Multiword Expressions in Russian Thesauri RuThes and RuWordNet

Natalia Loukachevitch
Lomonosov Moscow State University,
Bauman Moscow State Technical University
Moscow, Russia
louk_nat@mail.ru

German Lashevich
Kazan Federal University
Kazan, Russia
design.ber@gmail.com

***Abstract*—We present the types of multiword expressions included into the thesaurus of Russian language RuThes. Many of these expressions may look like compositional expressions but have specific relations that can be useful in applications. The relation system of the RuThes thesaurus allows natural description of relations between an expression and its components if necessary. Transforming the RuThes knowledge into the Princeton WordNet structure for creating Russian wordnet (RuWordNet), we transfer also all the described expressions into the new resource and propose to automatically introduce additional relations for their better representation.**

I. Introduction

One of the difficult issues in developing application-oriented resources such as wordnets or information-retrieval thesauri is the inclusion of concepts (synsets or descriptors) based on the senses of multiword expressions, for example noun compounds [1], [2], [3], [4]. Two main questions are usually discussed here: what the principles of inclusion of multiword expressions (especially compositional or semi-compositional ones) are and what types of relations should connect a multiword expression and its components in the thesaurus structure.

A lot of approaches were proposed for automatic extraction of idioms, collocations or multiword terms in numerous works, which should help to find multiword expressions appropriate for inclusion in lexical or terminological resources [5], [6], [7], [8], [9], [10], [11], [12]. Some works are devoted to automatic identification of idiomatic expressions in contrast to compositional expressions [13], [14]. Several datasets for non-compositionality detection have been created [13], [15]. These datasets contain expressions ordered from non-compositionality to compositionality according to human judgments.

However, developers of computational resources should have clear guidelines for the introduction of phrases into their resources, which can possibly depend on the goal of the resource under development. Thus, special instructions on introducing multiword terms exist for developing information-retrieval thesauri [16]. Developers of wordnet-like thesauri, a very popular type of resources, discuss the problem of introducing multiword expressions in their resources in several works [2], [4], [12], [17], [18].

The thesaurus of the Russian Language RuThes [19] has a specific approach in describing and including multiword expressions [21], [22]. This approach is supported with the relational structure of RuThes, which is quite different from WordNet-like resources. The specific feature of working with multiword expressions in RuThes is not to check them from the point of view of compositionality but to assess if the introduction of a multiword expression adds new useful information that does not follow from expression's components. The RuThes relations provide the possibility of describing the links between concepts corresponding to multiword expressions and concepts of its component words, if these semantic relations really exist.

Lately, the new Russian wordnet, RuWordNet [20], was semi-automatically generated from RuThes [23]. RuWordNet also contains many multiword expressions including those that look like compositional but can be useful for natural language applications. In this paper we present classes of multiword expressions described in RuThes and discuss the way to describe these expressions and appropriate links in RuWordNet.

The structure of the paper is as follows. In the second section we consider related work formulating rules for including phrases in thesauri intended for natural language processing or information retrieval applications. In the third section the general structure, system of relations and rules of inference in RuThes are described. The fourth section presents the principles of including phrases in RuThes. In the fifth section we describe the current state of RuWordNet and propose types of relations for the description of compositional multiword expressions.

II. Related work

Multiword expressions comprise a broad scope of phrases including idiomatic expressions, noun compounds, technical terms, proper names, verb-particle and light verb constructions, institutionalized phrases, and others [24], [25]. For some of these constructions such as idioms, it is evident that they should be included in computational lexicons. But for many of other expressions, it is not simple to make a decision about the necessity of their inclusion. Therefore, in various works [24], [25], [2], [4], additional criteria of phrases description are considered:

- high frequency,

- high calculated scores of association measures between phrase components,

ISBN 978-952-68397-8-3 (PDF)

- expressing a concept which is not lexicalized in one language, but is lexicalized in another language,

- synonym with a lexical unit in the same language,

- their components are highly polysemous,

- the modifier in a complex nominal has been reduced or has no semantic referentiability,

- truly complex nominals define a new subtype of the entity denoted by the head of the compound and others.

Manuals and standards on information-retrieval thesaurus development provide detailed principles for multiword term selection. For example, the American standard on the construction of monolingual thesauri Z39.19 considers such principles as (Z39.19 2005):

- frequency in domain-specific texts and importance for the domain community (literary warrant),

- splitting the parts would lead to ambiguity or loss of meaning,

- one component of a phrase is too vague,

- meaning of the compound term as a whole is not the sum of the meanings of its parts etc.

It is supposed that wordnets have to include so called lexicalized concepts as synsets [26]. However, developers of wordnets [1] stress that boundaries of lexicalization are very difficult to draw. Besides, there is a necessity to include non-lexicalized phrases into wordnets. First, non-lexicalized entities are often necessary for establishing relations between synsets of a non-English wordnet and English wordnet in cases of lexical gaps. Second, the inclusion of more multiword expressions into wordnets is considered as the possibility to introduce more syntagmatic information, describe additional information for better word sense disambiguation procedures.

To describe non-lexicalized free expressions in Italian wordnet Multiwordnet, Bentivogli and Pianta [2] proposed to use a special data structure called phraset. Phrasets can be added in correspondence with empty or non-empty synsets, for example:

Eng-synset: toilet_roll

Ita-synset: GAP

Ita-phraset: rotolo_di_carta_igienica.

Developers of Basque wordnet also add special syntagmatic concepts in their wordnet but mark them with the special label IXALEX.

To describe the inner structure of a multiword expressions described as a phraset, Bentivogli and Pianta proposed to use a composed-of relation between the described multiword expression and their components including their word sense specification if possible.

The composed-of relation does not describe semantic structure of a multiword expression. Therefore for description of multiword entities in Basque wordnet, Agirre et. al. suppose to use INVOLVED relations from the relation inventory of EuroWordNet: The INVOLVED relation is used to encode data on arguments or adjuncts lexicalized within a meaning of a 2nd order entity such as involved_theme, involved_instrument relations and others.

Maziarz et al. [4] try to formulate the procedural definition of multiword lexical units should be included in Polish wordnet (plWordNet) to make lexicographers apply these principles consistently. In the first experiment the authors provide linguists with the following definition of multi-word lexical unit - "an expression built from more than one word, associated with a definite meaning somehow stored in one's mental lexicon and immediately retrieved from memory as a whole". Then they ask 14 linguists to classify multiword expressions using this definition into three classes: multiword lexical unit (Yes – 1), not multiword lexical unit (No – -1), and don't know (0). They conclude that if to have a group of 5-7 linguists then it is possible to obtain multiword lexical units for inclusion to a wordnet with the appropriate agreement. But this approach is too expensive.

III. RuThes thesaurus

The thesaurus of Russian language RuThes [21], [22] is 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. 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 senses correspond to the concept, so called ontological synonyms.

Ontological synonyms of a concept can comprise words belonging to different parts of speech; language expressions relating to different linguistic styles, genres; idioms and even free multiword expressions, for example, synonymous to single words): красный, краснота, красный цвет (*red, redness, red color*).

The relations in RuThes are only conceptual, not lexical (as antonyms or derivational links in wordnets). They are constructed as more formal, ontological relations originated from traditional information-retrieval thesauri [16]. The set of conceptual relations includes:

- the class-subclass relation;
- the part-whole relation applied with the following restriction: the existence of the concept-part should be strictly attached to the concept-whole (so tree can grow in many places therefore concept $TREE$ cannot be directly linked to concept $FOREST$ with the part-whole relation, the additional concept $FOREST$ $TREE$ should be introduced);
- the external ontological dependence when the existence of a concept depends on the existence of another concept (in such a way forests depend on the existence of trees) [27]. In RuThes we denote this relation as association with indexes: asc_1 is directed to the main concept, asc_2 indicates towards the dependent concept;
- symmetric associations between much related concepts can be established.

Several properties are defined over RuThes relations. These properties give the possibility to make logical inference, in

particular, to find semantic relatedness between text entries that are not directly connected with each other by the thesaurus relations. These properties include:

- transitivity of class-subclass relations:

$subclass(X,Y) \wedge subclass(Y,Z) => subclass(X,Z).$

- transitivity of part-whole relations. It should be noted that the transitivity of part-whole relation is an often discussed issue in computational applications (see more detailed consideration in [22]) but the rules of establishing those relations in RuThes allows exploiting this property:

$whole(X,Y) \wedge whole(Y,Z) => whole(X,Z).$

Also the following inheritance rules are valid in RuThes:

- whole relations are inherited to subclasses:

$subclass(X,Y) \wedge whole(Y,Z) => whole(X,Z).$

- asc_1 relations are inherited to subclasses and parts:

$subclass(X,Y) \wedge asc_1(Y,Z) => asc_1(X,Z)$

$whole(X,Y) \wedge asc_1(Y,Z) => asc_1(X,Z).$

- asc relations are inherited to subclasses and parts:

$subclass(X,Y) \wedge asc(Y,Z) => asc(X,Z)$

$whole(X,Y) \wedge asc(Y,Z) => asc(X,Z).$

Considering all possible relation paths existing between two thesaurus concepts C_1 and C_2, it was supposed that those paths that can be reduced to a single relation with the application of the above-mentioned rules of transitivity and inheritance indicate semantic relatedness between concepts C_1 and C_2, so called semantic paths. Word and phrases presented as thesaurus entries assigned to the concepts C_1 and C_2 are also considered semantically related even if the length of the path is quite large (five and more relations). Such defined semantic similarity between words and phrases included in RuThes is used for query expansion in information retrieval, thematic text representation [28], representation of categories in text categorization [22], and automatic word sense disambiguation [29].

For representing multiword expressions, it can be supposed that the existence of a semantic path between concepts corresponding to a multiword expression and concepts of its component words means that the multiword expression preserves the senses of its component words.

The publicly available version of the RuThes thesaurus, RuThes-lite 2.0 comprises 115 thousand Russian words and expressions, can be seen at http://www.labinform.ru/ruthes/index.htm and can be obtained in form of xml files for noncommercial use.

IV. Multiword expressions in RuThes

The distinctive feature of RuThes is that it contains many multiword expressions. Experts are encouraged to introduce new multiword expressions into RuThes if they can substantiate their decision with the necessity to represent the expression in the thesaurus. The expert should show that adding the expression to the thesaurus gives useful information that does not follow from the component structure of this expression. Such information is usually expressed in form of additional thesaurus relations (or their deliberate exclusion), which enriches the thesaurus knowledge.

In fact, we shift the often discussed question on compositionality vs. non-compositionality of a multiword expression to the more visible question of adding information to a thesaurus. Usually various fixed and semi-fixed expressions, idioms are evident and do not require much argumentation for their inclusion into the thesaurus. But numerous noun and verb compounds are more difficult for analysis and the analysis of their usefulness can be easier than the abstract consideration of their non-compositionality.

Potential inclusion of compositional expressions into a thesaurus requires adequate description of links between them and their components allowing not to lose information about their semantic compositionality. If a compositional expression is attached to a concept C_{mwe} then concept C_{mwe} is usually described as a subclass of the concept C_H corresponding to the head word of the expression. The link to the concept C_D of syntactically dependent word in the expression is usually described as association relation (asc_1): $asc_1(C_{mwe}, C_D), asc_2(C_D, C_{mwe})$. For example, concept $AIRCRAFT\ INDUSTRY$ is described as follows: $asc_1(AIRCRAFT\ INDUSTRY, AIRCRAFT)$. These relations can be also represented not as direct relations of concepts but can be inherited according to the RuThes relation properties.

The published version of RuThes contains about 52 thousand multiword expressions.

The employed principles of introducing multiword expressions into RuThes can be subdivided as follows:

- absence of meaningful relations between an expression and senses of component words,
- synonym to own component word,
- additional relationships to other single words and multiword expressions.

In next subsections we will consider these principles in a more detailed way.

TABLE I. High frequent non-compositional expressions in a news collection

Expression	Comments
круглый стол (round table)	совещание (meeting)
юридическое лицо (legal person)	правовой статус организации (legal status of organization)
точка зрения (point of view)	мнение (opinion)
товарищеский матч (friendly competition)	вид соревнований, к друзьям отношения не имеет (subtype of competitions, no relation to friends)
ценная бумага (securities)	подвид собственности (subtype of property)
операционная система (operating system)	программа (software)
горячая линия (hot line)	линия связи (communication line)

A. Absence of relations between an expression and senses of component words

If an expression is an idiom, that is, its meaning is not the predictable sum of the meanings of its components, it is naturally represented in the form in thesaurus relations. We should do not describe relations to those component's concepts that have lost their meanings within the expression and we add relations for describing the additional thesaurus relations that follow from the meaning of the expression.

The Table I contains examples of idioms frequently used in a large collection of news articles. In RuThes, a separate concept corresponds to the expression круглый стол (*round table*) and it has no relations to concepts of words круглый (*round*) and стол (*table*). But it has the relations with concepts of meeting and discussion. The expression товарищеский матч (*friendly competition*) also has a special concept, this concept has no relations to the concepts denoting friends and friendly relationships.

TABLE II. HIGH FREQUENT MULTIWORD EXPRESSIONS IN A NEWS CORPUS SYNONYMOUS TO SINGLE WORDS

Expresison	Synonymous expression	Type of synonym
глава государства (head of state)	правитель (ruler)	Lexical unit
законодательное собрание (legislative assembly)	парламент (parliament)	Lexical unit
заработная плата (salary)	зарплата	Abbreviation
генеральный директор (director general)	гендиректор, гендир	Abbreviations
политическая партия (political party)	партия (party)	Component
компьютерная программа (computer program)	программа (program)	Component
принять участие (take participation)	участвовать (participate)	Derivation of component
программное обеспечение (computer software)	программа (software)	Derivation of component
боевые действия (combat actions)	бой (combat)	Derivation of component
высказать мнение (give an opinion)	высказаться	Derivation of component
большая часть (major part)	большинство (majority)	Derivation of component

B. Expression synonymous to own components

The next class of multiword expressions comprises those expressions which are elements of the same synset as its component word. We call them multisynonyms. The Table II contains examples of such expressions as политическая партия (*political party*), which is a quite frequent synonym to one of senses of its component word партия (*party*). Another example is компьютерная программа (*computer program*).

Multisynonyms can be also synonyms to derivations of their component words. The Table II contains the expression with light verb принять участие (*take participation*), which is a synonym to the verb участвовать (*participate*).

In creating RuThes, introduction of such multiword synonyms was especially encouraged, because the important feature of these expressions is that their components can be ambiguous, but the whole expression is unambiguous. Thus, if the expression is known and described in a thesaurus there are no problems with disambiguation of its components and with the semantic interpretation of the whole expression. In fact, these expressions can improve the recognition of own synsets. In addition, inclusion of such expressions in a synset often clarifies the sense of the synset. It is clear that introduction of these expressions does not require additional concepts.

Such multisynonyms are very common in Russian language. Currently, the published version of RuThes - RuThes 2.0 contains more than 13 thousand multiword synonyms. Numerous examples of multisynonyms can be found also in English. Examples of multisynonyms can be met in Princeton WordNet [26]. For example,

- the synset of *plant.1 (buildings for carrying on industrial labor)* contains multiword expression *industrial plant*,
- the synset of *platform.2 (a document stating the aims and principles of a political party)* includes multiword expression *political platform*,
- the synset of *park.5 (a lot where cars are parked)* contains expressions *car park* and *parking lot*.

TABLE III. HIGH FREQUENT MULTIWORD EXPRESSIONS GENERALIZING SINGLE WORDS

Expression	single words- hyponyms
акция протеста (action of protest)	забастовка, самосожжение (strike, self-burning)
глава государства (head of state)	президент, король (president, king)
взрывное устройство (explosive device)	бомба,ракета (bomb, missile)
федеральный закон (federal law)	кодекс, конституция (code, constitution)
телесное повреждение (body injury)	ожог, травма, пролежень (burn, trauma, bedsore)
медицинский работник (health professional)	врач, медсестра, (physician, nurse)

C. Additional relationships to other single words and multiword expressions

Some multiword expressions look like compositional ones but they have specificity in relations with other single words and/or expressions, which usually means that these expressions denote some important concepts, entities or situations. The following subtypes of these expressions can be considered:

- an expression is a synonym to a single word;
- an expression has a frequent abbreviation: генеральный директор – гендиректор, гендир (Table II);
- an expression has a synonymous expression and this fact is very difficult to infer from components: заснуть за рулем – заснуть во время движения (compare English expressions *falling asleep at the wheel* and *falling asleep while driving*);
- an expression generalizes several single words. Such expressions as *action of protest, head of state* (Table III) often look like compositional but they have a very important function of knowledge representation: they gather together similar concepts;

- an expression has relations that do not follow from its component words. The Table IV gives examples of such expressions. For example, it may be not very clear if the expression дорожное движение (*road traffic*) is compositional or not. But it has numerous relations with other expressions that cannot be inferred from its components.

TABLE IV. HIGH FREQUENT MULTIWORD EXPRESSIONS SEMANTICALLY RELATED TO OTHER EXPRESSIONS; ONE OR MORE COMPONENTS ARE NOT RELATED

Expression	Related expression	Type of relation
проезжая часть (traffic way)	автомобильная дорога (motor road)	whole
проезжая часть (traffic way)	дорожная полоса (road lane)	part
дорожное движение (road traffic)	левостороннее движение (left-hand traffic)	hyponym
	одностороннее движение (one-way traffic)	hyponym
выходной день (week end)	свободное время (free time)	hyperonym
питьевая вода (drinking water)	пресная вода (fresh water)	hyperonym

V. RUWORDNET: CURRENT STATE AND MULTIWORD EXPRESSIONS

To create RuWordNet, the conceptual net of RuThes was subdivided into three nets of synsets: nouns (including single nouns, noun groups, or preposition groups), adjectives (single adjective and adjective groups), and verbs (single verbs and verb groups) [22]. In such a way, the RuThes concepts comprising text entries of different parts of speech were transformed into several synsets. The divided synsets were linked with the relation of part-of-speech synonymy (POS synonymy). The Table V contains the statistics of RuWordNet synsets.

TABLE V. QUANTITATIVE CHARACTERISTICS OF SYNSETS IN RUWORDNET

Part of speech	Number of synsets	Number of unique entries	Number of senses
Noun	29296	68695	77153
Verb	7634	26356	35067
Adjective	12864	15191	18195

Synsets in RuWordNet have the following types of relations (Table VI):

- hyponym-hypernym relations. These relations were taken from the corresponding relations described in RuThes. In addition, the hyponym-hypernym relations can be obtained with transitive closure of RuThes relations: if a specific concept does not have a text entry belonging to a given part of speech (for example, adjective) but its child and parent have adjective text entries than the hyponym-hypernym relation is established between the child and the parent directly;
- part-whole relations (meronym-holonym). The part-whole relations from RuThes were semi-automatically transferred and corrected according to traditions of WordNet-like resources. The part-whole relations include the following subtypes: functional parts (*nostrils – nose*), ingredients (additives – substance), geographic parts (*Sevilia – Andalusia*), members (*monk – monastery*), dwellers (*Moscow citizen – Moscow*), temporal parts (*gambit – chess party*), inclusion of processes, acitivities (*industrial production – industrial cycle*);
- instance-class relations established for nouns; at present, they describe relations between synsets of specific geographical entities and their geographical types (*Moscow – city*);
- antonymy relations established mainly between synsets denoting properties and states. These relations were extracted semi-automatically from association relations between properties and states described in RuThes;
- POS-synonymy relations mentioned before.

TABLE VI. RELATIONS IN RUWORDNET

Part of speech	Hypernym	Instance Class	Holonym	POS Synonymy	Antonyms
Noun	39155	1863	10010	18179	455
Verb	10440	0	117	7451	20
Adjective	17834	66	829	14139	457

Multiword expressions from RuThes-2.0 were also transfered to RuWordNet. But because of another system of relationships, many of the compositional multiword expressions described in RuThes could lose their links to the synsets of their components, which can be not appropriate for NLP applications. Thus, such lost connections that can be inferrable from the RuThes structure should be automatically added.

With this aim, we introduce the pair of relations "has_component" and "component_for". We suppose that the direct relations (synonyms, hyponyms, hypernyms, parts, wholes, and POS synonyms) should not be additionally described between the synsets of a multiword expression and its components. But all the relations that require inference including the transitivity of hyponym-hypernym relations, the use of the dependence relations of RuThes, transitivity of part-whole relations which is presupposed in RuThes, the inheritance of relations to hyponyms and parts should be additionally indicated. The Table VII presents quantitative characteristics of relations between synsets of multiword expressions and their components.

TABLE VII. RELATIONS OF MULTIWORD EXPRESSIONS WITH THEIR COMPONENTS IN RUWORDNET

Type of relations	Number	Comment
The same synset	13367	
POS-synonymy	6285	
Other direct relations RuWordNet	16279	
Direct relations from RuThes not included to RuWordNet	15677	Should be added
Inferred relations from RuThes	12513	Should be added

VI. CONCLUSION

In this paper we described various principles for including multiword expressions into thesauri intended for using in natural language processing and information-retrieval applications. We argue that the thesaurus structure gives the possibility

to present the semantic peculiarity of an expression when describing those their relations that do not follow from its components.

We presented the types of multiword expressions included into the thesaurus of Russian language RuThes. Many of these expressions may look like compositional expressions but they have specific relations that can be useful in applications.

The relation system of the RuThes thesaurus allows natural description of relations between an expression and its components if necessary. Transforming the RuThes knowledge into the Princeton Wordnet structure for creating Russian wordnet (RuWordNet), we transfer also all the described multiword expressions in the new resource and propose to automatically introduce additional relations describing links of an expression with its components.

ACKNOWLEDGMENT

This work is partially supported by the Russian Science Foundation (grant N16-18-02074) in part of the analysis of RuThes multiword expressions (sections I-IV) and by the Russian Foundation for Humanities (grant N15-04-12017) in parts devoted to development of the Russian wordnet RuWordNet (section V).

REFERENCES

[1] E. Agirre, A. Izaskun Aldezabal and E. Pociello, "Lexicalization and multiword expressions in the Basque WordNet", *in Proc. of Third International WordNet Conference*, 2006.

[2] L. Bentivogli and E. Pianta, "Extending Wordnet with Syntagmatic Information", *in Proc. of Second Global WordNet Conference*, 2004, pp. 47–53.

[3] P. Kedzia, M., Piasecki, M. Maziarz and M. Marcińczuk, "Recognising compositionality of multi-word expressions in the wordnet oriented perspective", *in Proc. of Mexican International Conference on Artificial Intelligence*, 2013, pp. 240-251.

[4] M. Maziarz, S. Szpakowicz and M. Piasecki, "A procedural definition of multi-word lexical units", *in Proc. of Recent Advances in NLP Conference RANLP-2015*, 2015, pp. 427-435.

[5] C. Bonia, M. Green, J. Preciado and M.Palmer, "An approach to take multi-word expressions", *in Proc. of the 10th Workshop on Multiword Expressions*, 2014, pp. 94-98.

[6] B. Dobrov and N. Loukachevitch, "Multiple Evidence for Term Extraction in Broad Domains", *in Proc. of Recent Advances in NLP Conference RANLP-2011*, 2011, pp. 710-715.

[7] N. Astrakhantsev, D. Fedorenko and D. Turdakov, "Methods for automatic term recognition in domain-specific text collections: A survey", *Programming and Computer Software* , 2015, pp. 336-349.

[8] A. Gelbukh and O. Kolesnikova, "Multiword Expressions in NLP: General Survey and a Special Case of Verb-Noun Constructions", in *Computational Linguistics: Concepts, Methodologies, Tools, and Applications*, 2014, pp. 178-198.

[9] S. Gries, "50-something years of work on collocations: what is or should be next...", *International Journal of Corpus Linguistics* vol. 18, 1, 2013, pp. 137-166.

[10] E.Tutubalina, P. Braslavski, "Multiple Features for Multiword Extraction: Learning to Rank Approach" *in Proc. of Conference on Computatilnal linguistics and Intellectual technologies Dialog-2016*, 2016, pp.782-793.

[11] P. Pecina, "Lexical association measures and collocation extraction", *Language resources and evaluation* vol. 44,1-2, 2010, pp. 137-158.

[12] M. Piasecki, M. Wendelberger, M. Maziarz, "Extraction of the Multiword Lexical Units in the Perspective of the Wordnet Expansion", *in Proc. of Recent Advances in NLP Conference RANLP-2015*, 2015, pp. 512-522.

[13] S. Reddy, D. McCarthy and S. Manandhar. 2011. "An empirical study on compositionality in compound nouns", *in Proc. of The 5th International Joint Conference on Natural Language Processing IJCNLP-2011*, 2011.

[14] S. Cordeiro, C. Ramisch, M. Idiart, A. Villavicencio, "Predicting the Compositionality of Nominal Compounds: Giving Word Embeddings a Hard Time", *in Proc. of the 54th ACL-2016* , Volume 1: Long Papers, 2016, pp.1986-1997.

[15] C. Ramisch, S. Cordeiro, L. Zilio, M. Idiart, A. Villavicencio, and R. Wilkens, "How naked is the naked truth? a multilingual lexicon of nominal compound compositionality", *in Proc. of the 55th ACL-2016*, Volume 2: Short Papers, 2016.

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

[17] G. Lebani and E. Pianta, " Encoding Commonsense Lexical Knowledge into WordNet", *in Proc. of Global WordNet Conference, GWC-2012*, 2012.

[18] V. Vincze and A. Almási, "Non-lexicalized concepts in wordnets: A case study of English and Hungarian", *in Proc. of the 7th Global WordNet Conference GWC-2014*, 2014, pp. 118-126.

[19] Ruthes Thesaurus, Web: http://www.labinform.ru/pub/ruthes/

[20] RuWordNet Thesaurus, Web: http://ruwordnet.ru/, http://www.labinform.ru/pub/ruwordnet/index.htm

[21] N. Loukachevitch and B. Dobrov, "RuThes Linguistic Ontology vs. Russian Wordnets", *in Proc. of Global WordNet Conference, GWC-2014*, 2014, pp. 154-162.

[22] N. Loukachevitch, B. Dobrov, " The Sociopolitical Thesaurus as a resource for automatic document processing in Russian", *Terminology*, 21(2), 2015, 237-262.

[23] N. Loukachevitch, G. Lashevich, A. Gerasimova, V. Ivanov and B.Dobrov. "Creating Russian WordNet by Conversion", *in Proc. of Conference on Computational linguistics and Intellectual technologies Dialog-2016*, 2016, pp.405-415

[24] N. Calzolari, Ch. Fillmore, R. Grishman, N. Ide, A. Lenci, C. MacLeod and A. Zampolli, " Towards Best Practice for Multiword Expressions in Computational Lexicons", *in Proc. of LREC - 2002*, 2002, pp.1934-1940.

[25] I. Sag I, T. Baldwin, F. Bond, A. Copestake, and D. Flickinger," Multiword expressions: A Pain in the Neck for NLP", *In Proc. of CICLING 2002*, 2002.

[26] G. Miller, " Nouns in WordNet", *In: Fellbaum, C (ed) WordNet – An Electronic Lexical Database*, The MIT Press, 1998, pp.23-47.

[27] N. Guarino and C. Welty, "Evaluating ontological decisions with OntoClean", *Communications of the ACM*, V.45 (2), 2002, pp. 61-65.

[28] N. Loukachevitch and A. Alekseev, "Summarizing News Clusters on the Basis of Thematic Chains", *in Proc. of Language Resources and Evaluation Conference, LREC-2014*, 2014, pp. 1600-1607.

[29] N. Loukachevitch and I. Chetviorkin, "Determining the most frequent senses using Russian linguistic ontology RuThes", *in Proc. of the Workshop on Semantic resources and Semantic Annotation for Natural Language Processing and the Digital Humanities at NODALIDA 2015*, 2015, pp. 21–27.