Russian-Tatar Socio-Political Thesaurus: Methodology, Challenges, the Status of the Project

Alfiya Galieva, Tatarstan Academy of Sciences

amgalieva@gmail.com

Olga Nevzorova, Tatarstan Academy of Sciences & Kazan Federal University

onevzoro@gmail.com

Dilyara Yakubova, Tatarstan Academy of Sciences & Kazan Federal University

suleymanovad@gmail.com

Abstract

This paper discusses the general methodology and important practical aspects of implementing a new bilingual lexical resource – the Russian-Tatar Socio-Political Thesaurus that is being developed on the basis of the Russian RuThes thesaurus format as a hierarchy of concepts viewed as units of thought. Each concept is linked with a set of language expressions (words and collocations) referring to it in texts (text entries). Currently the Russian-Tatar Socio-Political Thesaurus includes 6,000 concepts, while new concepts and text entries are being constantly added to it.

The paper outlines main challenges of translating concept names and their text entries into Tatar, and describes ways of reflecting the specificity of the Tatar lexical-semantic system.

1 Introduction

At the present time, terminology resources for various domains and languages are being actively developed. For low-resource languages like Tatar, terminology compilation may be a challenge in many respects:

- because of the insufficient degree of terminology development that leads to absent relevant designations for special concepts in a language;
- because of lacking special sources for assembling terminology;
- because of lacking coherent terminology management at the terminology development stage as well as at the stage of usage, etc.

Automatic generation of domain-specific ontologies and thesauri may in many cases be a highly challenging task even for languages that possess

special lexicographic resources. For the Tatar language, there is a relatively small amount of such resources, and most of available ones leave a lot to be desired.

The Russian-Tatar Socio-Political Thesaurus is structured in the same way as the RuThes thesaurus for Russian as a network of concepts and their lexical entries with basic semantic relations mapped between them (Loukachevitch and Dobrov, 2002a). The Tatar part of the Thesaurus is mainly being compiled by manually translating Russian terms represented in the Russian RuThes, into Tatar. Also, the Tatar language specific concepts and their lexical entries are added, so each part of the Thesaurus – the Russian and the Tatar one – represents a unique language-internal system of lexicalizations. At the same time, the languages are interconnected so that it is possible to go from the concepts and words in one language to corresponding items in the other.

The project of developing the Russian-Tatar Socio-Political Thesaurus is aimed at compiling the whole body of modern Tatar vocabulary related to the following basic domains: state government, economy, social life, justice, warfare, culture, religion; all of these items have Russian equivalents (correspondences) and are represented at the conceptual and lexical entries levels that are hierarchically arranged. This work is of great scientific and practical importance, since the Thesaurus is aimed at embracing all basic concepts and their lexical entries used in the modern Tatar-language socio-political domain. Socio-political vocabulary designates phenomena of social and political life and reflects basic topics raised in the society as well as its values. Being extremely sensitive to all kinds of innovations, socio-political vocabulary undergoes unceasing changes – new lexical items emerge

while others become outdated relatively quickly. Thus, socio-political lexicons should be compiled on the data of constantly updated text collections.

The remainder of this paper is organized as follows. Section 2 outlines related works, and considerable attention is paid to the existing domain-specific thesauri. Section 3 gives a brief overview of the methodology used in the Thesaurus design and terminology collecting and translation. Section 4 presents main challenges encountered and ways to overcome these difficulties. Section 5 is devoted to one of the most interesting phenomena in Tatar socio-political discourse – a great number of domain-specific synonyms; the main causes of emerging synonymy are disclosed and structural types of synonyms are distinguished. Section 6 sketches out some technical aspects of implementing the new thesaurus and the current state of the project. Finally, Section 7 lists the conclusions and outlines the future work.

2 Related Works

The world of thesauri is not homogeneous. International organizations develop thesauri as means for document indexing and retrieval. Most of them ([UNESCO thesaurus](#), [Eurovoc](#)) are multilingual and obey international standards, such as [ISO 25964-1:2011](#). A different approach is implemented when a thesaurus is created as a lexical resource to be used in applications of computational linguistics and natural language processing. Such is [Princeton WordNet Thesaurus](#), which gave start to many projects on creating national wordnets. Most important dictionary publishers create their own thesauri ([Harper Collins](#), [Oxford](#)), both monolingual and multilingual, while some are being developed with the contribution of volunteers ([Open Thesaurus](#)) and therefore take much more time to complete. Some are developed manually and are more precise – although obviously more expensive than those developed automatically; some have open access, others not.

Another important distinction to be made is among general and domain-specific thesauri. As general thesauri give little coverage of the rich vocabulary of particular subject areas, domain-specific thesauri are required ([Manning et al., 2008](#)). They present an important instrument for extracting semantic information and term sense disambiguation ([Maynard and Ananiadou, 1998](#)).

Not all thesauri easily fall into one or the other group, like [GEMET](#) General Multilingual Environmental Thesaurus that was developed as an indexing, retrieval and control tool for the European Topic Centre on Catalogue of Data Sources and the European Environment Agency, although according to its creators it “was conceived as a “general” thesaurus, aimed to define a common general language, a core of general terminology for the environment” ([About GEMET](#)). Among domain-specific thesauri we can mention [APA Thesaurus of Psychological Index Terms](#) by the Association of Psychologists of America and [AAT Art & Architecture Thesaurus](#) by the Getty Research Institute. Both are aimed at making searching within their domain easier and more successful, but while the first one has no free online version, the second one is an open resource available as Linked Open Data. [IATE Interactive Terminology for Europe Thesaurus](#) has a different objective: it was created mainly for terminological and translation purposes.

As for the socio-political domain, several thesauri can be mentioned. One is the [Gesis Thesaurus for the Social Sciences](#) developed by Leibniz Institute for the Social Sciences for the content-oriented search by keywords. This multilingual resource (English, German, French and Russian) contains about 12,000 keywords and includes all the social science disciplines. The search may be carried out using alphabetic or systematic lists, and the user receives the following information on the term: descriptor, broader term, narrow term, combination and notation. Currently this Thesaurus is available in SKOS-format as Linked Open Data. Another example is the [European Language Social Science Thesaurus](#), which is developed by UK Data Archive University of Essex and intended to be both an information retrieval and a terminological tool. It is currently available in nine languages and includes over 3,000 terms for most languages.

Important thesauri in the socio-political domain were created by RuThes developers. One is Thesaurus on Sociopolitical life that includes more than 62 thousand terms, words and proper names, more than 27 thousand concepts and 102 thousand conceptual relations ([Loukachevitch and Dobrov, 2002](#)). It presents “a hierarchical net of concepts constructed specially as a tool for different applications of automatic text processing. It contains a lot of terms from economic, financial, political, military, social, legislative, cultural and other spheres” ([Loukachevitch and Dobrov, 2002](#)). Another one is

bilingual Russian-English Socio-Political Thesaurus that includes more than 32 thousand concepts, 78 thousand Russian terms and 85 thousand English terms (Ageev et al., 2006). For its creation, the developers combined methods of construction of information retrieval thesauri, the development of wordnets for various languages and ontology research.

3 Methodology of Designing the Thesaurus

Defining the principles for modeling a bilingual resource presupposes establishing the following: 1) the format of the vocabulary entry; 2) the volume and quality of the lexical material; 3) methods for identifying and resolving conflict situations arising in the process of translation (Galieva et al., 2017).

The Russian-Tatar Thesaurus is being developed in the format of the RuThes thesaurus. In the design of the Tatar part of the Thesaurus, the conceptual structure of RuThes is mainly preserved, i.e. the Tatar component is based on the list of concepts of RuThes. RuThes basic units are of conceptual nature rather than of linguistic one – they represent units of thinking that can be associated with several synonymous language expressions. Every concept should have distinctions from the related concepts; concepts are independent of contextual environment and should be expressed as specific sets of associated language expressions – text entries (Loukachevitch and Dobrov, 2002a).

The conceptual structure of RuThes determines the direction (sequence) of developing concepts in the Tatar part, and the basic structure of RuThes conceptual relations is preserved here as well. In cases when semantic relations between Tatar concepts do not correspond to those between the units of RuThes, these relations are established anew or revised.

The general methodology of creating the Tatar part of the Thesaurus includes the following steps.

1. Search for equivalents (corresponding terms) that are actually used in Tatar as translations of Russian terms.
2. Adding new concepts representing topics that are important for the socio-cultural life of the Tatar society but are not represented or represented insufficiently in the original RuThes (the list of required vocabulary is

compiled, concept names and lexical entries are distinguished and arranged according to the RuThes structure).

3. Revising the relations between concepts considering the place of new concepts in the hierarchy of the existing ones, and if necessary, adding new concepts of intermediate level.

During the first stage, RuThes concept names and text entries are translated into Tatar. The sources of translation include bilingual general-purpose dictionaries, special socio-political dictionaries, Tatar texts on socio-political topics, and available Russian-Tatar parallel texts.

The main requirements to Tatar concept names are as follows:

- clarity of meaning;
- use of neutral and common lexical units for concept names;
- brevity of denomination;
- selecting the most precise, the most frequent and officially used denomination from several options if there are any.

The requirement of the concept name clarity supposes that a concept must be intuitively understood from its designation.

Due to the discrepancy between grammatical, semantic and lexical systems of the Russian and Tatar languages, the structure and length of concept names may significantly differ. In particular, because of lacking one-word terms in Tatar, in some cases we have to give their explanatory translations.

When designating concepts, ambiguous words may be used in individual sense, so we provide necessary comments in brackets to identify the meaning of the concept. Because of the differences between the word semantic structures of the languages, those comments and specifications often differ in the Russian and Tatar parts of the Thesaurus. Table 1 shows some peculiarities of use of Russian and Tatar polysemous words as concept names.

An important task is to represent the language specific features of Tatar in the Thesaurus both on contextual and lexical levels.

When compiling the list of Tatar terms, we use data of available corpora built for the Tatar language by now, each in open access:

Example of concept name in Russian	Example of concept name in Tatar
СУД (ГОСУДАРСТВЕННЫЙ ОРГАН) 'court of law (state institution)'	СУД 'court of law'
РУКОВОДСТВО (УЧЕБ- НОЕ ПОСОБИЕ) 'guidance, manual'	КУЛЛАНМАЛЫК 'guidance, manual'
СТАДО 'herd'	КӨТҮ (МАЛ КӨТҮЕ) 'herd (flock of cattle)'
СТОИМОСТЬ 'cost'	КЫЙММӨТ (ЧЫГЫМНАР КҮЛӘМЕ) 'cost (an amount of expenses)'

Table 1: Examples of ambiguous concept names: use of specifying comments.

1. The [Corpus of Written Tatar](#) (CWT) compiled by researchers of Kazan Federal University, Russia;
2. “[Tugan Tel](#)” [Tatar National Corpus](#) (TNC) developed at the Institute of Applied Semiotics of Tatarstan Academy of Sciences.

These Tatar corpora are constantly filled with new texts, and the bulk of socio-political texts in them is recent (unlike those in fiction texts), so we may argue that the corpora represent the actual usage of Tatar.

4 Challenges

The main challenges of the current project concern acquiring lexical data and representing all Tatar terms in actual use.

1. Search for equivalents (corresponding terms) that are actually used in Tatar as translations of Russian terms.

Search for translation equivalents and adequate correspondences practically used in texts in many cases becomes a laborious and time-consuming task because available bilingual Russian-Tatar dictionaries of the general lexicon (for example, [Ganiev, 1997a](#)) and special lexicons ([Amirov, 1996](#); [Ganiev, 1997b](#)) are outdated, and new lexicons (for example, [Tagirova and Timerkhanov, 2014](#)) do not contain the required items.

The obsolescence of dictionaries becomes apparent in two main aspects:

- 1) the dictionaries are incomplete and do not contain entries reflecting present-day socio-political

items widely used in Russia (such as *political correctness*, *lines of communication*, *framework agreement*, *taxable* and many others);

- 2) the dictionaries do not contain target language (Tatar) translation units in actual use, containing instead those that had been actively used in the Soviet era.

Consequently, available bilingual Tatar-Russian dictionaries, including special socio-political dictionaries, may be used in developing the Russian-Tatar Thesaurus with a certain proviso, because of the containing obsolete lexical data and lacking lexical items of current interest in many cases.

This circumstance significantly slows down the translation process and compels us to look through a huge amount of media texts and texts of official documents manually in search for appropriate Tatar terms.

2. Adding new concepts representing topics that are important for the socio-cultural life of the Tatar society but are not represented or represented insufficiently in the original RuThes.

We add single concepts and new branches of concepts related to issues of political, religious and cultural life of the Tatar society and reflecting peculiarities of the history of the Tatar people (for example, Islam-related concepts, concepts related to the socio-political structure of Oriental societies, Tatar ethnographic phenomena, etc.).

Table 2 represents this part of the work displaying the new EID AL-FITR concept designating an important religious holiday celebrated by Muslims worldwide; Eid Al-Fitr marks the end of Ramadan — the month of Islamic fasting.

Adding single new concepts often leads to revising relations between the existing ones and adding new intermediate concepts.

3. Revising relations between the concepts while considering the place of new concepts in the hierarchy of the existing ones.

This can be illustrated by an example. Inserting the new 'EID AL-FITR' concept required some changes in the RuThes concepts hierarchy; we entered the following concepts of the higher level:

- OFFICIAL HOLIDAY OF TATARSTAN REPUBLIC;
- MUSLIM HOLIDAY.

Besides, the existing CALENDAR DATE concept was supplemented by those of lower level:

- HIJRI CALENDAR DATE;
- GREGORIAN CALENDAR DATE.

Concept	Relationship with other concepts		Text entries of the concept
	Concepts above	Associated concepts	
УРАЗА ГАЕТЕ 'Eid Al-Fitr'	Above: ДИНИ БӘЙРӘМ 'religious holiday'	Association 1: УРАЗА 'Muslim fasting'	ураза гаете, ураза бәйрәме, рамазан гаете, рамазан бәйрәме
	Above: ТАТАРСТАН РЕСПУБЛИКАСЫНЫҢ РӘСМИ БӘЙРӘМЕ 'official holiday of the Republic of Tatarstan'	Association 2: РАМАЗАН 'Ramadan'	
	Above: МӨСЕЛМАН ДИНЕ БӘЙРӘМЕ 'Muslim Holiday'	Association 3: УРАЗА ФИТЫРЫ 'Zakat al-Fitr'	
	Above: ҺИЖРИ КАЛЕНДАРЬ ДАТАСЫ 'Hijri calendar date'		
	Above: КАЛЕНДАРЬ ДАТАСЫ 'calendar date'		

Table 2: The new EID AL-FITR concept and its relationships with other concepts.

Adding the last two concepts seemed necessary due to principal dissimilarities between the European solar (Gregorian) calendar and the Islamic lunar calendar. Due to the fact that a year of the Islamic calendar is 10 or 11 days shorter than the solar year, dates of Muslim holidays shift forward every year as related to the European calendar. So, linking the concepts denoting Muslim holidays to the common CALENDAR DATE concept is insufficient, and therefore we are to mark-up the calendar type.

5 Synonymy in Tatar Socio-Political Vocabulary

Another challenge is related to the abundance of specific synonyms in the Tatar socio-political domain. This issue is quite interesting and deserves special attention.

Synonymy is one of fundamental concepts in linguistics that manifests itself at different language levels. A prevailing point of view in linguistics is that synonymy shows a high degree of language development, and that absolute (total) synonymy occurs infrequently because “yet even if near-synonyms do name one and the same thing, they name it in different ways: they present different perspectives on a situation (Divjak, 2010).

Available attitudes range from totally rejecting the idea of absolute (total) synonymy in language

to claiming extreme rareness of them (Ulmann, 1967; Lyons, 1995). The situation with Tatar socio-political vocabulary in many respects refutes this thesis about rareness of absolute synonymy.

A great number of emerging total synonyms in Tatar occurred under the influence of a combination of intralinguistic and extralinguistic factors. The Tatar culture is located at the intersection of the Occidental and Oriental civilisations, which leads to active lexical borrowing both from the Arab-Muslim cultural area and the European cultural area; vocabulary borrowing from European languages is taking place through mediation of the Russian language, and, certainly, a huge amount of words and constructions are taken from Russian. Besides, a significant part of synonyms are built on the basis of Turkic and Tatar lexical material.

Therefore, words of different origin (Turkic and Tatar, Arabic and Persian, Greek, Latin, English and Russian) related to the same referent coexist in Tatar, which engenders a great number of synonyms at the single word level.

Table 3 illustrates redistribution of absolute synonyms of European and Arabic origin in corpus collections (for nouns the number of Nominative case forms, and not the number of lemmas, is given).

It is noteworthy that all words of Arabic origin presented in Table 3 entered Tatar active lexicon

Lexeme	English translation, PoS	Origin of the word	Number in the CWT	Number in the TNC
экономик	Economic (adj.)	Greek	621	676
икътисади		Arabic	27,351	22,274
политик	Political (adj.)	Greek	1,005	1,536
сәяси		Arabic	25,011	24,489
республика	Republic (noun)	Latin	316,667	258,433
жөмһүрият		Arabic	2,479	1,631
информацион	Informational (adj.)	Latin	858	807
мәгълүмати		Arabic	7,076	689
суд	Court of law (noun)	Russian	15,725	14,243
мәхкәмә		Arabic	8,718	8,101

Table 3: Distribution of synonyms coming from different languages (according to corpus data)

relatively recently, in the late 80's; nevertheless, they succeeded to push out the words of European origin with the same meaning that had been actively used in the language of the Soviet era.

Often words of Tatar (Turkic) origin formed of different stems by means of Tatar derivational affixes may compete among themselves. Table 4 represents quantitative distribution of synonymous words referred to the BUSINESSMAN concept, in corpus collections, and the number of their most frequent derivatives.

Lexeme	English translation	Number in the CWT	Number in the TNC
эшкуар	businessman	8,606	11,280
эшмәкәр		22,671	13,876
кәсәбәче		122	101
эшкуарлык	entrepreneurship (business activity)	2,641	3,852
эшмәкәрлек		6,454	4,207
кәсәбәчелек		67	54

Table 4: Tatar origin synonyms meaning 'businessman' and 'entrepreneurship (business activity)'

Currently most Tatar socio-political terms are formed by calquing (component-by-component translating) corresponding Russian terms. Diverse groups of specialists with dissimilar world-views and ideological guidelines develop Tatar terminology; in designing terminology they may principally be oriented at different cultural spaces dominated by different languages:

- Tatar and Turkic vocabulary;
- Arabic and Persian vocabulary;
- international Greek, Latin or English vocabulary;
- Russian vocabulary, etc.

Dissimilar preferences of terminology creators lead to the use of different designations for the same entity.

Some peculiarities of lexical, derivational and grammatical systems of the Tatar language also lead to originating of a great number of synonyms. On the level of multiword terms and phrases, lexical synonymy is complicated by the factor of differing grammatical structures in use.

For example, in Turkic languages the following grammatical patterns of noun phrases are regularly corresponding (abbreviations: N – noun, ADJ – adjective, PL – plural affix, POSS_3 – possessive affix, 3d person, NMZL – nominalizer):

ADJ + N and *N + N*, *POSS_3*:

икътисадый кризис (ADJ + N), *икътисадо кризисы* (N + N, POSS_3)
'economic crisis'.

Such regular correspondences multiply the number of grammatical variants of multiword terms.

For example, we found in the media and official texts 9 noun phrases to designate the term *joint-stock company*; these are formed by 3 core nouns meaning '*company*' and are based on different grammatical patterns, see Table 5. All such synonyms are regarded as text entries of the same concept.

Since the linguistic situation in the Republic of Tatarstan is unstable, parallel denominations can be used for a wide range of phenomena, including some official names of departments and state structures. So, a separate task is to compile the most complete list of synonymous denominations related to the same concept in order to fix them as lexical entries for the Thesaurus, because lists of synonyms in available dictionaries are incomplete and in many cases limited to one word.

Gram-matical model of the term	Term variants	Frequency of use in CWT
N,NMZL + N.POSS_3	акционерлык жәмгыяте, акционерлык ширкәте, акционерлык оешмасы	4,320 13 13
N, PL + N.POSS_3	акционерлар жәмгыяте, акционерлар ширкәте, акционерлар оешмасы	274 2 1
ADJ + N	акционер жәмгыять, акционер ширкәт, акционер оешма	97 4 14

Table 5: Texts entries of the concept JOINT-STOCK COMPANY

6 Status of the Project

Currently the Russian-Tatar Socio-Political Thesaurus contains 6,000 concepts and constantly grows. For the Thesaurus, special software and web interface have been designed.

The Thesaurus system is implemented in the format of a web application as a special site <http://tattez.antat.ru/> and contains open and closed parts. The closed part is available to authorized users who have the authority to edit the Thesaurus; basic functions of adding new concepts and lexical entries, mapping relations between concepts, and referring to the data of Russian-Tatar dictionary are provided. An authorized user can also compare the conceptual structures of the Russian and Tatar parts of the Thesaurus, using a special markup reflecting the statuses of completeness and problems in description, and can also use convenient visualization tools with filtering by various parameters.

The main user functions in the open part of the system provide viewing the data, search and navigation through the concepts and text entries of the Thesaurus.

7 Conclusion

The Russian-Tatar Socio-Political Thesaurus is built as a hierarchical network of concepts and lexical entries from the most general to the most specific conceptual categories.

On the grounds of our research of Tatar socio-political terminology within the Thesaurus building project we can conclude that a certain Tatar terminology management is required. It may be descriptive (describing how terms are used in documents and media texts) and prescriptive (or even

normative) to document compilers (prescribing what terms must be used in standard work and official documentation and how they must be used). Therefore, developing the bilingual Russian-Tatar Thesaurus that would cover all existing variants of terms might be an important step to achieving this, on condition of providing users with an effective access to information.

Coexistence of different centres of building Tatar terms and lack of coordination between them leads to instability and imbalance of terminology when diverse synonymous terms and terminological variants of different origin and structure may be in actual use. Therefore, an important task is to search for all synonymous denominations and variants to fix them all as text entries of corresponding concepts. Further extension of the Thesaurus requires development of methods of extracting domain-specific terms from texts and corpus data.

Acknowledgements

The reported study was funded by Russian Science Foundation, research project № 16-18-02074.

References

- AAT Art & Architecture Thesaurus. <http://www.getty.edu/research/tools/vocabularies/aat/>
- About GEMET - GEneral Multilingual Environmental Thesaurus. <https://www.eionet.europa.eu/gemet/about?langcode=en>
- Alfiya M. Galieva, Alik V. Kirillovich, Natalia V. Loukachevitch, Olga A. Nevzorova, O.A., and Dzhavdet Sh. Suleymanov. 2017. *Sozdanie russko-tatarskogo tezaurusa po obshchestvenno-politicheskoy tematike: obshchie principy i aspekty realizatsii* "Development of Russian-Tatar Socio-Political Thesaurus: General Principles and Aspects of implementation". In *Nauchno-tekhnicheskaya informatsiya. Seriya 2: Informatsionnye processy i sistemy*, 2:20-28.
- APA Thesaurus of Psychological Index Terms. <http://www.apa.org/pubs/databases/training/thesaurus.aspx>
- Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to information retrieval*. Cambridge University Press
- Collins English Thesaurus. <https://www.collinsdictionary.com/dictionary/english-thesaurus>
- Corpus of Written Tatar. <http://search.corpus.tatar/en>

- Dagmar Divjak. 2010. *Structuring the lexicon: A clustered model for near-synonymy*. Vol. 43. Walter de Gruyter.
- Diana Maynard and Sophia Ananiadou. 1998. *Term Sense Disambiguation Using a Domain-Specific Thesaurus*. In Proceedings of 1st International Conference on Language Resources and Evaluation (LREC), Granada, Spain.
- ELSST European Language Social Science Thesaurus. <https://elsst.ukdataservice.ac.uk/>
- Eurovoc. The EU's Multilingual Thesaurus. <http://eurovoc.europa.eu/>
- F.I. Tagirova, A.A. Timerkhanov (ed.). 2014. *Russko-tatarskiy slovar aktual'noy leksiki* "Russian-Tatar dictionary of actual vocabulary". Kazan.
- Fuat A. Ganiev. (ed.). 1997a. *Russko-tatarskiy slovar* "Russian-Tatar Dictionary. Moscow", Insan.
- Fuat A. Ganiev. (ed.). 1997b. *Russko-tatarskiy obshchestvenno-politicheskiy slovar* "Russian-Tatar Socio-Political Dictionary". Kazan.
- Fuat A. Ganiev. (ed.). 2005. *Tatarskiy tolkoviy slovar* "Tatar Explanatory Dictionary". Kazan.
- GEMET Thesaurus. <https://www.eionet.europa.eu/gemet>
- Gesis Thesaurus for the Social Sciences. <http://www.gesis.org/en/services/research/tools/social-science-thesaurus/>
- IATE Interactive Terminology for Europe Thesaurus. <http://iate.europa.eu/SearchByQueryLoad.do;jsessionid=VFzZw0dUdNLutE-TAg3EV5RA7t1P7QpalbDS1MJzgdU3hyk-bPp9P7!559570392?method=load>
- International Organization for Standardization. Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval. <https://www.iso.org/standard/53657.html>
- John Lyons. 1995. *Linguistic Semantics*. Cambridge University Press.
- Kafil F. Amirov. (ed.). 1996. *Russko-tatarskiy yuridicheskiy slovar* "Russian-Tatar legal dictionary". Kazan.
- M. Lynne Murphy. 2003. *Semantic Relations and the Lexicon*. Cambridge University Press.
- Mikhail Ageev, Boris Dobrov, and Natalia Loukachevitch. 2006. *Socio-Political Thesaurus in Concept-Based Information Retrieval*. In: Peters C. et al. (eds) Accessing Multilingual Information Repositories. CLEF 2005. Lecture Notes in Computer Science, vol. 4022. Springer, Berlin, Heidelberg.
- Natalia Loukachevitch and Boris Dobrov, 2002a. *Development and Use of Thesaurus of Russian Language RuThes*. In Proceedings of workshop on WordNet Structures and Standardisation, and How These Affect WordNet Applications and Evaluation. (LREC 2002) / Dimitris N. Christodoulakis, Gran Canaria, Spain: 65-70.
- Natalia Loukachevitch and Boris Dobrov, 2002b. *Evaluation of Thesaurus on Sociopolitical Life as Information-Retrieval Tool*. In M.Gonzalez Rodriguez, C. Paz Suarez Araujo (Eds.) The Third International conference on Linguistic Resources and Evaluation (LREC-2002). – Vol.1 – 2002, Gran Canaria, Spain.
- Open Thesaurus. <http://openoffice-es.sourceforge.net/thesaurus/>
- Oxford Thesaurus. <https://en.oxforddictionaries.com/>
- Roy Harris. 1973. *Synonymy and Linguistic Analysis*. Basil Blackwell.
- RuThes Linguistic Ontology. http://www.labinform.ru/pub/ruthes/index_eng.htm
- Stephen Ullmann. 1967: *Semantics, An Introduction to the Science of Meaning*. Basil Blackwell. Oxford
- Thesaurus for the Social Sciences. <http://www.gesis.org/en/services/research/tools/social-science-thesaurus/>
- "Tugan Tel" Tatar National Corpus. <http://corpus.antat.ru/>
- UNESCO Thesaurus. <http://vocabularies.unesco.org/browser/thesaurus/en/>
- Wordnet. A lexical database for English. <https://wordnet.princeton.edu/>