Interfacing the Lexicon and the Ontology in a Semantic Analyzer

Igor Boguslavsky
Universidad Politécnica de Madrid
Institute for Information Transmission Problems of the Russian Academy of Sciences
igor.m.boguslavsky@gmail.com

Leonid Iomdin
Institute for Information Transmission Problems of the Russian Academy of Sciences
iomdin@iitp.ru

Victor Sizov
Institute for Information Transmission Problems of the Russian Academy of Sciences
sizov@iitp.ru

Svetlana Timoshenko
Institute for Information Transmission Problems of the Russian Academy of Sciences
nyrestein@gmail.com

Abstract

We discuss the possibility to link the lexicon of an NLP system with a formal ontology in an attempt to construct a semantic analyzer of natural language texts. The work is carried out on the material of sports news published in Russian media.

1 Introduction

Many Semantic Web applications need a much deeper semantic analysis of the text than is used today. Not only should the ontology elements be extracted from the textual data but also it is important to interpret the text in terms of the ontology. It is essential that IE and QA systems should be able to discover semantic similarity between the texts if they express the meaning in different ways. Cf. synonymous sentences (1) – (3):

(1) *Real Madrid and Barcelona will meet in the semi-finals on Thursday*.

(2) *The semi-final match between Real Madrid and Barcelona will take place on Thursday.*

(3) *The adversary of Real Madrid in the semi-finals on Thursday will be Barcelona.*

If we wish to extract the meaning from the text irrespective of the way it is conveyed, we should construct a semantic analyzer capable of producing identical semantic structures for sentences (1)-(3), or at least semantic structures whose equivalence can be easily demonstrated.

The problem becomes much more difficult if text understanding includes access to text-external world knowledge. For example, sentences (1)-(3) describe the same situation as (4).

(4) *The semi-finals on Thursday will see the champion of the UEFA Champions League 2008-2009 and the team of Manuel Pellegrini*.

To account for this synonymy, the system should know that it was the football club *Barcelona* who won the UEFA Champions League in 2008-2009, and that Manuel Pellegrini is the coach of *Real Madrid*. This implies that linguistic knowledge should be linked with ontological resources. The creation of a semantic analyzer of this type goes far beyond the task of assigning ontological classes to words occurring in the text. It requires a powerful wide-coverage linguistic processor capable of building coherent semantic structures, a knowledge-extensive lexicon, which contains different types of lexical information, an ontology, which describes objects in the domain and their properties, a repository of ground-level facts, and an inference engine.

Proceedings of the 6th Workshop on Ontologies and Lexical Resources (Ontolex 2010), pages 67–76,
Beijing, August 2010

A project NOVOFUT aiming at the development of a semantic analyzer of this type for Russian texts has started at the Institute for Information Transmission Problems of the Russian Academy of Sciences. It covers the domain of news about football. There are several reasons for this choice of domain. First, the news texts are written primarily for the general public, so that their understanding does not require specialized expert knowledge. This is a major advantage since it significantly facilitates the acquisition of the ontology. Second, the language typical of sports journalism is rich enough, which makes its interpretation linguistically non-trivial. Last but not least, sports enjoy enormous public interest. There are many sports portals publishing multifarious information on the daily (and sometimes hourly) basis and visited by a lot of people. Enhanced Question-Answering and Information Extraction in this domain are likely to attract many users.

The NOVOFUT semantic analyzer reuses many types of resources created or accumulated by the team in previous work. In this paper we focus on the static resources used by the analyzer – the lexicon and the ontology. The plan of the presentation is as follows. In Section 2 we discuss related work. In Section 3 we will briefly describe the linguistic processor we build on and its lexicon. Section 4 outlines a small-scale ontology developed for the project. The correlation between natural language words as presented in the lexicon and the ontology is the main concern of Section 5. In Section 6 the interface between the ontology and the lexicon is discussed. Future work is outlined in Section 7.

2 Related work

The link between the ontologies and NL texts is investigated in two directions – "from the ontology towards NL texts" and "from the texts towards the ontology". In the first case written texts are used as a means for ontology extension and population. To name but a few authors, McDowell and Cafarella (2006) start from an ontology and specify web searches that identify in the texts possible semantic instances, relations, and taxonomic information. In (Schutz and Buitelaar 2005) an interesting attempt is made to extract ontological relations from texts. (Buitelaar et al. 2008, Magnini et al. 2006, Maynard & al. 2006) are further advances in the direction of ontology population.

Finding NL equivalents to ontological elements and be monolingual or multilingual. A metamodel for linking conceptual knowledge with its lexicalizations in various languages is proposed in (Montiel-Ponsoda et al. 2007).

The second direction research starts from NL texts and tries to interpret them in terms of the ontology. In most cases, this takes the form of marking the text with ontological classes and instances. A typical example is (Sanfilippo et al. 2006). One should also mention the work based on the Generalized Upper Model (GUM), which is meant for interfacing between domain models and NL components (Bateman et al. 1995)

Our work belongs to this second direction, but our aim is not limited to finding ontological correlates to words. In many aspects we were inspired by the ontological semantic approach developed in the Mikrokosmos framework (cf. Nirenburg and Raskin 2004). We share many of its postulates and concrete solutions. In particular, semantic analysis of the text should be based on both linguistic and extra-linguistic knowledge. Linguistic knowledge is implemented in language grammars and dictionaries, while extra-linguistic knowledge is comprised in an ontology, which enumerates concepts, describes their properties and states interrelationships between them, and a fact repository which accumulates ground-level facts, such as, in our case, the data about concrete players, teams and matches. To a large extent, the ontology serves as the semantic language for meaning representation.

At the same time, there exist some differences between our approaches determined by the linguistic model adopted. Our work is based on the Meaning ⇔ Text theory (Mel'čuk 1974, 1996). In particular, we make extensive use of lexical functions, which constitute one of the prominent features of this theory. Thanks to lexical functions it turns out possible to reduce a wider range of synonymous sentences to the same semantic

structure, and in many cases, improve the performance of search engines (see e.g. Apresjan et al. 2009). Another difference between the Mikrokosmos approach and ours concerns the fact that the Mikrokosmos ontology is written in a specific in-house formalism. Our emphasis is on using as far as possible standard ontology languages (OWL, SWRL), in order to obtain interoperability with a wide and ever growing range of semantic web resources and inference engines.

3 The ETAP-3 Linguistic Processor and its Lexicon.

The multifunctional ETAP-3 linguistic processor, developed by the Computational Linguistics Laboratory of the Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, (see e.g. Apresjan et al. 2003), is the product of decades of research and development in the field of language modelling.

At the moment, ETAP-3 consists of a number of options, including

1) a rule-based machine translation system working both ways between Russian and English (plus several prototypes for other languages – French, German, Spanish, Korean and Arabic);

2) a system of synonymous and quasi-synonymous paraphrasing of sentences;

3) an environment for deep annotation of text corpora, in which SynTagRus, the only corpus of Russian texts tagged morphologically, syntactically (in the dependency tree formalism), and lexically was created, and

4) a Universal Networking Language (UNL) module, responsible for automatic translation of natural language text into a semantic interlingua, UNL, and the other way around.

The ETAP-3 processor is largely based on the general linguistic framework of the Meaning ⇔ Text theory by Mel'čuk. An important complement to this theory was furnished by the theory of systematic lexicography and integrated description of language proposed by Jurij Apresjan (2000).

One of the major resources used in ETAP-3 is **the combinatorial dictionary**. It offers ample and diverse data for each lexical entry. In particular, the entry may list the word's syntactic and semantic features, its subcategorization frames, as well as rules (or reference to rules) of a dozen types, which make it possible to describe peculiar behavior of individual words and exceptions to general rules in a complete and consistent way. Many dictionary entries contain information on **lexical functions,** to be discussed below in some detail.

The entry of the combinatorial dictionary has a number of zones, one of which provides the properties of the word that are manifested in the given language, while all the other zones contain information on the match between this word and its equivalent in a particular language. For example, the EN zone in the Russian combinatorial dictionary entry contains information on the translational equivalents of the respective Russian word into English. One field (TRANS) gives the default single-word translation (or several such translations) of this word in English. Other fields contain less trivial translation rules, or references to such rules.

A newly introduced ONTO zone offers information underlying the match between the Russian word and its counterparts in the ontology.

4 Ontology of football.

The ontology we are working with focuses in the first place on football. It contains information on teams, players, football field, sport events, and their properties. However, we want it to be extendable to other sports as well. That is why some classes are more general than would be needed for football alone. For example, instead of having one class `FootballPlayer`, the ontology has a more general class `Sportsman`, of which `FootballPlayer` is a subclass. An equivalence restriction states that `FootballPlayer` is a `Sportsman` whose SportType is `football`. In this way, sportsmen doing different types of sports can be treated by the ontology in a uniform way.

The football ontology is written in SWRL, which is OWL augmented with rules (Horrocks et al. 2004). In compiling it, we used some existing ontologies dealing with foot-

ball (http://www.lgi2p.ema.fr/~ranwezs/ontologies/soccerV2.0.daml). As usual, properties of classes are inherited by the subclasses. For example, the `Match` class is a subclass of `SportEvent`, which in its turn is a subclass of `Event`. `Match` inherits from `Event` the properties of having definite `Time` and `Place`. From `SportEvent` it inherits the fact that its participants should be `Sport-Agents`. Its own properties are: the number of participants is 2 (as opposed to championships, which have more) and it has a definite sport type (as opposed to Olympics, which involve many sport types). A subclass of `Match` is `Derby`, in which both participants should be from the same city or region. This property is implemented by means of a SWRL rule (cf. below). Another rule assigned to `Match` states that if its sport type is football (or any other team sport), then its participants should be teams and not individual sportsmen, as is the case in tennis or chess. `Sportsman` is a subclass of two classes: `Person`, from which it inherits the property of having a name and a birth date, and `SportAgent`, which includes also `Team` and from which it inherits the property of having a definite sport type and a coach.

5 Correlation between the words and the elements of the ontology.

As mentioned above, the ontology plays a two-fold role. On the one hand, it is a repository of domain-specific knowledge, and on the other hand, it is a semantic metalanguage used for representing the meaning of natural language texts. All meaningful natural languages elements (words, grammatical constructions, morphological features) must be interpreted in ontological terms. This makes the correlation between the lexicon and the ontology far from trivial. In this section, we will present several typical situations and illustrate them with examples.

5.1 One-to-one correspondence between NL words and ontology elements.

The simplest situation occurs when a word directly corresponds to an ontology element – a class, an individual, a relation, an attribute or its value. Example:

(5) *Real Madrid pobedil Arsenal so sčetom 3:1* 'Real Madrid defeated Arsenal 3 to 1'.

Here *Real Madrid* and *Arsenal* are individuals – instances of the `Team` class, the verb *to defeat* corresponds to the `WinEvent` class, and numbers *3* and *1* are values of attributes `scoreWinner` and `scoreLoser`, respectively. In the semantic structure (SemS) classes are represented by instances supplied by a unique numeric identifier. SemS for sentence (5) looks as follows:

```
hasWinner(WinEvent01, Real-
Madrid)&hasLoser(WinEvent01,
Arsenal)& scoreWinner
(WinEvent01,3)&
scoreLoser(WinEvent01,1)
```

5.2 One ontology element – several words ("multi-word concepts").

This is a very typical situation, especially in the area of terminology. For example, in ordinary language *željtaja kartočka* 'a yellow card' is a free noun phrase that simply denotes a card whose colour is yellow. In the sports domain, it is a single concept that refers to one of several possible punishments a referee can mete out for a rules infraction. Therefore, it is represented as one element in the ontology. Some other examples of multi-word sport concepts: *uglovoj udar* 'corner', *svobodnyj udar* 'free kick', *pravyj poluzaščitnik* 'right tackle'.

5.3 One word – several ontological elements.

Many words that can be interpreted in terms of ontological elements do not correspond to any single class, relation or instance. Their definition consists in a configuration of these elements. Most often, it is a class with some of the properties instantiated. In principle, often there are two options: one can either postulate two different classes (e.g.

Sportsman and FootballPlayer as its subclass), or only one class (Sportsman) and represent the *football player* as a Sportsman whose SportType property is football. There is no general solution to this alternative. In general, it is desirable to obtain a parsimonious ontology and refrain from introducing new classes, if one can define a concept in terms of existing classes and their properties. However, if a concept has important properties of its own, it is preferable to present it in the ontology as a separate class. In our example, FootballPlayer has Specialization which other sportsmen do not have (goalkeeper, forward, back, etc.) For this reason, it is postulated as a separate class of the ontology, with the indication of its equivalence to the anonymous class "Sportsman and hasSportType football".

An interesting and typical case are adjectival modifiers to nouns of the type *ispanskij* 'Spanish', *francuzskij* 'French', *moskovskij* 'of Moscow', and the like. Usually, dictionaries provide a very general definition of such words. For example, the COBUILD English dictionary gives only one meaning for the adjective *Spanish*: 'belonging or relating to Spain, or to its people, language or culture'. However, in real life situations this word is often interpreted by the speakers in a more specific way, according to the context. Ontological description should try, as far as possible, to take contextual factors into account and make explicit the specific interpretation of the modifier in each particular case. Sometimes, it can be done by means of rules that refer to ontological categories. For example, the meaning of the adjective *ispanskij* 'Spanish' mentioned above, when applied to geographical objects (*rivers, cities, mountains, roads,* etc.), narrows down to (hasLocation Spain). If this adjective modifies a noun denoting an industrial or agricultural product (*car, wine, olive oil,* etc.), it is rather interpreted as (producedIn Spain). We will hardly understand the phrase *Spanish wine* as denoting the wine located in Spain. Textual objects (*songs, literature, poetry,* etc.) move the adjective towards denoting the Spanish language: (inLanguage Spanish). Of course, these rules are not always sufficient for disambiguation. If an object falls into more than one category, several interpretations are possible. In particular, a book is both a textual and a consumer object. Therefore, *a Spanish book* can be interpreted as a book written in Spanish, and as a book published in Spain.

In many cases, adjectives serve as arguments of nouns. The semantic role of this argument may be different for different nouns (cf. (6-8)), and even for the same noun (cf. (9-11)):

(6) *presidential decree* – 'the president issued a decree': hasAgent(decree,president);

(7) *presidential elections* – 'somebody elects the president': hasObject(elect,president);

(8) *Babylonian invasion* – 'Babylon invaded some city or country': hasAgent(invade,Babylon); but not 'some city or country invaded Babylon': hasObject(invade,Babylon);

(9) *economic advisor* – 'advises in the area of economics': hasTopic(advisor,economics);

(10) *American advisor*: hasNationality(advisor,USA);

(11) *presidential advisor* – 'advises to the president': hasAddressee(advisor,president).

5.4 A word is interpreted in ontological terms but does not have any fixed ontological equivalent.

There is a large class of words that denote individuals which are in a certain relation to other individuals: *brother, sister, uncle, wife, friend, enemy, classmate, co-author, co-regent, coeval, adversary, ally,* etc. Of course, these words can be easily represented as ontology properties: hasBrother(John, Bill), hasSister(John, Mary). However, such representation does not reveal the meaning of the concepts. Being a brother of somebody means being a male and having common parents with this person. This meaning shares the second component ('having common parents') with the property of being a sister and

differs from it in the first component ('being a male'). Such a definition of meanings requires the use of variables. This is the point where the OWL expressive capacity is insufficient and one has to recur to SWRL rules:

```
Person(?person1)&Gender (?person1,male)&hasParent(?person1, ?person3) &hasParent (?person2,?person3)→ brother(?person1, ?person2)

Person(?person1)&Gender (?person1,female)&hasParent(?person1, ?person3)& hasParent (?person2,?person3)→ sister(?person1,?person2)
```

In a similar way one can define the concept of *adversary* (in sports), as used for example in sentence (3) above. *Adversary of Z* is someone different from Z who plays in the same match as Z:

```
SportAgent(?agent)& Match(?match)& hasParticipant(?match,?agent)& hasParticipant(?match,?z)& differentFrom(?agent,?z) → adversary(?agent,?z)
```

Among the words that require variables for their ontological definition are not only relational nouns. There are many other words that cannot be translated into ontological categories without claiming identity (or difference) of the properties of some individuals. Here are some examples from the football domain.

A *derby* is a match whose participants are from the same city or region. Our ontology defines the concept of *derby* as follows:

```
hasParticipant(?match, ?participant1)& hasParticipant (?match, ?participant2)&differentFrom (?participant1,?participant2) &hasLocation (?participant1, ?location) &hasLocation (?participant2,?location) → derby(?match)
```

Pobednyj gol ('decisive goal') is a goal which was scored when both teams had equal score and which was the last goal in the match. However, since having no subsequent goals cannot be expressed in SWRL we will convey this idea by saying that the goal brought the victory in the match to one of the teams. We will need the following classes and properties:

`GoalEvent`, with the properties: `hasAgent`, `atMinute`, e.g. *on the tenth minute*, `inMatch`, `hasResult` (inherited from the more general class `Event`).

`SituationInMatch` (the score at a given moment), with the properties: `inMatch`, `atMinute`, `scoreParticipant1`, `scoreParticipant2`.

`WinEvent`, with the properties: `hasWinner`, `hasLoser`.

`Team`, with the property `hasPart`, to be filled by instances of `Sportsman`.

Besides that, we need the property `timeImmediatelyBefore`, inherited by moments of time from `Time`.

We will describe the situation by means of two rules. Rule (12) says that the goal that brought a victory can be called a decisive goal. Rule (13) complements this description by saying that if a goal brings a victory, the winner is the team whose player scored it and this goal was scored at the moment when both teams had equal score.

(12)
```
GoalEvent(?goal)&WinEvent (?victory)& hasResult(?goal,?victory) → decisiveGoal(?goal)
```

(13)
```
hasResult(?goal,?victory) &hasAgent(?goal,?player)& hasPart (?team,?player)&atMinute (?goal,?min0)&inMatch (?goal, ?match)& timeImmediatelyBefore(?min1,?min0)& SituationInMatch (?situation)&inMatch (?situation,?match)& atMinute (?situation,?min1) → hasWinner(?victory,?team) &scoreParticipant1(?situation, ?n) &scoreParticipant2(?situation,?n).
```

5.5 Ontology and Lexical Functions.

A lexical function (LF), in the Meaning ⇔ Text theory (Mel'čuk 1996), has the basic properties of a multi-value mathematical

function. A prototypical LF is a triple of elements {***R,*** X, Y}, where ***R*** is a certain general semantic relation obtaining between the argument lexeme X (the keyword) and some other lexeme Y which is the value of ***R*** with regard to X (by a lexeme in this context we mean either a word in one of its lexical meanings or some other lexical unit, such as a set expression). Here are some examples for the $Oper_1$ and $Oper_2$ functions: $Oper_1$ (*control*) = *exercise* (*control*), $Oper_1$ (*research*) = *do* (*research*), $Oper_1$ (*invitation*) = *issue* (*an invitation*), $Oper_1$ (*doubt*) = *have* (*doubts*), $Oper_1$ (*defeat*) = *suffer* (*a defeat*), $Oper_1$ (*victory*) = *gain* (*a victory*), $Oper_1$ (*campaign*) = *wage (a campaign),* $Oper_2$ (*control*) = *be under* (*control*), $Oper_2$ (*analysis*) = *undergo* (*an analysis*), $Oper_2$ (*invitation*) = *receive* (*an invitation*), $Oper_2$ (*resistance*) = *encounter* (*resistance*), $Oper_2$ (*respect*) = *enjoy* (*respect*), $Oper_2$ (*obstacle*) = *face (an obstacle)*.

Y is often represented by a set of synonymous lexemes Y_1, Y_2, …., Y_n, all of them being the values of the given LF ***R*** with regard to X; e. g., Magn (*desire*) = *strong / keen / intense / fervent / ardent / overwhelming*. All the LF exponents for each word are listed in the lexicon.

LFs have a strong potential for advanced NLP applications. Apresjan *et al.* (2007) shows how LFs can be used in parsing, machine translation, paraphrasing. In parsing, LFs are used to resolve or reduce syntactic and lexical ambiguity. The MT system resorts to LFs to provide idiomatic target language equivalents for source sentences in which both the argument and the value of the same LF are present. The system of paraphrasing automatically produces one or several synonymous transforms for a given sentence or phrase by means of universal LF-axioms; for example: *He respects* [X] *his teachers – He has* [$Oper_1(S_0$ (X))] *respect* [S_0 (X)] *for his teachers – He treats* [Labor1-2(S_0 (X))] *his teachers with respect – His teachers enjoy* [$Oper_2(S_0(X))$] *his respect*. It can be used in a number of advanced NLP applications ranging from machine translation to authoring and text planning.

In ontologically-oriented semantic analysis different LFs are reflected in different ways.

An LF corresponds to an ontological class.

Many LFs represent bundles of words that are semantically identical or very close and therefore can serve as representatives of this common meaning. We illustrate this with two closely related LFs (Apresjan *et al*. 2008).

The meaning covered by $LiquFunc_0$ is 'to cause to cease to exist or to be taking place'. This concept corresponds, in particular, to the following English verbs: *to stop (the aggression), to lift (the blockade), to dispel (the clouds), to demolish (the building), to disperse (the crowd), to avert (the danger), to cure (the disease), to close (the dispute), to break up (the family), to annul (the law), to dissolve (the parliament), to denounce (the treaty), to bridge (the gap).* Another LF of the Liqu family – $LiquFact_0$ – refers to a different kind of elimination. It means 'to cause to cease functioning according to its destination'. When somebody *closes the eyes*, they do not cease to exist, they only stop functioning. Some more examples: *shut down (the factory), stop (the car), land (the airplane), depose (the king), switch off (the lamp), neutralize (the poison), empty (the bucket).*

These LFs, along with several dozen others, play a significant role not only in text understanding and generation. They contribute in an interesting way to one of the crucial functions of ontologies – inference of implicit knowledge. Important inference rules can be easily formulated in terms of LFs: if the blockade is lifted (=$LiquFunc_0$), it does not exist any more. Another example of the LF-based inference (this time it is LF $Real_1$): *He fulfilled* (= $Real_1$) *the promise to buy a bicycle* → *He bought a bicycle.*

It should be emphasized that, given a lexicon which contains LF data (which is the case of our ETAP dictionary), the acquisition of this part of the ontology is straightforward.

An LF generates an ontological relation.

This case can be illustrated by support verbs of the Oper-Func-Labor family that attach one of the arguments to the noun. For example, in sentence *Father gave me an advice* the subject of the $Oper_1$-support verb *to give* (*father*) is the Agent of *advice*, while in *The proposal received much attention* the subject

of the $Oper_2$-support verb *to receive* (the proposal) is the Object of *attention*. Other examples of $Oper_1$ and $Oper_2$ were given in 5.5 above. Some examples of other LFs of this family:

$Func_1$: *(fear) possesses (somebody), (rumour) reaches (somebody), (the blame) falls on (somebody) / (the blame) lies with (somebody), (control) belongs to (somebody), (responsibility) rests with (somebody).*

$Func_2$: *(proposal) consists in (something), (criticism) bears upon (something), (revenge) falls upon (somebody).*

$Labor_{1-2}$: *keep (something) under (control), submit (something) to (analysis), meet (somebody) with (applause), put (somebody) under (arrest), hold (somebody) in (contempt), bring (something) into (comparison with something), take (something) into (consideration).*

An LF has no ontological correlate.

This is the case of $Func_0$. This LF neither denotes a concept, nor attaches an argument to a concept. It only duplicates the meaning of its keyword and has no correlate in the SemS. For example, in sentence (2) above the phrase *the match took place* (= $Func_0$) is only represented by the concept `Match`. Other examples of $Func_0$: *(the snow) falls, (the wind) blows, (the danger) exists, (the war) is on, (changes) occur.*

6 Lexicon ↔ Ontology interface.

For the purposes of semantic analysis, the Russian dictionary and the ontology are linked in the same way as dictionaries of different languages are linked in Machine Translation options of the ETAP-3 system. As noted in Section 2, if the system performs translations from language *L* to language *L′*, all dictionary entries of *L* contain a special zone (ZONE: L′) where all translation variants of the given word into *L′* are recorded. The semantic analysis option uses the ONTO zone of the Russian dictionary. In this zone, two types of information may be written:

- **Default translation**. This is a one-word equivalent of the given word, which is used if no translation rule is applicable.

For example, Russian *komanda* 'team' has the `Team` class as its ontological counterpart. This is written in the ontological zone of *komanda* as follows:

ZONE: ONTO
TRANS: `Team`

Names of ontological individuals are also often translated by default.

- **Translation rules.** A rule is written every time one needs to carry out an action which does not boil down to the default translation.

Let us give several examples of translation rules written in the ONTO zone of the Russian lexicon. We will not give their formal representation and restrict ourselves to explaining what they are doing in plain words.

Pobeditel' 'winner' is a SportAgent (i.e. a sportsman or a team) that won some contest:

```
SportAgent(?x)&WinEvent(?y)&
hasWinner(?y,?x).
```

Phrases of the type *komanda NN* 'team of NN' (where NN is a proper human name in the genitive case) are translated in four different ways depending on the ontological information assigned to NN.

(a) If NN is the name of a player, the phrase is represented as 'the team of which NN is a player': *komanda Arshavina* 'Arshavin's team' = `Team(?team) &hasPart(?team,Arshavin).`

(b) If NN is the name of a coach, the phrase is represented as 'the team of which NN is the coach': *komanda Pellegrini* 'Pellegrini's team' = `Team(?team) &hasCoach(?team,Pellegrini)`

(c) If NN is the name of a captain, the phrase is represented as 'the team of which NN is the captain': *komanda Iraneka* 'Iranek's team' = `Team(?team)& hasCaptain(?team,Iranek)`

(d) If NN is neither a player, nor a coach, nor a captain, the phrase is represented as 'the team of which NN is a fan': *komanda Ivana* 'Ivan's team' = `Team(?team)& hasFan(?team,Ivan)`

It is well-known that genitive noun phrases (or phrases "N1 of N2" in English) are very vague semantically, and their interpretation is very much dependent on the context. This example shows that even within the

part/whole interpretation such a phrase, paradoxically, has two opposite varieties: either N2 is part of N1, as in *the team of Arshavin/Arshavin's team*, or N1 is part of N2, as in *the leg of the table*.

The following examples involve the property `hasLocation`, which characterizes both sport events (*The match took place in Madrid*), and sport agents (*the Ukrainian sportsman, a London club*).

Frequently, a football match is played in a location, such that one of the teams is from that location while the other is not. This situation can be represented by the following SemS:

(14)

```
Match(?match)&hasLocation(?match,?place)&hasParticipant
(?match, ?team1)&hasParicipant(?match,?team2)&differentFrom(?team1,?team2)&hasLocation(?team1,?place)&¬hasLocation(?team2,?place)
```

In the natural language this situation can be viewed from different angles and denoted by different words.

Xozjaeva 'home team' denotes a team that plays a match in a place it is from, the adversary being from a different place. *Gosti* 'visitors' is a team that plays a match in a location different from the place it is from, the adversary being the home team. *Prinimat'* 'to receive' means to play a match being a home team, to host it. *Igrat' v gostjax* lit.'to play being guests' means to play a match away.

Although all these words correspond to the same situation (14), their translation rules cannot be identical. The rules should not only introduce SemS (14), but also assure correct amalgamation of this SemS with SemSs of other words. In particular, the rule for *prinimat'* 'receive' should guarantee that in (15) Real Madrid instantiates variable `?team1` of (14), and Barcelona – variable `?team2`.

(15) *Real Madrid prinimal Barcelonu* 'Real Madrid hosted Barcelona'

The rule for *gosti* 'visitors' should see to it that in (16) `hasWinner` property of `WinEvent` be filled by variable `?team2` of (14):

(16) *Gosti vyigrali 3:1* 'the visitors won 3 to 1'

This is assured due to marking `?team2` in the *gosti* 'visitors' rule as the head element of SemS (14). Naturally, in the *xozjaeva* 'home team' rule the same role is assigned to `?team1`.

7 Future work.

In the continuation, it is planned to enlarge both the ontology and ONTO zone of the Russian lexicon. We are investigating the possibility of merging our small football ontology with some existing larger upper level ontology. The difficult task will be to unify our semantic rules with the axioms of this ontology.

A second direction of our future activity is connected with another component of the semantic analyzer, which we did not touch upon in this paper. It is the set of semantic rules which are not incorporated into the lexicon due to their general character. This component also requires significant enhancement.

An important extension of this work consists in introducing an inference component based on the SWRL rules.

Acknowledgement

This study has received partial funding from the Russian Foundation for Humanities (grant No. 10-04-00040a), which is gratefully acknowledged.

References

Apresjan, Ju. D. Systematic Lexicography. Oxford University Press, 2000, XVIII p., 304 p.

Apresjan, Jury, I. Boguslavsky, L. Iomdin, A. Lazursky, V.Sannikov, V.Sizov, L.Tsinman. ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the MTT // *MTT 2003, First International Conference on Meaning – Text Theory (June 16-18 2003).* Paris: Ecole Normale Supérieure, 2003. P. 279-288.

Apresjan, Jury, Igor Boguslavsky, Leonid Iomdin and Leonid Tsinman. Lexical Functions in Actual NLP Applications // Selected Lexical and Grammatical Issues in the Meaning–Text Theory. In honour of Igor Mel'čuk. (Ed. by Leo Wanner). John Benjamins, Studies in Language Companion Series 84. 2007. P. 199-230.

Apresjan, Ju.D., P.V. Djachenko, A.V. Lazursky, L.L. Tsinman. O kompjuternom uchebnike russkogo jazyka. [On a computer textbook of Russian.] *Russkij jazyk v nauchnom osveshchenii*. 2008, No. 2 (14). P. 48-112.

Apresjan Ju., I. Boguslavsky, L.Iomdin, L.Cinman, S.Timoshenko. Semantic Paraphrasing for Information Retrieval and Extraction. In: T.Andreasen, R.Yager, H.Bulskov, H.Christiansen, H.Legind Larsen (eds.) Flexible Query Answering Systems. Proceedings, 8th International Conference, 2009. Lecture Notes in Computer Science 5822. pp. 512-523

Bateman J., B. Magnini and G. Fabris. The Generalized Upper Model Knowledge Base: Organization and Use. In: Towards Very Large Knowledge Bases, pp. 60-72, IOS Press. 1995.

Buitelaar P., Ph. Cimiano, A.Frank, M. Hartung, S.Racioppa. (2008). "*Ontology-based Information Extraction and Integration from Heterogeneous Data Sources*." In: International Journal of Human-Computer Studies, 66(11).

Horrocks, I., P.F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof and M. Dean (2004). *SWRL: A Semantic Web Rule Language Combining OWL and RuleML*. // W3C Member Submission 21 May 2004.

Magnini B., Emanuele Pianta, Octavian Popescu, and Manuela Speranza. (2006). "*Ontology Population from Textual Mentions: Task Definition and Benchmark.*" In: Proceedings of the Ontology Population and Learning Workshop at ACL/Coling 2006.

Maynard D., Wim Peters, and Yaoyong Li. (2006). "*Metrics for Evaluation of Ontology-based Information Extraction.*" In: WWW 2006 Workshop on Evaluation of Ontologies for the Web (EON 2006)

McDowell Luke K., Michael Cafarella. Ontology-driven Information Extraction with OntoSyphon. In: Proceedings of the 5th International Semantic Web Conference (ISWC 2006). Volume 4273 of LNCS., Athens, GA, Springer (2006) 428 – 444

Mel'čuk, Igor. *Opyt teorii lingvisiticheskix modelej «Smysl ↔ Text»* [The theory of linguistic models of the Meaning ↔ Text type]. Moscow, 1974 (2nd edition 1999).

Mel'čuk, Igor. Lexical Functions: A Tool for the Description of Lexical Relations in Lexicon. *L. Wanner (ed.), Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam, 1996, 37-102.

Montiel-Ponsoda, E., Aguado de Cea, G. y Gómez-Pérez, A. 2007 "Localizing ontologies in OWL. *From Text to Knowledge: The Lexicon/Ontology Interface. WS 2. The 6th International Semantic Web Conference*."

Nirenburg, Sergei, and Victor Raskin. *Ontological Semantics.* The MIT Press. Cambridge, Massachusetts. London, England, 2004.

Sanfilippo A., Tratz S., Gregory M., Chappell A., Whitney P., Posse Ch., Paulson P., Baddeley B., Hohimer R., White A. Automating Ontological Annotation with WordNet. In: Sojka P., Key-Sun Choi, Ch. Fellbaum, P. Vossen (Eds.): GWC 2006, Proceedings, pp. 85–93.

Schutz, A. and P. Buitelaar. RelExt: A Tool for Relation Extraction from Text in Ontology Extension. In: Y. Gil et al. (Eds.), ISWC 2005, LNCS 3729, pp. 593–606, 2005.