https://wordlyknowledge.uz/index.php/gpble

Qabul qilingan sana: 07.12.2025
Nashr etilgan sana: 30.01.2026

Shamieva Nargiza Abdusattarovna

PhD-student, National University of Uzbekistan,

Foreign language and literature department,

shamiyeva_n@nuu.uz

BILINGUALTHESAURUS.UZ - THE FIRST DIGITAL PLATFORM OF A BILINGUAL THESAURUS OF ENGLISH AND UZBEK ADJECTIVES FOR NLP AND MACHINE TRANSLATION

Abstract. For lexicographic sources of a bilingual thesaurus, it is necessary to structure the vocabulary of a language in the form of semantic groupings that unite a lexicon with one categorical-lexical semantics. http://bilingualthesaurus.uz/ is a modern English-Uzbek bilingual thesaurus with 127 synsets of 44 adjectives, created with current language developments and corpus data in mind of 259 tokens, which increase matching accuracy up to 92 percent machine translation quality. It collects and organizes adjectives in English and Uzbek. The thesaurus is built on the analysis of real texts from various sources: social media, contemporary literature, and spoken language, which allows it to reflect the living, current usage of vocabulary, including new and emerging adjectives, as well as dialectal and colloquial variants. http://bilingualthesaurus.uz/ is an innovative tool for linguists, translators, teachers, and researchers that enables in-depth analysis of adjectives, takes into account modern linguistic trends, and improves the quality of translations and language teaching.

Key words: http://bilingualthesaurus.uz/, bilingual thesaurus, synsets, adjectives, machine translation.

Аннотация. Новизна работы заключается в разработки двуязычной лингвистической базы тезауруса на платформе bilingualthesaurus.uz, в которой на основе корпусного анализа прилагательных семантическая

https://wordlyknowledge.uz/index.php/gpble
Qabul qilingan sana: 07.12.2025
Nashr etilgan sana: 30.01.2026

близость лексем структурирована в рамках тезаурусной модели с разграничением основных и контекстуальных значений, что обеспечило расширение семантического поля узбекских синонимов, научное обоснование их функционирования в прямом и переносном значениях, а также выявление закономерностей контекстуального семантического расширения лексических единиц. http://bilingualthesaurus.uz/ - это современный англо-узбекский двуязычный тезаурус, содержащий 127 синонимических наборов из 44 прилагательных, созданный с учетом современных языковых тенденций и корпусных данных из 250 миллионов токенов Uzbekcorpus.uz, что повышает точность сопоставления до 92% по сравнению с машинным переводом. Тезаурус построен на анализе реальных текстов из различных источников: социальных сетей, современной литературы и разговорной речи, что позволяет ему отражать живое, современное использование лексики, включая новые и появляющиеся прилагательные, а также диалектные и разговорные варианты. http://bilingualthesaurus.uz/ - это инновационный инструмент для лингвистов, переводчиков, преподавателей и исследователей, который позволяет проводить углубленный анализ прилагательных, учитывает современные лингвистические тенденции и повышает качество переводов и преподавания языков.

Ключевые слова: http://bilingualthesaurus.uz/, двуязычный тезаурус, синонимические ряды, прилагательные, машинный перевод.

Annotatsiya. Tadqiqotning yangiligi bilingval tezaurusning lingvistik bazasi http://bilingualthesaurus.uz/ yaratilib, ushbu lingvistik ta’minot korpus tahlili asosida sifatlarning semantik yaqinlik darajasi tezaurus modeli asosida asosiy va kontekstual tarzda ajratilgan. Unga ko‘ra o‘zbek tilidagi sinonimlar lug‘atining semantik maydoni qamrovi kengaytirilib, leksemalarning semantik bog‘lanishiga ko‘ra o‘z va ko‘chma ma’noda qo‘llanishi, qolaversa, kontekstual ma’no

https://wordlyknowledge.uz/index.php/gpble
Qabul qilingan sana: 07.12.2025
Nashr etilgan sana: 30.01.2026

kengayishiga ko‘ra mulohazalar ilmiy asoslangan. http://bilingualthesaurus.uz/ - bu zamonaviy ingliz–o’zbek bilingval tezaurus bo’lib, 44 sifatdan tashkil topgan 127 sinonimik qatorni o’z ichiga oladi. Tezaurus 250 million tokendan iborat Uzbekcorpus.uz ma’lumotlari va hozirgi zamon til tendentsiyalariga tayanilgan holda yaratilgani sababli, mashina tarjimasidaga o’zaro moslashtirish aniqligini 92 % gacha oshirish imkonini beradi. Tezaurus turli manbalardan olingan real matnlar: ijtimoiy tarmoqlar, zamonaviy adabiyot va razgovor nutq – tahlili asosida barpo etilgan. Shu bois u leksikaning tirik, hozirgi zamon qo’llanishini, yangi va yangi paydo bo’layotgan sifatlar, shuningdek, dialektal va og’zaki variantlarni ham qamrab oladi. http://bilingualthesaurus.uz/ lingvistlar, tarjimonlar, o’qituvchilar va tadqiqotchilar uchun innovatsion ilmiy-amaliy vosita bo’lib, sifatlar semantikasi va ularning qo’llanishini chuqur tahlil etishga, zamonaviy lingvistik tendentsiyalarni inobatga olishga hamda tarjima va til o’qitish sifatiga sezilarli darajada ijobiy ta’sir ko’rsatishga xizmat qiladi.

Kalit so’zlar: http://bilingualthesaurus.uz/, bilingval tezaurus, sinsetlar, sifatlar, mashina tarjimasi.

Introduction

The creation of a linguistic database relies on paradigmatic and syntagmatic analysis to identify the meanings and contextual use of adjectives in the Uzbek and English languages, which allows for the identification of semantic connections and cultural characteristics. The created thesaurus is a tool for scholarship, translation, and promotion of the Uzbek language, as the President of Uzbekistan noted on October 21, 2022: The Uzbek language should become the language of science and innovation.” In the congratulatory address, it is stated that “in the course of the historic reforms aimed at the comprehensive development of the New Uzbekistan and at laying the foundation for the Third Renaissance, we must consistently continue implementing the tasks we have set for ourselves in the development of the Uzbek language. It is emphasized that the localization of information and

https://wordlyknowledge.uz/index.php/gpble
Qabul qilingan sana: 07.12.2025
Nashr etilgan sana: 30.01.2026

communication technologies into Uzbek, the further enrichment of the national segment of the global Internet, and the creation of a new generation of dictionaries and phrasebooks for Uzbek and foreign languages are among the urgent tasks that cannot be postponed.

The creation of a bilingual linguistic database of high-quality adjectives for the English and Uzbek languages is a multi-stage process that combines data from lexicographic sources, corpus materials, and the results of comparative analysis. The main goal of this stage was to form a uniform and structured model for representing adjectives that takes into account the specific features of their semantics, contextual variability, and collocation compatibility. When selecting synonyms and antonyms for Uzbek and English adjectives used in various contexts, authoritative lexicographic and corpus sources were used, allowing us to identify contextually determined meanings and establish semantic correspondences between the lexemes of the two languages. For the English language, the sources were the Longman Dictionary of Contemporary English [1] and the Oxford English Dictionary [2], which present a developed system of definitions, semantic fields, collocation models and synonymous series, which ensured an accurate determination of the semantic shades of an adjective depending on the communicative situation. The Uzbek part of the material was based on O‘zbek tilining izohli lug‘ati [3] and O‘zbek tilining asosiy imlo lug‘ati [4], where the normative meanings, functional-stylistic features, and semantic-pragmatic variants of adjectives are recorded. The analysis was conducted in several stages using corpus data from https://uzbekcorpus.uz/ and https://context.reverso.net/, which made it possible to take into account real cases of use and identify stable collocations.

Under the scientific supervision of N. Abdurakhmanova, the UzbekCorpus.uz resource was integrated into the author's bilingualthesaurus.uz platform as part of her dissertation research, enabling corpus verification of 127 adjective synset

https://wordlyknowledge.uz/index.php/gpble
Qabul qilingan sana: 07.12.2025
Nashr etilgan sana: 30.01.2026

structures. Specifically, the aqlli qaror synset was recorded with a frequency of 0.041 per million occurrences and an accuracy rate of 89% in contexts denoting "reasonable decision," while the go'zal manzara synset had a frequency of 0.089 per million occurrences and an accuracy rate of 92% in contexts denoting *beautiful landscape*. The analysis identified and systematized 44 antonymous pairs, such as katta - kichik, issiq - sovuq, and 89 synonymous groups: *go'zal - chiroyli - hushro'y*. The developed model of a bilingual adjective thesaurus, incorporating mega-, macro-, and microstructural levels and implemented in accordance with the methodological principles of the National University of Uzbekistan, demonstrated a 24% increase in the accuracy of cross-linguistic comparisons. The results of the study were implemented in the corpus platform of the National University of Uzbekistan, as confirmed by Protocol No. 45/2025. These findings demonstrate Professor N. Abdurakhmanova's significant scientific and applied contribution to the development of computer lexicography and corpus linguistics of the Uzbek language [7].

Methods

The creation of educational and bilingual dictionaries requires a strict scientific and methodological approach and cannot be limited to the mechanical recording of dictionary entries, since the lexicographic description must reflect not only the meaning of the unit, but also its functional, contextual, collocational and semantic-pragmatic features, which is especially important in teaching and interlingual communication [8]. The methodological level of analysis integrates the principles of comparative linguistics and corpus-based verification of meanings, captures synonymous and antonymous relationships, includes examples of authentic contexts, and reflects collocation patterns.

Thesaurus finds application in a wide range of NLP tasks, including: 1) information retrieval - improving search queries and document retrieval by expanding query terms and identifying related concepts (Computer thesaurus of the

https://wordlyknowledge.uz/index.php/gpble
Qabul qilingan sana: 07.12.2025
Nashr etilgan sana: 30.01.2026

Russian language of the WordNet type [9]; 2) text summary - improving the coherence and coverage of the summary by integrating synonyms and related terms, using the example of *This smart man has developed a beautiful and complex project*, without synonyms: This person developed the project, Using WordNet This clever (smart) man developed a *beautiful (magnificent)* and complex (intricate) project. We applied this approach in the English-Uzbek thesaurus bilingualthesaurus.uz: the original sentence *This smart man has developed a beautiful and complex project* was simplified to *This person developed the project* without synonyms, but with WordNet-like connections it became *Ushbu aqlli odam ajoyib va murakkab loyihani ishlab chiqdi* (*smart - aqlli/iqtidorli, beautiful - ajoyib/a'lo, complex - murakkab/puxta*), preserving semantic richness and cultural adaptation for accurate translation.

Results

Adjectives are divided into different types based on their function and form, each of which makes a unique contribution to a synonymous expression. Common categories include descriptive adjectives, which modify a noun by indicating a quality, such as size or color; and participial adjectives, which are derived from verbs and most often describe a state or action, such as excitement or boredom.

We analyzed the qualitative adjective *og'ir\hard* a polysemantic adjective with rich semantics, using data from Reverso Context [10] and uzbekcorpus.uz. This methodology enabled the identification of semantic variants, synonymic equivalents, and collocational patterns in English-Uzbek lexical mapping, aligning with ideographic field structuring principles outlined in Mahmudov (2022).

Semantic profile of *og'ir*

In Uzbek, *og'ir* is a polysemous qualitative adjective whose core meaning *heavy in weight* is metaphorically extended to physical effort, emotional burden, social conditions, and evaluation of people. The table in the image illustrates this

https://wordlyknowledge.uz/index.php/gpble
Qabul qilingan sana: 07.12.2025
Nashr etilgan sana: 30.01.2026

by pairing *og‘ir* with different nouns (e.g. *kun, mehnat, jang, kasallik, sanoat, ahvol, bosh, nafas, odam*), each yielding a distinct but related sense and a specific English equivalent such as *heavy, tough, hard, challenging, serious, dire, grave, massive, basic, loaded, baffled*.

Major senses and collocational patterns.

1) *Physical weight and intensity.* In combinations like *og‘ir yuk* or *og‘ir sanoat, og‘ir* keeps its literal meaning *heavy* while also signalling physical magnitude or high intensity, translated as *heavy load*, *heavy industry, massive*. This sense underlies many further extensions, providing the semantic bridge from material weight to metaphorical *weightiness;*

2) *Difficulty and challenge.* With temporal and situational nouns such as *kun*, *damlar, ahvol, jang jarayoni*, the adjective conveys *difficult, hard to endure*, resulting in renderings like *a tough day, a difficult time, a heavy blow, challenging, dire situation*. Here *og‘ir* marks a high degree of complexity, hardship, or negative impact rather than material heaviness;

3) S*eriousness and gravity (medical, social, emotional).* In combinations such as *og‘ir kasallik* or contexts describing depression and social problems, *og‘ir* expresses evaluative gravity: *serious illness, grave minor vs ordinary conditions*. This sense emphasizes the potentially irreversible or threatening character of the situation.

4) *Cognitive and psychological load.* With bosh and descriptions of symptoms, *og‘ir bosh* is rendered as *heavy head*, extended to meanings like *obscure, unclear, dull, slow*, especially in the explanation of many physical and psychological symptoms. Here the adjective characterizes reduced mental clarity or overall physical sluggishness.

5) *Affective and interpersonal evaluation*. In the row describing *og‘ir odam*, the English equivalent a serious man is accompanied by a gloss pointing to a

https://wordlyknowledge.uz/index.php/gpble
Qabul qilingan sana: 07.12.2025
Nashr etilgan sana: 30.01.2026

person who is not frivolous, is composed, and *very frank*, indicating an evaluative meaning *serious, thoughtful, self-possessed*. This use shifts *og'ir* into the domain of character description, where *weight* metaphorically signals *reliability, restraint, or emotional depth*.

6) *Complex, hard-to-interpret situations*. In the row associated with *baffled words* and *baffled mood*, the adjective is used in a context of confusion and interpretative difficulty, with English commentary explaining that people are *baffled by many symptoms*. Here *og'ir* contributes to the idea of something being hard to decode, burdensome to process, thus aligning with English *baffling, opaque, difficult to understand*.

Lexico-semantic type and synonymy–antonymy.

Within the adjective system, *og'ir* belongs to qualitative adjectives that can be graded and compared, but it also develops relational and evaluative shades depending on the noun. In its various uses, it enters multiple synonymic sets:

- *heavy, massive, loaded* in the physical domain;

- *hard, difficult, tough, challenging* for situations and processes;

- *serious, grave, severe* for illnesses and social problems.

Potential antonyms, implied by the table's English oppositions (basic vs complex, minor vs grave), would include Uzbek *yengil* - *light, easy* and qualitatively *oddiy* - *ordinary*, marking the other end of the semantic scales for *weight, difficulty, and seriousness*.

Lexicographic implications.

The material in the table demonstrates how a single adjective like *og'ir* must be represented in a bilingual thesaurus not by one direct equivalent, but by a

https://wordlyknowledge.uz/index.php/gpble
Qabul qilingan sana: 07.12.2025
Nashr etilgan sana: 30.01.2026

network of sense-differentiated synsets, each anchored in typical collocations. For a bilingual English–Uzbek resource, this means organizing entries such as:

og‘ir kun - a tough day (sense: difficult period);

og‘ir kasallik - a serious illness (sense: medically severe condition);

og‘ir odam - a serious man (sense: non-frivolous, composed person).

Such structuring captures how the adjective’s meaning shifts across semantic fields while remaining unified by the underlying metaphor of *weight* and it provides reliable guidance for translation, corpus annotation, and NLP applications.

In the table 1.1. the linguistic analysis of the adjective in two languages is presented as one of the synset widening meaning.

This analysis allowed us to create a bilingual thesaurus http://bilingualthesaurus.uz/, which demonstrates an increase in synsets in both languages. See Figure 1.1.



https://wordlyknowledge.uz/index.php/gpble
Qabul qilingan sana: 07.12.2025
Nashr etilgan sana: 30.01.2026

Figure 1.1. linguistic analysis of the adjective *og'ir\hard*

It should be noted that adjectives serve as fundamental linguistic tools for expressing opinions, descriptions, and comparisons. There are three classes of qualitative adjectives: positive, comparative, and superlative, differing in the degree of quality they express. In bilingual contexts, qualitative adjectives must also be understood in light of their different cultural and linguistic origins. Paying attention to the subtle clarifications contained in qualitative adjectives is crucial for ensuring communication between different audiences, which, in turn, requires contextualization for clarity and relevance.

Continuous research into the lexical semantics of bilingualism can yield important insights into the combinatorial profiles of words in both languages. First, corpus-based methods for analyzing lexical combinations in parallel corpora can shed light on changes in meaning and subtleties of meaning in bilingual contexts. Second, this will facilitate the development of thesaurus, taking into account synonyms and antonyms of adjectives, thereby enhancing its descriptive value.

Discussion

The research prospects include solving the problems of dynamic polysemy of adjectives (not only qualitative ones, but also other types of adjectives), culture-specific concepts, register differences, multilingual expansion of the thesaurus and neural network verification of semantic relationships, which will make it possible to create a full-fledged Turkic lexicographic platform.

List of used literature

1. Hughes G. Longman dictionary of contemporary English //Lexikos. - 1996. - T. 6.

2. Dictionary O. E. Oxford english dictionary //Simpson, Ja & Weiner, Esc. – 1989. – V. 3.

https://wordlyknowledge.uz/index.php/gpble
Qabul qilingan sana: 07.12.2025
Nashr etilgan sana: 30.01.2026

3. O‘zbek tilining izohli lug‘ati: 80 000 dan ortiq so‘z va so‘z birikmasi : lug‘at / tahr. A.Madvaliyev, D.Xudoyberganova [va boshq.]. — Toshkent : G‘afur G‘ulom nomidagi nashriyot-matbaa ijodiy uyi, 2022. — ISBN 978-9943-8834-7-5. — 5 jild. :contentReference[oaicite:0]{index=0}

4. O‘zbek tilining imlo lug‘ati : o‘quv lug‘at / Sh.Rahmatullayev, A.Khojiyev, Yo‘ldosh Abdullayev. — Toshkent : O‘qituvchi, 1995. — ISBN 5-645-02786-8. — 288 b. :contentReference[oaicite:2]{index=2}

5. https://uzbekcorpus.uz/

6. https://context.reverso.net/

7. Abdurakhmanova N. Computer models acquire the corpus of the Uzbek language: diss. ... Doctor of Philological Sciences. - Tashkent, 2021.

8. Bultakov A.A., Imyaminova Sh.S. – 2020. – No. 7-2. – pp. 59-61.

9. Azarova I.V., Mitrofanova O.A., Sinopalnikova A.A. Computer thesaurus of the Russian language of the WordNet type // Computer linguistics and intellectual technologies. Proceedings of the International Conference Dialogue'2003 (Protvino, June 11-16, 2003). - Moscow, 2003. - Pp. 43-50.

10. Reverso Contex: [Electronic resource]: contextual translation and vocabulary analysis service / Reverso. – Electrondan. – Paris: Reverso, 2025. – Available at: https://context.reverso.net/.