RuThes Linguistic Ontology vs. Russian Wordnets

Natalia Loukachevitch
Research Computing Center
of Lomonosov Moscow State
University
louk_nat@mail.ru

Boris Dobrov
Research Computing Center
of Lomonosov Moscow State
University
dobrov_bv@mail.ru

Abstract

The paper describes the structure and current state of RuThes – thesaurus of Russian language, constructed as a linguistic ontology. We compare RuThes structure with the WordNet structure, describe principles for inclusion of multiword expressions, types of relations, experiments and applications based on RuThes. For a long time RuThes has been developed within various NLP and information-retrieval projects, and now it became available for public use.

1 Introduction

Since its appearance Princeton WordNet has attracted a lot of attention of researchers and other specialists in natural language processing and information retrieval (Fellbaum, 1998). National wordnets for many languages in the world were initiated.

For developing a wordnet for a new language, several approaches can be applied. The first approach is based on automated or manual translation of Princeton WordNet (Balkova et al., 2008; Linden and Carlson, 2010). The second approach consists in creating of a wordnet from scratch using language-specific dictionaries and corpora (Climent et al., 1996; Azarowa, 2008; Kunze and Lemnitzer, 2010). This approach often implies the modification of the initial set of Princeton WordNet lexical relationships, introduction and justification of new relations, which usually requires additional time-consuming efforts (Maziarz et al., 2013; Pedersen et al., 2012).

At least three attempts to create a Russian wordnet are known. RussNet (Azarowa, 2008) began to be developed from scratch and at this moment continues to be quite small (not more than 20 thousand synsets). Two other Russian wordnets were generated using automated translation (Gelfenbeyn et al., 2003; Balkova et al., 2008). The former one is publicly available (`http://wordnet.ru/`) but represents the direct translation from Princeton Wordnet without any manual revision. The webpage of the latter one ceased to exist.

The structure of Princeton WordNet (and other wordnets) is based on sets of partial synonyms – synsets, organized in hierarchical part-of-speech-based lexical nets according mainly to hyponymy-hypernymy relations. Every part-of-speech net has its own system of relations.

Wordnets are often referred as linguistic or lexical ontologies (Magnini and Speranza, 2002; Veale and Hao, 2007), synsets of WordNet are often considered as lexicalized concepts. However, wordnets are mainly intended to describe lexical relations, what is quite different from the primary aim of ontologies to describe knowledge about the world, not about the language (Buitelaar et al., 2009; Nirenburg and Raskin, 2004). This difference reveals, for example, in the above mentioned division of wordnets to different part-of-speech subnets, because a part of speech cannot be a divisive feature in construction of ontologies.

In this paper we will consider the structure and current state of Thesaurus of Russian language (linguistic ontology) RuThes, which for a long time has been developed within various NLP and information-retrieval projects (Loukachevitch and Dobrov, 2002), and now it is prepared to become available for public use. In this resource we attempted to create a linguistically motivated ontology (not a lexical net), based on the denotational part of lexical senses and concept-based (not lexical) relations. At present, RuThes comprises more than 158 thousand unique words and expressions, which are structured into 53.5 thousand concepts.

The structure of this paper is as follows. Section 2 is devoted to the comparison of units in ontologies, wordnets and information-retrieval thesauri. In Section 3 main components of RuThes are considered. In Section 4 we describe several applications and the evaluation of RuThes. At last in Section 5 we describe our licensing policy for RuThes distribution.

2 Units in Ontologies, Wordnets and Information-Retrieval Thesauri

Ontologies are often considered as logical theories, which should be independent of natural language (Buitelaar et al., 2009; Smith, 2004). The general recommendations on the ontology concepts (classes) are usually described as follows (Noy and McGuinness, 2001; Nirenburg and Raskin, 2004):

- one needs to distinguish the concept and its name, i.e. synonyms do not represent different classes, synonyms are just different names of the concepts
- a concept should be distinctly different from its parent and from the concepts at the same level (sibling concepts).

However, to use ontologies in natural language processing, concepts of ontologies should be associated with language expressions and structures. In (Maedche and Zacharias, 2002; Buitelaar et al., 2006; Buitelaar et al., 2009) special models for linking natural language expressions and ontological entities are proposed.

From another point of view, an ontology cannot be fully independent of natural language. Ch. Brewster and colleagues (Brewster et al., 2005) stress that people manipulate concepts through words. In all known ontologies the words are used to represent concepts. Therefore, phenomena that are not verbalized, cannot be modeled. Y. Wilks (Wilks, 2008) asserts that the symbols in representation languages are fundamentally based on the natural language.

WordNet was created as a lexical rather than ontological resource (Fellbaum, 1998). However, over time, the growing importance of the ontological research, as well as the similarity of the WordNet noun hierarchy with an ontology became apparent (Miller and Hristea, 2006).

At the same time there exist a lot of deficiencies of WordNet descriptions from the ontological point of view (Guarino, 1998). Numerous examples of confusion between a concept and its names can be found in WordNet (Loukachevitch, 2009). Separate synsets are introduced for different ways of naming the same entities including the support of specific hierarchies for different parts of speech, for description of old and new names of the same entities, specific word usage in different dialects of the language or text genres (*moke - donkey*, *nose - nozzle*) etc. This is due to the fact that the basic relation in WordNet is the synonymy, based on the principle of substitution of one for another in sentences (Fellbaum, 1998). Some of new wordnets enhance the diversity of lexical relations between words to describe mainly their derivational links (Azarowa, 2008; Derwojedowa et al., 2008; Maziarz et al., 2013; Bosch et al., 2008).

However, it was supposed in (Edmonds and Hirst, 2002; Hirst, 2009) that a fine-grained hierarchy is inappropriate as a model for the relationship between the senses of near-synonyms in a lexicon for any practical use in tasks such as machine translation and other applications. They assert that, "what is required is a very coarse-grained conceptual hierarchy so that whole clusters of near-synonyms are mapped to a single node: their core meaning".

If to look at information-retrieval thesauri as representative sources of the terminology and domain knowledge one can see that most standards and guidelines for information-retrieval thesauri construction highlight the connection between the terms and concepts of a subject field (ISO 2788-1986, 1986; Z39.19, 2005). So the American standard Z39.19 points out that a term is one or more words referring to a concept (Z39.19, 2005). A concept is considered as a unit of a thought, regardless of the terms that express them.

Creating RuThes as an ontology with concept-based (not lexical) relations, we assumed that the concept-oriented approach to the lexical knowledge representation gives the possibility of better matching between languages (Edmonds and Hirst, 2002), more natural connection with domain terminologies, which are inherently concept-based (Z39.19, 2005); and more reliable logical inference based on current ontological research (Masolo et al., 2003; Guarino, 2009; Guizzardi, 2011).

3 RuThes linguistic ontology

RuThes Thesaurus of Russian language can be called a linguistic ontology for natural language processing, i.e. an ontology, where the majority

of concepts are introduced on the basis of actual language expressions.

In construction of RuThes we combined three different methodologies:

- methods of construction of information-retrieval thesauri (concept-based units, a small set of relation types, rules of multi-word expression inclusion)
- development of wordnets for various languages (language-motivated units, detailed sets of synonyms, description of ambiguous expressions)
- ontology research (concepts as main units, strictness of relation description, necessity of many-step inference).

RuThes is a hierarchical network of concepts. Each concept has a name, relations with other concepts, a set of language expressions (words, phrases, terms) whose meanings correspond to the concept.

3.1 RuThes units

In RuThes, a unit is presented not by a set of similar words or terms, as it is done in the WordNet thesaurus, but by a concept – as a unit of thought, which can be associated with several synonymic language expressions. Every concept should have distinctions from related concepts, which are independent from context and should be expressed in specific set of relations or associated language expressions – text entries.

Words and phrases, which meanings are represented as references to the same concepts of the thesaurus, are called ontological synonyms. Ontological synonyms can comprise:

- words belonging to different parts of speech (*stabilization, stabilize, stabilized*) – therefore the number of RuThes concepts is approximately 2.5 times less than in a wordnet-like resource of the same size. Text entries are provided with part-of speech information;
- language expressions relating to different linguistic styles, genres;
- idioms and even free multiword expressions (for example, synonymous to single words).

Each concept should have a clear, univocal and concise name. Such names often help to express, delimit the denotational scope of the concept. Besides, such names facilitate the analysis of the results of natural language processing.

Name of a concept can be:

- one of unambiguous text entries;
- an unambiguous multiword expression;
- a pair of synonyms that uniquely identifies the concept;
- an ambiguous word with a relator similar to those used in traditional information retrieval thesauri (Z39.19, 2005).

If necessary, a concept may have a gloss, which is not a part of the concept name.

Language expressions that may give rise to a separate concept in RuThes belong not only to the general vocabulary, but also can be terms of specific subject domains within the broad scope of social life (economy, law, international relations, politics, transport, banks, etc.), so-called *socio-political domain (*Loukachevitch and Dobrov, 2004).

This is due to the fact that many professional concepts, terms, and slang of these domains penetrate easily into the general language, and can be widely discussed in mass media. Besides, such a scope of concepts facilitates the application of RuThes in specialized subdomains of the broad socio-political domain. Examples of such concepts in RuThes include: *EMERGENCY LOAN, TAX EXEMPTION, IMPORT TAX, DEMOGRAPHIC INDICATOR* etc.

In fact, we subdivide the whole scope of RuThes concepts to:

- **General Lexicon** comprising concepts that can be met in various specific domains. In this, General Lexicon approximately corresponds to the Factotum domain in the Wordnet domain set (Gonzalez et al., 2012; Bentivogli et al., 2004),
- and **Socio-political Thesaurus** containing thematically oriented lexemes and multiword expressions as well as domain-specific terms of the broad sociopolitical domain.

After a concept has been introduced, an expert searches for all possible synonyms or derivative synonyms (that is derivates preserving the sense of an initial word), single words and phrases that can be associated with this concept. For example, a concept *ДУШЕВНОЕ СТРАДАНИЕ (wound in the soul)* has more than 20 text entries including such as: *боль, боль в душе, в душе на-*

*болело, душа болит, душа саднит, душевная пытка, душевная рана, душевный недуг, наболеть, рана в душе, рана в сердце, рана души, саднить (*several English translations may be as follows: *wound, emotional wound, pain in the soul* etc.).

At present RuThes includes 53.5 thousand concepts, 158 thousand unique text entries (75 thousand single words), 178 thousand concept-text entry relations, more than 215 thousand conceptual relations.

3.2 Multiword expressions in RuThes

One of difficult issues in wordnet development is inclusion of synsets based on senses of multiword expressions, for example noun compounds (Bentivogli and Pianta, 2004; Agirre et al., 2006; Kunze and Lemnitzer, 2010). Two main questions are usually discussed here: what are the principles of inclusion of multiword expressions (especially compositional or semi-compositional ones) and what types of relations should connect a multiword expression and its components in the wordnet structure.

In RuThes introduction of concepts based on multiword expressions is not restricted and even encouraged if (and only if) this concept adds some new information to knowledge described in RuThes.

Such additional information may be subdivided into several types.

A concept denotes an important entity. So in our Russia-oriented resource *ПРЕЗИДЕНТ РОССИЙСКОЙ ФЕДЕРАЦИИ* (*Russian President*) is an example of such a concept. Another variant of the same issue is the existence of important parts or participants for an entity or event. So, for *АРЕНДА* (lease) concept, such additional concepts as *АРЕНДНАЯ ПЛАТА* (*lease payment*), *АРЕНДНЫЙ ДОГОВОР* (*lease agreement*), *АРЕНДНОЕ ИМУЩЕСТВО* (*leasehold property*) are introduced, because they present important issues of lease services. At the same time concept *АРЕНДНЫЙ ДОГОВОР* (lease agreement) is an important subtype of concept *ГРАЖДАНСКО-ПРАВОВОЙ ДОГОВОР* (*legal agreement*).

A new concept has relations that do not follow from the component structure of an underlying multiword expression. This is a reason to introduce concept *ИЗБРАНИЕ ПАПЫ РИМСКОГО (papal election)* - it has a relation to concept *КОНКЛАВ (papal conclave)*. Another example is concept *ТЮНИНГ АВТОМОБИЛЯ (car tuning)* having relations to concepts *АВТОСЕРВИС (auto service)*.

A new multiword-based concept has a text entry that is not motivated by the component structure of a basic expression, for example, concept *ЗАСНУТЬ ЗА РУЛЕМ* describes also an "interesting" synonym *заснуть во время движения (*compare English expressions *falling asleep at the wheel and falling asleep while driving)*. Also this concept has an "interesting" relation to concept *ДОРОЖНО-ТРАНСПОРТНОЕ ПРОИСШЕСТВИЕ (road accident)*.

At last, an important additional factor, which can stimulate inclusion of a concept to the thesaurus, is the ambiguity of components of an unambiguous phrase, such as *положение дел (state of affairs)*.

3.3 RuThes relations

RuThes relations are of conceptual nature, not lexical ones. It is not a simple task to choose an appropriate set of relations for such a broad and diverse scope of concepts. RuThes has a small set of conceptual relations consisting of three main relations that are also applicable to a lot of various domains (Dobrov and Loukachevitch, 2006) and describe the most important links of a concept.

The first relation is the traditional hyponymic (taxonomic) relation. To establish such relations we apply additional tests similar to ones used in ontology development. The tests are directed to avoid incorrect use of taxonomic relations and not to mix them up with other types of relations, because errors in relation types degrade logical inference Gangemi et al., 2001).

We consider role-type relations as especially dangerous ones when a role concept (such as *EMPLOYEE*) is located as a parent concept for a type (as *PERSON*) (see discussion about roles and related problems in (Guarino, 1998; Gangemi et al., 2001; Fellbaum, 2002)). Therefore establishing the taxonomic relationship we also check the fulfillment of the following principle: every instance of a child concept should be at the same time the instance of a parent concept (*not every person is an employee*).

The second conceptual relation used in RuThes is the part-whole relation. The part-whole relations can be applied in various domains, exist in diverse forms. Therefore in computer resources different approaches representing these relations can be taken (Winston et al., 1987; Guarino, 2009; Sowa,

2000). So, for example, the tradition to describe part-whole relations in wordnets differs considerably from the guidelines of information-retrieval thesauri construction (Z39.19, 2005; Fellbaum, 1998).

In RuThes we use the generalized part-whole relation, which means that besides traditional types of part-whole relations (physical parts, process parts), relations between the following types of entities can be considered as part-whole relations:

- an attribute and its bearer,
- a role or a participant of a situation and the situation: *investor - investing, player - playing* (compare (Loebe, 2007)),
- entities and situations in the encompassing sphere of activity: *industrial plant - industry, tennis racket - tennis, tennis player - tennis*. So these subtypes of part-whole relations in RuThes play the role of so-called WordNet domains, which were introduced to alleviate "tennis problem" – the lack of relations between synsets involved to the same situation or domain (Bentivogli et al., 2004; Gonzalez et al., 2012)

and several others.

In such a broad scope RuThes part-whole relations are close to so called *internal relations* (parthood, constitution, quality inherence, and participation) as described in (Guarino, 2009).

At the same time RuThes part-whole relations have a very important restriction: a concept-part should be related to its whole during normal existence of its instances: so called *inseparable parts* or *mandatary wholes* (Guizzardi, 2011). From this point of view, *TREE* concept is not described as part of *FOREST* concept, because trees can grow in many places, not only in forests.

Thus, the inference mechanism can rely on the chain of part-whole relations so we use the transitivity of such restricted part-whole relations (Guizzardi, 2011).

Let us see examples of the transitivity chain of part-whole relations:

- *(whole (**ACCUSED PERSON**, PUBLIC PROSECUTION),*
- *whole (PUBLIC PROSECUTION , JUDICIAL TRIAL),*
- *whole (JUDICIAL TRIAL, **JUDICIAL PROCEEDINGS**))*
- → *whole (**ACCUSED PERSON**, **JUDICIAL PROCEEDINGS**)*

The next relation in RuThes ontology is unsymmetrical association asc_1-asc_2, which represents *external dependence* in ontological terms (Gangemi et al., 2001; Guarino, 2009).

This relation is established between two concepts C_1 and C_2 when two requirements are fulfilled:

- neither taxonomic nor part-whole relations can be established between C_1 and C_2 in RuThes linguistic ontology,
- the following assertion is true: C_2 exists means C_1 exists (necessarily existent entities are excluded from consideration).

These two conditions mean that concept C_2 (dependent concept) externally depends on C_1:

$asc_1 (C_2, C_1) = asc_2 (C_1, C_2)$

Examples of dependent concepts for *AUTOMOBILE* concepts are as follows:

- asc_2 *(AUTOMOBILE, AUTOMOTIVE INDUSTRY*): concept *AUTOMOTIVE INDUSTRY* exists only if concept *AUTOMOBILE* exists;
- asc_2 *(FOREST, TREE)* concept *FOREST* exists only if concept *TREE* exists.

Relations of ontological dependence are applicable in various domains, therefore they are usually used in top-level ontologies (Sowa, 2000; Masolo et al., 2003; Grenon, 2003). Besides in (Kumar and Smith, 2004) authors discuss the importance of such a relation for the biology domain: cell movement cannot exist without cells. It is the first time when such relations are basic relations for a linguistic ontology.

An additional advantage of using this relation in linguistic ontologies consists in its usefulness for description of links between a concept based on the sense of a compositional multiword expression and concepts corresponding to the components of this multiword expression.

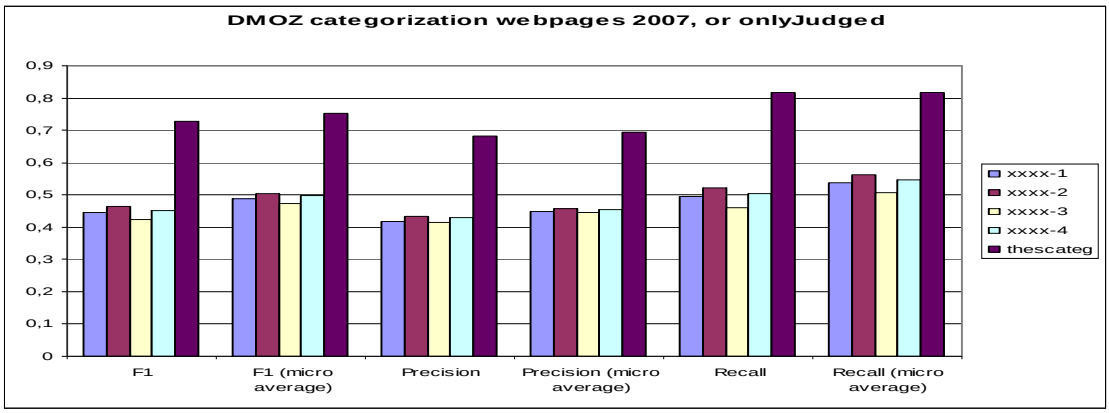

Fig 1. F1-measure, precision and recall of text categorization systems at ROMIP 2007.

So a multiword-based concept (for example, *AUTOMOBILE RACING*) is described as a dependent concept and its component concept (*AUTOMOBILE*) as a main concept. This allows us to introduce concepts based on various types of multiword expressions as described in section 3.2 and establish their necessary relations.

To conclude this section, we would like to stress there exists the similarity between all above-mentioned relations, which determines their considerable importance in concept description. These relations are established when concept instances or concepts themselves should coexist, what means that using these relations, we describe the most inherent (and, therefore, reliable) relations of concepts.

4 Testing RuThes in Automatic Document Processing

RuThes lingustic ontology provides the detailed coverage of single words, expressions and senses of contemporary written Russian (mainly, news articles, laws and official documents). The quality of descriptions originates from several sources.

First, since 1996 RuThes was used in various projects with governmental bodies and commercial organizations (in such applications as conceptual indexing in information-retrieval systems, knowledge-based text categorization, automatic summarization of single and multiple documents, question-answering etc.) and every project gave us the possibility to improve descriptions of lexical senses, to reveal useful expressions.

Second, 200 thousand words in a dictionary form (so called lemmas) ordered in decreasing frequency were extracted from the document frequency list of information-retrieval system RUSSIA (`www.uisrussia.msu.ru/`), in which contemporary Russian legal documents and newspaper materials are stored (2 million documents). The contemporary usage of these lemmas (distinct from proper names) was checked out during ten years of work mainly in news collections of online news services.

In combination with other techniques we applied RuThes in tasks of Russian Information Retrieval Evaluation Seminar (ROMIP) (Dobrov et al., 2004). So in 2007 we tested our knowledge-based text categorization system in ROMIP text categorization evaluation (Ageev et al., 2008a). The task was to automatically classify documents of 1.5 mln. webpages using 247 categories (Russian part of DMOZ categories `www.dmoz.org`). The training collection included 300 thousand documents with DMOZ category labels.

For every category, we created a Boolean expression over a relative small number of "supporting" concepts of the thesaurus. After that initial Boolean expressions were expanded on the basis of properties of the thesaurus relations. Final Boolean expressions usually include much more disjunctive and conjunctive components, sometimes in hundreds times more. Thus, these expanded Boolean descriptions of categories were used in automatic categorization of documents.

For example, Music category was described with single concept $MUSICAL\ ART_Y$, where `Y` means full expansion to lower levels of the hierarchy including hyponyms, parts and dependent concepts. So the full Boolean expression for this category looks like a disjunction of more than

400 concepts: *ADAGIO* ∨ *ACCORDION* ∨ … ∨ *ORCHESTRA* ∨ …).

The aim of our experiment was to obtain the best results of text categorization by minimal human efforts. The given system of 247 categories was described during eight hours by two knowledge engineers (overall time) (Ageev et al. 2008a). Fig. 1 demonstrates the performance of the created categorization system (*thescateg*) in comparison to machine learning approaches (SVM-based runs).

It is possible to see that the results of the knowledge-based system are considerably better. In our opinion, the achievement of such results is due to large volumes of knowledge described in RuThes and its consistent representation. Besides, in this evaluation machine learning approaches should process a highly inconsistent training collection because DMOZ manual labels were provided for the whole websites, but the contents of specific pages from these sites could be quite different from title pages.

In fact, more than twenty knowledge-based text categorization systems were implemented on the RuThes basis.

At last, Socio-political thesaurus (see section 3.1) is used as a search and visualization tool in several information-retrieval systems. Also in experiments the usefullness of Socio-political thesaurus for processing of long information-retrieval queries and as a basis for text clustering was proved (Ageev et al., 2008b; Dobrov and Pavlov, 2010).

5 Publication of RuThes

At present, RuThes thesaurus is partially involved in several commercial projects with other organizations and therefore it cannot be published as a whole. But the interest in a large thesaurus of Russian language is considerably growing therefore we decided to publish RuThes partially.

The first publicly available version of RuThes (RuThes-lite) contains around 50 thousand words and expressions and is available from `http://www.labinform.ru/ruthes/index.htm`. The next version including 100 thousand text entries will be published in the beginning of 2014. We distribute RuThes-lite as free for non-commercial use (Attribution-NonCommercial-ShareAlike 3.0 Unported license).

6 Conclusion

In this paper we presented RuThes linguistic ontology. This resource has been developed for a long time (more than fifteen years) and was used as a resource in various applications of NLP and information retrieval such as conceptual indexing, semantic search, query expansion, automatic text categorization and clustering, automatic summarization of a single document and multiple documents.

Now we decided to provide public access to RuThes and in this paper we described its structure and current state. We hope that this resource, having the broad and detailed lexical and terminological coverage of contemporary Russian news articles and official documents, will facilitate development of NLP techniques and research for Russian language.

7 Acknowledgements

The work is partially supported by Dmitrii Zimin Dynastia Foundation with financial support of Yandex founders.

References

Mikhail Ageev, Boris Dobrov, Pavel Krasilnikov, Natalia Loukachevitch, Andrey Pavlov, Alexey Sidorov, and Sergey Shternov. 2008a. UIS RUSSIA at ROMIP-2007: Search and classification. *In Proceedings of Russian Seminar on Information-Retrieval Methods ROMIP 2007-2008* (In Russian).

Mikhail Ageev, Boris Dobrov, Natalia Loukachevitch, and Sergey Shternov. 2008b. UIS RUSSIA at ROMIP-2008: Search and classification of legal documents. *In Proceedings of Russian Seminar on Information-Retrieval Methods ROMIP 2007-2008* (In Russian).

Eneko Agirre, Izaskun Aldezabal, and Eli Pociello. 2006. Lexicalization and multiword expressions in the Basque WordNet. In *Proceedings of Third International WordNet Conference*, Jeju Island (Korea):131-138.

Irina Azarowa. 2008. RussNet as a Computer Lexicon for Russian. *In Proceedings of the Intelligent Information systems IIS-2008*: 341-350.

Valentina Balkova, Andrey Suhonogov, and Sergey Yablonsky. 2008. Some Issues in the Construction of a Russian WordNet Grid. *In Proceedings of the Forth International WordNet Conference*, Szeged, Hungary:44-55.

Luisa Bentivogli and Emanuele Pianta. 2004. Extending wordnet with syntagmatic information. *In Pro-*

ceedings of Second Global WordNet Conference:47-53.

Luisa Bentivogli, Pamela Forner, Bernardo Magnini, and Emanuele Pianta. 2004. Revising WordNet domains hierarchy: semantics, coverage, and balancing. *In Proceedings of COLING 2004*, Geneva, Switzerland:101-108.

Sonya Bosch, Christiane Fellbaum, and Karel Pala. 2008. Enhancing WordNets with Morphological Relations: A Case Study from Czech, English and Zulu. *In Proceedings of the Fourth Global WordNet Conference*:74-90.

Christopher A. Brewster, Jose Iria, Fabio Ciravegna, and Yorick Wilks. 2005. The Ontology: Chimaera or Pegasus. *In Proceedings Dagstuhl Seminar Machine Learning for the Semantic Web*: 89-101.

Paul Buitelaar, Philipp Cimiano, Peter Haase, and Michael Sintek. 2009. Towards Linguistically Grounded Ontologies. The Semantic Web: Research and Applications. *In Proceedings of the European Semantic Web Conference*. Springer Verlag, LNCS 5554:111-125.

Paul Buitelaar, Michael Sintek, and Malte Kiesel. 2006. A lexicon model for multilingual/multimedia ontologies. *In Proceedings of the 3rd European Semantic Web Conference (ESWC06).*

Salvador Climent, Horacio Rodriguez, and Julio Gonzalo. 1996. Definitions of the links and subsets for nouns of the EuroWordNet project. Deliverable D005, EuroWordNet, LE2-4003, Computer Centrum Letteren, University of Amsterdam.

Magdalena Derwojedowa, Maciej Piasecki, Stanisław Szpakowicz, Magdalena Zawisawska, and Bartosz Broda. 2008. Words, Concepts and Relations in the Construction of Polish WordNet. *In Proceedings of GWC-2008*:162-177.

Boris Dobrov, Igor Kuralenok, Natalia Loukachevitch, Igor Nekrestyanov, and Ilya Segalovich. 2004. Russian Information Retrieval Evaluation Seminar. In *Proceedings of LREC-2004*:1359-1362.

Boris Dobrov and Batalia Loukachevitch. 2006. In Development of Linguistic Ontology on Natural Sciences and Technology." In *Proceedings of LREC-2006.*

Boris Dobrov and Andrey Pavlov. 2010. Basic line for news clusterization methods evaluation. *In Proceedings of Russian Conference on Digital Libraries RCDL-2010:* 287-295 (in Russian).

Philip Edmonds and Graeme Hirst. 2002. Near-synonymy and lexical choice. *Computational linguistics,* 28 (2):105-144.

Christiane Fellbaum. 1998. WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.

Christiane Fellbaum. 2002. Parallel Hierarchies in the Verb Lexicon. *In Proceedings of 'The Ontologies and Lexical Knowledge bases' workshop*. Las Palmas, Spain:27-31.

Aldo Gangemi, Nikola Guarino, Claudio Masolo, and Alessandro Oltramari. 2001. Understanding Top-Level Ontological Distinctions. *In Proceedings of IJCAI 2001 workshop on Ontologies and Information Sharing*:26-33.

Ilia Gelfenbeyn, Artem Goncharuk, Vlad Lehelt, Anton Lipatov, and Viktor Shilo. 2003. Automatic translation of WordNet semantic network to Russian language. *In Proceedings of International Conference on Computational Linguistics and Intellectual Technologies Dialog-2003.*

Aitor Gonzalez, German Rigau, and Mauro Castillo. 2012. A graph-based method to improve wordnet domains. *Computational Linguistics and Intelligent Text Processing*, Springer, LNCS-7181: 17-28.

Pierre Grenon. 2003. Spatio-temporality in Basic Formal Ontology: SNAP and SPAN, upper-level ontology, and framework for formalization. PART I. *IFOMIS Report 05/2003*.

Nicola Guarino. 1998. Some ontological principles for designing upper level lexical resources. *In Proceedings of First International Conference on Language Resources and Evaluation*: 527-534.

Nicola Guarino. 2009. The ontological Level: Revisiting 30 years of Knowledge Representation. *Conceptual Modeling: Foundations and Applications*. Springer-Verlag Berlin, Heidelberg: 52-67.

Giancarlo Guizzardi. 2011. Ontological foundations for conceptual part-wholes relation: the case of collectives and their parts. *Advanced Information Systems Engineering*, Springer CAiSE, LNCS 6741:138–153.

Graeme Hirst. 2009. Ontology and the Lexicon. In: Staab S., Studer R. (eds.) *Handbook on Ontologies in Information Systems*: 269-292.

ISO 2788-1986. 1986. Guidelines for the establishment and development of monolingual thesauri.

Anand Kumar and Barry Smith. 2004. The ontology of blood pressure: a case study in creating ontological partitions in biomedicine.

Claudia Kunze and Lothar Lemnitzer. 2010. Lexical-Semantic and Conceptual relations in GermaNet. *In Storjohann P (ed) Lexical-semantic relations: Theoretical and practical perspectives,* 28:163-183.

Krister Linden and Lauri Carlson. 2010. Finnwordnet — wordnet på finska via översättning. *LexicoNordica — Nordic Journal of Lexicography*, 17:119–140.

Frank Loebe. 2007. Abstract vs. Social Roles: Towards a general theoretical account of roles. *Applied Ontology*, v2 (2):127-158.

Natalia Loukachevitch and Boris Dobrov. 2002. Development and Use of Thesaurus of Russian Language RuThes. *In Proceedings of workshop on WordNet Structures and Standartisation, and How These Affect WordNet Applications and Evaluation*. LREC-2002:65-70.

Natalia Loukachevitch and Boris Dobrov. 2004. Sociopolitical Domain as a Bridge from General Words to Terms of Specific Domains. *In Proceedings of Second International WordNet Conference GWC-2004*:163-168.

Natalia Loukachevitch. 2009. Concept Formation in Linguistic Ontologies. Conceptual Structures: Leveraging Semantic Technologies. *In Proceedings of ICCS-2009*. Springer Verlag, LNAI-5662:2-22.

Claudio Masolo, Stefano Borgo, Aldo Gangemi, Nicola Guarino, and Alessandro Oltramari. 2003. WonderWeb Deliverable D18: Ontology library (final). *Technical report*, Laboratory for Applied Ontology, ISTC-CNR, Trento, Italy.

Alexander Maedche and Valentine Zacharias. 2002. Clustering Ontology-based Metadata in the Semantic Web. *In Proceedings PKKD-2002*:342-360.

Bernardo Magnini and Manuela Speranza M. 2002. Merging Global and Specialized Linguistic Ontologies. *In Proceedings of OntoLex*:43-48.

Marek Maziarz, Maciej Piasecki, and Stanisław Szpakowicz. 2013. The chicken-and-egg problem in wordnet design: synonymy, synsets and costitutive relations. *Language Resources & Evaluation* .

George Miller and Florentina Hristea. 2006. WordNet Nouns: Classes and Instances. *Journal of Computational linguistics*, 32(1):1-3.

Sergey Nirenburg and Viktor Raskin. 2004. Ontological Semantics. Cambridge, MIT Press.

Natalia F.Noy and Deborah McGuinness. 2001. Ontology Development 101: A Guide to Creating Your First Ontology. *Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880.*

Bolette Pedersen, Lars Borin, Markus Forsberg, Krister Linden K., Heili Orav, and E´ırikur Rognvaldsson. 2012. Linking and Validating Nordic and Baltic Wordnets. A Multilingual Action in META-NORD. *In Proceedings of GWC-2012*: 254-259.

Barry Smith. 2004. Beyond Concepts: Ontology as Reality Representation. *Proceedings of International Conference on Formal Ontology and Information Systems FOIS-2004*.

John Sowa. 2000. Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks Cole Publishing Co., Pacific Grove, CA.

Tony Veale and Yanfen Hao. 2007. A context-sensitive framework for lexical ontologies. *Knowledge Engineering Review*, 23(1):101-115.

Yorick Wilks. 2008. The Semantic Web: Apotheosis of annotation, but what are its semantics? *IEEE Intelligent Systems*, 23(3):41-49.

Morton Winston, Roger Chaffin, and Douglas Herrmann. 1987. A taxonomy of part-whole relations. Cognitive science, 11(4):417-444.

Z39.19. 2005. Guidelines for the Construction, Format and Management of Monolingual Thesauri. NISO.