Using of Minimal Recursion Semantic representation for linguistic information

Kanishcheva O. V.
National Technical University “Kharkov Polytechnic Institute”, olya-kanisheva@rambler.ru

In this article will discuss current approaches to the representation of linguistic information. Particular attention is paid Minimal Recursion Semantic (MRS) and the Robust Minimal Recursion Semantic (RMRS) as one of the most promising directions. Shows the use of Minimal Recursion Semantic (MRS) for the Russian language.

INTRODUCTION

The choice of a formal language to represent the linguistic knowledge is quite a challenge, since this formalism to describe not only natural language, but it should be fairly easy to implement. The development of natural-language systems, one way or another, faced with the problem of representation, storage, and interpretation of linguistic information. In this case, the representation of the data includes the tools and the formalisms used for presentation, methods of storage in the processing and interpretation of the data system [1].

There are different classification approaches to the representation of linguistic data. In this paper we shall use the following classification:

1) Approaches based on the markup (markup-based), in which additional information is stored directly in the text in the form of additional markup (HTML, SGML or XML).

2) Approaches based on the annotations (annotation-based), in which information is stored separately and contains references to the source code.

3) Approaches based on abstractions (abstraction-based), in which the text is stored only as part of a data structure, which in turn represents all information in the form based on a specific formalism.

4) Approaches, in which there are no restrictions on the representation of the data.

In this paper we consider a more detailed approach based on abstractions, because they are based on evidence establishing the structure, which is a common means to represent the linguistic information. Many formalisms for analysis using them.

Of particular interest are formalisms such as Minimal Recursion Semantic (MRS) and the Robust Minimal Recursion Semantic (RMRS). The main idea of this formalism is to convert a nested structure in the flat. Thus, the nested structure of the attributes (or predicates) can be transformed in a variety of structures (which can be combined symbols of conjunction). The formalism is an extension of RMRS MRS, which is the main difference lies in the fact that the structures of several signs are divided into binary predicates.

MINIMAL RECURSION SEMANTIC

Minimal Recursion Semantic [3], this formalism is used for semantic representation of data by means of elementary predicates. It is widely used, especially for linguistic theories (HPSG). Robust Minimal Recursion Semantic [4] is a variant of MRS. While the MRS, in foreign practice were used for manual processing grammar HPSG, RMRS is also suitable for use in surface analysis methods of textual information, including the fragmentation of phrases and stochastic analyzers that operate without a detailed glossary.

MRS – semantic representation, which uses first-order predicate logic. This is not a semantic theory based on logical formulas. MRS reduces the computational complexity for the construction of linguistic structures that preserves value for the target language.

Almost MRS was performed in English Resource Grammar, broadcoverage HPSG grammar using MRS as its semantic

representation. Another application of this formalism can be found in machine translation, statistical analysis, dialog systems, information retrieval, ontologies [2].

THE PURPOSE OF THIS WORK

The purpose of this study is to examine this formalism to represent data as Minimal Recursion Semantic. Show it to use the Russian language.

USING MINIMAL RECURSION SEMANTIC FOR SUBMISSION DATA IN RUSSIAN

An MRS representation consists of a triple, as shown in (1). This section explains all three elements and their purposes. There are, furthermore, two important notations of how to present MRSs, the standard way and as MRS graphs, which are both introduced below.

< hook , EP bag , handle constraints > (1)

The first element is the hook of the structure. It is important during the semantic of composition of complete MRSs. The second element is the EP bag. It is a set of predicates that describes the lexical and some relational semantic information contained in a sentence. The last element is a set of handle constraints that specify certain scopal relations of the elements in the EP bag.

At the heart of an MRS representation is a set of elementary predications (EP) called the EP bag. EPs are basic relations, similar to predicates in first-order logic. They normally correspond to a single lexeme, often referred to by its lemma. Every EP is marked by a label, has a relation name and a certain number of arguments, depending on the arity of the predicate. (2) shows the general notation of an EP.

label: relation(argument0, ..., argumentn) (2)

(3) presents an EP bag for the example sentence *Каждый человек вероятно любит природу*.

EP bag:
l1: каждый q (x1; h1; h2, h3),
l3: человек n (x1),
l4: вероятно adv (e1; h4), (3)
l5: любит v (e2; x1; x2),
l6: природу n (x2).

Handle constraints: h1 =q l3, h3 =q l5. Relations that describe lexical words start with an underscore, followed by the lemma of the word, followed by another underscore and the part-of-speech information. Optionally, a last underscore can separate the part-of-speech from a number that constitutes an additional sense distinction among words with the same lemma and part-of-speech.

The logical conjunction operator ^ is given a special status in the MRS formalism [5, 6]. In natural language it is generally used for composing semantic expressions, while the other logical connectives (disjunction _, etc.) only contribute to the semantics when they are lexically licensed. Also, they appear in more restricted contexts. As a consequence, EP conjunctions are made implicit by using identical labels for all members of the conjunction. Our phrase in (3) is constructed using identical labels, but note that implicit conjunctions are versatile in their potential usage.

Prepositional phrases, for example, are constructed in the same way, labeling the preposition EP with the same label as the EP it is attached to.

There are different types of variables that are used in MRS [7]. Table 1 lists all of them. Variables can have features attached to them that can carry morphological information. For example, nominal variables can have values for person, number and gender, while event variables carry tense and mood.

Table 1. Different types of variables used in the context of MRS

Variable	Usage
a	anchors uniquely identify an EP (only in RMRS)
l	labels "tag" one or more EPs
h	holes are arguments slots for embedding other EPs
x	nominal variables are introduced by nouns and adjectives
e	event variables are introduced by verbal and adverbial EPs
u	used to mark unspecified obligatory arguments
i	used to mark unspecified optional arguments

Every EP has characteristic arguments that get introduced depending on the part-of-speech.

For nouns and adjective, the first argument is always a nominal variable that stands for the nominal object.

Holes can be seen as empty slots for other EPs. By equating the holes with EP labels, a predicate logic formula with embedded predicates can be created. Such linkings are referred to as configurations or scope-resolved MRSs that represent the individual linguistic readings for a sentence described by an MRS. Possible configurations for the predicates in (3) are shown in Figure 1. The MRS itself, however, is a flat representation and avoids embedding. Moreover, it is possible configurations that can be constructed by equating holes and labels underspecified concerning the scope relations and stands for the set of all.

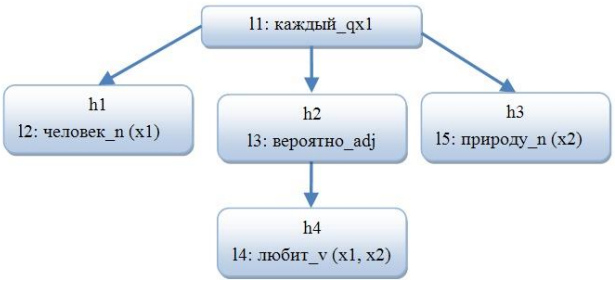

Figure 1 – Configuration for the EPs

CONCLUSIONS

This paper demonstrates the use of Minimal Recursion Semantic (MRS) for the Russian language. This study is promising, since the use of multilingual language resources structures MRS, makes the data more useful for further deep and surface processing. In the future, plans to use the formal-semantic representation to generate new knowledge, with the assistance of the algebra of predicates and predicate operations [8, 9].

RMRSs have already been used in systems for question answering, information extraction, email response, creative authoring and ontology extraction.

REFERENCES

[1] Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : учеб. пособие / Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. – М.: МИЭМ, 2011. – 272 с.

[2] Azarova I., Sinopalnikova A. Adjectives in Russnet // International WordNet Conference, GWC 2004 Brno, Czech Republic, January 20-23, 2004. P. 251-259.

[3] Copestake, Ann, Dan Flickinger, Carl Pollard and Ivan A. Sag. Minimal Recursion Semantics: an Introduction. In preparation, 1999.

[4] Ann Copestake. Applying Robust Semantics. In Proceedings of the 10th Conference of the Pacfic Assocation for Computational Linguistics (PACLING), pages 1-12, Melbourne, Australia, 2007.

[5] Copestake, Ann, Dan Flickinger, Ivan A. Sag and Carl Pollard (2005) 'Minimal Recursion Semantics: An introduction', Research on Language and Computation, 3(2–3), 281-332.

[6] Copestake, Ann, Dan Flickinger, Ivan A. Sag and Carl Pollard (2005) 'Minimal Recursion Semantics: An introduction', Research on Language and Computation, 3(2–3), 281–332.

[7] Copestake, Ann and Dan Flickinger (2000) 'An open-source grammar development environment and broad-coverage English grammar using HPSG', Proceedings of the Second conference on Language Resources and Evaluation (LREC-2000), Athens, Greece, pp. 591–600.

[8] Partee, Barbara H. and Vladimir Borschev (1998). Integrating lexical and formal semantics: Genitives, relational nouns, and type-shifting. In Proceedings of the Second Tbilisi Symposium on Language, Logic, and Computation, ed. Robin Cooper and Thomas Gamkrelidze, 229-241. Tbilisi: Center on Language, Logic, Speech, Tbilisi State University.

[9] Kanishcheva O. V. Using the algebra predicate and predicate operations for the description of natural – language relations in tasks for representations knowledge / O. V. Kanishcheva // XII International PhD Workshop OWD 2010, 23-26 October 2010. – Gliwice and Wisla, 2010. – P. 67-69.