Intelligent Information Systems 2008
ISBN 978-83-60434-44-4, pages 341–350

RussNet as a Computer Lexicon for Russian

Irina V. Azarowa

Saint-Petersburg State University, Saint-Petersburg, Russia

Abstract

In this paper we present two aspects of a wordnet-type lexicon RussNet for the Russian language: the peculiarities of its structure and its usage for computer text analysis. The information resources for RussNet structuring is discussed, the lexico-statistic corpus-based parameters and their interpretation analysed. The basic relations for wordnet dictionaries are considered.

Keywords: wordnet dictionary, semantic relations, corpus-based statistics, computer lexicon

1 Introduction

The computer thesaurus RussNet[1] developed at the Department of Applied Mathematic Linguistics of Saint-Petersburg State University was planned as a computer lexicon for the text analyser for Russian (Azarova *et al.*, 2005). This thesaurus belongs to the large group of wordnet-dictionaries constructed for languages of various types. The prototype – the Princeton WordNet (PWN) was initially designed to demonstrate the human cognitive behaviour embedded into a lexicon (Fellbaum, 1998). Later this model was widely used for automatic text processing (Voorhees, 1998). A number of wordnets for European languages were constructed according to the PWN prototype (Vossen, 1998; Tufiş *et al.*, 2004). Initially national thesauri were aligned to specially chosen notions – Interlingual Index (ILI), later — to PWN 2.X. That allows the use of an English "interim language" for multilingual information retrieval and other tasks.

For any wordnet project, the key issue is to what degree the project would follow the general principles of previous "best practice", thus from the beginning of RussNet project we aimed at conceptual conformity rather than automatic extension or generation of the original PWN. Below we will discuss the reproduced features of wordnet technique as well as additional components oriented to RussNet usage in the computer text analyser Rus4IR.

2 RussNet Structure

We interpreted RussNet as a network of semantic relations between synsets – lexicalised notions expressed by synonymic words and collocations in Russian. The metaphorical notion of a "lexical space" was a variant of lexical representation in

[1] `http://project.phil.pu.ru/RussNet`

native speakers' mind. It was pointed out that WordNet was constructed in a field of psycho-lexicology (Miller *et al.*, 1990). As a result, psycholinguistic methods and decisions were widely used in PWN 1.5. In addition, sorting word meanings according to their frequency was accomplished on the basis of a semantically tagged corpus. In recent projects the major decisions were made with the help of modern dictionaries (monolingual or bilingual) and corpus investigations.

We use several "second order" resources in RussNet project: explanatory and synonymy dictionaries and thesauri for modern Russian, and the *Russian word association dictionary*, Moscow 2002 (RWA). For various reasons, it was impossible to transfer the data in those resources into RussNet: some of them are outdated, and describe word senses in a heterogenous manner. To substantiate our technique, we compiled a balanced corpus of modern texts, which is used as a "first order" resource. Corpus texts are dated from 1985 up to now, so as to present the contemporary period of Russian without excessive communist and Soviet rhetoric. The corpus includes 21 million word forms. The texts in the corpus are restricted in length (about 200 word forms on average), so that frequencies of concrete words were not distorted by thematic references. The majority of the texts (60%) are articles on various topics from newspapers and magazines, covering subject diversity of the common Russian language (Azarova and Sinopalnikova, 2004b). The corpus is managed by Bonito[2], a program developed by our Czech colleagues.

The corpus statistics is an important part of data structuring in RussNet, because statistical characteristics of words may be used in automatic text analysis as well. The most important point of RussNet technique is corpus-based lexicostatistics. Below we specify its main theses. It was shown that lexical statistics on a balanced text corpus is closely similar to RWA information, though word association data reflect relative relevance of language phenomena in comparison to absolute corpus frequencies (Sinopalnikova, 2004). So, our method is a combination of the original PWN methodology with that applied in EuroWordNet.

We utilize the morpho-syntactic parser Rus4IR to process Russian word forms. The parser has a derivational module that may attach regular derivates to the existing synsets by specified semantic links, but it checks actual entries first.

2.1 The Lexical Inventory

Any wordnet project is characterised by synset quantity, the number of words and semantic relations, and so on. These characteristics show the "state of the art" in the wordnet project, though the process of updating the lexicon appears to be endless, as a comparison of synsets in PWN 1.5 and PWN 2.3 demonstrates.

Filling up the RussNet collection of synsets is accomplished by working out separate lexical groups. The procedure for capturing their structure includes several steps: (1) choosing the relevant lexical units (LUs) from 1000 most frequent words of a certain part of speech [POS] (Sharoff, 2002), (2) extending the number of LUs by paradigmatic associations from RWA, (3) providing conceptual analysis of corpus contexts for chosen LUs.

The third step requires a more detailed description. The corpus texts are

[2] `http://nlp.fi.muni.cz/projects/bonito`

marked morphologically. Though lemmatization is ambiguous, the Bonito corpus manager allows us to remove evident mistakes and improve the context output for the specified lemma. Next, the whole assemblage of contexts is reduced to a random sample of 100-150 contexts, which are marked up manually according to the definitions in the classic 4-volume Russian explanatory dictionary MAS[3]. If some meaning in the sample context has no dictionary match, the list of meanings is updated. After counting frequencies of different meanings in the sample, the actual list of meanings for a particular LU is reorganized according the frequencies in the sample. The meanings with zero frequencies are considered potential, and not included in RussNet. The frequencies of meanings for a LU are extrapolated onto the total distribution of corpus contexts. They are measured in *instances per million* [IPM] – the number of occurrences per million word forms in the corpus.

We experimented with a sample size required for sense identification and observed that 100 random contexts gave the same proportion of senses as sets with 1000 and more contexts (Azarova *et al.*, 2005), with an error lower than 0.5%. The comparison of training sets of 25 and 200 sentences by (Leacock and Chodorow, 1998), which were used for disambiguation procedure, showed similar results. Context proportions of rare meanings may be distorted considerably, but these senses are hardly subject to ordering by frequency. The marked-up corpus indicates that 50-70% of the appearances of most polysemous word are dominated by one sense. Other senses appear progressively less frequently, and the threshold for frequency ordering is 5%. It is impossible in practice to enumerate all rare senses – some of them appear once in the corpus (per 21 million word forms). The corpus-based principle for choosing LUs and their senses is oriented toward further automatic text analysis: we list those words and collocations which cover from 85% to 95% of texts. We put into RussNet the most common expressions in respect to the text style or the typical way of expressing the content.

In this manner we prepared 24 lexical groups. *Nouns*: Abstract, Animal, Artefacts, Body parts, Group, Food, Location, Person, Time. *Verbs*: Body functions, Change, Cognition, Competition, Contact, Emotion, Modality/Volition, Motion, State. *Adjectives*: Descriptive/Size, Emotional assessment, Human features. *Adverbs*: Degree, Subject's state, Location. The total number of synsets in these groups is 3329, LUs – 3860.

Regular transposition of meanings from one POS to another allows us to structure in a similar way abstract nouns after verbal and adjectival patterns, descriptive adverbs after descriptive adjectives, relational adjectives after noun structures. The total number of transpositions is 2386. Moreover, other productive derivates may be attached semi-automatically to the existing synsets (Azarova *et al.*, 2002). It appeared, however, that only a smaller part of them actually appear in the corpus texts, which contradicts corpus-based choice of LUs. Therefore, we put actualized derivates into RussNet; others are considered potential and may be attached by a derivational module of Rus4IR parser. We have approximately 12,500 synsets in RussNet. Table 1 shows the proportions of synsets, word-sense pairs and average polysemy in PWN and RussNet.

As can be easily seen from the table data, RussNet has fewer noun synsets and

[3] `http://feb-web.ru/feb/mas/mas-abc/default.asp`.

TABLE 1: The proportions of synsets and word-sense pairs for various POS in PWN and RussNet

	Nouns		Verbs		Adjectives		Adverbs	
	PWN	RN	PWN	RN	PWN	RN	PWN	RN
Synsets	69%	44%	12%	35%	16%	16%	3%	5%
Word-sense pairs	70%	30%	12%	52%	15%	14%	3%	4%
Average polysemy	1.23	1.2	2.17	2.6	1.44	1.6	1.24	1.4

noun LUs in comparison with PWN.

In order to supply a wordnet dictionary with terminological LUs, it is necessary to process terminological subcorpora (Marinelli and Spadoni, 2005; Faure and Nédellec, 1998). There are a lot of terminological synsets in PWN2.X, cf. definitions of *star* 'a celestial body of hot gases that radiates energy derived from thermonuclear reactions in the interior' in contrast with 'any celestial body visible (as a point of light) from the Earth at night ', which are present as different meanings. Both aspects are expressed in one definition in MAS. Terminological structuring may differ from the common language one, because the former represents the scientific point of view, and latter – the "naive picture of the world" (Apresjan, 1995).

2.2 Synset Structure

It is a synonymic relation that plays the major role in any wordnet structure. The criterion of synonymy in PWN (Miller *et al.*, 1990) was based on symmetric substitutability of one word form for another, though it was clear that in some cases it would be impossible definitely to approve or reject the alleged synonyms. So, the granularity of synonyms is subjective. That is illustrated convincingly by contradictions in synonymic dictionaries.

Take as an example *to star* in PWN 2.1. The definitions of the first two meanings look similar: 'feature as the star' (3 occurrences from tagged texts), 'be the star in a performance' (no occurrence). In the English-Russian dictionary (ed. by Ju. Apresjan, Moscow, 1994) they are interpreted as nuances of the third meaning in a full list of 6 glosses. What is the reason for splitting one meaning into two, and omitting the other 4?

In order to reduce subjectivity of synonym grouping in the wordnet technique, automatic procedures are invented based on a similar distribution of words (Derwojedowa *et al.*, 2008) or semantic similarity – so-called Word-Sketches (Kilgariff *et al.*, 2004). We observe statistical parameters of synonyms. Usually, there is one dominant synonym in a synset; this LU is stylistically neutral. Other synonyms are marked: they have obviously lower frequency and inconsistent distribution in various parts of the corpus. Synonyms with low frequencies constitute the subordinate part of a synset. They are not ordered by frequencies. We may detect asymmetry in substitutability for subordinate synonyms: they are commutated freely in a domain specific to the subordinate LUs. Outside this scope, however, inserting a subordinate LU into the context looks strange or impossible, as was shown in (Azarova and Sinopalnikova, 2004a) for adjectives *большой* 'big' and *крупный*

'large, large-scale'. In Apresjans's definition of synonymy (Apresjan, 1995) substitutability may be valid only in one context, though words are synonyms on the condition of similarity of their senses.

We add a requirement of homogeneous context markers for synset members, though the syntagmatic data were considered external to wordnet structure in PWN. The procedure of contexts from the random sample matching to meanings is supported by description of grammatical and semantic features of the context. If they are recurrent, they are reliable arguments for choosing some meaning from a list. We call them active and passive valencies (see section 2.4). These characteristics are not the result of introspection, but generalisation of context features.

We show in (Azarova and Sinopalnikova, 2004b) that context evidence for a dubious group of synonyms may have a complex structure. We considered two emotional verbs *сердиться* 'to be angry' and *злиться* 'to be irritated', which are described in *New Explanatory Synonymic Dictionary* (Apresjan, 1999) as synonymous. The former refers to the negative mood of a human with a high social status, the latter – to the same state of a human with a low social status or some analogous state of an animal. Context markers in the corpus do not substantiate the restriction of the semantic type for the subject of the state. However, 30% of contexts for *сердиться* include a syntactic group (a prepositional phrase or a sentence) denoting the rational motive of a negative mood. In 35% contexts the verb *сердиться* appears with negation or modal qualification ('don't be angry', 'was caused to be angry', and so on). The usual designation of an inchoative for this state was *рассердиться* 'to become angry'. The recurrent context markers for *злиться* are somewhat different. 25% of contexts include the group denoting someone (even oneself) or something that irritates without description in terms of a cause and effect. This verb appears in a row with such verbs as *кричал* 'shouted and was irritated'. The syntactic construction *начать злиться* 'to start to be irritated' is used besides the standard inchoative *разозлиться* 'to become irritated'. These context features are difficult to group compactly, but they show that *сердиться* has a statistically unstable valency for the rational motive of a negative mood, while *злиться* is used primarily as a designation of an uncontrolled negative emotional state. Moreover, it is a much more intense feeling than *сердиться*. These verb meanings are so called "near synonyms". They form two separate groups, and they have their own intensive forms (feelings in the extreme degree: *беситься* 'to be furious' and *гневаться* (no exact English equivalent, translated in dictionaries 'to be angry').

2.3 Semantic Relations

The structure of "lexical space" is provided by semantic relations, which usually connect synsets, though may as well link synset members – LUs. The most powerful semantic relation is hyponymy connecting superordinate and subordinate synsets. It structures noun synsets into semantic trees. Some trees are traditional for semantic or ideographic descriptions: *человек* 'person', *растение* 'plants', *животные* 'animals', some are rather strange: *части тела* 'body parts', *пища* 'food', *местоположение* 'location', etc. The relation is asymmetrical and transitive, therefore, it may be used for information retrieval – it

permits to generalise a meaning. In the text, hyponyms may be replaced by their hypernym. This feature is used for the one-way substitution test vs. reciprocal substitutions for synonyms as a special test for making decisions.

Hyponymic synsets may have several hypernyms, though this is not assessed as typical. Some structures have multiple superordinate synsets on the regular basis. For example, different designations of persons (*man, leader, teacher* etc.) are incorporated into the tree *человек* 'person' and another tree *причина* 'cause' (causal agent, causal agency). From our point of view, it is better to unite elementary trees into "superstructures", e.g., *одушевленные* 'animate or living', which includes *человек* 'person' and *животные* 'animal'. We describe the semantic attributes of valencies in terms of semantic trees, groups of trees, and subtrees.

It is not quite clear how many levels of hyponymic hierarchy are necessary for a wordnet: PWN structuring of the animal and plant trees are very similar to the biological taxonomy with 9 or even more levels. Otherwise, there are a lot of flat structures of hyponyms, especially on the first level of hierarchy, e.g., subordinates of *человек* 'person'. The flat hyponymy structure mixes groups with the different base for classification: 'adult', 'adventurer'... 'blond'... 'homunculus', etc. Flat structures may be split into clusters. Unfortunately, they may overlap, e.g., *роща* 'grove, usually consisting of birches', *бор* 'a coniferous forest, usually of considerable size'.

Another semantic relation, meronymy, is also asymmetric and transitive. It may organize synsets into hierarchical structures similar to hyponymic, e.g., *части тела* 'body parts' or *часть* 'part'. It usually produces network structures when it connects synsets from hyponymic trees. The useful dependence between these hierarchical relations is the inheritance of meronyms in hyponymic trees from the "entity" group: the synset *дерево* 'tree' has parts (meronyms) *ствол* 'trunk', *крона* 'crown' etc. which are inherited by its hyponyms: *берёза* 'birch', *клён* 'maple' and so on, and this synset has inherited parts from its superordinate: *корень* 'root', *лист* 'leaf'.

Verbal hyponymy is troponymy. It forms semantic trees as well. Some of them are rather traditional for thesaurus description: *движения* 'motion', *эмоциональные* 'emotion', *восприятия* 'perception', some are similar to an aggregation: *изменения* 'change'.

Attributive senses (adjectives and adverbs) were not supposed to be organised into united structures in PWN, but a number of small flat groups. Nevertheless, an adjectival hierarchy is constructed after the noun pattern in GermaNet. We consider that attributive synsets are in some degree organized according to the noun "entity" type (relative attributes), in part they follow verbal pattern (e.g. emotional adjectives), and descriptive adjectives have their own structure.

We associate a position in the hierarchy with statistical parameters of synset dominants: LUs with frequency above 60 IPM occupy the first two levels of the semantic tree. Synsets with low frequency dominants have no hyponyms. The background for this procedure is the famous inverse correlation between the frequency of a lexical unit and its content due to Mikołaj Kruszewski and Jerzy Kuryłowicz: "the less frequent a lexical item is, the richer content it has".

Another direction of our study is the formal procedure for semantic groupings. We used partially the method proposed in (Leacock and Chodorow, 1998) aimed at

disambiguation of different meanings of a verb *serve* using unambiguous morphologically marked-up contexts. We received reassuring results (Azarova *et al.*, 2008) for training the distribution patterns of traditional verbal groups: communication, cognition, stative, motion, emotion, possession, contact, modal, creation, perception. We calculated the appropriate width of the distribution window for verbs [-3, +5] and chose the relevant tag inventory. Now nouns and adjectives are being investigated in a similar fashion. We expect to use morphological distribution patterns for automatic classification of contexts so as to facilitate the lexicographic work and acquire objective data concerning the set of semantic trees. In case of success, the patterns may be applied to automatic text analysis.

We will not list other typical semantic relations used in RussNet, but we will mention the atypical ones. Thus, the conversion relation is distinguished from antonymy, e.g., *дочь* "'daughter' connected by an *antonymy* relation with *сын* 'son' and a *conversion* relation with *мать* 'mother' and *отец* 'father'.

Various derivational semantic relations (Azarova *et al.*, 2002) are used: *derivational synonymy, derivational antonymy, derivational hyponymy, derivational role* relation, and so on. For example, derivational synonyms are usually informal derivates, complimentary synonyms denote sympathy and love (*баба* 'grandmother' ⇒ *бабуля*) or appraisal (*современный* 'modern' ⇒ *супер-современный*). Non-complimentary synonyms designate people and objects with a slight pejorative nuance (*дама* 'a woman' ⇒ *дамочка*) or a strong one (*старик* 'an old man' ⇒ *старикашка*). These relations connect LUs with regular derivates, which may be potential, that is, have no occurrences in the corpus texts. They may be produced in the derivational module of the parser, and associated by derivational semantic relations with corresponding LUs of actual synsets (Azarova *et al.*, 2005).

The aspect pairs of verbs are usually interpreted as grammatical forms or very close synonyms, the difference of their meanings being the "aspect semantics" of the perfective verb. We observed in the procedure of verbal contexts analysis that there exists frequency dissimilarity of context proportions for aspect pairs. Some meanings are predominantly realised by perfective verbs, others – by imperfective ones. Such discrepancies are characteristic of imperfective verbs denoting activity, which are often *imperfectiva tantum*: *учиться* 'study', *работать* 'work', *служить* 'serve', *заботиться* 'care', and so on. Perfective verbs designate actions, which may be a realized "quantum" of activity (Azarova and Ushakova, 2007). We incorporate into RussNet two "semantico-grammatical" relations: *perfect_pair* and *imperfect_pair*, that connect "activity" synsets to "action" synsets, thus producing partially parallel structures (Azarova *et al.*, 2006).

2.4 Valency Frames

It is stressed in many papers that a lexical meaning contains a syntagmatic component (Apresjan, 1995), regularly expressed in contexts representing it. One or several valencies of a synset form its valency frame. Valency description is a generalisation of recurrent context features in the corpus. There are two main types: active and passive valencies. Active valencies are often used in syntactic theories. They characterise the predicted syntactic positions, usually for verbs. In the RussNet technique (Azarova *et al.*, 2006), valency features are collected in the corpus

contexts, thus producing the basis for their description.

Let us consider as an example the verb *направиться* 'to make one's way towards something'. The dictionary defines its first meaning (17 IPM) "to move in some direction". The position of the action subject is filled by a designation of a human being or a group of people in 98% of contexts. Moreover, the direction of movement is specified in 92% of contexts by a prepositional group (*к* + NP dative or *в* + NP accusative). Thus, this verb has two active valencies. When context markers have high frequencies of realisation in the contexts, they are considered to represent an obligatory valency. The first valency position is filled by LUs from the *person* tree or the *people* subtree of RussNet or pronouns replacing these words. The second valency is expressed by words from the *entity* group (*artefact, natural object, location, person* etc.). The location of a movement trajectory is specified in a small number of contexts (10%). It is not characteristic of this particular verbal sense, but of any predicate expressing action or motion. Such context features are considered occasional; they are not registered as valencies. An optional valency is a context feature which is between this highly probable appearance and highly improbable one. We assign threshold values for distinguishing obligatory from optional, and optional from occasional. It is 66% and 33% respectively.

The passive valency generalises the peculiarities of word usage in some sense. For example, a noun *гусеница* 'caterpillar' has two meanings: a wormlike larva and a caterpillar track. The first meaning is expressed by an animate noun, the second – inanimate. In some cases, it is possible to specify the group of head (syntactically dominant) words for some sense. For example, when attached to verbs of speech, the Russian word form *в лицо* means 'without ceremonies'. This does not happen when this prepositional group occurs with verbs of perception or manipulation with objects denoting "into the face".

The semantic specification of a valency is similar to explanations in dictionaries, for example, *катиться* 'to move rolling (about round objects)', though this description is too narrow for an actual corpus context, when humans and stones are rolling.

Morpho-syntactic specification of a valency is an enumeration of typical grammatical forms (or one particular form) of valency realisation in corpus contexts.

We expect that active valencies may be partially similar in a particular semantic tree. We evaluated matching of the valency parameters in 3 semantic trees (Azarova *et al.*, 2006) in order to find out inheritance of valency "parts" in this structure in a manner similar to meronym transmission in noun semantic trees. The results are not definitive yet.

The valencies may be useful for disambiguating polysemous words (Azarova *et al.*, 2005). It is difficult to assess the contribution of valencies in automatic analysis on the full scale, because RussNet is not finished yet. However, it can be shown that they may help reduce the number of syntactic and lexical interpretations. For example, the phrase *Я был знаком с тобой...* 'I was familiar to you' or 'with you I was a sign'). It has two grammatical structures: the word form *знаком* may be lemmatised as (1) a predicative form (singular, masculine) of an adjective *знакомый* 'familiar' or (2) an ablative singular of a noun *знак* 'a sign'). The adjective *знакомый* has an optional valency, grammatically expressed by the preposition *с* 'with' and the ablative of a noun, which is filled by LUs from

the semantic tree *person* (in the sense 'familiar to somebody') or *object* (in the sense 'familiar with something'). The personal pronoun *тобой* (2nd person, singular, ablative) is equal to the occurrence of a noun LU from the tree *person*. The matching of a phrase grammatical structure and a valency specification allows us to choose the first sense (synset) of *знакомый* 'familiar to'. The optional valency of the third meaning *знак* 'sign') is a genitive form of abstract nouns 'a sign of something' (*знак остановки* 'the stop sign') is not validated in the context, so gives us no evidences for the choice of interpretation with a predicate noun *знак*. Thus, the verification of valency frames against the context markers observed in texts provides the tool for disambiguation of grammatical structures and alternative synsets.

3 Results and Future Work

We now outline automatic text analysis with RussNet as a computer lexicon (Azarova *et al.*, 2005). RussNet provides a network structure of the lexicon in the system. Lemmas are identified in the text by morpho-syntactic modules of Rus4IR parser. They are projected onto the RussNet network defining the number of lexico-semantic variants of analysis. If no other steps of analysis are implemented, it is possible to use the most frequent meaning (or 3 top frequent meanings). Due to frequency ordering of meanings, the valid interpretation would be prevailing over an invalid one. The more advanced approach foresees the grouping of synsets extracted from the text into lexical chains (Voorhees, 1998), which partially disambiguate them. The further step to intelligent analysis is a representation of sentence structures in the form of head-modifier pairs (Koster, 2004). Some pairs match terminological items for information retrieval. RussNet or other wordnet is used for identification of synonyms. The full-scale syntactic analysis allows us to exploit valency frames for lexical and syntactic disambiguation (Azarova *et al.*, 2005). Semantic relations of the chosen synset and the top node of the hyponymy tree provide semantic interpretation and the system of inference.

The usage of RussNet as a computer lexicon in the automatic text analysis is still in the initial phase. The technique of wordnet construction as a natural linguistic ontology makes it clear to what extent it is possible to extract an adequate language model from accessible linguistic data.

References

Y. D. Apresjan (1995), *Lexical semantics. The means of synonymy in language*, Moscow.

Y. D. Apresjan, editor (1999), *New Explanatory Synonymic Dictionary. Issue 1*, Moscow.

I. Azarova, O. Mitrofanova, A. Sinopalnikova, M. Yavorskaya, and I. Oparin (2002), RussNet: Building a Lexical Database for the Russian Language, in *Workshop on WordNet Structures and Standardisation, and how these affect Wordnet Application and Evaluation*, pp. 60–64, Las Palmas de Gran Canaria.

I. Azarova and A. Ushakova (2007), The lexical group of verbs denoting activity in

RussNet, in *Bulletin of the Saint-Petersburg State University*, pp. 331–340, Saint-Petersburg.

I. V. Azarova, V. L. Ivanov, and E. A. Ovchinnikova (2006), RussNet Valency Frame Inheritance in Automatic Text Processing, in *Proceeding of Dialog'2006*, pp. 18–25, Moscow.

I. V. Azarova, V. L. Ivanov, E. A. Ovchinnikova, and A.A. Sinopalnikova (2005), RussNet as a Semantic Component of the Text Analyser for Russian, in *Proc Third International WordNet Conference*, pp. 19–28, Brno.

I. V. Azarova, A. S. Marina, and A. A. Sinopalnikova (2008), Verification of Valency Frame Structures by means of Automatic Context Clustering in RussNet, in *Proc Fourth International WordNet Conference*, pp. 35–43, Szeged.

I. V. Azarova and A. A. Sinopalnikova (2004a), Adjectives in RussNet, in *Proc Second International WordNet Conference*, pp. 251–259, Brno.

I. V. Azarova and A. A. Sinopalnikova (2004b), Corpus Statistics for RussNet structuring, in *Proc Corpus Linguistics 2004*, pp. 5–15, Saint-Petersburg.

M. Derwojedowa, M. Piasecki, S. Szpakowicz, M. Zawisławska, and B. Broda (2008), Words, Concepts and Relations in the Construction of Polish WordNet, in *Proc Fourth International WordNet Conference*, pp. 162–177, Szeged.

D. Faure and C. Nédellec (1998), A corpus-based conceptual clustering method for verb frames and ontology acquisition, in *Proc Workshop on Adaptating Lexical and Corpus Resources to Sublanguages and Applications*, pp. 5–12, Granada.

Ch. Fellbaum, editor (1998), *WordNet. An Electronic Lexical Database*, MIT Press, Cambridge.

A. Kilgariff, P. Rychly, P. Smrž, and D. Tugwell (2004), The Sketch Engine, in *Proc EUROLEX-2004*, pp. 105–116, Lorient.

C. H. A. Koster (2004), Head/Modifier Frames for Information Retrieval, in *Proc CICLing-2004*, pp. 420–432, Springer LNCS 2945, Seoul.

C. Leacock and M. Chodorow (1998), Combining Local Context and WordNet Similarity for Word Sense Identification, in Ch. Fellbaum, editor, *WordNet: An Electronic Lexical Database*, pp. 265–283.

R. Marinelli and G. Spadoni (2005), Some Considerations in Structuring a Terminological Knowledge Base, in *Proceedings of the Third International WordNet Conference*, pp. 217–223, Brno.

G. A. Miller, R. Beckwith, Ch. Fellbaum, D. Gross, and K. Miller (1990), Introduction to WordNet: An On-line Lexical Database, in *International J. of Lexicography 3 (4)*, pp. 235–244.

S. Sharoff (2002), The frequency dictionary for Russian, http://www.comp.leeds.ac.uk/ssharoff/frqlist/frqlist-en.html.

A. A. Sinopalnikova (2004), Word Association Thesaurus As a Resource for Building WordNet, in *Proc Second International WordNet Conference*, pp. 199–205, Brno.

D. Tufiş, D. Cristea, and S. Stamou (2004), BalkaNet: Aims, Methods, Results and Perspectives: A General Overview, in *Romanian J. on Information Science and Technology, Special Issue on BalkaNet, 7(2-3)*.

E. M. Voorhees (1998), Using WordNet for Text Retrieval, in Ch. Fellbaum, editor, *WordNet: An Electronic Lexical Database*, pp. 285–303.

P. Vossen, editor (1998), *EuroWordNet: A Multilingual Database with Lexical Semantic Network*, Kluwer, Dordrecht.