
- Dey, A. K. (2001). Understanding and using context. *Personal and Ubiquitous Computing*, 5(1). Special issue on Situated Interaction and Ubiquitous Computing.
- Flycht-Eriksson, A. (1999). A survey of knowledge sources in dialogue systems. In *Proceedings of IJCAI'99 workshop on Knowledge and reasoning in practical dialogue systems*, pages 41–48, Stockholm, Sweden.
- Grosz, B. and Sidner, C. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.
- Hobbs, J. R. (1985). On the coherence and structure of discourse. Technical Report Report No. CSLI-85-37, Center for the Study of Language and Information, Stanford University.
- Jameson, A. (2001). Modelling both the context and the user. *Personal and Ubiquitous Computing*, 5:29–33.
- Jurafsky, D. and Martin, J. (2001). *Speech and Language Processing*. Prentice Hall.
- Larsson, S. and Traum, D. R. (2000). Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural Language Engineering*, 6:323–340.
- Melin, H. (2001). ATLAS: A generic software platform for speech technology based applications. *TMH-QPRS, Quarterly Progress and Status Report*, 42.
- Müller, C., Großmann-Hutter, B., Jameson, A., Rummer, R., and Wittig, F. (2001). Recognizing time pressure and cognitive load on the basis of speech: An experimental study. In *UM2001, User Modeling: Proceedings of the Eighth International Conference*.
- Pakucs, B. (2003). SesaME: A Framework for Personalised and Adaptive Speech Interfaces. In *Proceedings of EACL-03 Workshop on Dialogue Systems: Interaction, Adaptation and Styles of Management*, Budapest, Hungary.
- Seward, A. (2000). A tree-trellis n-best decoder for stochastic context-free grammars. In *Proceedings of 6th International Conference on Spoken Language Processing*, Beijing, China.
- Turunen, M. and Hakulinen, J. (2000). Jaspis - a framework for multilingual adaptive speech applications. In *Proceedings of 6th International Conference of Spoken Language Processing (ICSLP 2000)*, Peking, China.