

A Knowledge-rich Lexicon for Bulgarian*

Kiril Simov

Linguistic Modelling Laboratory, Bulgarian Academy of Sciences, Bulgaria
kivs@bultreebank.org

Abstract. In contrast to the morphological and syntactic processing, the semantic annotation, based on ontology, is still underdeveloped for Bulgarian. On the other hand, the prerequisites for an ontological annotation are already available. These are as follows: a morpho-syntactic tagger for Bulgarian with more than 95% accuracy; a dependency parser with more than 84% accuracy; a general chunker and a named entity grammar. Therefore, the next logical step is the semantic annotation. As a minimal set of semantic resources we consider the following ones:

- a lexicon for Bulgarian aligned to an upper ontology as a mechanism to cover the common lexica in domain texts, and aligned to domain ontologies to cover domain terminology;
- a corpus, annotated with ontology information in order to train machine learning component for automatic word sense disambiguation;
- an annotation grammar for Bulgarian, based on syntactic knowledge of Bulgarian and conceptual information from the ontology.

In this paper, we will focus on the description of the lexicon.

1 Introduction

Semantic Annotation (Tagging) is a natural further development in the area of language resources after the creation of morphologically and syntactically annotated corpora. The importance of Semantic Annotation became a hot topic within the initiative for creation of Semantic Web. Although much work is already done in the area, the term "semantic annotation" is not yet well defined - see [8] and citation therein. In our work we consider the text as consisting of two types of information: (1) ontological classes and relations, and (2) world facts. The ontological part determines generally the topic and the domain of the text. We call the corresponding "minimal" part of ontology implied by the text ontology of the text. The world facts represent an instantiation of the ontology in the text (here higher order entities like beliefs, claims, etc. are also included). Both types of information are called uniformly 'semantic content of the text'. Both components of the semantic content are connected to the syntactic structure of the text. Any (partial) explication of the semantic content of a text will be called semantic annotation of the text. Defined in this way, the semantic annotation could contain also some pragmatic information and actual world knowledge.

In order to support this kind of semantic annotation we rely on a knowledge-rich lexicon to determine the content of the semantic annotation. The lexicon is aligned to an upper ontology which covers the general meanings of the lexical items. In addition to the upper ontology the lexicon might be aligned to domain ontologies in order to support more precise domain annotation. In the paper a special focus is put on the role of the regular polysemy and metonymy. They are encoded as special patterns extracted from a semantically annotated corpus and reflecting the conceptual structure of the ontology. The lexicon is also connected to an annotation grammar which establishes a relation between the ontology and the text. In this paper we will not discuss the grammar and the annotation process.

The structure of the paper is as follows: the next section discusses the structure of a domain ontology, its connection to an upper ontology; the third section provides a model of ontology-to-text relation which is a motivation for the creation of the a knowledge-rich lexicon of Bulgarian;

* This work is partially supported by LTfLL project (Language Technology for Lifelong Learning - IST-212578).

the next section discusses the extensions of the ontology-to-text relation with respect to general lexica and coverage of some phenomena; the fifth section compares our work with some other works; and the last section concludes the paper.

2 The Structure of the Domain Ontology

Independently from the methodology for ontology creation, the end result has the following structure:

- Domain layer. At this layer we have the real domain concepts and relations representing the main notions in the domain. These concepts and relations are used in solving different tasks such as representation of domain knowledge, representation of common conceptualization for information exchange in the domain, semantic annotation of domain texts, etc.
- Upper layer. The alignment of the domain layer to an upper ontology is an obligatory step in each ontology creation methodology. This alignment ensures several properties of the domain ontology: (1) consistency with the design of the upper ontology; (2) inheritance of the knowledge represented in the upper ontology.
- Middle layer. This layer contains concepts and relations which are not part of the upper or the domain layers, but play important role for the alignment between them.
- Language layer. It is supposed that the domain ontology (together with middle and upper layers) is language independent, formalized in some ontology representation language. In practise such an ontology needs has to be aligned to some language resources. This is necessary in order the ontology to be presented to users who do not know much of ontology and to support analysis of texts. As a minimum it is necessary to have a lexicon aligned to the concepts and the relations in the ontology.

We have used this structure of the ontology in three European projects - LT4eL, AsIsKnown and LTfLL. In each of them we have used as an upper ontology DOLCE Ontology [10] for several reasons: (1) it is constructed on rigorous basis which reflects the OntoClean methodology [6]; (2) it is represented in OWL-DL; (3) the authors of the ontology provide us comments and help on the alignment of the domain ontology to DOLCE. For the middle layer we have used OntoWordNet [4] - a version of WordNet aligned to DOLCE. OntoWordNet facilitates the alignment between the upper ontology and domain layer. This is ensured by providing more understandable concepts (more specific and closer to the domain) and the mapping between the concepts is easier. In the middle layer we include from OntoWordNet only those concepts that are necessary to support the alignment between the domain layer and the upper layer. The domain layer is created for each domain. The result of three layers is a domain ontology with a better structuring of the concepts and relations. Also relations and axioms are inherited from DOLCE to the domain layer.

Language layers in each of the projects were created on the basis of the model of the ontology-to-text relation presented in the next section.

3 Ontology-to-Text Model

In this section we represent the two main components that define the ontology-to-text relation. These components are: lexicon and concept annotation grammar.

The lexicon plays twofold role in our architecture. First, it interrelates the concepts in the ontology to the lexical knowledge used by the grammar in order to recognize the role of the concepts in the text. Second, the lexicon represents the main interface between the user and the ontology. This interface allows for the ontology to be navigated or represented in a natural for the user way. For example, the concepts and relations might be named with terms used by the users in their everyday activities and in their own natural language (e.g. Bulgarian). This could be considered as a first step to a contextualized usage of the ontology in a sense that the ontology could be viewed through different terms depending on the context. For example, the color

names will vary from very specific terms within the domain of carpet production to more common names used when the same carpet is part of an interior design. Thus, the lexical items contain the following information: a term, contextual information determining the context of the term usage, grammatical features determining the syntactic realization within the text. In the current implementation of the lexicons the contextual information is simplified to a list of a few types of users (producer, retailer, etc). With respect to the relations between the terms in the lexicon and the concepts in the ontology, there are two main problems: (1) there is no lexicalized term for some of the concepts in the ontology, and (2) there are lexical terms in the language of the domain which lack corresponding concepts in the ontology, which represent the meaning of the terms. The first problem is overcome by writing down in the lexicon also non-lexicalized (fully compositional) phrases to be represented. These different phrases or terms for a given concept are used as a basis for construction of the annotation grammar. Having them, we might capture different wordings of the same meaning in the text. The picture below shows the mapping varieties. It depicts the realization of the concepts (similarly for relations and instances) in the language. The concepts are language independent and they might be represented within a natural language as form(s) of a lexicalized term, or as a free phrase. In general, a concept might have a few terms connected to it and a (potentially) unlimited number of free phrases expressing this concept in the language¹. Some of the free phrases receive their meaning compositionally regardless of their usage in the text, other free phrases denote the corresponding concept only in a particular context. In our lexicons we decided to register as many free phrases as possible in order to have better recall on the semantic annotation task. In case of a concept that is not-lexicalized in a given language we require at least one free phrase to be provided for this concept. The following picture shows the mapping from the ontology to the lexicon:

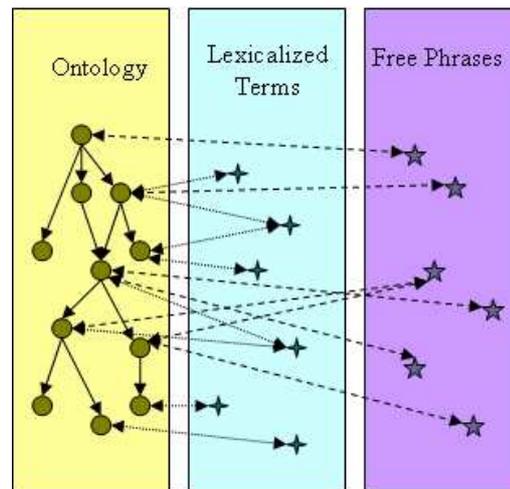


Fig 1. Ontology-to-Lexicon Relation.

The picture depicts the realization of the ontological concepts in a natural language. The concepts are language independent and they might be represented within a natural language as form(s) of a lexicalized term (or item), or as a free phrase. In general, a concept might have a few terms connected to it and a (potentially) unlimited number of free phrases expressing this concept in the language. Some of the free phrases receive their meaning compositionally regardless their usage in the text, other free phrases denote the corresponding concept only in a particular context.

In order to solve the second problem (missing concept for a lexical item) we modify the ontology in such a way that it contains all the important concepts for the domain. However, this solution

¹ The presence of free phrases in the lexicon is also motivated by the fact that the lexicalization is not a discrete feature. There are many different degrees of lexicalization. Thus the free phrases are the extreme end of the scale.

requires a special treatment of the "head words" in the lexicons, because such phrases allow bigger freedom with respect to their occurrences in the text. Variability is a problem even with respect to the lexicalized cases and the idea is to represent the most frequent (based on a corpus) variants for each concept. The specific solutions for the lexical terms without appropriate concept in the ontology are the following:

More detailed classes in the ontology. In cases where it is possible, we are creating more specific concepts in the ontology. For example, the concept of 'shortcut' in the domain of Computer Science for End Users, is denoted by different lexical items in English depending on the operating system, because each operating system (MS Windows, Linux, etc) as a rule introduces its own terminology. When the notion is borrowed in other languages, it could be borrowed with different granularity, thus, we introduce more specific concepts in the ontology in order to ensure correct mapping between languages.

More complex mapping exists between the ontology and terms in some language. Our initial idea was that each meaning of a lexical item in any language is mapped to exactly one concept in the ontology. If for some lexical item this one-to-one mapping is not appropriate or it requires very complicated changes in the ontology, we realize a mapping based on ontology expressions instead of a single concept. This mechanism allows us to keep the ontology simpler and more understandable, and to handle cases that do not allow appropriate mappings. Currently, such cases are not detected in domains for which we applied this model.

We could summarize the connection between the ontology and the lexicons in the following way: the ontology represents the semantic knowledge in form of concepts and relations with appropriate axioms; and the lexicons represent the ways in which these concepts can be realized in texts in the corresponding languages. Of course, the ways in which a concept could be represented in the text are potentially infinite in number, thus, we could hope to represent in our lexicons only the most frequent and important terms and phrases. Here is an example of an entry:

```
<entry id="entry-34">
  <owl:Class rdf:about="http://www.asisknown.org/AIKHT#CarpetOWN">
    <rdfs:comment>a piece of thick heavy fabric (usually with nap or pile)
used to cover a floor</rdfs:comment>
    <rdfs:subClassOf>
      <owl:Class rdf:about="http://www.asisknown.org/AIKHT#FloorCovering"/>
    </rdfs:subClassOf>
  </owl:Class>
  <def>a piece of thick heavy fabric (usually with nap or pile)
used to cover a floor</def>
  <termg lang="en">
    <term shead="1">carpet</term>
    <term>carpeting</term>
    <term>rug</term>
    <term type="nonlex">textile floor covering</term>
    <def>a piece of thick heavy fabric (usually with nap or pile)
used to cover a floor</def>
    <gramline>reference to finite state grammar</gramline>
  </termg>
</entry>
```

Each entry of the lexicons contains the following types of information: (1) information about the concept from the ontology which represents the meaning for the terms in the entry; (2) explanation of the concept meaning in English; (3) a set of terms in a given language that have the meaning expressed by the concept; and (4) relation to grammar rules. The concept part of the entry provides minimum information for formal definition of the concept. The English explanation of the concept meaning facilitates the human understanding. The set of terms stands for different wordings of the concept in the corresponding language. One of the terms is the representative for the term set.

Note that this is a somewhat arbitrary decision, which might depend on frequency of term usage or specialist's intuition. This representative term will be used where just one of terms from the set is necessary to be used, for example as an item of a menu. In the example above we present the set of English terms for the concept 'carpet'. One of the terms is non-lexicalized - attribute **type** with value "nonlex". The first term is representative for the term set and it is marked-up with attribute **shead** with value "1". The elements **gramline** provide links to linguistic features of the terms like lemmatized variants of the terms, implementation as regular expressions to be compiled as finite state automata, etc.

Here we present a (part of) DTD for the lexicon:

```
<!ELEMENT OntoLexicon (entry+)>

<!ELEMENT entry
      ((owl:Class|rdf:Description|rdf:Property), def, termg+)>

<!ELEMENT def (#PCDATA)>

<!ELEMENT termg (term+,def?,gramline*)>
<!ATTLIST termg
      lang (bg|cs|de|en|fr|hu|it|mt|nl|pl|pt|ro|ru) # REQUIRED
>

<!ELEMENT term (\#PCDATA)>
<!ATTLIST term
      type (lex|nonlex)          "lex"
      shead (1|0)                "0"
      gram CDATA                 #IMPLIED
>

<!ELEMENT gramline (#PCDATA)>
```

The lexicon consists of entries. Each entry consists of a concept, relation or instance (partial) definition, followed by a definition of the concept content in English and one or several term groups. Each term group represents all the available lexical terms or free phrases for the corresponding concept (relation or instance) in a given natural language (determined by the attribute lang). Optionally, the term group for a given language could contain a definition of the content of the concept in that language. Each term represents a normalized form of the term. Additionally, we could state whether: the term is a lexicalization of the concept in the language or it is a free phrase (attribute type); the term is representative for the concept in the language (the attribute shead) or not; and which grammar rules recognize this term (related to the concept (relation or instance) of the entry) in text. The format of the currently implemented grammars is given below.

The second component of the ontology-to-text relation, the concept annotation grammar, is ideally considered as an extension of a general language deep grammar which is adopted to the concept annotation task. Minimally, the concept annotation grammar consists of a chunk grammar for concept annotation and (sense) disambiguation rules. The chunk grammar for each term in the lexicon contains at least one grammar rule for recognition of the term. As a preprocessing step we consider annotation with grammatical features and lemmatization of the text. The disambiguation rules exploit the local context in terms of grammatical features, semantic annotation and syntactic structure, and also the global context such as topic of the text, discourse segmentation, etc. Currently we have implemented chunk grammars for several languages. We have implemented a very simple disambiguator which uses an unigram model.

For the implementation of the annotation grammar we rely on the grammar facilities of the CLaRK System [13]. The structure of each grammar rule in CLaRK is defined by the following DTD fragment:

```

<!ELEMENT line (LC?, RE, RC?, RM, Comment?) >
<!ELEMENT LC (#PCDATA)>
<!ELEMENT RC (#PCDATA)>
<!ELEMENT RE (#PCDATA)>
<!ELEMENT RM (#PCDATA)>
<!ELEMENT Comment (#PCDATA)>

```

Each rule is represented as a line element. The rule consists of a regular expression (RE) and a category (RM = return markup). The regular expression is evaluated over the content of a given XML element and could recognize tokens and/or annotated data. The return markup is represented as an XML fragment which is substituted for the recognized part of the content of the element. Additionally, the user could use regular expressions to restrict the context in which the regular expression is evaluated successfully. The LC element contains a regular expression for the left context and the RC for the right one. The element Comment is for human use. The application of the grammar is governed by XPath expressions which provide additional mechanism for accurate annotation of a given XML document. Thus, the CLaRK grammar is a good choice for the implementation of the initial annotation grammar.

The creation of the actual annotation grammars started with the terms in the lexicons for the corresponding languages. Each term was lemmatized and the lemmatized form of the term was converted into regular expression of grammar rules. Each concept related to the term is stored in the return markup of the corresponding rule. Thus, if a term is ambiguous, then the corresponding rule in the grammar contains reference to all concepts related to the term.

The following picture depicts the relations between lexical items, grammar rules and the text:

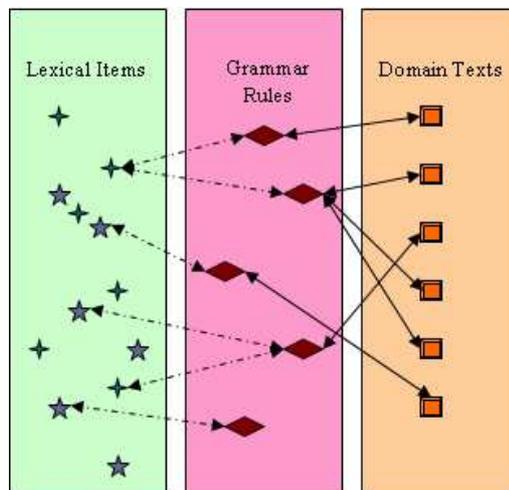


Fig 2. Lexicon-to-Text Relation.

The relations between the different elements of the models are as follows. A lexical item could have more than one grammar rule associated to it depending on the word order and the grammatical realization of the lexical item. Two lexical items could share a grammar rule if they have the same wording, but they are connected to different concepts in the ontology. Each grammar rule could recognize zero or several text chunks.

The relation ontology-to-text implemented in this way provides facilities for solving different tasks, such as ontology search (including crosslingual search), ontology browsing, ontology learning.

In order to support multilingual access to semantic annotated corpus we have to implement the relation for several languages using the same ontology as starting point. In this way we implement a mapping between the lexicons in these languages and also comparable annotation of texts in them.

We have been using the relations between the various elements for the task of ontology-based search. The connection from ontology via lexicon to grammars is relied on for the concept annotation of the text. In this way we established a connection between the ontology and the texts. The relation between the lexicon and the ontology is used for definition of user queries with respect to the appropriate segments within the documents. The annotation of texts in different languages on the basis of the same ontology could facilitate the definition of similarity metrics between such texts.

4 A Knowledge-rich Lexicon of Bulgarian

The main problem with the model of the ontology-to-text relation, described in the previous section, is the fact that the annotation of domain texts with domain concepts is very sparse. For example, in the domain of Computer Science for End Users we have annotated 8 concepts within 100 tokens (with 14.8 tokens per sentence = 1.19 concepts per sentence at average). This sparse annotation blocks possibilities for using better methods for word sense disambiguation. This holds when the lexical items in the domain lexicon are ambiguous among themselves or with respect to the general lexica. For example, the concepts 'key-of-keyboard', 'key-of-database' and 'key-for-door' have the same wording in English and the last one is not from the domain ontology.

We consider two solutions to this problem: (1) better annotation grammar, and (2) Interaction with general lexica. The first can be done by exploiting coreferential relations and lexical chains. The second via connection to lexicons like WordNet. In order to benefit from these solutions, we have to tune them to the model of ontology-to-text relation. First, in order to construct lexical chains and coreferential relations in which the domain terms in the text to participate we need these terms and the surrounding general lexica to share their semantic annotation. In order to ensure this we have to align the general lexica with appropriate semantic information.

Ideally, each meaning of the general lexicon has to be presented in the ontology in order to use the model of ontology-to-text relation from the previous section. Unfortunately such an ontology does not exist yet. Thus, we have to use a smaller ontology and to change the implementation of the ontology-to-text relation.

From our experience within the projects mentioned above we can conclude that there exist a relatively stable upper and middle part of each of the domain ontologies. Thus, for the creation of an appropriate lexical resource for semantic annotation we consider as a first step the building of an upper-middle layer ontology which to provide the necessary semantic information for the tasks of word sense disambiguation. In our case this is a mixture of DOLCE and the upper part of OntoWordNet. Such an ontology can be used for several tasks: (1) representation of general meaning of lexical items in a language; (2) basis for construction of domain ontologies and lexicons.

In the previous model we have used `equality` relation between the conceptual information in the ontology and the meaning of the corresponding lexical items. In this new lexicon this will not be possible because there will be no enough concepts in the ontology. Thus, the first difference from the previous model is that we will allow also the relation `subsume` to be used. The lexicon entry for each lexical item will specify what the relation is between the meaning of the lexical item and the corresponding concept. The requirement for the mapping via `subsume` relation is as follows: the concept that is used with in the ontology to be the most specific one available.

In addition to the mapping to the ontology we want to represent also information necessary for some of the more important phenomena for the task of word sense disambiguation: polysemy, metonymy² and verb representation. The first two phenomena - polysemy and metonymy are treated in similar way. First of all, the word senses are represented in the ontology. Thus, the lexical representation is done via appropriate mappings to corresponding concepts in the ontology.

² The treatment of metaphorical uses are recorded as separate entries in the lexicon.

Let us consider the case of metonymy. In general, metonymy is defined as a trope in which one entity is used to stand for another associated entity³. Thus, we can consider metonymy to be encoded via a composition of ontology relations encoded in the lexicon. For example, let us suppose that we have to annotate the sentence "She was wearing stripe." First we annotate 'stripe' as a kind of a **property** and as such it is connected to 'cloth' via **property-of** relation and 'cloth' is annotated as **material** and it is connected to 'clothing' via the **made-of** relation. The concept 'clothing' is of the relevant type for the object of the verb 'to wear'. Thus, the understanding of the sentence is something like: "She was wearing a clothing made from a textile with a stripe design." The composition of the corresponding relations is stored in the lexical entries for the corresponding lexical items. In the case of metonymy this is a better option, because the possible patterns are (potentially) infinite in number. Representing each metonymy usage as a separate meaning will result in many strange meanings for the lexical items. In this way we separate the most frequent metonymy uses as inference patterns and the actual inference during the analysis of the discourse where the lexical item is used metonymically. Similarly we treat the polysemy. The different meanings are represented in the ontology as different concepts and these concepts are connected via appropriate relations. The main difference here is that for each of the meaning we construct a separate lexical entry. Thus, always during the analysis of the text we have to disambiguate between these senses. In some cases more than one of the senses are visible via one usage of the lexical item. For example, in the sentence "This large book is very interesting." the word 'book' is used simultaneously as a **physical object** selected by 'large' and as an **information object** selected by 'interesting'.

Encoding of verbs is also very important for the task of semantic annotation. We assume that the appropriate information is also represented in two ways: (1) in the ontology each verb is connected to concept representing the event related to the meaning of the verb. In the ontology all the participants (irrespectively whether they are considered as arguments, adjuncts, etc.) are represented as such via appropriate relations; (2) the linguistic behavior is encoded in the lexicon as a set of frames. These frames determine the role of each participant in the a given event.

The actual lexicon is under construction. It is based on several machine-readable dictionaries: a Morphological Dictionary, a Valence dictionary and an Explanatory Dictionary of Bulgarian. The selection of the lexical items is on the basis of construction of the lexicon aligned to the upper and middle parts of the ontology where we encoded about 3000 lexical entries. The rest of of lexical items are selected on the basis of their ranking in a large Bulgarian corpus (72 million running words from BulTreeBank text archive). The ranks are calculated via automatic morphosyntactic analysis of the corpus and then lemmatization. For each lemma we consider the frequency in the corpus and in how many documents the lemma occurs.

5 Discussion

The need of a knowledge-rich lexicon of Bulgarian is motivated by the need to introduce more world knowledge in the semantic analysis of the text. As it was mentioned in [7], the most lexical relations necessary to determine the semantic content of the lexical items are non-classical in contrast the classical ones, i.e. **hyponymy**, **meronymy**, **antonymy**. The non-classical relations are specific for some classes of meanings, i.e. **made-of**, **used-for**, etc. In our case we assume that these relations are represented in the ontology. Thus, they are formally defined, can be used in inference process and can be used for representation of some language phenomena like polysemy, metonymy, etc.

From point of view of the complexity and precision of ontology according to Nicola Guarino ([5]) we have the following classification of ontologies:

- Lexicon: *Machine Readable Dictionaries; Vocabulary with NL definitions*
- Simple Taxonomy: *Classifications*
- Thesaurus: *WordNet; Taxonomy plus related-terms*

³ <http://www.sil.org/linguistics/GlossaryOfLinguisticTerms/WhatIsMetonymy.htm>

- Relational Model: *Light-weight ontologies; Unconstrained use of arbitrary relations*
- Fully Axiomatized Theory: *Heavy-weight ontologies.*

The classification starts with less formal and knowledge-poor ontology — simple lexicons and ends with heavily constrained theories about the world. Our attempt is to move the current semantic lexicons from the level of thesaurus to the level of light-ontologies (as a minimum).

Our approach gains in many respects from such works as WordNet [3], EuroWordNet [14], SIMPLE [9]. The mapping between the language specific lexicons was facilitated by the ontology. Our model shares common features with other lexicon models: with WordNet-like ([3]; [14]) lexicons we share the idea of grouping lexical items around a common meaning and in this respect the term groups in our model correspond to synsets in WordNet model. The difference in our case is that the meaning is defined independently in the ontology. With SIMPLE model [9] we share the idea to define the meaning of lexical items by means of the ontology, but we differ in the selection of the ontology which in our case represents the domain of interest, and in the case of SIMPLE reflects the lexicon model. With the LingInfo model ([1]; [2]; [12]) we share the idea that grammatical and context information also needs to be presented in a connection to the ontology, but we differ in the implementation of the model and the degree of realization of the concrete language resources and tools. At the end we would like to mention the work on Ontology Semantics ([11]) which is very similar to our model except that we use existing ontologies like DOLCE and we allow for an incremental construction of the lexicon.

6 Conclusion

In this paper we presented a further developed model for ontology-to-text relation connecting the conceptual information in an ontology to the lexical items and grammatical rules for realization of this information in texts. We started with domain ontologies and lexicons and then extended the model and their coverage to general lexica. The model represents also phenomena like polysemy, metonymy, verbal frames. The resulting lexicon will ensure better semantic annotation of texts. Our future goals are to implement a system for automatic word sense disambiguation, metonymy usage discovery. Also, the lexicon together with the ontology could be used for the creation of domain ontologies and lexicons.

References

- [1] Buitelaar P., Declerck Th., Frank An., Racioppa St., Kiesel M., Sintek M., Engel R., Romanelli M., Sonntag D., Loos B., Micelli V., Porzel R., Cimiano Ph. (2006) LingInfo: Design and Applications of a Model for the Integration of Linguistic Information in Ontologies. In: *Proceedings of OntoLex06, a Workshop at LREC*, Genoa, Italy.
- [2] Buitelaar, P., Sintek, M., and Kiesel, M. (2006). A Lexicon Model for Multilingual/Multimedia Ontologies In: *Proceedings of the 3rd European Semantic Web Conference (ESWC06)*, Budva, Montenegro.
- [3] Fellbaum Chr. (1998). Editor. *WORDNET: an electronic lexical database*. MIT Press.
- [4] Gangemi, A., Navigli, R., and Velardi, P. (2003). The OntoWordNet Project: extension and axiomatization of conceptual relations in WordNet. In: *Meersman R, et al. (eds.), Proceedings of ODBASE03 Conference*, Springer.
- [5] Guarino N. (2000) *Invited Mini-course on Ontological Analysis and Ontology Design*. First Workshop on Ontologies and lexical Knowledge Bases - OntoLex 2000. Sozopol, Bulgaria.
- [6] Guarino, N., and Welty, C. (2002). *Evaluating Ontological Decisions with OntoClean*. Communications of the ACM, 45(2): 61-65.
- [7] Morris J. and Graeme Hirst Gr. (2004). Non-Classical Lexical Semantic Relations. In: *Proceedings of the HLT Workshop on Computational Lexical Semantics*. Boston, Massachusetts, USA. pp 46–51.
- [8] Kiryakov At., Popov B., Terziev Iv., Manov D., and Ognyanoff D. (2005). *Semantic Annotation, Indexing, and Retrieval*. Elsevier's Journal of Web Semantics, Vol. 2, Issue 1.
- [9] Lenci A., Busa F., Ruimy N., Gola El., Monachini M., Calzolari N., Zampolli A., Guimier E., Recourcé G., Humphreys L., von Rekovsky U., Ogonowski A., McCauley Cl., Peters W., Peters Iv., Gaizauskas R., Villegas M. (2000). *SIMPLE Work Package 2 - Linguistic Specifications, Deliverable D2.1*. ILC-CNR, Pisa, Italy.
- [10] Masolo, C., Borgo, S., Gangemi, A., Guarino, N., and Oltramari, A. (2002). *Ontology Library (final)*. WonderWeb Deliverable D18, December 2003. <http://www.loa-cnr.it/Publications.html>.
- [11] Nirenburg S. and Raskin V. (2004). *Ontological Semantics*. MIT Press.
- [12] Romanelli, M., Buitelaar, P., and Sintek, M. (2007). Modeling Linguistic Facets of Multimedia Content for Semantic Annotation. In: *Proceedings of SAMT07 (International Conference on Semantics And digital Media Technologies)*, Genova, Italy, pp 240-251.
- [13] Simov, K., Peev, Z., Kouylekov, M., Simov, A., Dimitrov, M., Kiryakov, A. (2001). CLaRK - an XML-based System for Corpora Development. In: *Proceedings of the Corpus Linguistics 2001 Conference*. Lancaster, UK.
- [14] Vossen P. (1999). Editor. *EuroWordNet General Document. Version 3, Final, July 19, 1999*. <http://www.hum.uva.nl/ewn>