Natural Language Semantics and Problem of Layers

Roman Mouček

University of West Bohemia, Department of Computer Science and Engineering
Pilsen, Czech Republic

E-mail: moucek@kiv.zcu.cz

Keywords: natural language semantics, abstract layer, artificial intelligence, linguistics, software engineering

Abstract: Introduction

The computerized processing of natural language semantics and a consequent model of real word concepts is one of the main unsolved tasks of informatics. This problem is partly solved e.g. in applications of artificial intelligence like human-computer dialogue systems or in applications of text classification. The most impressive idea seems to be an effort to realize a semantic web, which should replace the nowadays approach to information search.

However, all these ideas and application prototypes suffer from the enormous complexity and ambiguity of natural language semantics. As the results, the existing applications rely more on processing of morphology and syntax and take into account only a narrow part of reality. An effort to build a general ontology and to divide the processing of semantics into separate tasks faces to a lot of important problems and no model introduced till now can be considered as the suitable one.

The main goal of this paper is to introduce one of the essential problems connected with processing of natural language semantics – an effort to divide natural language semantics into several layers and process them separately.

Layers of semantics

There is no doubt that to process a natural language semantics means to process a complex system, which is closely connected with human thinking and reasoning. The large complexity of this system and our limitations in understanding to this system naturally lead to attempts to identify a set of layers in semantics, which could be processed separately to some extent. The identification of layers depends on the point of view of different fields dealing with problem of semantic processing.

Cognitive sciences differentiate between semantic, episodic, procedural, emotional etc. memories [1], which "define" different kinds of semantic information. This approach seems to be an interesting one for the possible computer processing. If there exists an episodic memory, which depends on context and autobiographical events, there is probably no sense to try to model this kind of semantic information. The number of possible inputs, which influence the content of this kind of memory, corresponds to whole life experience.

If we accept and idea that episodic memory is the essential one for human reasoning, then we should answer the question if it is useful to model the other types of memories, which have some relationship to semantics. Then only the semantic memory (context independent memory) seems to be suitable for modeling by computer tools.

Linguistics has introduced different types of meaning layers. It works e.g. with conceptual and collocation layers of semantics. Our research group tried to use and adapt some of the linguistic methods, but finally we found them out as inapplicable. On the other side, some linguists maintain that a simple and complex language sign is in fact inseparable and any effort to divide it is unreasonable.

The artificial intelligence also introduced several approaches how to divide semantics in layers. The human-computer dialogue systems usually cope with the sequence of the voice signal recognition (the output is the word lattice), linguistic analysis (syntactic and semantic analysis, the output is a linguistic formalism), the dialogue manages (dialogue history and the general context) and a speech synthesizer. The very important task is a building of ontology, very often in two abstract layers - the general one and the domain one. The final result of this approach is then that the human-computer dialogues systems operate only in narrow domains and the final semantic representation and interpretation is reduced to selection of key words relevant to elaborated domain.

Besides the problem with cooperation of different ontologies (and the necessity to reconstruct them in the case of change in the elaborated domain) there is an essential problem with formalisms. No formalism describes the natural language semantics in a way that enables to process the semantics automatically. The formalisms cannot be mapped to the natural semantics not to omit a number of expressions possible used even in a small domain and also not to produce a number of utterances without any sense.

If the methods of artificial intelligence are suitable for very specific situations, software engineering provide the methods and tools for partial semantic modeling of real world. The phases of life cycle, object oriented analysis and object oriented design provide a framework how to map the semantics of real world to the structure similar to concepts of semantic memory. Then the basic model consists of number of typed concepts, mutual associations and iterations. The principle of encapsulation provides a mechanism for interpretation of typed concepts and their features.

Conclusion

Because of similarity between the structure of semantic memory and the process of real world modeling introduced with software engineering we decided to change our attitude to modeling of natural language semantics. The used formalism is limited only by small number of basic concepts. The model is based on detection of associations and their modeling (no specific ontology is used), connection of semantic representation and interpretation under principle of encapsulation and strategy of maintaining and control of large number of small objects.

References

[1] F. Koukolík: *Já (o vztahu mozku, vědomí a sebeuvědomování)*, Charles University in Prague, Karolinum, Czech Republic, 2003.