

# Creating an English-Uzbek Bilingual Thesaurus of Frequently Used Adjectives in Uzbek Corpus

Nilufar Abdurakhmonova

Head of Department of Computational and applied linguistics  
National University of Uzbekistan  
Tashkent, Uzbekistan  
Nilufarabdurakhmonova4@gmail.com

Nargiza Shamieva

Department of Foreign language and Literature  
National University of Uzbekistan,  
Tashkent, Uzbekistan  
shamievanargiza5@gmail.com

**Abstract**—Linguistic materials must be updated and expanded continuously due to the dynamic nature of language evolution. In order to fill the void in current lexical resources, this project attempts to produce a bilingual thesaurus of new and emerging adjectives from English to Uzbek. This research will create a thorough thesaurus to aid in translation, language acquisition, and cross-cultural communication by looking at existing usage patterns and contextual applications. This project taking into account how social media and technology have impacted new adjectives as well the global culture shift. The thesaurus will be accompanied by text data from a range of sources — online platforms, modern literature and spoken discourse for example — to reflect actual usage in context. In addition, this research will cover regional aspects of English and Uzbek so thesaurus will consider dialectal variations or colloquial language. The end product will act as an invaluable resource not only for linguists and translators, but also teachers in adjusting their teaching materials to reflect linguistic changes.

**Keywords**—*adjectives, semantic types, structure, corpus analysis, manifestation, thesaurus*

## I. INTRODUCTION

The state of the language is refreshed by new words and idioms every moment. Among others, adjectives are needed for a more expressive vocabulary and to bring subtle thoughts across. There, considering the importance of adjectives, there is an immediate call for a bilingual thesaurus from English to Uzbek with a view to novel and emerging adjectives for linguists, translators, teachers, and students. The paper demonstrates techniques and obstacles within the development of such a thesaurus, specifically for a useful resource that enhances multilingual understanding and proficiency. This thesaurus is intended to bridge an important gap by supplying adjectives that reflect contemporary and changing ideas, allowing for a more detailed, accurate communication between English and Uzbek. In this age of globalization and technology, languages are always being embellished with new descriptive words that reflect the changes and innovations and shifts in social attitudes and perceptions. For translators, a good, current adjective base is vital to the preservation of meaning and tone when passing from one language to another. Speaking of educational, a bilingual thesaurus with contemporary adjectives will provide teachers and students alike with a more useful tool with which to interact with the language of today, in both tongues. Also, the thesaurus facilitates greater cultural exchange, so that the user can learn not just the literal translations, but the cultural

implications behind words and when to use them. The challenges in developing such a resource lie in the identification and translation of adjectives that may not have direct equivalents in both languages, requiring creative linguistic solutions. In order to surmount these hurdles, the procedure entails working with native speakers, linguists, and subject matter gurus to certify the precision and applicability of the entries. This project is part of a greater movement to keep language vital in a world that is growing smaller and smaller, where languages must adapt as the world changes. Also, the thesaurus is a great teaching tool that could enhance language learning through the provision of adjectives that are currently in use for the learner. As this type of thing continues to expand, it will need frequent updates to keep up with the innovative use of language, that way it will always be a useful and necessary tool for those speaking both languages.

## II. LITERATURE REVIEW

The language develops by accepting new terms and idioms that are added to it once and for all. Especially, adjectives have a big role in enriching one's descriptive language and leading toward complex thoughts. Thus, the urgent problem for linguists, translators, teachers, and students is to create a bilingual thesaurus with a focus on new and developing adjectives from English into Uzbek. It surveys methods and problems concerned with constructing such a thesaurus, which will help multilingual understanding and competence. [1].

In fact, adjectives form the definite requirement of any language to express specificity and make it expressive. They embellish the language by giving specificity to and changing nouns. Adjectives must be translated appropriately in multilingual contexts. [2].

Unique difficulties in bilingual lexicography include contextual variances, colloquial idioms, and cultural nuances. To guarantee accurate and meaningful translations, these criteria must be carefully taken into account while creating a thesaurus [3].

Unlike previous research, which found at least two ways to evaluate the usefulness of a thesaurus. One is to apply it to practical tasks, such as information retrieval and sentence understanding. Another is to construct hierarchies from real data, e.g., a huge corpus, and compare them to the hierarchies in a hand-coded thesaurus [4].

This is consistent with the findings in Uzbek corpus creating in the field which is accommodating in machine translation [5].

Speaking of the study that is focused on creating a bilingual thesaurus of new and emerging adjectives from English to Uzbek, D. Sharipova's research *Semantic Changes in Uzbek Adjectives (2015)* seems to lay a solid ground for our work and also for the understanding of how the adjectives in Uzbek language are changing. Sharipova emphasizes the importance of continuously updating linguistic materials to reflect language evolution, aligning with our project's goal of addressing the gap in current lexical resources. Her analysis of semantic shifts in Uzbek adjectives highlights the dynamic ways in which adjectives adapt to cultural and societal changes, which is crucial for our thesaurus. Our project builds upon Sharipova's work by not only looking at the semantic shifts in Uzbek, but also the cross-linguistic borrowing of new English adjectives that have arisen due to globalization, social networking, and technological progress. By including this work on semantic development, our bilingual thesaurus will not only show the linguistic changes in Uzbek, but how these changes interact with world linguistic trends, bettering translation, language learning, and cross-cultural communication [6].

For our research on building a bilingual thesaurus of new and developing adjectives from English to Uzbek, Rahmatullaev's work, the dynamics of adjective meanings in Uzbek (2018) is very pertinent. Rahmatullaev discusses the semantic shifts of adjectives in the Uzbek language and the necessity for linguistic tools to be able to track these shifts. His theory of meaning construction is very much related to the goal of your project, which is to collect not only the conventional adjectives, but also the unique and rising adjectives under the impact of the world. We find Rahmatullaev's work on the development of adjectives to be very useful theoretically to our research especially when we discuss how social media, technology, and global culture has contributed to the creation and use of new adjectives. By building on his analysis of meaning shifts, our thesaurus will offer a contemporary and practical resource that goes beyond the historical evolution of adjectives to include modern linguistic phenomena. This tie-in makes our lit review much stronger because we've established our research as a continuation of Rahmatullaev's work but with an emphasis on cross-linguistic applicability and the technological aspect of L1 acquisition [7].

N.Z. Abduraxmonova's dissertation, *Linguistic Support of the Program for Translating English Texts into Uzbek (2018)*, offers valuable insights into the development of linguistic resources for translation between English and Uzbek, making it directly relevant to our project on creating a bilingual thesaurus of new and emerging adjectives. Abduraxmonova's work deals with that linguistic support needed for translation, especially with simple sentences, and stresses the need for linguistic materials to be constantly updated for the continuing changes in language, which is something our project is all about, capturing those adjectives that are so dynamic and constantly evolving. Abduraxmonova's work provides a practical foundation for our project, as it demonstrates the need for robust linguistic tools that facilitate accurate translation between the two languages. While her focus is on simple sentences, our project extends this by addressing the nuanced translation of adjectives, especially those shaped by modern influences like social media, technology, and global culture shifts. That

way, we can tie our research in with Abduraxmonova's, and show how both works add to the overall development of linguistic tools for translation and intercultural communication, while our research narrows the scope to adjectives, but in a larger, more modern lexical field. This connection not only supports our literature review, but places our project as a supplementary and indispensable arm of current translation aid [8].

Development of Russian-Turkic Parallel Corpora and MT Systems" (2020) is very relevant to our Uzbek-English Bilingual Thesaurus of New Emerging Adjectives" project. It's called the TurkLang-7 and it's going to be doing some work with creating parallel corpora and machine translation (Machine Translation) systems for the Russian-Turkic languages-just goes to show that collecting linguistic data can be very useful for communication between languages. That would tie in nicely with our plans to build a bilingual thesaurus, since both projects are aimed at improving linguistic resources to facilitate translation and language learning between any language pair. The work is really built upon the authors' use of parallel corpora and their emphasis on constructing large linguistic databases that accurately represent current trends of usage. Similar to how the TurkLang-7 project bridges linguistic gaps between Russian and Turkic languages, your thesaurus aims to bridge the gap between English and Uzbek, particularly in the area of novel and emerging adjectives influenced by global trends such as social media and technology. Additionally, their focus on machine translation highlights the growing relevance of computational tools in linguistic projects, suggesting that our thesaurus could potentially benefit from integration with such technologies in the future. Incorporating Kubedinova's findings into our literature review situates our research within the broader context of efforts to improve multilingual resources, particularly in Turkic language contexts. Our project extends their work by focusing on adjectives and offering a more specialized resource for translation, language learning, and cultural exchange between English and Uzbek speakers [9].

### III. METHODOLOGY

Gathering appropriate mostly used adjectives from Uzbek corpus on recently created and popular English adjectives is the first stage in the thesaurus creation process. The corpus has been collected from a range of recent texts, such as:

- Posts on social media;
- literary works;
- new books and publications;
- online news items;
- proverb collection.

On the basis of the Uzbek corpus, we can analyze the number and combinations of adjectives Fig. 1.

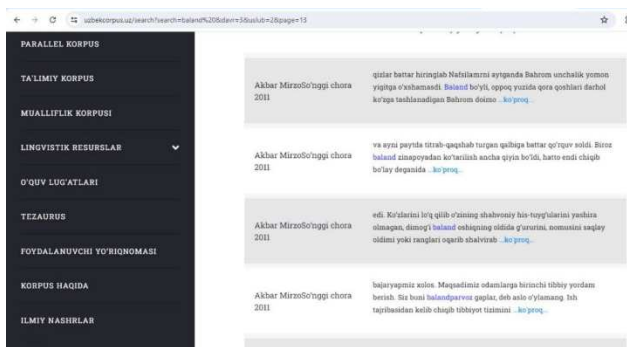


Fig. 1. An abstract from uzbekcorpus.uz where the samples of the adjective BALAND were taken in the context of various Uzbek texts.

To compile a large dataset of adjectives, linguistic corpora and automated web scraping methods will be used.

With the help of the corpus instrument according to the Ingram model, the words that are coming before or after the necessary adjective from concordance have been selected. A corpus analysis of the gathered data will be performed to find often occurring and contextually relevant adjectives. For frequency analysis, collocation extraction, and contextual analysis, programs like AntConc and Sketch Engine will be utilized [10]. The next stage is to translate the adjectives into Uzbek after they have been identified. Proficiency in language is necessary, but this procedure also calls for cultural adaptation to guarantee that the translations make sense in the Uzbek context. To improve the translations, native speakers and skilled translators will work together.

Compiling the thesaurus is the last step. Semantic categories will be used to categorize the adjectives, and each entry will contain the following:

- The English adjective;
- Its Uzbek equivalents;
- Model phrases in both languages;
- Cultural quirks and contextual notes.

The thesaurus will be hosted on a digital platform that will enable easy access and frequent updates.

#### IV. DISCUSSION AND RESULTS

During the course of the research, several difficulties were encountered:

• *Cultural Nuances:* A few English adjectives have no exact Uzbek equivalents, necessitating imaginative and sensitive translations.

• *Contextual Variations:* The meaning of adjectives might vary depending on the situation, thus it's important to provide thorough example sentences to demonstrate their use.

• *Data Volume:* Significant computational resources and experience were needed to manage and analyze enormous datasets. Below are examples of such adjectives, including their Uzbek equivalents, example sentences, and contextual notes, which are taken from REVERSO Context (English version) and the uzbekcorpus.uz, which contain a huge amount of Uzbek linguistic sources, for the Uzbek sentences Table I.

TABLE I. Example Entries for a Bilingual Thesaurus:

BALAND\HIGH (given adjective)			
New formed synonym	Collocating noun	Collocating noun	New formed synonym
Uzbek		English	
<i>Baland</i>	<i>musiqatovush</i>	<i>loud voice</i>	<i>booming</i>
<i>Yuqori</i>	<i>harorat</i>	<i>high temperature</i>	<i>elevated</i>
<i>Novcha</i>	<i>yigit</i>	<i>Tall guy</i>	<i>towering</i>
<i>Osmono'par</i>	<i>bino</i>	<i>High building</i>	<i>High-rise</i>
<i>Uzun</i>	<i>terak</i>	<i>Towering poplars</i>	<i>soaring</i>
<i>muvaffaqiyatli</i>	<i>natijako'rsatkich</i>	<i>a great balance</i>	<i>successful</i>
<i>Qimmatbaho</i>	<i>narh</i>	<i>The rising cost</i>	<i>precious</i>
<i>Tez</i>	<i>tezlik</i>	<i>maximum recorded speed</i>	<i>Top</i>
<i>Mashhur, hurmatli</i>	<i>nom</i>	<i>well-known name</i>	<i>famous</i>
<i>erishib bo'lmaydigan</i>	<i>dor</i>	<i>an unattainable dream</i>	<i>impossible</i>
<i>Qattiq, kuchli</i>	<i>ovoz</i>	<i>loud voice</i>	<i>big</i>
<i>izzatli</i>	<i>obro', hurmat</i>	<i>a respected position</i>	<i>valued</i>
<i>Jasur</i>	<i>qad, bosh</i>	<i>a brave man</i>	<i>Courageous</i>
<i>o'ziga ishongan</i>	<i>shasht</i>	<i>fearless</i>	<i>Intrepid</i>
<i>mag'rur</i>	<i>dimog'</i>	<i>proud</i>	<i>arrogant</i>

<i>ko'tarilgan</i>	<i>harorat</i>	<i>High temperature</i>	<i>hot</i>
--------------------	----------------	-------------------------	------------

Synonymous Uzbek-English adjective collocation

**English Example Sentence:** His loud voice drew my attention [11].

**Uzbek Example Sentence:** Guzar tomondan baland muzika ovozi yangraydi. O'.Hoshimov. Dehqonning bir kuni [12].

**English Example Sentence:** In few cases, there is also a high temperature [11].

**Uzbek Example Sentence:** U joyda harorat ortadi, chunki yuqori haroratdabegona jon egalaritezroq halok bo'ladi. I.Sulton. Genetik [12].

**English Example Sentence:** That tall guy on the first floor [11]?

**Uzbek Example Sentence:** Yigitning ko'ziga uzoqdan novcha bo'yli bir kishi ko'rindi [12].

**English Example Sentence:** This is another tall building in the city of Mumbai [11].

**Uzbek Example Sentence:** 2012-yil 30-martda bino 310 metr balandlikka ko'tarildi va hozirda u Yevropadagi oltinchi eng baland osmono'par bino hisoblanadi [12].

**English Example Sentence:** The lanes lined with apple orchards and towering silver poplars are what make the city famous [11].

**Uzbek Example Sentence:** Yo'lning ikki cheti baland teraklar bilan qoplangan, atrof manzarasi esa chiroyli deyishga arzigulik edi [12].

**English Example Sentence:** We think that's a great balance [11].

**Uzbek Example Sentence:** Prezident yordamchisi Saida Mirziyoyeva "Muvaffaqiyatli islohotlar – yuqori natijalar" xalqaro anjumanida ishtirok etdi [12].

**English Example Sentence:** The rising cost of petrol is the cause [11].

**Uzbek Example Sentence:** Ya'ni 1-iyundan e'tiboran yakuniga yetishi kerak bo'lgan eng baland chakana narx uchun belgilangan muddat yana ikki oy davomida ayrim o'...[12].

**English Example Sentence:** Seventy-two km/h is the maximum recorded speed [11].

**Uzbek Example Sentence:** O'zbekiston Respublikasining o'ziga xos tabiiy-iqlim va ekspluatatsiya sharoitida yuqori tezlikda harakatlanishni tashkil etishga ta'sir etuvchi omillar [12].

**English Example Sentence:** Probably the most well-known name on this list [11].

**Uzbek Example Sentence:** ... boshga Ota o'g'li Ota nomini baland Rostini aytsang, [Tora Mirzayev, Asqar Musoqulov, Bahodir Sarimsoqov «o'zbek xalq maqollari» 2005'] [12].

**English Example Sentence:** 'Atlantis' is a lost continent, an unattainable dream, a place and time beyond the laws and concepts, an unforgettable childhood [11].

**Uzbek Example Sentence:** Osilsang, baland dorga Yiqilsang, maydonga, [Tora Mirzayev, Asqar Musoqulov, Bahodir Sarimsoqov «O'zbek xalq maqollari» 2005'.] [12].

**English Example Sentence:** His loud voice drew my attention [11].

**Uzbek Example Sentence:** Endi u dunyoda huru g'ilmonlar bilan ayshini surib yurgandir. Jalilning baland ovozda aytgan bu gaplari Asadbekka quvvat berganday sergaklantirdi.Tohir Malik Shaytanat 1-kitob1989 [12].

**English Example Sentence:** Joachim Descartes intended René to take up a respected position in the legal profession, just as his elder brother had done [11].

**Uzbek Example Sentence:** ...bo'lib qolgandir», deb o'ylab, hujjatlarni boshliqqa uzatdi: — Kollektivda obro'si baland ekan-u, kollektiv bu baobro' odamning o'limidan bexabar qolibdi-da, a [12]?

**English Example Sentence:** But only a brave man could have written this book [11].

**Uzbek Example Sentence:** Erkin Vohidov Yaxshidir achchiq haqiqat ozodligi uchun bosh ko'tarishga da'vat etadi: So'yla, So'yla, qadding ko'tarib baland [12].

**English Example Sentence:** Our fearless leader has fearlessly lost it [11].

**Uzbek Example Sentence:** unchalikka «Bu galgi guldastraning kattaligiga qaraganda, oshiq yigitning shashti ancha balandga o'xshaydi. Bir kecha xoli qo'ysak, nima Malika qizlar atay g'ashiga..., Akbar Mirzo So'nggi chora 2011 [12].

From the above examples, we can identify the following synonymous series of adjectives in combination with different nouns Fig. 2.

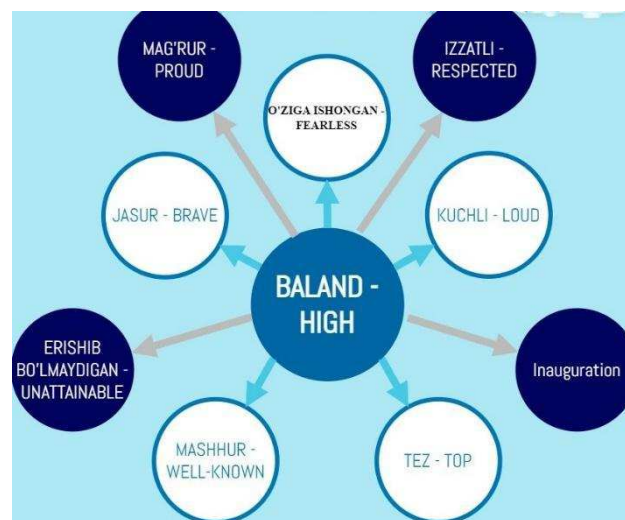


Fig. 2. Synonymous line of Uzbek-English adjectives.

In order to tackle these issues, the subsequent remedies were put into practice:

- *Collaborative Translation:* Using a number of native speakers and translators ensured accurate and culturally appropriate translations.

- *Advanced Corpus techniques:* Processing big datasets quickly was made possible by the use of powerful corpus analysis techniques.

- **User feedback:** The thesaurus's usability was enhanced and entries were refined by taking user feedback into account.

The following are some possible effects of this multilingual thesaurus's creation:

- *Better Quality of Translation:* Precise and contextually relevant translations support excellent bilingual communication.

- *Language Learning:* This thesaurus acts as a great tool for Uzbek and English learners by helping in vocabulary development and comprehension of its usage.

- *Cultural Exchange:* The thesaurus promotes communication and understanding across cultural divides with the help of bridging language gaps.

## V. RECOMMENDATION

In the future, this research can be utilized to develop and support more sophisticated cross-linguistic studies in the adjective-noun relationship including the English, Russian, and Uzbek languages, hence improving machine translation systems. By using these structural similarities, the constraints of these machines regarding under-represented languages, like Uzbek, could be addressed, and their application in multilingual non-dominant environments could be enhanced.

Finally, I would like to express my gratitude to Professor Nilufar Abdurakhmanova of the National University of Uzbekistan for all of her invaluable contributions to the field of computational linguistics with reference to the Uzbek corpus. She has made the advancements required for more inventive and in-depth research in this area possible, including enabling the success of this kind of study.

## VI. CONCLUSION

The study underlines the necessity of generating new lexical resources to correspond with the dynamic changing of linguistic needs. The bilingual thesaurus of new and emerging adjectives from English to Uzbek is considered an important resource in order to enhance multilingual competence and cultural awareness for educators, learners, and translators. Further development will be done on the thesaurus in the future and research into its uses in different linguistic and pedagogical contexts. With the constant growth of interglobal communication, a bilingual thesaurus must be open to new and upcoming adjectives in order to maintain current and meaningful language tools. Technology, society, and the culture are all expanding at such a rapid pace that lexical databases must be constantly revised to accommodate the ways in which language molds itself to modernity. A complete thesaurus between English and Uzbek would allow teachers to give students the ability to analyze and discuss contemporary texts and contexts more thoroughly. This site not only aids in translation work but also helps promote cross-cultural comprehension, for the use of this site will allow for the most subtle nuances of meaning to be preserved in the translation of languages. The growth of vocabulary in both languages is essential to the development of a high level of language fluency which will enable the user to successfully maneuver intricate phrases in various academic, professional, and social situations. Not only that, but the addition of newly formed adjectives allows students to express themselves in a way to describe new experiences, especially in the fields of technology, social justice, and

ecological change. This thesaurus is under continual investigation as to its effectiveness and hopefully will shed some light on how to best design bilingual lexical tools for various user populations such as students, translators and language professionals. Also, the thesaurus will be a prototype for other thesaurus in other language pairs, and will therefore help in the overall goal of language diversity and accessibility in the world. This is a living project, and it is clear that linguists, educators, and technologists must work together to see that the thesaurus continues to be a live document growing with the languages it represents. In the end, this bilingual thesaurus can have a key part in the field of language pedagogy and intercultural communication, with users having the ability to articulate themselves more eloquently in English and Uzbek.

## REFERENCES

- [1] T. Atkins and M. Rundell, "The Oxford Guide to Practical Lexicography," Oxford University Press, 2008, pp. 256-259. ISBN: 9780199277711. DOI: 10.1093/acprof/9780199277711.001.0001. "The Oxford Guide to Practical Lexicography" [Online]. Available: [https://www.academia.edu/21938330/The\\_Oxford\\_Guide\\_to\\_Practical\\_Lexicography](https://www.academia.edu/21938330/The_Oxford_Guide_to_Practical_Lexicography).
- [2] R. Carter and M. McCarthy, Cambridge Grammar of English: A Comprehensive Guide, Cambridge University Press, 2006, pp. 654-658. ISBN: 9780521674393. DOI: 10.1017/CBO9780511486401. [Online]. Available: <https://www.scrip.org/>.
- [3] R. R. K. Hartmann, "Interlingual Lexicography: Selected Essays on Translation Equivalence, Contrastive Linguistics and the Bilingual Dictionary," Cambridge Scholars Publishing, 2007, pp. 102-105. ISBN: 9781847183557. [Online]. Available: <https://www.semanticscholar.org/>.
- [4] K. Kanzaki, Q. Ma, E. Yamamoto, and H. Isahara, "Unlike previous research, which found at least two ways to evaluate the usefulness of a thesaurus," in Proceedings of the International Conference on Language Resources and Evaluation (LREC), Genoa, Italy, May 2006, pp. 2230-2235. ISBN: 9782951740860. DOI: 10.1145/1164853. [Online]. Available: <https://aclanthology.org/volumes/L08-1/>.
- [5] N. Abdurakhmonova, I. A. Shakirovich, and K. N. S. O'G'Li, "Morphological analyzer (morfoAnalyse) Python package for Turkic language," Science and Education, vol. 3, no. 4, pp. 45-50, 2022. [Online]. Available: [https://www.researchgate.net/publication/365106575\\_MorphUz\\_Morphological\\_Analyzer\\_for\\_the\\_Uzbek\\_Language](https://www.researchgate.net/publication/365106575_MorphUz_Morphological_Analyzer_for_the_Uzbek_Language).
- [6] D. Sharipova, "Semantic Changes in Uzbek Adjectives," Tashkent University Press, 2015, pp. 120-125. ISBN: 9789943023456.
- [7] R. Rahmatullaev, "The Dynamics of Adjective Meanings in Uzbek," Tashkent University Press, 2018, pp. 98-103. ISBN: 9789943045670. [Online]. Available: [https://www.researchgate.net/publication/348749918\\_Comparative\\_analysis\\_of\\_adjectives\\_in\\_English\\_and\\_Uzbek\\_languages](https://www.researchgate.net/publication/348749918_Comparative_analysis_of_adjectives_in_English_and_Uzbek_languages)
- [8] N. Z. Abduraxmonova, "Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aftoref." (2018).
- [9] L. Kubedinova, A. Khusainov, D. Suleymanov, R. Gilmullin, N. Abdurakhmonova. "First Results of the TurkLang-7 Project: Creating Russian-Turkic Parallel Corpora and MT Systems. Proceedings of the Computational Models in Language and Speech Workshop" (CMLS 2020) co-located with 16th International Conference on Computational and Cognitive Linguistics (TEL 2020) .2020/11: 90-101. CMLS 2020 Proceedings (CEUR-WS).
- [10] P. Baker, A. Hardie, and T. McEnery, A Glossary of Corpus Linguistics, Edinburgh University Press, 2006, pp. 78-81. ISBN: 9780748621361. DOI: 10.3366/Edinburgh/9780748621361.001.0001. [Online]. Available: <https://journals.openedition.org/>
- [11] Reverso Context. [Online]. Available: <https://context.reverso.net/>.
- [12] Uzbek National Corpus. [Online]. Available: <https://uzbekcorpus.uz/>.