

Lexicographic Study of Synonymy: Clarifying Semantic Similarity between Words

Gulnara Gimaletdinova, Liliia Khalitova, Valery Solovyev, Vladimir Bochkarev

Kazan Federal University,
Russia

gim-nar@yandex.ru

Abstract. The problem of determining semantic similarity between words affects the understanding of synonymy and creates obstacles to the work of lexicographers. The study was carried out as a part of a larger research project on expert assessment of synonymic rows in RuWordNet thesaurus (a WordNet-like thesaurus for the Russian language). The aim of this study is to analyze RuWordNet thesaurus and compare it with classical dictionaries of Russian synonyms. For this purpose, the authors singled out entry words (adjectives $N = 68$ and verbs $N = 117$) and their analogues (adjectives $N = 558$ and verbs $N = 1410$) from the New Explanatory Dictionary of Russian Synonyms by Yu. Apresyan (NEDS). An analogue is viewed as a word whose meaning essentially intersects with the general meaning of a given synonymic row, although it lacks the needed semantic similarity that could indicate the presence of synonymy or near-synonymy (Apresyan). The quantitative analysis based on the breadth-first search (BFS) algorithm estimated the distance between each pair entry word→analogue. The quantitative method revealed that the analogues described in NEDS correlate with the hyponyms and hyperonyms in RuWordNet which contributes to the study of near-synonymy. The qualitative method (observation and linguistic interpretation) was used to analyze pairs entry word→analogue which showed the longest distance; such words were 52 adjectives and 15 verbs. First, the meanings of entry words and analogues were checked against two Russian language thesauri, then, their representation in the tree graph of RuWordNet was traced. The analysis revealed inaccuracies concerning the similarity between certain words. The recommendations for further improvement of RuWordNet were given.

Keywords. Synonymy, semantic similarity, near-synonyms, RuWordNet.

1 Introduction

As early as the second half of the 19th century, structuralists defined synonyms as lexical items that share the same, or similar, meanings based upon contextual factors [19]. However, to date, many issues of synonymy, in particular the main problem of determining the semantic similarity between words are still under discussion. A number of researchers point out that there is no need to analyze identical words with similar meanings, since the very nature of a language sign implies a need for differentiation (semantic, stylistic, etc.) rather than semantic similarity [6, 7, 11].

There is a number of lexicographical studies where the following terms are used to denote different degrees of semantic similarity: synonyms, near-synonyms, analogues, hyponyms and hyperonyms [2, 10, 12]. There is no agreement in the understanding of what semantic similarity is [1, 9, 28]. Therefore, lexicographic dictionaries and ontologies/thesauri present synonyms and near-synonyms differently.

In the Russian language, a modern thesaurus RuWordNet¹ [18] is considered to be one of the most successful resources of synonyms for automatic word processing. The Russian language thesaurus RuWordNet was created on the basis of the automated transformation of RuThes thesaurus² into the well-known WordNet format³.

RuWordNet thesaurus has noun, adjectival and verbal sets of synonyms (synsets) organized in accordance with RuThes concepts. The thesaurus contains 111.5 thousand words and expressions of the Russian language including 29297 synsets of nouns, 12865 synsets of adjectives and 7636 synsets of verbs.

¹ <https://ruwordnet.ru/ru>

² <http://www.labinform.ru/pub/ruthes/>

³ <https://wordnet.princeton.edu>

RuWordNet establishes hyponym-hyperonymic (genus–species) and antonymic relationships, as well as the relationships of instance–class, part–whole, reason, logical sequence and subject area (domain) [17].

Although there are studies that concern verification and assessment of RuWordNet [4, 18, 23], in this article the authors propose the clarification of the semantic proximity of words in the thesaurus based on quantitative and qualitative methods.

The following specific research question was addressed in the study: How can inaccuracies in RuWordNet thesaurus be revealed and explained by clarifying the degree of semantic similarity between lexical items?

The authors see the main contributions of the research in the following:

1. Using quantitative method (breadth–first search (BFS) algorithm) the authors provided an analysis of semantic similarity between words (adjectives and verbs) in RuWordNet thesaurus.
2. Applying a qualitative method (observation and linguistic interpretation) the authors clarified semantic similarity between the words with similar meaning and gave recommendations concerning the revealed deficiencies.

2 Related Work

In a broad sense, synonymy implies identity, generality and is manifested at various linguistic levels of the language, but mainly in vocabulary. Although the concept of synonymy is well known, the exact criteria for synonymy are still a subject of controversy.

This is associated with the difficulties both in determining the criteria for distinguishing synonyms from non–synonyms and in practical application.

Semantic similarity, in its turn, is an intuitively clear criterion, but difficult to define. Moreover, questions of semantic similarity between words are important for different lexicographic descriptions, in particular for making synonymic dictionaries, thesauri and lexical databases [26, 22, 27].

The inventory of synonymic rows and the description of semantic similarity between synonyms depend on how accurately and objectively the distance between words is measured [16].

2.1 Theoretical Questions of Synonymy and Semantic Similarity between Words

Following [6], in linguistics the division of synonyms into absolute, propositional/cognitive, and near–synonyms (quasi–synonyms/plesionyms) has been adopted.

Absolute synonyms are lexical elements that can be used interchangeably in all contexts, since they express an absolute identity of meanings and must share all semantic and syntactic properties with each other [7]. However, absolute synonymy, if it exists at all, is quite rare [7, 12, 25]. The constant development and change of natural languages lead to semantic changes when one of the words of the synonymic series becomes obsolete or develops another semantic function. In this regard, absolute synonyms are a rare occurrence [6]. Absolute synonymy is limited mainly to dialectal variations and technical terms [12].

To propositional (cognitive) synonyms it is customary to attribute lexical elements that have ‘certain common semantic properties’, expressing paradigmatic relationships and designating identity in the composition of individual phrases or whole sentences [6]. For example, the nouns violin and fiddle, while not absolute synonyms in the English language, in the sentence He is turning his violin/fiddle demonstrate the use of cognitive synonyms [7]. Cognitive synonyms have common propositional meanings, but differ in the degree (presence/absence) of expressivity. Cognitive synonyms are so similar in meaning that when interchangeable in a certain context, they cannot be differentiated either denotatively or connotatively [6].

To designate near–synonyms, the terms ‘plesionyms’ [6] and ‘almost synonyms’ [12, 25] are used. Cruse contrasts near–synonyms to cognitive synonyms because they express different truth values in a given context [6]. Near–synonymy is the most complex notion, since near–synonyms serve to denote the same concept, while they do not allow substitution in the same contextual use. For example, the adjectives handsome and pretty denote the concepts of external attractiveness, but the first is mainly used to describe males, and the second — females⁴ [7].

Thus, near–synonyms are not completely interchangeable, but differ in shades of designation, connotation, implicativity, accent or register. On the whole, researchers note that from a linguistic point of view, the distinctive properties of near–synonyms are of more interest than their general semantic features [25].

⁴It should be noted that in contemporary English the adjective handsome is frequently used with reference to both males and females.

Cruse carefully details the main types of lexical relationships, noting that paradigmatic relationships occur between elements that can replace each other in the same context, while syntagmatic relationships occur between elements that might be used in the same context. Hyponymic relationships include the notion of Z in X and Z in Y, cognitive synonymy can be expressed as X exactly equivalent to Y, while near-synonymy — X is similar in meaning to Y [6].

In Russian linguistics the studies show no agreement in defining absolute synonymy and the terms 'absolute synonym', 'exact synonym' [2, 14] or 'complete synonym' [14, 15] are used. Absolute synonyms are opposed to incomplete synonyms, or quasi-synonyms/near-synonyms [2]. The principles of distinguishing between absolute synonyms and near-synonyms were defined by Apresyan as follows: 1) a completely identical meaning; 2) the same valency, the number of actants and role structures; 3) reference to the same part of speech [2]. Near-synonyms should have the last two characteristics, but not necessarily completely coincide in meaning, i.e. being rather similar in interpretation, they have differences in the denotative and significative layers of meaning [14].

In most studies two main types of near-synonymic differences are distinguished: hyperonymic (inclusion of meanings, cf. to hurt — to bruise) and hyponymic (intersection of meanings, cf. to bruise — to itch — to ache) [2]. Thus, the relationship between synonymy and hypo-hyperonymy is established through the phenomenon of near-synonymy, that is, the existence of inaccurate, 'approximate' synonyms [8]. Moreover, possible incompatibility of near-synonyms is associated with a hyponymic correlation: mother — father, to go — to run, to ask — to order [14].

2.2 Description of Semantic Similarity between Words in RuWordNet Thesaurus

The hierarchy of all words and phrases in RuWordNet is based on the system of RuThes concepts. The concepts are directly related to the semantic meanings of words and expressions of the Russian language.

As an information retrieval resource for automatic word processing, the thesaurus describes words and their similarity using the principles different from other linguistic, in particular, lexicographic sources [17].

A RuThes concept is a word, a set expression or a free phrase, the meaning of which might be presented by means of a number of ontological synonyms. For example, if we trace the adjectival synset *veselyy* 1 'jolly' the ontological synonyms are the noun *veselye* 'joy' and

the phrase *vesolye nastroyeniye* 'cheerful', they are links to one and the same RuThes concept: *veselye*, *vesolye nastroyeniye* 'joy, cheerful'. The description of semantic similarity between words in RuWordNet is made by means of near-synonyms, words that are close in meaning, but are related to other RuThes concepts.

So, near-synonyms of the adjective *veselyy* 1 are the hyperonyms *radostnyy* 'glad' and *mazhorny* 'in high spirits' referring to the more general RuThes concepts: *chuvstvo radosti* 'feeling of joy' and *khorosheye nastroyeniye* 'good mood' respectively. Other near-synonyms are the hyponyms *igrivyy* 'playful' and *shalovlivyy* 'frolicsome' related to the same RuThes concept: *igrivyy*, *shalovlivyy* 'playful, frolicsome'.

Sets of synonyms (synsets) presented in RuWordNet include words, word combinations and set phrases, and these are the main structural elements of the thesaurus. The concept of synonymy used by RuWordNet developers is based on the criterion that two expressions are synonymous if replacing one of them with another in a sentence does not change the truth value of this statement. In cases where a word has several meanings, it is included in several different synsets [17]. The similarity between the words with similar meaning in the tree graphs of RuWordNet can be presented as a path length from node 1 to node 2 [29]. In this research, the path length will be measured in steps.

The distance between the words and phrases that are given in the same synset is equal to 0 (zero) steps: e.g. the adjectives *bezlyudnyy* 'uninhabited' and *pustynnyy* 'deserted'. The distance between the adjectives *bezlyudnyy* 'uninhabited' and *pustoy* 'empty' and *bezlyudnyy* 'uninhabited' and *malolyudnyy* 'poorly populated' equals to 1 (one) step indicating to hyperonymic and hyponymic relationships accordingly. Moving along the hierarchy of other near-synonyms more steps can be traced, for example, the path length between the words *bezlyudnyy*→*malolyudnyy*→*uyedinyennyy* 'uninhabited'→'poorly populated'→'secluded'.

A number of approaches to measuring similarity between concepts have been taken in previous studies. Wu and Palmer's [29] edge-counting approach suggests measuring the semantic relations between concepts by calculating the lowest super-ordinate depth. According to this approach, the similarity of the concepts increases when the depth of the lowest superordinate of the two concepts becomes deeper. Another method of measuring semantic similarity proposed by Resnik [21] is the information-based approach which measures semantic similarity between two concepts in a taxonomy

and considers conceptual similarity in terms of class similarity of noun synsets.

2.3 Description of Semantic Proximity in the New Explanatory Dictionary of Russian Synonyms

The New Explanatory Dictionary of Russian Synonyms by Yu. D. Apresyan [3], referred to hereafter as NEDS, is a dictionary of a fundamentally new 'active' type because it suggests a detailed description, explanation and use of near-synonyms.

Apresyan clarified the criterion of semantic similarity giving a comparison of analytical interpretation of words. As noted earlier in Section 2.1, despite their semantic similarity, near-synonyms can differ in conceptual content, register, estimated content, compatibility, etc. Moreover, the values of many near-synonyms differ in several parameters, depending on the context.

In NEDS the description of synonymic rows is enriched with analogues. According to Apresyan, an analogue is a word whose meaning essentially intersects with the general meaning of a given synonymic row, although it lacks the needed semantic similarity that could indicate the presence of synonymy or near-synonymy.

In NEDS almost all synonymic rows are extended with analogues, which, according to Apresyan [3], would broaden and clarify the meaning of a particular synonym. For example, the synonymic row *obeshchat'* 1, *davat'* (chestnoye) slovo 1, *sulit'* 1.1, *klyast'sya*, *obyazyvat'sya* 'promise, give a word, bode, give an oath, pledge' has a detailed analytical description of the differences in use of these synonyms and at the end of the dictionary entry a list of analogues is given including 17 items such as *zaveryat'*, *garantirovat'*, *predskazyvat'* 'assure, guarantee, predict' and others.

Addressing the issues of near-synonyms and analogues is relevant for a number of reasons. Firstly, this is due to the increasing interest in ideographic descriptions of the 'active' type, in particular, in connection with the construction of linguistic thesauri [17], which strive to describe all hyponymic, hyperonymic and other relations between words that are close in meaning rather than to structure synonyms. Secondly, the study of near-synonymy is relevant in the view of the increasing interest in describing the linguistic picture of the world and systemic phenomena in vocabulary [2].

3 Data and Related Methodology

The research presents the analysis of RuWordNet thesaurus which was compared with the NEDS by Apresyan [3], in terms of semantic similarity between words. The present study focuses on two parts of speech, adjectives and verbs, because nouns are analyzed within a separate research work.

The study included three main stages.

At the first stage, the authors recorded all entry words that are adjectives (N=68) and verbs (N=117) and registered all lexical items deemed by Apresyan to be analogues⁵. The data collected at this stage was used for the further linguistic interpretation and comparative analysis of near-synonyms (analogues in NEDS and hyperonyms/hyponyms in RuWordNet), which enables to juxtapose semantic relations between near-synonyms in RuWordNet and NEDS.

At the second stage, the quantitative method was used to measure the distance between the analogues and the corresponding adjectival and verbal synsets in RuWordNet. For this purpose, a special computer programme compiled for the project was used to determine the lexical distance between each of the analogues and the corresponding lexical synset in RuWordNet. The distance between these words was measured in steps from 1 to 6.

The programme considers the network of semantic relations as an undirected graph, the vertices of which are words and word combinations, and the edges are their semantic relations. The degree of semantic similarity between two words is evaluated by finding the length of the shortest path in the graph connecting the vertices corresponding to these words [24]. The breadth-first search algorithm was used to find the length of the shortest path [5].

Firstly, the incidence matrix was constructed. To find the length of the shortest path from the word A (entry word) to the word B (analogue), neighborhoods of increasing radius for the word A were successively built. The calculations stopped when the vertex B fell into the resulting neighborhood for a certain radius value. The results of the quantitative analysis were summarized and presented in Excel tables, from which all the possible combinations for each pair of words (entry word→analogue) and the distance between them (number of steps) could be traced.

⁵A comparatively small number of adjectives and verbs analyzed in the research is explained by the explanatory type of NEDS which contains only 354 entries 'representing the basic groups of antropocentric lexica of Russian' [3].

The programme showed that the distance between the analogues and the entry words ranged from 1 to 6 steps. The full observation and the results of statistical analysis are given in Section 5.

Thus, the programme considered the principle of constructing thesauri which presents the relationship between words in synsets as a hierarchy (i.e. tree graph) indicating synonymic, hyperonymic and hyponymic relationships.

At the third stage, we focused on the items with the largest lexical distance between the analogue and synset which were 5 and 6 steps. The programme revealed 52 adjectives and 15 verbs and these lexical items were analysed more closely and checked against two Russian language thesauri [13, 20]. The qualitative method — observation and linguistic interpretation of lexical items — was used to verify the analogues' representation in NEDS and their distribution in the tree graph of RuWordNet.

So, the quantitative and the qualitative methods allowed the authors to verify the meanings of particular adjectives and verbs in RuWordNet, to reveal some deficiencies concerning the similarity between near-synonyms and to give recommendations for further improvement of the thesaurus.

4 Results

We recorded all entry words presented in NEDS, 68 adjectives and 117 verbs and counted their analogues. The number of analogues for 68 adjectives included 558 items and for 117 verbs 1410 items. The data was presented in the form of tables⁶.

To find the distance between the entry word (Adj N=68 and Verb N=117 verbs) and each analogue (Adj N=558 and Verb N=1410 verbs), all possible combinations were measured by the computer programme and presented in the form of Excel tables. For adjectives the total number of possible combinations was 10837, for verbs 138505. The programme found the length of the shortest path between the word A (entry word) to the word B (analogue). Figure 1 shows the possible combinations for the entry word *svoystvennyj* 'intrinsic'.

The programme also showed the distance in steps ranging from 1 to 6 between each pair 'entry word → analogue' (columns D–I, where column D is equal to one step and column I to six steps correspondingly). For example, the distance between the pairs of adjectives *svoystvennyj* → *kharakternyj* 'intrinsic → characteristic'

⁶<https://kpfu.ru/kompleksnyj-analiz-struktury-i-soderzhaniya-366287.html>

Table 1. Semantic similarity between entry words and analogues in RuWordNet

Part of speech	Number of entry words	Number of analogues	Distance (in steps)	Raw number	Percentage
Adjectives	68	558	1	171	30.65
			2	150	26.88
			3	122	21.86
			4	63	11.29
			5	44	7.89
			6	8	1.43
Verbs	117	1410	1	392	27.80
			2	347	31.70
			3	386	27.38
			4	170	12.06
			5	13	0.92
			6	2	0.14

and *svoystvennyj* → *tipichnyj* 'intrinsic → typical' is equal to 1 step, for the pair *svoystvennyj* → *vrozhdennyj* 'intrinsic → inherent' is equal to 2 steps, while the distance between *svoystvennyj* → *spetsificheskij* 'intrinsic → peculiar' and *svoystvennyj* → *spetsifichnyj* 'svoystvennyj → specific' is equal to 4 steps (see Fig. 1).

The results of measuring the semantic similarity of adjectives and verbs are shown in Table 1.

The data presented in Figure 2 show that for both adjectives and verbs, the number of examples with the distance equal to 1–4 steps present the majority (for adjectives 90,68%, for verbs 98,94%). This proves that RuWordNet thesaurus and NEDS describe the semantic similarity between words almost equally. However, the examples with the distance of 5–6 steps (for adjectives 9,32%, for verbs 1,06%) revealed the discrepancies in the representation of semantic similarity between words in the thesaurus and NEDS. So, these cases were subjected to a qualitative expert analysis (see Section 5).

5 Discussion and Recommendation

Due to the broad approach to the issues of synonymy in linguistics and lexicography, in particular regarding semantic similarity between words, it is difficult to distinguish between the phenomena of synonyms,

A	B	C	D	E	F	G	H	I
55	415	'свойственный'	'характерный'	[]	[]	[]	[]	[]
55	416	'свойственный'	'типичный'	[]	[]	[]	[]	[]
55	417	'свойственный'	'органический'	'естественный'	'отличительный'	'специфический'	[]	[]
55	417	'свойственный'	'органичный'	'естественный'	'отличительный'	'специфический'	[]	[]
55	417	'свойственный'	'прирожденный'	'разумный'	'отличительный'	'специфический'	[]	[]
55	417	'свойственный'	'прирожденный'	'харизматический'	'отличительный'	'специфический'	[]	[]
55	418	'свойственный'	'органический'	'естественный'	'отличительный'	'специфичный'	[]	[]
55	418	'свойственный'	'органичный'	'естественный'	'отличительный'	'специфичный'	[]	[]
55	418	'свойственный'	'прирожденный'	'разумный'	'отличительный'	'специфичный'	[]	[]
55	418	'свойственный'	'прирожденный'	'харизматический'	'отличительный'	'специфичный'	[]	[]
55	419	'свойственный'	'прирожденный'	'врожденный'	[]	[]	[]	[]
55	420	'свойственный'	'неотъемлемый'	[]	[]	[]	[]	[]

Fig. 1. The combinations of paths for the adjective *svoystvennyy* 'intrinsic' in Excel table

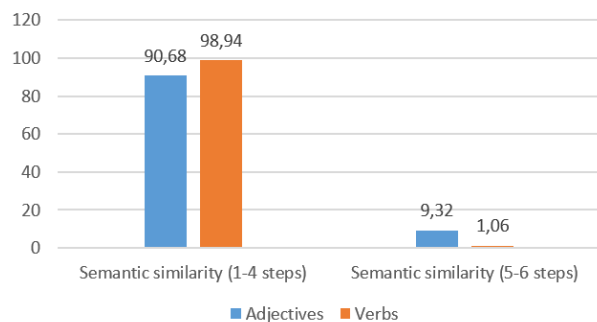


Fig. 2. Degree of similarity of adjectives and verbs in RuWordNet

near-synonyms and analogues. In the present study the authors tried to apply a quantitative computer-based method to measure the degree of semantic similarity between words.

The quantitative analysis based on the breadth-first search algorithm showed that the analogues (in Apresyan's terminology) described in NEDS correlate with the hyponyms and hyperonyms presented in RuWordNet.

The programme also measured the distance between such words in steps equal 1–6. The comparative analysis of RuWordNet thesaurus and NEDS proved that the words with the proximity equal to 1–4 steps made the majority. The examples below illustrate some results of calculations.

One step: *bol'shoy*→*gigantskiy* 'big→giant'; *zhalovat'sya*→*stonat'* 'complain→moan'.

Two steps: *bystryy*→*shustryy* 'quick→nimble'; *ugrozhat'*→*zapugivat'* 'threaten→intimidate'.

Three steps: *gluppyy*→*nesposobnyy* 'stupid→unable'; *tsenit'*→*lyubit'* 'appreciate→love'.

Four steps: *populyarnyy*→*scandal'nyy* 'popular→scandalous'; *khvastat'sya*→*gordit'sya* 'boast→be proud'.

The pairs 'entry word→analogue' whose distance was 5–6 steps were subjected to qualitative expert analysis. The raw number of such words was 52 (9,32%) for adjectives and 15 (1,06%) for verbs (see Section 4). We were particularly interested, firstly, why the words with a high degree of semantic similarity (analogues according to NEDS) show semantic distance equal to 5–6 steps in RuWordNet and, secondly, whether it might identify any deficiencies in RuWordNet. The meanings of words were analyzed and checked against to two Russian language thesauri [20, 13] and compared with RuThes concepts in RuWordNet.

The analysis revealed 27 adjectives (51,9%) and 8 verbs (53,3%) the meanings of which should be clarified. We stated three main reasons of semantic distance due to inaccuracies: 1) the exclusion of certain meanings of polysemantic words in RuWordNet (16 adjectives and 5 verbs); 2) the absence of indirect (figurative) meanings in RuWordNet (4 adjectives and 3 verbs); 3) no stylistic marking in RuWordNet (7 adjectives). However, the analysis of other cases (adjectives N=25, verbs N=7) indicated that the semantic similarity shown in RuWordNet corresponds to the explanation given in the Russian language thesauri. For recommendations and examples see Table 2.

Table 2. Recommendations for the improvement of RuWordNet thesaurus

Examples of pairs 'entry word→analogue' of adjectives and verbs in NEDS	Distance between words in RuWordNet (in steps)
1. Recommendation: Add meanings to RuWordNet in accordance with meanings given in Russian language thesauri	
vorchat'→gryzt' 'grumble→nag'	5
ladit'→szhit'sya 'get along→get used to smb'	5
nakazyvat'→sekvestrovat' 'punish→sequester'	5
naprorochit'→sglazit' 'prophecy→jinx'	5
skromnichat'→plakat'sya 'be too modest→whinge'	5
bezlyudnyy→nezhiloy 'deserted →uninhabited'	6
vinovaty→obvinyaemyy 'guilty→accused'	5
gotovy→namerevat'sya 'ready→intend'	5
gromkiy→znamenityy 'pompous→famous'	5
gromkiy→proslavlennyy 'pompous→famed'	5
gromkiy→priznannyy 'pompous→recognized'	5
gromkiy→obshchepriznannyy 'pompous→generally recognized'	5
gromkiy→imenityy 'pompous→eminent'	5
konfliktnyy→zadiristy 'conflict→cocky'	5
malen'kiy→strochnoy 'small→lowercase'	5
malen'kiy→karmanny 'small→pocket'	6
dalekiy→zakholustnyy 'far→ provincial'	5
dalekiy→ periferiyyny 'far→peripheral'	5
dal'novidnyy→prozorlivyy 'far-sighted→penetrating'	5
pustoy→nezhiloy 'empty→ uninhabited'	5
sleduyushiy→nizhesleduyushchiy 'following→following after'	5
2. Recommendation: Add indirect (figurative) meanings to RuWordNet in accordance with meanings given in Russian language thesauri	
dosazhdat'→terebit' 'annoy→pick at'	5
privoryat'sya→perekrasit'sya 'pretend→change colour, repaint'	5
ugadat'→prosech 'guess→catch on, understand'	5
gromkiy→preslovuty 'pompous→well-known'	5
ogromnyy→volchii 'huge→wolfish (appetite)'	5
ogromnyy→l'vinyy 'huge→lion's (share)'	5
otchetlivyy→chekanny 'clear→chased (step)'	5
3. Recommendation: Consider style and register	
gotovy→ne proch 'do not mind'	5
gostepriimnyy→otkrytyy 'hospitable→open'	5
konfliktnyy→zabiyaka 'conflict→bully'	5
konfliktnyy→zadira 'conflict→teazer'	5
ogromnyy→sobachiy 'huge→doggy (cold)'	5
izvestnyy→khvalenny 'famous→vaunted'	5
sovmestnyy→sobornyy 'joint→collective'	5

The qualitative analysis allowed the authors to give specific recommendations concerning particular adjectives and verbs and the ways direct and indirect meanings might be represented in RuWordNet. Some examples are given below.

We singled out a pair of adjectives *gromkiy*→*znamenityy* 'pompous→famous' which distance according to a BFS algorithm is equal to 5 steps. The meanings were verified in the Russian language thesauri. In RuWordNet a RuThes concept

for *gromkiy* is *pompous* in figurative meaning while in Russian thesaurus by Ozhegov there is one more figurative meaning 'widely known, publicised' [20] which is not presented in RuWordNet. The authors recommend to add this meaning which might change the semantic similarity between the adjectives *gromkiy* and *znamenityy* in the described meaning.

We analyzed the pair of verbs *ugadat'*→*prosech* 'guess→catch on, understand' which distance is equal to 5 steps. In NEDS, the verb *prosech* is presented in

an indirect and stylistically marked (slang) meaning as an analogue to the verb *guess* [3]. In RuWordNet, the verb *prosech* is presented only in its direct meaning cut through (RuThes concept: to cut a hole). We suggest to add the figurative meaning of the verb *prosech* ‘catch on, understand’ to the existing RuWordNet synset *dogadat’sya — smeknut’ — soobrazit’* ‘realize — get the clue — grasp’ (RuThes concept: realize) which is hyponymic to the verb *guess* (RuThes concept: guess, (realize by guessing)). In this case, semantically close verbs *guess* and *catch on, understand* in RuWordNet will reveal hyponymic relationships with the distance equal to 1 step. Therefore, we recommend to add the figurative meaning of the verb *prosech* ‘catch on, understand’ to RuWordNet.

Similarly, we analyzed the pair of verbs *dosazhdat’* → *terebit’* ‘annoy → pick at’ which distance is equal to 5 steps. Presently, the verb *terebit’* ‘fumble’ in RuWordNet is given only in its direct meaning (RuThes concept: fumble, pull). We recommend to add a figurative meaning to RuWordNet that might change the semantic distance between these verbs.

6 Conclusions

In this research, the authors analyzed the ways words with similar meanings are presented in RuWordNet thesaurus (hyponyms and hyperonyms) and NEDS (analogues). The quantitative method (a BFS algorithm) was used to measure the semantic similarity between these words and revealed the distance between them in steps. The applied quantitative method indicated that the analogues described in NEDS correlate with the hyponyms and hyperonyms in RuWordNet which contributes to the study of near-synonymy.

The qualitative method allowed the authors to identify deficiencies in RuWordNet. The observation and linguistic analysis of the near-synonyms enabled to point out certain shortcomings and propose changes to RuWordNet to improve the deficiencies. The generalizations concern both direct and indirect (figurative) meanings and style (register). The study revealed that certain meanings of lexical items were missed and should be added to RuWordNet.

Acknowledgements

The reported study was funded by the Russian Foundation for Basic Research (Grant No. 18-00-01238).

References

1. **Adamska-Salaciak, A. (2013).** Equivalence, Synonymy, and Sameness of Meaning in a Bilingual Dictionary. *International Journal of Lexicography*, Vol. 26, No. 3, pp. 329–345. DOI: 10.1093/ijl/ect016.
2. **Apresyan, Y. D. (1995).** *Izbrannyye trudy v 2 tomakh. Tom 1. Leksicheskaya semantika. Cinonimicheskiye sredstva yazyka.* Russian Academy of Sciences, Moscow.
3. **Apresyan, Y. D., editor (2003).** *Novyi Ob’yasnitel’nyy Slovar’ Sinonimov Russkogo Yazyka. 2e izd. Shkola “Iazyky slav’anskoy kultury”, Moscow.*
4. **Bochkarev, V. V., Solovyev, V. D. (2019).** Properties of the network of semantic relations in the Russian language based on the RuWordNet data. *Journal of Physics: Conference Series*, Vol. 1391, pp. 012052. DOI: 10.1088/1742-6596/1391/1/012052.
5. **Cormen, T., Leiserson, C., Rivest, R., Stein, C. (2009).** *Introduction to Algorithms.* 2nd edn. MIT Press, Cambridge, MA.
6. **Cruse, D. (1986).** *Lexical semantics.* Cambridge University Press, Cambridge.
7. **Cruse, D. A. (2004).** *Meaning in Language: An Introduction to Semantics and Pragmatics.* Oxford University Press.
8. **Denisova, E. S., Shumilova, A. A. (2016).** Generation and perception of near-synonyms in a linguistic consciousness of preschoolers and primary school children (according to the experiment). *Siberian Journal of Philology*, Vol. 4, pp. 193–199. DOI: doi:10.17223/18137083/57/17.
9. **Dimarco, C., Hirst, G., Stede, M. (1993).** The semantic and stylistic differentiation of synonyms and near-synonyms. AAAI Technical Report SS-93-02.
10. **Divjak, D. (2010).** Structuring the Lexicon. De Gruyter Mouton. DOI: doi:10.1515/9783110220599.
11. **Divjak, D., Gries, S. T. (2006).** Ways of trying in Russian: clustering behavioral profiles. *Corpus Linguistics and linguistic theory*, Vol. 2, No. 1, pp. 23–60. DOI: doi:10.1515/CLLT.2006.002.
12. **Edmonds, P., Hirst, G. (2002).** Near-synonymy and lexical choice. *Comput. Linguist.*, Vol. 28, No. 2, pp. 105–144. DOI: 10.1162/089120102760173625.
13. **Efremova, T. (2006).** *Sovremennyye tolkovyye slovar’ russkogo yazyka v 3 tomakh.* Russkiy yazyk, Moscow.

14. **Kobozeva, I. M. (2000).** *Lingvisticheskaya semantika: Uchebnoye posobiye.* Editorial URSS, Moscow.
15. **Krongauz, M. A. (2005).** *Semantika: uchebnik dlya studentov lingvisticheskikh fakul'tetov vysshikh uchebnykh zavedeniy.* Akademia, Moscow.
16. **Levitsky, V. V. (2012).** *Semasiologiya.* 2nd edn. Nova Kniga, Vinnitsa.
17. **Loukachevitch, N. (2011).** *Tezaurusy v zadachakh informatsionnogo poiska.* Moskovskiy universitet, Moscow.
18. **Loukachevitch, N., Lashevich, G. (2016).** Multi-word expressions in Russian thesauri RuThes and RuWordNet. *Proceedings of the AINL FRUCT 2016*, pp. 66–71.
19. **Lyons, J. (1968).** *Introduction to Theoretical Linguistics.* Cambridge University Press. DOI: 10.1017/CBO9781139165570.
20. **Ozhegov, S. I. (1992).** *Tolkovyy slovar' russkogo yazyka.* Izdatel'stvo "Az", Moscow.
21. **Resnik, P. (1999).** Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *J. Artif. Int. Res.*, Vol. 11, No. 1, pp. 95–130.
22. **Richardson, R., Smeaton, A. F., Smeaton, A. F., Murphy, J., Murphy, J. (1994).** Using WordNet as a knowledge base for measuring semantic similarity between words. Technical report, *Proceedings of AICS Conference.*
23. **Solovyev, V., Gimaletdinova, G., Khalitova, L., Usmanova, L. (2020).** Expert assessment of synonymic rows in RuWordNet. **van der Aalst, W. M. P., Batagelj, V., Ignatov, D. I., Khachay, M., Kuskova, V., Kutuzov, A., Kuznetsov, S. O., Lomazova, I. A., Loukachevitch, N., Napoli, A., Pardalos, P. M., Pelillo, M., Savchenko, A. V., Tutubalina, E.,** editors, *Analysis of Images, Social Networks and Texts*, Springer International Publishing, Cham, pp. 174–183.
24. **Steyvers, M., Tenenbaum, J. B. (2005).** The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, Vol. 29, No. 1, pp. 41–78. DOI: https://doi.org/10.1207/s15516709cog2901_3.
25. **Storjohann, P. (2009).** Plesionymy: A case of synonymy or contrast? *Journal of Pragmatics*, Vol. 41, pp. 2140–2158.
26. **Sun, K., Huang, Y., Liu, M.-C. (2011).** A wordnet-based near-synonyms and similar-looking word learning system. *J. Educ. Technol. Soc.*, Vol. 14, pp. 121–134.
27. **Wan, S., Angryk, R. (2007).** Measuring semantic similarity using wordnet-based context vectors. *2007 IEEE International Conference on Systems, Man and Cybernetics*, pp. 908–913.
28. **Wang, S., Huang, C.-R. (2017).** Word sketch lexicography: new perspectives on lexicographic studies of Chinese near synonyms. *Lingua Sinica*, Vol. 3, pp. 11. DOI: 10.1186/s40655-017-0025-4.
29. **Wu, Z., Palmer, M. (1994).** Verbs semantics and lexical selection. *Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics, ACL '94, Association for Computational Linguistics, USA*, pp. 133–138. DOI: 10.3115/981732.981751.

*Article received on 25/04/2021; accepted on 14/06/2021.
Corresponding author is Valery Solovyev.*