

# Bulgarian Language Resources for Ontology-Based Semantic Search

Kiril Simov, Petya Osenova  
Linguistic Modeling Department,  
Institute for Parallel Processing,  
Bulgarian Academy of Sciences  
25A Acad. G. Bonchev St.  
Sofia 1113, Bulgaria  
{kivs|petya}@bultreebank.org

## Abstract

This paper presents the language resources, which would facilitate the ontology-based semantic search. Some of these resources are language independent, such as the domain ontology. Some depend on the specific language: terminological lexicons, annotation grammars, sense disambiguation rules, gold standard corpus.

Here we focus on the Bulgarian resources constructed in two domains for supporting the semantic annotation and subsequently – the semantic search.

## Keywords

Ontology, Ontology-based Lexicon, Semantic Annotation, Semantic Search, Bulgarian, Annotation Grammars

## 1. Introduction

For many tasks related to Natural Language Processing and Semantic Web the semantic search in multimedia documents is a prerequisite. But reliable automatic semantic annotation systems do not exist for many languages. Their creation depends in many respects on the construction of gold standard corpora and related language resources. In two European projects - LT4eL<sup>1</sup> (Language Technology for eLearning) (see [7]) and AsIsKnown<sup>2</sup> (A Semantic-Based Knowledge Flow System for the European Home Textiles Industry) - we have created new or adapted already existing language resources to support ontology-based semantic search within domain specific documents. The sets of documents are different for these projects - learning objects on Computer Science for non-specialists in the case of LT4eL, and fashion textile magazines in the case of AsIsKnown. Here we mainly present the language resources for Bulgarian, which are a prerequisite for creating of such corpora and for their subsequent applications. It is necessary to stress that our work within both projects is by rule multilingual (including at least two languages), since the ontologies which we developed have to be shared among our partners, and also the results have to be

distributed to a wider audience. In both cases we adapt an ontology development methodology which is inherently incremental. We start as a first step with the creation of corpora containing representative texts within the domains, annotated with semantic information. As an obligatory part, the initial corpora contain also definitions of the meanings of the terms (or phrases) identified within the corpora. On the basis of these corpora we construct a first version of the ontologies. They are equipped with terminological lexicons and annotation grammars. The lexicons are the mediators between the ontologies and the end users. The grammars mediate the link between ontologies and domain texts. Then the grammars are used for re-annotation of the corpora. The result is verified manually and, as a result, new concepts are added to the ontologies, the lexicons are extended with new lexical items, the grammars are made more precise. In this methodology we follow the ideas of [18] which stresses the importance of semantically annotated corpora for ontology evaluation and development.

As a first step to the semantic annotation we consider the standard linguistic pre-processing of documents, such as tokenization, morphosyntactic annotation, chunking, named entity recognition. This pre-processing has been done in accordance with the approach we have developed within Bulgarian treebank [1]. Needless to say, the tools had to be adapted to the domain-based task. The adaptation required addition of new rules and vocabulary. Special attention was paid to the new senses of words and terms in the domain texts, since their common meaning happens to be also frequently used in such texts.

Thus, the ideal situation for an adequate ontology-based semantic annotation would be the interaction between the domain ontology, the terminological lexicon(s) and the annotation grammars. The ontology represents the relevant concepts in a domain and the relations between them. The ontology is connected to the lexicons. The concepts and the relations in the ontology represent the semantic elements which we annotate the documents with. For each concept/relation in the lexicon we represent lexicalized terms and non-lexicalized free phrases. Thus, our lexicon comprises lexicalized terms as well as free paraphrases of

---

<sup>1</sup> <http://www.lt4el.eu/>

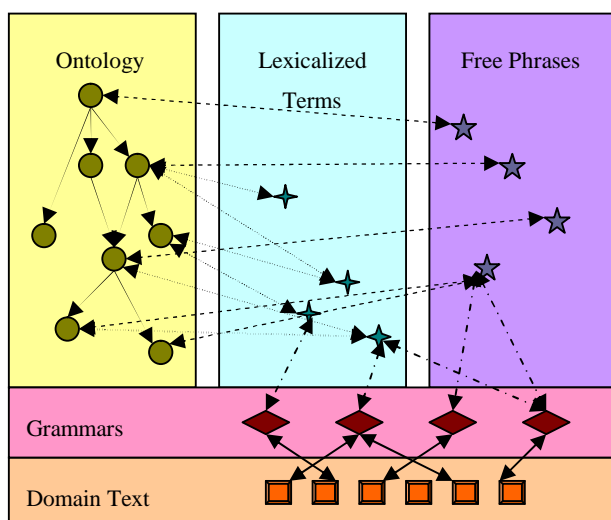
<sup>2</sup> <http://www.asisknown.org/>

the concept in question. The grammars reflect the mapping between the terms/phrases in the lexicon and their realization in the text.

The structure of the paper is as follows: in the next section we discuss our language model; then the creation of the initial corpora in some domain is described. In Section 4 the general idea behind the domain ontologies is discussed. Section 5 presents the methodology of constructing terminological lexicons. In Section 6 the regular grammars are presented as an annotation tool, which connects the ontological concepts via lexicons with the text segments. Section 7 describes the interaction among all the components with respect to their improvement. The last section concludes the paper and outlines the future work.

## 2. Linguistic Model

In this section we present briefly the linguistic model of the annotation process adopted for the tasks within the two projects. We assume that the ontology is the repository of the lexical meaning of the language. Thus, we have started with a concept in the ontology and we searched for lexical items and non-lexical phrases that convey the content of the concept. There are two possible problems here: (1) there is no lexical item for some of the concepts in the ontology, and (2) there are lexical items in the language without a concept representing the meaning of the lexical item in the ontology. The first problem is overcome by allowing in the lexicon also non-lexical (fully compositional) phrases to be represented. The second problem is solved by extension of the ontology. The lexicon items are then mapped to grammars. These grammars relate the lexicon to the text. This mapping is necessary as much as lexical items and phrases from the lexicons allow for multiple realizations in the text and require some additional linguistic knowledge in order to disambiguate between different meanings of some lexical item or phrase. The following figure depicts the elements of the model.



We have been using the relations between the different 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. On this basis a multilingual search strategy was implemented within LT4eL project [16].

Our approach gains in many respects from such works as WordNet [11], EuroWordNet [12], SIMPLE [13]. In spite of the fact that we employ the experience from these projects (mapping to WordNet and Pustejovsky's ideas in SIMPLE), we also suggest an alternative for the connection between the ontology and the lexicons. Our model is very close to LingInfo model (see [15] and [8]) not only with respect to the mapping of the lexical items to concepts, but also with respect to the other language processing tools we connect to the ontology – the concept annotation grammars and concept disambiguation tools. As to WordNet and EuroWordNet, we differ in the direction of the workflow, i.e. we start from ontology to the lexicon, not vice versa. From SIMPLE we differ in using a domain ontology instead of a general linguistic ontology. From LingInfo model we differ in the fact that the three components (ontology, lexicon and grammars) are represented by different models. For detailed discussion see [17].

In the next sections we present the realization of the model for Bulgarian within the framework of the two projects.

## 3. Creation of the Initial Corpora

For most tasks in NLP a creation of the so-called gold corpora is required. Here we will discuss the content and format of the domain corpora as well as the pre-processing steps.

In the domain of Computer Science for non-specialists we gathered learning objects in Bulgarian in several sub-domains, such as Web technology, Microsoft products. The size of this corpus is 207 000 tokens. In the domain of textile we gathered texts from Bulgarian fashion magazines of interior design such as “Домът” (The home), etc. It comprises 50 000 tokens. Some of the materials have also an English translation.

In both domains the original files came in various formats, such as HTML, Word, PDF. However, the markup language of processing is XML. Thus, an appropriate converter is always needed to ensure the transfer of the text data. Moreover, the tasks of both projects are to work with a dynamic database of documents. Also, the appropriate DTDs were constructed with respect to the level of processing. When in appropriate format for further processing, the files underwent several pre-steps: token

segmentation, morphological tagging, morphosyntactic disambiguation, chunking, named-entity recognition, sentence delimitation. All these tasks were performed automatically or semi-automatically, re-using the Basic language resource kit (BLARK) of BulTreeBank Project (see Technical Report 1<sup>3</sup>). Additionally, a lemmatizer was added to connect the wordforms with the base forms.

Depending on the goals of both projects, next steps of processing showed some difference. In LT4eL project the texts were also annotated manually with keywords and definitions. This was needed for testing the automatic keyword extractor and glossary finder for Bulgarian. In case of AsIsKnown Project, we manually constructed a Treebank of definitions taken from the lexicons and standards. This Treebank was used as a source for extraction of knowledge from syntactic patterns as well as testing the automatic tools against it. The nouns in the definitions were connected to appropriate concepts in the domain ontology (including synsets from OntoWordNet). Then appropriate relations were defined over the syntactic structure of the trees within the treebank.

The main problems we faced when performing the preprocessing were of two kinds: 1) due to the quality of the collected domain texts, and 2) due to the specificities of the domain. Concerning the first problem, the most annoying error that caused noise to the tokenizer was the usage of Latin letters in Cyrillic words, and vice versa. As for the second problem, we had to handle the domain specific multiwords, and to extend the dictionary with domain specific terms or meanings. For example, in LT4eL domain there are sequences like *TCP/IP*, *JPEG/SPIFF* or *XHTML 1.0*. Sometimes the translation from English also puzzled the standard Bulgarian tokenizer. For example, complex segments like: *Не-Е-Число* (not-is-number) or *запиши-веднъж-чети-много-пъти* (write-once-read-many-times).

## 4. Domain Ontology

We explore two domain ontologies, created by us within the two projects. As it was mentioned above, the domain of LT4eL Project is “Computer Science for Non-Computer Scientists”. To put it in a more detailed way, it covers topics like operating systems; programs; document preparation – creation, formatting, saving, printing; Web, Internet, computer networks; HTML, websites, HTML documents; email, etc. The domain of AsIsKnown Project is “Home Textile”. It covers topics like carpets, curtains, damasks, cloths, etc. and their related features. This ontology represents features related to both the production of home textile and the usage of it. The main application of the LT4eL ontology has to do with the indexing of Learning Objects within the domain in order to support search for

learning topics. The AsIsKnown ontology supports two main tasks: (1) unification of different conceptualizations in the domain; and (2) annotation of articles from fashion magazines in the area of interior design. The first task is necessary for the interchange of information between the users of AsIsKnown system. Here they need to synchronize their internal conceptualization to the conceptualization of the ontology. The second task is necessary for the analysis of trends (one of the services supported by AsIsKnown system). For trend analyses the search for concept co-occurrences plays an important role.

In general, the domain ontology is supposed to consist of the main concepts in a certain domain. It has to be connected to an upper ontology, from which to inherit the appropriate properties. For the two projects we chose the DOLCE upper ontology - [2], [3]. The mapping to it was done via OntoWordNet ([10]) and Wordnet 2.0 ([11]). The mapping to DOLCE ensures that the domain ontologies satisfy the requirements of OntoClean methodology (see [9]). The mapping to WordNet facilitates also a mapping to SUMO (see [14]). However, domain ontologies are still a resource that cannot be taken for granted, and hence, a strategy is needed to construct such an ontology. Within the two projects we used the following ways of deriving a good network of concepts: in LT4eL we annotated learning object on the required topic with keywords. For each of these keywords we gathered definitions in English from the web. We also provided Bulgarian definitions via translation or search on the web. Afterwards concepts were created on the top of the meanings, explicated by the definitions. In the case of AsIsKnown at the beginning terminological lexicons and standards in the textile domain were used. Then, magazines were also processed to get the terms on the tendencies of interior design.

From the above considerations it became clear that the resources we needed concerning Bulgarian language were keyword lists and terms in the domains. For the distribution of Bulgarian keywords within LT4eL project see [4]. As for AsIsKnown, we initially worked preferably with English sources. However, some Bulgarian interior design magazines were also processed.

Here we must note that the ontology is meant to be a language-independent resource. As such, it relies on the correct presentation of the various meanings. In a multilingual context the ontology has to handle the discrepancy between the conceptualization of the knowledge and its lexicalization in languages. Sometimes there is no lexicalized term for a concept in some language, but the concept itself should be presented in the ontology. In such a case we either select a term naming from a list of terms (when available), or we construct a name of meaningful words that altogether do not form a good expression in the language in question. We use English as the typical lingua franca among languages. But the name of

---

<sup>3</sup> At: <http://www.bultreebank.org/TechRep/BTB-TR01.pdf>

the concept might be in any natural language, because what really matters is the definition. For example, in the ontology there is a concept **AuthorityFile** with the meaning: *A file of authorized cataloging entries for names, subjects, and series. The file gives the correct form of these entries, and, in the case of series, decisions on shelving and access within the library.* However, in Bulgarian there is no lexicalised term, and a free phrase is used: *файл с права* ('file with rights').

Some problems arise also from the fact that the definitions, in general, reflect various aspects of the concept, but very often they are either too general, or too specific. Sometimes there are not enough domain definitions per ambiguous concepts, which leads to availability of beyond-domain interpretations only. In these cases the intervention of the ontology creator is required, and he/she adds the domain meaning.

## 5. Terminological Lexicons

The terminological lexicons can be viewed as lists of the main keywords in a certain domain. An important drawback of the standard lexicons is that they consist of lexicalized terms. However, in real domain texts there are much more paraphrases of the existing lexicalized term(s), or even worse - there are free phrases to a notion, which is not lexicalized in the language. Thus, an extension of the lexicons is needed, which to capture the synonymy beyond lexicalization and to suggest other phrasal units when a one-word term is missing.

Thus, the entries in the lexicon should be viewed as lists of various wordings of one and the same concept. This approach becomes highly relevant in real search scenarios, because users usually write expressions according to their level of knowledge on the topic. These expressions might be terms or some free phrases. Of course, it is impossible to predict all the wordings which correspond to a concept. For that reason, we first concentrated on the most frequent ones. For example, for the concept **AlphanumericDisplay** with meaning: *a display that gives the information in the form of characters (numbers or letters)* in Bulgarian are given 9 alternatives, some of which reflect the spelling variation (*буквеноцифров дисплей, буквено-цифров дисплей, символен дисплей, буквеноцифров монитор, буквено-цифров монитор, символен монитор, буквеноцифров екран, буквено-цифров екран, символен екран*).

We also used the definitions as source for extending the lexicon (and the ontology). For example, if we have the term **AudioConference** with the meaning: *An electronic meeting in which participants in different locations use telephones or audio conferencing equipment to interactively communicate with each other in real time,* then we could add also the concept **ElectronicMeeting** as a superconcept.

The generalized structure of the lexicons is as follows: (1) a leading term which constitutes the meaning for all the term wordings within the entry. This term usually ensures the mapping to the relevant concept; (2) explanation of the concept meaning in lingua franca (usually it is English, but in fact it might be any natural language); and (3) a set of terms in a given language that have the meaning expressed by the leading term.

The idea of having a leading term when more expressions per concept are available is a tricky issue. The criteria for choosing such a term might be numerous: it occurs more often than the others, it is accepted to be a standard term in the special literature, etc. This strategy is justified in cases like the following: for the concept **AlphabetCharacter** with a meaning: *the conventional characters of the alphabet used to represent speech* the direct translation in Bulgarian *знак от азбука* ('sign of alphabet') is not as natural as the word *буква* ('letter'). The same non-naturalness holds for the expressions *буквен символ* ('letter symbol'), *буквен знак* ('letter sign'). On the other hand, sometimes there are two or more competing terms, and the only difference is that one of them has a native origin, while the other is transliterated from another language (e.g. in Bulgarian *абрeвиатура* and *съкращение* for *abbreviation* in English). In such cases the ordering in a term group is accidental.

## 6. Annotation Grammars

In order to connect the relevant text chunks with the ontological concepts, regular grammars are used. We call them concept annotation grammars or simply annotation grammars. In general the annotation grammars comprise two interconnected steps: (1) concept annotation step and (2) disambiguation step. In this initial phase of development of annotation grammars for our ontologies we perform the first step by the cascaded regular grammar in the CLARK system, and the second step by the constraint facilities in the same system. Thus, the grammars encode the lexicon in the Regular expression area, and the corresponding ontological concepts in the Return mark-up area. The competing expressions get the same semantic interpretation.

We define the regular grammars as reflections of the expressions in the lexicon. The quality of the grammar predefines the coverage and precision of the annotation, and hence – the efficiency of the search. Since at the beginning the ontology and the lexicons are not complete, the grammars assign all the possible mappings of the concepts to the detected text fragments. In order to improve the situation we rely on the longest match strategy of the grammar engine. For example, first the two-word sequence of 'personal computer' is recognized before annotating the occurrences of the superconcept 'computer'. The disambiguation is performed by two methods - rule-based, where manually defined rules are implemented, and

statistical, where the tools learn from the corpus and assign the correct concept automatically. For the moment we are working on the rule-based method. The statistical method will be implemented later when the gold standard corpora for Bulgarian are checked and verified.

One of the main problems with the initial development of the annotation grammars is their interaction with the general chunk grammars we already have developed within Bultreebank project [1]. The chunk grammars cover the simple noun phrases, complex named entities, the verbal phrases, prepositional phrases and clauses in sure positions. The concept annotation grammars at this stage are chunk grammars related to the concepts within the ontologies. The problems arise in the places where these two types of grammars interact with each other. The main cases are when the annotation grammars recognize smaller chunks in comparison to the general chunk grammars, and when the annotation grammars recognize bigger chunks on the basis of the sure semantic information. In the first case we have to decide whether the smaller chunk is really a representative of the concept, additionally modified in the context, or it is a not representative of the concept because the bigger chunk is the representative of a different concept. The second case is often related to fixed phrases which contain a subchunk related to a concept. We have handled the problem by an extension of the annotation grammars with rules for bigger chunks which are not related to the concepts in the ontologies. The case of bigger chunks, recognized by the annotation grammar, is solved by ordering of the applications of the grammar modules. Thus we have three layers of grammars: (1) grammars for chunks of fixed phrases or concepts outside of the annotation grammars, (2) annotation grammars themselves, and (3) general annotation grammar. The application of the general annotation grammars is not obligatory for the ontology-based semantic search task, but the result of it is very useful for finding new concepts missing in the ontology, for finding new phrases missing in the lexicons and for relation annotation in future.

## 7. Mutual improvement strategy

As it was mentioned above, the creation of domain ontologies depends on the available domain texts and the terminological lexicons or standards. However, these prerequisites are never perfect – either we get a small set of texts which are often noisy, or the lexicons present only some terms, but not their variety. For that reason we chose a strategy for a mutual gradual improvement of all these resources. Let us see how it works in more detail below.

In the first run, the annotation grammars encode not only the real corresponding concepts per term expression, but also two more possibilities: ERASE and EXTENDED. After the grammar annotation, the constraint manager is executed. It supports the manual concept selection in cases

of ambiguity. Actually, two constraints are provided: Constraint 1 (*Select Each Concept*) or Constraint 2 (*Select Ambiguous-only Concept*). Constraint 1 stops at each recognized term despite being ambiguous or non-ambiguous. The option ERASE is chosen when a concept was assigned to a common word, not a term. The option EXTENDED is chosen when a concept is recognized partially. This option covers two basic cases: occurrence of general vs. specific notions (e.g. *Internet* vs. *Wireless Internet*), or notions that can be expressed by a single word as well as multiwords (*disc* vs. *hard disc*; *user* vs. *end-user*). There is a third option, which is incorporated into both constraints, namely – adding a correction over a concept. This happens when the term is used in a broader or a narrower sense, which lacks in the assigned concept (e.g. *Insert* concept in narrowing sense of the term *Paste*, and in broader sense of the term *Insert*). The usage of Constraint 1 is recommended at the beginning of the annotation process, when the annotation grammar is not considered to be very precise, and when its automatically compiled versions rely only on lemmas.

All these repairing techniques (although subjective and depending on the annotator) lead to the improvement of the regular grammar, which assigns the concepts. Consequently, it means improvement to the ontology and the lexicon.

The usage of Constraint 2 can be relied on at a later stage, when the annotation grammar has been improved at least to some extent (as a result of the previous constraint). This constraint does not introduce artificial ambiguity choices. It stops only at real ambiguities in the texts. For example, the term ‘word’ might be assigned two concepts depending on the context: either common words (*WordLang*), or elements of computer memory (*WordMemory*).

## 8. Conclusions

In this paper we presented the current state of the language resources which are necessary for an efficient semantic search. We concentrated especially on Bulgarian language resources, such as domain golden corpora, terminological lexicons, annotation grammars. The interaction among all these components (together with the domain ontology) was described in more detail. From our experience in two domains we may conclude that the resources depend on each other, and hence - develop in parallel. After taking into account the error analysis from the annotators, the resources raise their coverage and precision. The added value of the Semantic search becomes even better in a multilingual environment. In both projects we work in such a context, which gives us the opportunity to compare language models and evaluation results.

In future we plan to apply the model to general lexica in order to cover the most frequent meanings in Bulgarian.

We already started the annotation of the Bulgarian treebank in this direction. Also, we are planning the extension of the lexicons and the grammars with respect to the relations represented in the ontologies. We envisage such an extension to facilitate the task of sense disambiguation and also the development of deeper syntactic analysis.

The implementation of the lexicons and the grammars is done within CLaRK System – [5], [6]. The ontology is processed by Protège System (<http://protege.stanford.edu>).

## Acknowledgements

This work has been partially supported by two European projects: LT4eL (Language Technology for eLearning) project of the European Commission under the Information Society and Media Directorate, Learning and Cultural Heritage Unit (FP6-027391) and AsIsKnown (A Semantic-Based Knowledge Flow System for the European Home Textiles Industry) project of the European Commission under IST-2004-2.4.7 - Semantic-based Knowledge and Content Systems (FP6-028044).

## References

- [1] Kiril Simov, Petya Osenova, Alexander Simov, Milen Kouylekov. 2005. Design and Implementation of the Bulgarian HPSG-based Treebank. In Erhard Hinrichs and Kiril Simov, editors, *Journal of Research on Language and Computation, Special Issue*, Kluwer Academic Publishers, pages 495-522.
- [2] Claudio Masolo, Stefano Borgo, Aldo Gangemi, Nicola Guarino, Alessandro Oltramari and Luc Schneider. 2002. The WonderWeb Library of Foundational Ontologies. WonderWeb Deliverable D17, August 2002. <http://www.loa-cnr.it/Publications.html>.
- [3] Claudio Masolo, Stefano Borgo, Aldo Gangemi, Nicola Guarino and Alessandro Oltramari. 2002. Ontology Library (final). WonderWeb Deliverable D18, December 2003. <http://www.loa-cnr.it/Publications.html>.
- [4] Lothar Lemnitzer and Paola Monachesi. 2007. Keyword extraction for metadata annotation of Learning Objects. In *Proceedings of Workshop on Natural Language Processing and Knowledge Representation for eLearning Environments. RANLP 2007*, Borovets, Bulgaria.
- [5] Kiril Simov, Zdravko Peev, Milen Kouylekov, Alexander Simov, Marin Dimitrov, Atanas Kiryakov. 2001. CLaRK - an XML-based System for Corpora Development. In: *Proc. of the Corpus Linguistics 2001 Conference*. Lancaster, UK.
- [6] Kiril Simov, Milen Kouylekov, Alexander Simov. Cascaded Regular Grammars over XML Documents. In: *Proc. of the 2nd Workshop on NLP and XML (NLPXML-2002)*, Taipei, Taiwan.
- [7] Paola Monachesi, Lothar Lemnitzer, Kiril Simov. Language Technology for eLearning. *Proceedings of EC-TEL 2006, in Innovative Approaches for Learning and Knowledge Sharing, LNCS 0302-9743*, pp. 667-672. 2006
- [8] Paul Buitelaar, Michael Sintek, Malte Kiesel A Lexicon Model for Multilingual/Multimedia Ontologies In: *Proceedings of the 3rd European Semantic Web Conference (ESWC06)*, Budva, Montenegro, June 2006.
- [9] Nicola Guarino and Christopher Welty. 2002. "Evaluating Ontological Decisions with OntoClean." *Communications of the ACM*, 45(2): 61-65.
- [10] Aldo Gangemi, Roberto Navigli, Paola Velardi. The OntoWordNet Project: extension and axiomatization of conceptual relations in WordNet. Meersman R, et al. (eds.), *Proceedings of ODBASE03 Conference*, Springer, 2003.
- [11] Christiane Fellbaum. 1998. Editor. WORDNET: an electronic lexical database. MIT Press.
- [12] Piek Vossen (ed). EuroWordNet General Document. Version 3, Final, July 19, 1999, <http://www.hum.uva.nl/~ewn>
- [13] Alessandro Lenci, Federica Bosa, Nilda Ruimy, Elisabetta Gola, Monica Monachini, Nicoletta Calzolari, Antonio Zampolli, Emilie Guimier, Gaëlle Recourcé, Lee Humphreys, Ursula Von Rekovsky, Antoine Ogonowski, Clare McCauley, Wim Peters, Ivonne Peters, Robert Gaizauskas, Marta Villegas. 2000. SIMPLE Work Package 2 - Linguistic Specifications, Deliverable D2.1. ILC-CNR, Pisa, Italy.
- [14] Ian Niles and Adam Pease. Towards a standard upper ontology. In: FOIS '01: *Proceedings of the international conference on Formal Ontology in Information Systems*, ACM Press (2001)
- [15] Paul Buitelaar, Thierry Declerck, Anette Frank, Stefania Racioppa, Malte Kiesel, Michael Sintek, Ralf Engel, Massimo Romanelli, Daniel Sonntag, Berenike Loos, Vanessa Micelli, Robert Porzel, Philipp Cimiano LingInfo: Design and Applications of a Model for the Integration of Linguistic Information in Ontologies. In: *Proc. of OntoLex06, a Workshop at LREC, Genoa, Italy, May 2006*.
- [16] Eelco Mossel. 2007. Crosslingual Ontology-Based Document Retrieval. In *Proceedings of Workshop on Natural Language Processing and Knowledge Representation for eLearning Environments. RANLP 2007*, Borovets, Bulgaria.
- [17] Kiril Simov and Petya Osenova. 2007. Applying Ontology-Based Lexicons to the Semantic Annotation of Learning Objects. In *Proceedings of Workshop on Natural Language Processing and Knowledge Representation for eLearning Environments. RANLP 2007*, Borovets, Bulgaria.
- [18] Christopher Brewster, Harith Alani, Srinandan Dasmahapatra, and Yorick Wilks. 2004. Data Driven Ontology Evaluation. *International Conference on Language Resources and Evaluation*, Lisbon, Portugal.