SEMANTIC HIERARCHY AND DIALOGUE SYSTEM

Roman Mouček,
Miloslav Konopík

Department of Computer Science and Engineering,
University of West Bohemia, Univezitní 8, Pilsen, Czech Republic
moucek@kiv.zcu.cz, konopik@kiv.zcu.cz

Abstract: Semantic analysis is one of the difficult tasks, which has to be solved during development of the computerized dialogue system. There is a question how to compose the meaning of user utterance to achieve a meaning form, which is understandable to the dialogue system. Then a relevant answer to the user can be generated. The meaning of the user utterance is composed using also a semantic hierarchy in a specific domain. This semantic hierarchy serves as a base of the process of semantic representation and interpretation strategies.

Keywords: semantic analysis, semantic hierarchy, semantic role, semantic concept, dialogue system, semantic interpretation.

1. INTRODUCTION

The computerized information dialogue systems are ones of the most practical tasks, on which the theoretical knowledge background of the semantic analysis of natural spoken language can be experimentally verified. However, the first and the most important question is still if computer can really understand to human being in some way. Is it possible even if we know that understanding of two human beings is often full of communication problems and consequent explanations? Let suppose that the answer is yes, to continue in our thoughts and to face the other forthcoming problem: how to do it. In other words, how to make the human-computer interaction, which is close to human spoken language, easy and pleasant for usage by humans and also still suitable for computer processing. Or, how to represent and to process semantics of natural language in computer, which is based on von-Neumann conception?

2. ABSTRACT LEVELS OF SEMANTIC INFORMATION

Let suppose that a language expression of our thoughts can be considered as an abstract level describing our inner brain processes and world perception. This language expression is on one hand redundant and on the other hand it suffers from incompleteness and limited means to transfer all the possible information. A user of computerized information dialogue system (Fig. 1) produces such information, an utterance, which is further processed. The abstract level 1 represents a sequence of input spoken utterance rewritten as a sentence. Depending on the situation, this representation can be easily understood by another human being. However,

it can be also fully misinterpreted because of loss of prosodic and nonverbal information. Is this conversion of spoken utterance to the written sentence an applicable process or can it fully or partially disable the utterance comprehension? The research in the field of neuroscience shows that partly different regions of human brain are responsible for processing of spoken and written information. On the other side, we are not able to process the spoken utterance in the linguistic module of dialogue system directly, the conversion of spoken form to the written one is necessary.

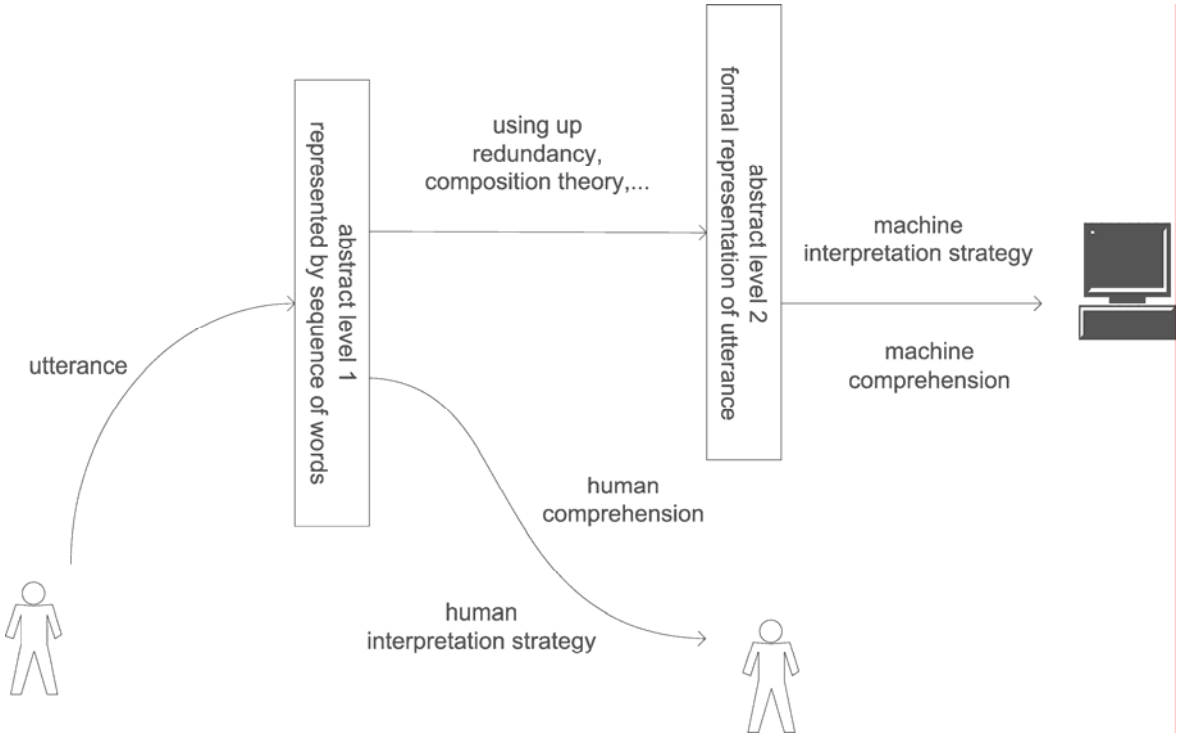

Fig. 1. Abstract levels of semantic representation

Let believe that the next processing of written sentence is still meaningful. Then we have to formalize the written sentence to build a semantic representation of the sentence and to make a semantic interpretation possible (abstract level 2). The formalized written sentence (abstract level 2) has to follow two criterions:

- it has to be finally transferable to the programming language,
- it has to contain all the necessary information.

These two criterions are often in conflict. The process of formalization is always connected with loss of information and we have to rely on redundancy of natural language (some words in a word sequence are not important to process them – they can be omitted in the process of understanding (e.g. *good morning*)). The techniques used in the process of formalization (like compositional theory, theory of microsituations, syntactic-semantic patterns, etc.) and their applicability to the computerized dialogue system are in detail described in (Mouček, 2004).

Then the abstract level 2 includes:

- existence of semantic representation of written sentence (a set of semantic concepts, e.g. *vehicle*, *station*, *bus*, etc...) – it can formally solved by the means of object oriented programming languages (classes and their instances),

- existence of semantic hierarchy – a semantic concept on a higher abstract level covers a set of concepts on a lower abstract level (e.g. *vehicle* covers *bus*, *train*, *tram*…) – it can formally solved by the means of object oriented programming languages (inheritance, abstract classes, interfaces…),
- validity of compositional theory – a semantic concept (usually on a higher abstract level) is a composition of other semantics concepts on a lower level (e.g. *time* composed from semantic concepts *tomorrow* and *late afternoon*),
- existence of compositional rules, which are finally represented with the functions (methods) of programming language – the rules describe the process of semantic interpretation (Konopík, 2004),
- existence of domains – only words and sentence parts relevant to the domain are taken into account (domains also partially solve the problem of large vocabulary, user intense, overlapping of semantic concepts, etc.)

3. SEMANTIC ROLES, CONCEPTS AND HIERACHIES

The building of semantic hierarchies is closely connected with definition of semantic roles (Mouček, 2004) for segments of elaborated sentence (words or word collocations). The number and scope of semantic roles depend especially on the domain width and selected abstract level.

Example:

Sentence (in Czech):
Prosím vás, jak to jede zítra po ránu, no spíš dopoledne, do Koterova.

Sentence (in English):
Please, how does it go tomorrow early morning, rather late morning to Koterov.

The following semantic roles are identified and the semantic concepts built:
Crequest, Qall, Cgo, Ctime, Ctime, Ccontradiction, Ctime, Cdestloc

The abstract levels of semantic roles and consequently built semantic concepts are analyzed to find their position in the semantic hierarchies. The organization of these hierarchies is also determined by the selected abstract level, domain width, perception of the situation and also by needs of semantic interpretation process.

It seems that building a semantic hierarchy is a process fully dependent on a target application. Is there any possibility to build a context-independent part of this hierarchy?

Inspecting the possibility to build general hierarchies (in the domain of city public transport and hotel accommodation) only the set of the time concepts can be considered as reusable in various kinds of domains (Fig. 2). The part of semantic hierarchy of semantic concepts, which is built for the domain of city transport, is presented as a component structure in Fig. 2. The concepts presented as rectangles are the component concepts, while the concepts presented as ellipses are the atomic concepts. The structure of semantic concept inheritance is presented in (Mouček, 2004).

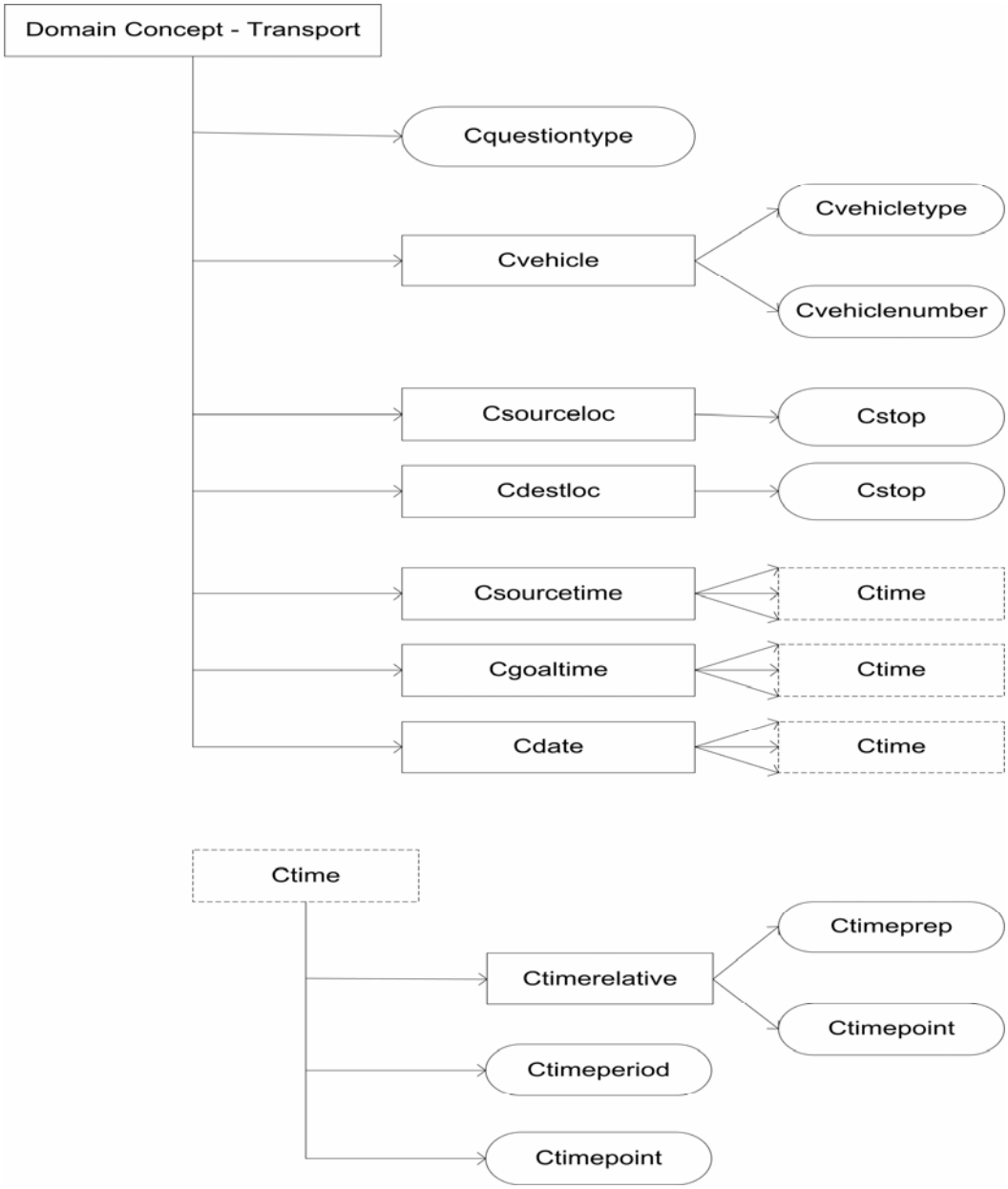

Fig 2. Component structure of city transport domain concept

4. RESULTS

Depending on the selected abstract level, building a semantic hierarchy within domain of city transport and domain of accommodation encountered serious problems. The most important are:

- meanings of semantic concepts are overlapping – it is hard to build a general semantic hierarchy, it is hard to decide the appropriate abstract level of meaning,
- compositional theory is not valid in general – there exist incoherent rules for semantic interpretation and rules built on experience with large corpus,
- meaning can be marked with a wrong semantic concept even in a small domain (e.g. *the bus of people came – the bus* marked as *vehicle,* but denotes *number*)
- the more complex sentences are not recoverable if the higher abstract level is proposed,
- the number of compositional rules can increase substantially,

- there appear circuits in the compositional process; the resulting instances form a general graph, which is hardly interpretable,
- the proposals of interpretation rules appear not to be general on the selected abstract level.

These problems turned out as the crucial to build a semantic hierarchy, and consecutively a domain concept representing a more complex user sentence, which could be successfully interpreted (domain concept is the main concept in the domain and covers all the semantic concepts and interpretation rules).

The solution to the problems described above is based on an existence of small domain expressing the clear intense of user with:

- appropriately selected abstract level of semantic roles,
- a small, mostly domain oriented semantic hierarchy,
- a small number of mostly domain oriented interpretation rules,
- a strong semantic prediction

5. CONCLUSION AND FUTURE WORK

The experiments performed on the corpus of spoken utterances have confirmed that building a domain independent semantic hierarchy would be a critical problem during construction of computerized dialogue system. A problematic selection of abstract level of semantic representation, overlapping of semantic categories and occurrence of incoherent rules for consequent semantic interpretation enables to build a real application, which is strictly domain dependent and uses up a small number of semantic concepts and interpretation rules and dialogue strategy with a strong semantic prediction. All in all, we know in general how to formally build semantic hierarchy and to propose and process interpretation rules. On the other side, we also know that meanings (semantic concepts) do not create a large, general semantic hierarchy (even a very small hierarchy is domain dependent) and the interpretation rules are dependent on these semantic concepts.

It means that we have to reconcile to the fact that a new application using the common scheme mentioned above also requires a specific proposal of semantic concepts, attributes of task concept and definition of interpretation rules. Then our future research will be based on working on more general processes rather then working on general ontology or general interpretation rules (it seems to be helpful to use an accessible computational power for them). The second way, more general, supposes a fast progress in neuroscience and new perspectives on the functioning of brain parts responsible for language and speech. Then a significant change in the approach to semantic modeling of natural language can be expected. Moreover, this change has to be probably accompanied with a significant change in used technologies.

REFERENCES

Konopík, M. (2004) Natural language recognition methods - design and implementation of semantic interpretation module (in Czech), *Thesis*, University of West Bohemia, Plzeň, Czech Republic.

Mouček, R. (2004). Semantics in Dialogue Systems, *PhD Thesis*, University of West Bohemia, Plzeň, Czech Republic.