# LANGUAGE UNDERSTANDING AND GENERATION IN CHINESE SPOKEN DIALOGUE SYSTEMS

**Bei Liu, LiMin Du, ZhiWei Fang, XianFang Wang**

Center for Speech Interaction Technology Research
Institute of Acoustics, Chinese Academy of Sciences
DSP BLDG, 17 Zhongguancun Rd., Beijing 100080, China
liub@iis.ac.cn, Dulm@iis.ac.cn

**Abstract**

This paper presents a new semantic description method by building a Semantic Concepts System (SCS), in which Semantic Patterns are used to describe and recognize the speech acts expressed by spoken utterances. By this means, users' intentions can be understood and elicited more directly and precisely. Moreover, Based on the domain knowledge related SCS, we can build high efficient language understanding models and generation models for different domains. This method provides solutions in building universal information service SDS platform, and has been applied to develop Beijing Railway Station Ticket Information SDS (BEST) and the Capital Airport Flights Information SDS (BAFI) successfully.

## 1. INTRODUCTION

Traditional language understanding methods emphasize complete syntactic analysis in which user inputs will be analyzed with strict grammars and logic rules ( Klatt 1980). But when dealing with loose-structured spoken languages, syntactic analysis always fails (Dowding et al. 1993) (Seneff 1992). To solve this problem, researchers prefer semantic-driven analysis rather than syntactic-driven analysis, for semantic analysis is more robust, and can generate proper system responses even by partial parsing instead of complete parsing (Mctear 2000). Although it is difficult to analyze complex sentences without syntactic confine, semantic analysis can do most of the job, because spoken language is mostly composed of simple sentences, and the ability of processing spontaneous inputs is more important than that latent defect (Zue and Glass 2000).

spontaneous Language possesses the dual property of presentational diversity and compositional simplicity. That is, although the users' utterances (textual or spoken ) can have rich variations, they are mainly composed of simple semantic units and only express a limited number of intentions that can be described by countable semantic patterns. Therefore, we put forth a SCS-based semantic description method, in which the SCS (Semantic Concepts System) is built up by collecting all the possible semantic expressions in a specific domain, then decomposing them into basic semantic concepts sequences and classifying them by intentions. These intention-labeled concepts sequences are defined as Semantic Patterns (SP) which can be used to describe and recognize most of the speech acts in a certain domain. SCS establishes good foundations to set up efficient language models used in SDS, such as Semantic Network – the semantic understanding model, speech Recognition Network – the decoding language model for recognizer and Response Pattern Sets – the generation model. Also SCS helps to analyze the semantics of users' utterances and their intentions correctly, then to generate more natural and proper system responses. Moreover, building SCS makes it easy to design the developing platform of Chinese SDS for information service domains.



## 2. THE SEMANTIC DESCRIPTION OF DIALOGUE SPEECH ACTS

If we want systems to precisely recognize the semantic meaning or intentions of users' utterances, then response properly, firstly the semantic meaning of the users' utterances must be understood correctly. In fact, users' utterances can be interpreted into a series of elementary semantic units which we call semantic concepts. For example, in ticket service domain, when a user says he want to buy some tickets from one place to another, sentences that express the same intention usually consist of the following semantic concepts: *Source, Destination, Depart_time, Arrive_time, Ticket_number, Ticket_kind et al*. Therefore, such an intention or semantic meaning can be described by a sequence or combination of related semantic concepts. For the convenience of studying similar domain-oriented speech acts, we define the following template of semantic concepts sequence to describe users' intentions:

*Intention:meta,sequence,{Concept [,Concept]*};*

Where, symbol "*" indicates this element will repeat for zero or more times. Since the concepts sequences can be ordered or unordered, they can be defined respectively as:

*Intention:meta,sequence,ordered,{Concept [,Concept]*};*

*Intention:meta,sequence,unordered,{Concept [,Concept]*};*

Concepts inside the square brackets are optional, while others are essential. If some concepts sequences are impossible to be integrated into one semantic pattern, then it is allowed to define several Semantic Patterns for one speech act or intention.

In order to build a computable SCS, the definition of semantic concepts, especially their particle sizes are very critical. If the particle size is too small, it will result in dramatic expansion of Semantic Pattern Sets, then large memory and time costs. However, if the particle size is too big, the number of patterns will decrease but new problem appears that the users' inputs can not be matched exactly with the patterns, then ambiguities will occur. One effective solution is that firstly try to decompose the utterances into primary concepts, then reduce the basic concepts with common properties to a upper-layer concept (whose particle size is bigger than the original ones), or if needed, continue decomposing concepts into several sub-concepts that will construct a new lower-layer (whose particle sizes are smaller than the original ones). Repeat this process until a suitable hierarchical model of the concepts is obtained.

In general, a specific semantic meaning or intention can be expressed by one or more Semantic Patterns that are unique to each other. So all the Semantic Patterns together with the hierarchy model of all the concepts, construct the SCS that can best describe all the speech acts within the task domain.

Then based on SCS, a semantic parsing rule set can be defined and used as the language understanding model of SDS, which is called Semantic Network (SN). SN is constructed as RTNs, which can effectively avoid repeated definition in the computing realization of the SCS, as well as save computing time and memory space (Wang 2002). In BEST, because the users' inputs within ticket service domain only expresses a limited number of speech acts and intentions that can be described by countable semantic patterns, so it only needs to definite a comparatively small number of rules to accomplish most of the domain tasks.

## 3 SCS BASED LANGUAGE GENERATION

Language generation is a very important aspect in building man-machine interfaces, and the quality of responses is the key factor in evaluating the performance of the whole dialogue system (Glass et al. 1994). BEST applies mix-initiative strategy, in which system and users can both control the dialogue process. The aim of system responses is to guide dialogue process positively, and give the user a satisfactory reply. Altogether, system responses generated in a dialogue system mainly include the following types:

(1) Queries and demands: instruct users to provide the necessary information items needed to accomplish dialogue tasks.
(2) Suggestions: Suggest and remind users what to do next, which helps to reduce system restrictions, or provide other choices when current query can not be handled.
(3) Clarifications and Confirmations: Prompt users to make a clarification or confirmation, when users' intentions are changed or the confidence value of the information items that users provided is below the threshold value.
(4) Negotiations: Negotiate with users when system needs extra restrictions to select the most suitable data from many candidates or to reduce the searching range of database.
(5) Direct answers: Answer users' queries properly according to the database searching results.

All the types of system responses can be generated by building corresponding system responses templates to construct the Response Pattern Set (RPS). Actually, SCS provides a good foundation to build RPS, because SCS provides most of the concepts sequences expressing all kinds of speech acts within task domain. Therefore, when system needs to generate a spoken reply to the user's queries, it can find the suitable semantic pattern templates by searching the RPS, then replace the query keywords or semantic variables in the templates with the results of database search, finally generate a complete response by connecting corresponding templates.

Then we give three kinds of response templates in RPS for example, which are showed below:

(1) *Query_SVN = response,{ String };*
(2) *Query_SVN_confirm = response,{ String$_1$ ,{ SVN, 1}, String2};*
(3) *Query_SVN_select = response,{ String$_1$, { SVN, 1}, or ,{ SVN, 2 }, String$_2$ };*

(1) is a query template for necessary information item that is denoted by *SVN*; (2) is a kind of confirmation template for the information item; Then (3) is a kind of clarification template to let user make a selection when the item has inconsistent values. *Strings* are the canned words in a template.

## 4. EXPERIMENTS

### 4.1 Experimental System Architecture

We use BEST as our experimental dialogue system. Since our aim is to test the computability and portability of SCS, the system input and output apply textual format. The system architecture is showed in Figure1.

In order to improve the portability of the system, domain-related sections are concentrated in a system module called Configure File, which includes Semantic Network (SN), Dialogue Control Strategy Set (DCSS) and system Response Pattern Set (RPS). Therefore only by rewriting the content of the Configure File according to different domain, can we realize different domain-oriented SDS, which embodies the portability of the SCS-based systems.

The system modules that will read Configure File are Language Understanding Module, Dialogue Management

Module, and Response Generation Module. Built on SCS, SN is the language model of understanding module and it mainly performs semantic analysis of the users' inputs in terms of Dialogue Context, and finally transforms the textual inputs into semantic descriptions. The semantic concepts in users' inputs are recognized to form a concepts sequence. Then by parsing this sequence with the rules in SN, user's intention can be recognized and filled into the Dialogue Context.

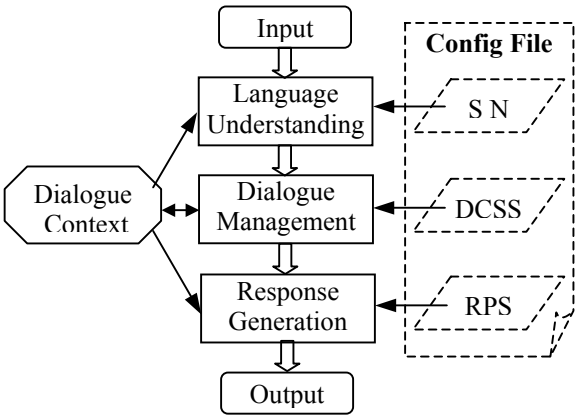


Figure 1: Architecture of the Experimental System

Users' intentions are associated with the dialogue control strategies in DCSS. Based on the recognized user intention, Dialogue Management Module will call corresponding control strategy which defines all the information items needed to complete a specific domain tasks. Then based on the parsing result of user input, system checks the evaluated information items and judges if they are sufficient. If not, or doubted, system will generate a query or confirmation or clarification responses in terms of the Dialogue Context and system expectations by using corresponding confirmation or query templates in RPS. In this case, the response templates are defined in the Configure File. When the information items are sufficient, they will be sent to the task processing module, in which system will complete this dialogue task, and generate a response according to the task processing status.

### 4.2 System Evaluation

The evaluation standard for SDS should be that systems can precisely recognize the semantic meaning or intentions of users' utterances, then reply users properly, and accomplish dialogue tasks through as small number of interactions as possible.

To test the system, we gathered 500 sample dialogues from Internet evaluation logs (between Jun. 2002 to Sep. 2002). And four parameters were used to evaluate the system (Simpson and Fraser 1993) (Glass 2000), and the evaluation results are illustrated in Table 1. The statistic data within domain (see the In Domain Stat. rows of Table 3) indicate that the system have comparatively high Semantic Units Understanding Rate (precision rate is 98.98%), high Task Success Rate (95.83%), and high Contextual Appropriateness (85%) which reflects the appropriateness of system responses while keeping fewer dialogue turns (average 4.11). But sometimes users' utterances may be beyond the task domain, which will result in the decrease of performance parameters (see the Total Stat. rows of Table 3). So we need to gather more Semantic Patterns of users' utterances and modify the concepts model to perfect the SCS. Now BEST has been published on Internet (http://www.iis.ac.cn/sds.asp) for users to evaluate, as well as to collect more practical data to improve the system performance.

| Evaluation Parameters | Semantic Units Understanding Rate | | |
|---|---|---|---|
| | Correct | Recall | Precision |
| Total Stat. | 79.17% | 81.38% | 92.79% |
| In Domain Stat. | **92.03%** | **95.83%** | **98.98%** |
| Evaluation Parameters | Task Success Rate | Average Number of Turns | Contextual Appropriateness |
| Total Stat. | 81.38% | 5.07 | 67% |
| In Domain Stat. | **95.83%** | **4.11** | **85%** |

Table 1: The Evaluation Result of BEST

## 5. ACKNOWLEDGEMENTS

This paper is a partial result of project G1998030505:The Research of the Theory and Platform for Chinese Spoken Dialogue System funded by the National Key Fundamental Research Project – Image, Speech, Natural Language Understanding and Information Mining.

## References

Dowding, J., et al. 1993. A Natural Language System for Spoken Language Understanding. *In Proc. ARPA Workshop on Human Language Technology*, 21-24.

Glass, J., Polifroni, J., and Seneff, S. 1994. Multilingual Language Generation Across Multiple Domains, *In Proc. ICSLP*, 983-976.

Klatt, D. H. 1980. Overview of the ARPA Speech Understanding Project. *In Trends in Speech Recognition*. Prentice-Hall, New Jersey.

Seneff, S. 1992. A Natural Language System for Spoken Language Applications. *Comput. Linguistics* 18(1):61-86.

Simpson, A., and Fraser, N. 1993. Black Box and Glass Box Evaluation of the SUNDIAL System. *In Proc. Eurospeech' 93*, 33-40. ESCA, Berlin.

Wang, X.F. (2002). Research on Language Understanding and Dialogue Management in Chinese Spoken Dialogue System. Doctor Dissertation, Institute of Acoustics, Chinese Academy of Sciences.

Zue, V. W. and Glass, J. R. 2000. Conversational Interfaces: Advances and Challenges. *In Proc. IEEE* 88 (8): 1166-1180.